Deutsches Zentrum für Luft- und Raumfahrt
German Aerospace Center

BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Temporal complexity reduction for modeling the sector coupling of gas and electricity

Isabelle Viole

Matriculation number 883033

Department VIII of Mechanical Engineering, Event Technology, Process Engineering

Beuth University of Applied Sciences Berlin

Master thesis submitted for the degree of

Master of Engineering (M.Eng.)

Master Programme **Maschinenbau Erneuerbare Energien**
(Mechanical Engineering - Renewable Energies)

Submission: 7 December 2020

Supervisors

Prof. Dr. P. Kohlenbach — Beuth University of Applied Sciences Berlin

M.Sc. W. Heitkötter — German Aerospace Center (DLR), Institute of Networked Energy Systems

Abstract

In the future, the sector coupling of power, heat, gas and mobility shall enable a stronger cross-sectoral integration of renewable energies and provide new flexibility options for the electricity sector, e.g. with seasonal storage of energy in the gas network. The optimal expansion of the grid and storage units within energy systems is determined by applying optimisation models. These are becoming increasingly complex and computationally more challenging, e.g. with the expansion of renewable energies and the coupling of sectors, the intermittent time series of renewable power generators and the state-of-charge of long-term energy storage systems have to be accounted for. Complexity reduction methods are widely used to reduce computing time while maintaining a reasonable accuracy. In recent research, a large number of methods to reduce temporal complexity emerged that are tailored to power-based energy system models. However, their applicability to sector-coupled models is still to be explored.
This thesis applies known methods that reduce the temporal complexity in systems with long-time storage options in *eTraGo*, an energy system model that couples electricity and gas. An innovative method for temporal complexity reduction, hierarchical clustering of consecutive hours with variable segment lengths, is integrated into *eTraGo*. The applicability of this method is evaluated against a previously integrated method. Both methods are applied in a scenario with significant Power-to-gas installations.
When these methods cluster the gas sector's input time series by itself, the deviation to the reference case gas time series is smaller than when the clustering algorithms are applied on the input time series of both sectors. When applied to the time series of both sectors, the novel method introduces less deviation than the previously implemented method.
This thesis found that the innovative method of hierarchical clustering of consecutive hours results in a better trade-off between reducing computing time and the error introduced to the annual costs of the whole system than the previously introduced method to reduce temporal complexity. However, both methods severely deviate by more than 50% from the reference case with regards to the electrolyzer expansion.

Kurzfassung

Die Kopplung der Energiesektoren Strom, Wärme, Gas und Mobilität soll in Zukunft eine stärkere sektorenübergreifende Integration von erneuerbaren Energien ermöglichen sowie neue Optionen für Flexibilität im Stromsektor schaffen, wie z. B. durch die saisonale Speicherung von Energie im Gasnetz. Um den optimalen Ausbau verschiedener Transport- und Speichermedien in Energiesystemen zu bestimmen, werden Optimierungsmodelle verwendet. Mit dem wachsenden Anteil von erneuerbaren Energien und der Kopplung von Sektoren wird diese Modellierung zunehmend komplexer und rechnerisch herausfordernder, da Zeitreihen von fluktuierenden erneuerbaren Stromerzeugern und die Speicherstände langfristiger Energiespeicher berücksichtigt werden müssen. Komplexitätsreduktionsmethoden sind gebräuchlich, um die Rechenzeit von Energiesystemmodellen zu reduzieren und gleichzeitig eine akzep-

table Genauigkeit beizubehalten. In wissenschaftlichen Arbeiten der letzten Jahre sind eine Vielzahl an Methoden zur Reduzierung der zeitlichen Komplexität entwickelt worden, die auf Energiesystemmodelle des Stromsektors anwendbar sind. Ihre Anwendbarkeit auf sektorgekoppelte Modelle ist jedoch noch weitgehend unerforscht.

Diese Masterarbeit integriert verschiedene Methoden zur zeitlichen Komplexitätsreduktion in *eTraGo*, einem gekoppelten Strom- und Gasmodell, und wertet deren Anwendbarkeit aus. Dabei werden Komplexitätsreduktionsmethoden ausgewählt, die auf Systeme mit langfristigen Speicherpotentialen anwendbar sind. Eine innovative Methode, das hierarchische Clustering von aufeinanderfolgenden Stunden mit variablen Segmentlängen, wird dafür in *eTraGo* implementiert. Die Anwendbarkeit dieser Methode wird mit einer vorab im Modell integrierten Methode verglichen. Beide Methoden werden in einem Szenario mit viel Power-to-Gas Kapazität angewandt. Wenn die Clustering-Algorithmen nur auf die Inputzeitreihe des Gassektors im Modell angewandt werden, so ist die Abweichung der daraus resultierenden Inputzeitreihen vom gesetzten Referenzfall niedriger als wenn die Algorithmen mit den Inputzeitreihen sowohl des Gas- als auch des Stromsektors eingesetzt werden. Wenn sie auf beide Sektoren angewandt werden, kann die innovative Clusteringmethode deutlich niedrigere Abweichungen in den Inputzeitreihen erreichen als die bereits zuvor im Modell integrierte Methode.

Die neu implementierte Methode führt im Rahmen des hier gesetzten Referenzfalls zu einem besseren Trade-Off zwischen Rechenzeitverkürzung und Abweichung der jährlichen Kosten des Gesamtsystems. Beide verglichenen Methoden resultieren jedoch in einer Abweichung von mehr als 50% im Elektrolyseurausbau.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations

CH_4	Methane
H_2	Hydrogen
BNetzA	Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, engl.: Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway
CAPEX	Capital expenditures
DCE	Duration curve error
DLR	Deutsches Luft- und Raumfahrtzentrum, engl.: German Aerospace Center
DSO	Distribution system operator
DVGW	Deutscher Verein des Gas- und Wasserfaches, engl.: German Association for gas and water applications
EHV	Extra-high voltage, i.e. 380 or 220 kV in Germany
ENTSO-E	European Network of Transmission System Operators for Electricity
ESM	Energy system model
GHG	Greenhouse gas
HV	High voltage, i.e. 60-110 kV in Germany
LOPF	Linear optimal power flow
nRMSE	normalized Root Mean Square Error
OPEX	Operational expenditures
P2G	Power-to-gas
RES	Renewable energy sources
segm/y	Number of segments per year
tday/y	Typical days per year
TSAM	Time series aggregation method

Acronyms

$\overline{Seg0}$, $\overline{Seg1}$, $\overline{Lin0}$, $\overline{Lin1}$	average over the z runs of $Seg0$, $Seg1$, $Lin0$, $Lin1$, respectively

eGo^n	energy Grid optimisation of n flexibilities
$Seg0_z$, $Seg1_z$, $Lin0_z$, $Lin1_z$	z^{th} run of $Seg0$, $Seg1$, $Lin0$, $Lin1$, respectively
eTraGo	electric Transmission Grid optimization
Lin0	*eTraGo* run with the linkage method applied on the input time series of the gas sector only
Lin1	*eTraGo* run with the linkage method applied on all input time series of the gas and power sectors
NEP 2025 B	Netzentwicklungsplan, engl.: Network Development Plan, Scenario B 2035, version from 2015
NEP 2035 C	Netzentwicklungsplan, engl.: Network Development Plan, Scenario C 2035, version from 2021
NEP 2035	Scenario of *eTraGo*, consisting of scenario B2-2035 [10] for the power-side and Scenario C 2035 [6] for the gas-side
open_eGo	open electricity Grid optimization
PyPSA	Python for Power System Analysis
Ref	Reference case run of *eTraGo*, no time series aggregation applied
Seg0	*eTraGo* run with the segmentation method applied on the input time series of the gas sector only
Seg1	*eTraGo* run with the segmentation method applied on all input time series of the gas and power sectors
tsam	time series aggregation module

***eTraGo* formula symbols [Unit]**

$\bar{L}_{n,P\to\mathrm{H_2}}$	Maximum installable electrolyzer capacity at bus n	$\mathrm{MW_{el}}$
$\bar{N}_{n,t,\mathrm{H_2}}$	Maximum hydrogen feed-in capacity in gas network at bus n at time step t	$\mathrm{MW_{th}}$
$\bar{N}_{n,\mathrm{H_2}}$	Maximum hydrogen feed-in capacity in gas network at bus n	$\mathrm{MW_{th}}$
ℓ	Index for line/transformer	–
$\eta_{P\to\mathrm{H_2}}$	Electrolyzer efficiency	$\mathrm{MW_{th}/MW_{el}}$
η_s^{char}	Efficiency of storage charging	–
η_s^{dis}	Efficiency of storage discharging	–
η_s^{self}	Self discharge efficiency of storage s	–
$[h_{n,s,t}]^+$	Positive part of storage dispatch n, s, t	$\mathrm{MW_{el}}$
$\tilde{L}_{n,P\to\mathrm{H_2}}$	Minimum installable electrolyzer capacity at bus n	€/$\mathrm{MW_{el}}$
c_ℓ	Annualised capital expenditures of line/transformer	€/$\mathrm{MW_{el}}$
$c_{n,r}$	Annualised capital expenditures of generator r at bus n	€/$\mathrm{MW_{el}}$

$c_{n,s}$	Annualised capital expenditures of storage s at bus n	€/$\mathrm{MW_{el}}$
$c_{P\rightarrow \mathrm{H_2}}$	Electrolyzer capital expenditures	€/$\mathrm{MW_{el}}$
$d_{\mathrm{H_2},n,t,th}$	Hydrogen load at bus n at time step t	$\mathrm{MW_{th}}$
$d_{\mathrm{H_2},t,th}$	Hydrogen demand at time step t	$\mathrm{MW_{th}}$
$d_{P\rightarrow \mathrm{H_2},n,t,pu}$	Per unit grid-oriented electrolyzer load at bus n at time step t	–
$E_{\mathrm{H_2}stor,t}$	Hydrogen storage state of charge at time step t	$\mathrm{MWh_{th}}$
$E_{\mathrm{H_2}stor}$	Hydrogen storage energy capacity	$\mathrm{MWh_{th}}$
$E_{P\rightarrow \mathrm{H_2},th}$	Annual hydrogen production	$\mathrm{MWh_{th}}$
F_ℓ	Branch capacity of line/transformer	$\mathrm{MW_{el}}$
$g_{\mathrm{H_2},n,t,th}$	Hydrogen produced at bus n at time step t	$\mathrm{MW_{th}}$
$g_{n,r,t}$	Dispatch of generator r at bus n at time step t	$\mathrm{MW_{el}}$
$G_{n,r}$	Generator capacity of generator r at bus n	$\mathrm{MW_{el}}$
$g_{P\rightarrow \mathrm{H_2},n,t}$	Electrolyzer dispatch at bus n at time step t	$\mathrm{MW_{el}}$
$g_{P\rightarrow \mathrm{H_2},t,th}$	Total hydrogen production by electrolysis at time step t	$\mathrm{MW_{th}}$
$H_{n,s}$	Rated capacity of storage s at bus n	$\mathrm{MW_{el}}$
$L_{n,P\rightarrow \mathrm{H_2}}$	Electrolyzer capacity at bus n	$\mathrm{MW_{el}}$
$l_{P\rightarrow \mathrm{H_2},n,t,th}$	Hydrogen production by electrolysis at bus m at time step t	$\mathrm{MW_{th}}$
$L_{P\rightarrow \mathrm{H_2},NEP}$	Total electrolyzer built capacity in *NEP 2035 C*	$\mathrm{MW_{el}}$
$l_{P\rightarrow \mathrm{H_2},t}$	Maximum electrolyzer load time series over all buses at time step t	$\mathrm{MW_{el}}$
$m_{\mathrm{H_2}}$	Index for collective hydrogen bus	–
n	Index for bus	–
$o_{n,r}$	Operational expenditures of generator r at bus n	€/$\mathrm{MW_{el}}$
$o_{n,s}$	Operational expenditures of storage dispatch s at bus n	€/$\mathrm{MW_{el}}$
$P^{char}_{s,t}$	Charging capacity of storage s at time step t	MW
$P^{dis}_{s,t}$	Discharging capacity of storage s at time step t	MW
P^{nom}_{s}	Nominal power of storage	MW
r	Index for generator/load technology	–
$SOC^{intra}_{s,t}/SOC^{intra}_{s,t+1}$	State of charge of storage s at time step t/$t+1$	MWh
s	Index for storage	–
t	Index for time step	–
$t_{\mathrm{max},s}$	Maximum amount of hours storage s can be discharged at full capacity	h
w_t	Weighting of time step t	–
$d_{\mathrm{H_2}stor,t,th}$	Hydrogen storage or discharge at time step t	$\mathrm{MW_{th}}$
$FLH_{P\rightarrow \mathrm{H_2},NEP}$	Electrolyzer full load hours in *NEP 2035 C*	h

Gas network storage estimation formula symbols [Unit]

Δm_i	Storable mass of gas i in gas network based on the pressure difference (estimation case 1)	kg
d_{DN}	Average nominal diameter of the high pressure gas pipes	m
$E_{ann,C1}$	Estimated annual energy storage capacity of methane-hydrogen mixture in the gas network based on the pressure difference (estimation case 1)	$\mathrm{TWh_{th}}$
E_{C1}	Energy storage capacity of methane-hydrogen mixture in the gas network based on the pressure difference (estimation case 1)	$\mathrm{TWh_{th}}$
$E_{dem,a}$	Gas demand in year a by energy	TWh
$E_{i,C1}$	Energy storage capacity of gas i in the gas network based on the pressure difference (estimation case 1)	$\mathrm{TWh_{th}}$
$E_{i,C2}$	Energy storage capacity of gas i in the gas network based on 5 vol% rule (estimation case 2)	$\mathrm{TWh_{th}}$
$H_{U,i}$	Lower heating value of gas i	$\mathrm{TWh_{th}}$
i	Index for type of gas	$-$
l_{gas}	Length of the high pressure gas network	$\mathrm{m^3}$
m	Mass	kg
p	Pressure	Pa
p_1	Pressure at point 1 in gas network	Pa
p_2	Pressure at point 2 in gas network	Pa
R_i	Specific gas constant	$\mathrm{Jkg^{-1}K^{-1}}$
T	Temperature	K
V	Volume	$\mathrm{m^3}$
$V_{dem,a}$	Gas demand in year a by volume	$\mathrm{m^3}$
V_{gas}	Volume of the high pressure gas network	$\mathrm{m^3}$
$V_{i,a}$	Demand of gas i in year a by volume	$\mathrm{m^3}$
Z	Compressibility	$-$
$Z_{1,i}$	Compressibility of gas i at p_1	$-$
$Z_{2,i}$	Compressibility of gas i at p_2	$-$

Other formula symbols [Unit]

$\eta_{H_2 \rightarrow CH_4}$	Methanation efficiency	$-$
$\eta_{P \rightarrow H_2 \rightarrow CH_4}$	Power-to-hydrogen-to-methane efficiency	$-$
$\eta_{P \rightarrow H_2 \rightarrow P}$	Power-to-hydrogen-to-hydrogen efficiency	$-$
$\eta_{H_2 stor}$	Hydrogen storage efficiency	$-$

μ_c	k-centroid of cluster c	–
$\overline{\mathbf{x}}_i$	Centroid of cluster i	–
$\overline{t_{ref}}$	Average computing time over reference case runs	–
$\overline{t_{TSAM}}$	Average computing time over runs of a TSAM applied	–
AOE	Average objective error	%
ATR	Average time reduction	%
c	Index for cluster	–
$D_{Ward}(i,j)$	Dissimilarity between pair of clusters i,j	–
i	Period in Figure 2.12	
i,j	Pair of clusters	–
k	Set number of clusters	–
N_g	Point in time after one period i in Figure 2.12	
$nRMSE$	Normalized root mean square error	%
t	Index for time step	–
T_t	Number of time steps t	–
x	n-dimensional vector $\{x_1, x_2, ..., x_n\}$	–
x_i	Vector x at i	–
x_t	Original state in Figure 2.12	
$x_{app,t}$	Approximated value at time step t	–
x_{app}	Approximated value	–
x_{ig}^{intra}	Intra-period state in Figure 2.12	
x_i^{inter}	Inter-period state in Figure 2.12	
$x_{ref,t}$	Reference value at time step t	–
x_{ref}	Reference value	–
x_t	Data point at time step t	–
y	n-dimensional vector $\{y_1, y_2, ..., y_n\}$	–
y_i	Vector y at i	–
$d_{\mathrm{E}}(x,y)$	Euclidean distance of two vectors x and y	–
$d_{\mathrm{M}}(x,y)$	Manhattan distance of two vectors x and y	–
$\overline{\lvert x_{ref}\rvert}$	Average of absolute reference values of all time steps t	–

Chapter 1

Introduction

The transition of primary energy production to use fossil-free sources is an essential cornerstone to significantly reduce greenhouse gas (GHG) emissions and to meet the 2 $°C$ target of the Paris Climate Agreement of 2015 [11]. To model the effects of policies or the expansion of certain technologies, energy system modeling is an important tool. It can also serve as a guide for decision-makers. With the rising number of optimisation variables in these models, complexity reduction methods become increasingly important, as they allow acceptable levels of computing times while maintaining much of the accuracy. While different methods of complexity reduction are being applied in power-only models [12], complexity reduction methods that can be used in sector-coupled models, i.e. models that also include other energy systems, such as the gas system, are not well researched to date. This work is dedicated to find, apply and evaluate temporal complexity reduction methods that are usable in sector-coupled energy system modeling of gas and power in Germany.
This thesis is conducted in cooperation with the German Aerospace Center (DLR) and aims to reduce the temporal complexity in the model *eTraGo* (electric Transmission Grid optimisation), an open-source optimisation model which couples the electricity and gas systems. In pursuing this goal, three research questions are investigated:

1. *Which methods for temporal complexity reduction over time can be used to efficiently reduce the computing time of the model?*
2. *To what extent are these methods applicable when the storage capacity of the gas network is taken into account? What adjustments are necessary for that?*
3. *How can these complexity reduction methods be used across sectors for both power and gas?*

This work is split into the following chapters:

2 **Theoretical background and methodology**: What role does the sector coupling of gas and electricity have in the German energy transition? How much gas and hydrogen can be stored within the German gas grid? How is this storage potential alongside with Power-to-Gas units implemented in *eTraGo*? Which different methods to reduce complexity in the temporal scope exist and which are applicable on the model used?

3 **Results and analysis**: Analyses results on both input and output values from applying the temporal complexity reduction methods in comparison to a set reference case.

4 **Resume and outlook**: Deals with the evaluation of the results of the various methods. Discusses the results and provides an outlook on future scientific work in the field of complexity reduction as well as needed work on complexity reduction within *eTraGo*.

Chapter 2

Theoretical background and methodology

Energy system models (ESMs) are used to provide analysis and insights on how energy currently is and will be extracted, converted, transported and utilized [7]. They are relevant for grid operators, policy-makers and investors alike, as they project information like the necessary expansion and the costs of energy systems. With the planned expansion of RES and sector coupling, ESMs also need to adjust for these changes. This increases their complexity significantly, as a larger spatial and temporal resolution is needed for the decentralized and intermittent sources of power such as solar and wind power generation as well as the decentralized Power-to-gas (P2G) facilities [7]. A method to keep these models computable is complexity reduction. This reduces the number of optimisation variables by decreasing the number of time steps and nodes for ESMs, thereby decreasing the models' complexity. The key trade-off when reducing complexity in a model is the conflict of objectives between lowering the computational effort and maintaining an acceptable modelling accuracy, i.e. with as little deviation from the non-complexity-reduced output data as possible. The quality reached in these two goals by a trade-off will be a key indicator to evaluate the different methods applied in this thesis in Section 3.
This chapter provides the theoretical background relevant to the research questions outlined in Chapter 1 and describes the methods applied to answer them in later sections. 2.1 describes the planned expansion of renewable energy generation and sector coupling in German policy, with a closer look at the coupling of the gas and power sector and P2G technology. Next, 2.2 expands on the theory of ESMs in general and the model applied in this work, *eTraGo*, in specific. It describes how *eTraGo* finds an optimal solution and how the gas sector is integrated into this formerly power-only model. After the validity of the model is discussed based on previously conducted studies on the power-only side of *eTraGo* and the selected input data sets, reasons why *eTraGo* needs to be reduced in its temporal complexity are presented. Once 2.3 provides a general overview on potential methods to reduce complexity in the temporal sphere, clustering algorithms are selected as the complexity reduction method for this thesis and 2.4 explains frequently used metrics and types of algorithms. 2.5 compares recent literature where authors applied temporal clustering algorithms on ESMs and discusses which of these methods can be applied to sector-coupled models. In this, an especial account is taken of the P2G electrolyzer and gas storage options that *eTraGo* comprises. Based on this comparison, two methods are selected and 2.6 describes how these methods are applied in *eTraGo*. This chapter closes with an explanation of the indicators that will be used to analyse the results in Chapter 3.

2.1 The expansion of renewable energies and sector coupling in Germany

Climate change is one of the main challenges of the 21st century, with the global mean temperature reaching 1.1 °C over pre-industrial global mean temperature in 2019 [13]. At the Paris climate conference in 2015, 197 parties, including Germany, negotiated an agreement to limit the rise of the global mean temperature in the current century to a maximum of below 2 °C above pre-industrial temperatures by reducing the emission of GHGs [11].

One key measure to reduce GHG emissions is to transition away from fossil energy sources. In Germany, with the phasing out of nuclear energy by the end of 2021, this task has to be fulfilled by renewable energy sources (RES) only. Looking at their current deployment, they had a share of 17 % of the final energy consumption in 2019 [14], which the Federal Government wants to raise to 30 % by 2030 [15]. However, the usage of renewable energies is distributed unevenly amongst energy sectors. While the electricity sector had a share of 42.1 % of power from RES, the share of renewable energy in the heating and mobility sector were 14.5 % and 5.6 %, respectively in 2019 [14]. This discrepancy can be particularly problematic when climate protection and renewable energy targets become more demanding in the future. Sectors that are more difficult to decarbonize will have to increase their usage of renewable energies proportionally more with increased climate targets. This is because sectors where it has been easier to implement RES, such as the electricity system, cannot contribute much more to GHG savings than they already do.

One way to address this problem is sector coupling. This concept envisages coupling the previously largely independent energy systems of electricity, gas, heat, and mobility. This can be done by transforming power into gas with P2G technology, power into heat with heat pumps, or using battery-driven cars [16] (see the scheme of a sector-coupled energy system in Figure 2.1). On the one hand, this enables the use of electricity from RES in the other sectors and thereby achieves a higher penetration of renewable energies throughout. On the other hand, synergistic effects such as using the gas network as a flexible buffer for the electricity sector make this concept interesting in terms of the security of power supply [17].

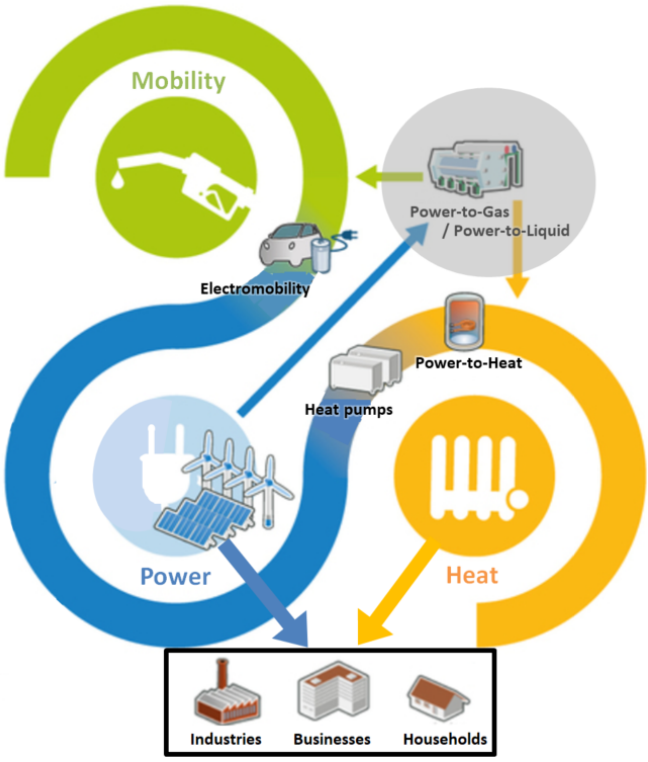

Figure 2.1: Scheme of sector-coupled energy system, translated and amended from [1].

The German government sees sector coupling as a relevant method for the country's energy transition and its goal to lower GHG emissions, as shown by mentions in the Climate Action Plan 2050 [16] and the Green Paper on Energy Efficiency [18]. The term itself can be defined differently, e.g. the *German Federal Ministry of Economics and Energy* explains, "electricity from renewable sources must be used efficiently for heat provisioning, the transport sector and industry" [18, p. 14].

The *German Association for Gas and Water applications* (DVGW) states, "sector coupling connects the electricity, heating and gas networks as well as the mobility sector"[1] [19]. For the scope of this work, the definition of the DVGW will be followed, as it highlights the role of gas for sector coupling, which will be discussed next.

2.1.1 A closer look: Sector coupling of power and gas

The scope of this thesis is the coupling of the power and gas sectors. Coupling these includes the conversion of electricity into gas, the storage of gas and the re-conversion of gas into electricity, as shown in Figure 2.2. Both the energy transition and renewable goals can benefit from the sector coupling of

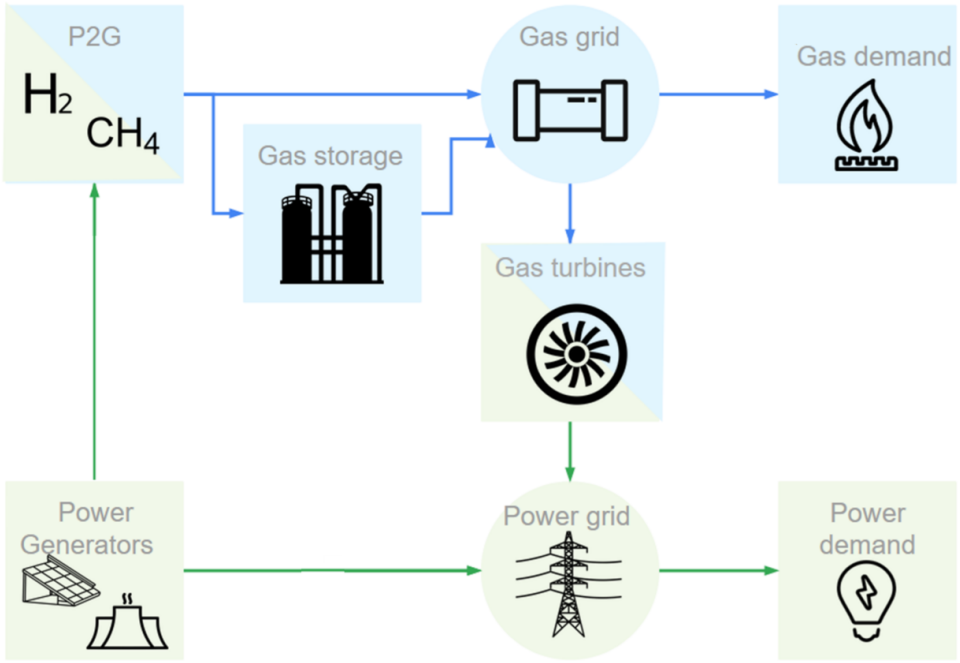

Figure 2.2: Sector coupling of the power and gas sector.

power and gas. One reason that is often discussed to transform power into gas with electrolyzers is excess power from intermittent RES. When the power generation from solar and wind generators exceeds the demand or cannot be transported to the consumers due to network congestions, these power plants need to be curtailed. Solar generation in this work means photovoltaic cells, as these are the main solar power generators in Germany. In 2019, about 1.1 % of all power generated was curtailed [20]. A study by the *Fraunhofer ISE* projects the curtailed power to increase to 6 $\mathrm{TWh_{el}}$ or a share of 1.6 % of the generated power by 2050, when RES contribute more to the total electricity production [21]. Other sources estimate this amount to reach $9 - 12\,\mathrm{TWh_{el}}$ in 2030 [22] [23]. While applying P2G to use this excess electricity is often discussed, some authors state that building electrolyzers for only excess power would not be economical, as the electrolyzer capacity would only be used during a few hours of extremely high supply from RES [24]. The capital costs of the electrolyzers (in this work set to 350000 €/MW_{el} [5]) would then have to be passed on to the hydrogen produced in these few hours.

[1] Quote translated from German, original version: "Die Sektorenkopplung (auch Sektorkopplung genannt) verbindet die Strom-, Wärme- und Gasnetze als auch den Mobilitätssektor miteinander." [19]

Electrolyzers could however also be operated more continuously, as a flexibility option for the power sector. Flexibility becomes increasingly important with the rising share of energy from intermittent RES. Photovoltaic systems and wind turbines depend on sun and wind supply, which makes their production ever-changing and prone to forecasting errors. This requires measures for flexibility. In 2019, these two technologies produced 71.1 % of all power from RES in Germany [14]. The expansion of RES in the power system together with the planned electrification of the heat and electricity sectors results in a higher need for flexibility in the power systems of the close future. P2G can also help to minimise the expansion of the power grid infrastructure, as the gas produced can be transported in the gas network and therefore less electricity burdens the power grid [25].

The gas sector can deliver flexibility to the power sector both in times when power generation falls below or exceeds the demand. In the case of supply exceeding demand, P2G plants can convert power into gas. Electrolyzers produce hydrogen (H_2) by splitting water into oxygen and hydrogen, using electricity. If the hydrogen is not to be used or stored as such, methanation can be carried out as a second step, in which carbon dioxide (CO_2) is added to obtain methane (CH_4) [26]. Note that all these transformation steps imply conversion losses. The generated gas can be fed into the gas network. Hydrogen is allowed to be fed into the gas network until it reaches a concentration of 5 vol% today due to the safety hazard that this highly explosive gas poses [27]. Applying this cap, the hydrogen storage capacity of the German gas grid is estimated at 3 to 15 $\mathrm{TWh_{th}}$ (see 2.2.4 for more detail). Methane can be fed into the grid without any percentage limits. Apart from gas storage within the gas network, gas can also be stored underground. These storage units were primarily constructed to compensate between the even supply of natural gas from the border import points and in-country production, and the temperature-dependent gas demand. In 2018, Germany had an underground storage capacity of $280.02\,\mathrm{TWh_{th}}$ working gas, of which $134.12\,\mathrm{TWh_{th}}$ were to be stored in caverns, the rest in pore storage and other storage units. This storage covered about 27.3 % of that year's gas demand [28], [29].

When the power demand exceeds the supply, non-intermittent power generators need to supply the power difference. Among conventional power plants, combined-cycle gas turbines are expected to be the generators of choice in the upcoming decades, as they have a comparatively short ramping time and low GHG emissions [30].

In addition to the need for flexibility within minutes and hours, scenarios with an even higher share of renewable energies face an issue known in German as the so-called *Kalte Dunkelflaute*[2], that describes a period of consistently cold temperatures on top of low wind speeds. This leads to a low power generation from intermittent RES, while demand is higher than average because of the typical seasonal demand for lighting and heating. Seasonal storage options need to be available to secure supply during these times. One option for this would be to store carbon-neutral gas, which was produced by RES and P2G, and to convert this back into electricity by gas-fueled power plants.

2.1.2 Estimating the gas storage capacity in the German gas network

Storing gas within the gas grid is possible because gases are compressible. In contrast to the power grid, which always needs to balance generation and demand at any point in time, the gas network can

[2] engl.: Cold and dark doldrums. Since this term is not generally used in English, the German term is applied, which often occurs in political discussions.

stand a difference between feed-in and feed-out at a point in time. The difference is accounted for by the high-pressure gas grid whose pressure may theoretically vary between 10^5 and 10^7 Pa [§ 3 Nr. 16 EnWG]. Medium and low-pressure pipes need to be kept at a more constant pressure and hence are not considered as storage options in this thesis [31].

Different numbers can be found for the energy storage potential of the gas network. The total gas storage potential based on pressure differences within the gas grid is stated to be $130\,\mathrm{TWh_{th}}$ by different gas associations/companies and the Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway (BNetzA) [32], [33], [34]. When enquiring at the BNetzA about the calculation for this storage amount, they could however not retrace where this number came from.

The total hydrogen storage potential within the gas grid at a hydrogen proportion of 5 vol% was estimated by two studies at $3\,\mathrm{TWh_{th}}$ [35] or $15\,\mathrm{TWh_{th}}$ [36].

As the specifications on the storage potential within the gas grid differ significantly, the potential storage capacities for gas and hydrogen in the German gas grid are estimated in this work as follows: Case 1 focuses on the storage potential of natural gas due to the possible pressure difference in the gas network, and case 2 looks at the overall storage potential of hydrogen within the gas grid. Both estimations are set for the year 2030, as this was the closest year to the later-on used scenario year 2035 for which input values could be found. Based on these estimations, a gas network storage will be included in the model *eTraGo* in Section 2.2.4.

Case 1 - Storage potential due to pressure difference

It is known that the high pressure pipes in Germany have a total length $l_{gas} = 1.22 \cdot 10^8\,\mathrm{m}$ [37]. Additionally, this estimation assumes:

1. The natural gas in the network consists of 95 vol.-% methane and 5 vol.-% hydrogen, which is the maximum hydrogen content allowed by the DVGW [27].
2. The pipes have an average nominal diameter of $d_{DN} = 0.86\mathrm{m}$ (calculated by weighted averaging data from [38]).
   From (1.) and (2.) the volume V of the gas network follows as:

$$V_{gas} = l_{gas}\pi\frac{{d_{DN}}^2}{4} \approx 7.087 \cdot 10^7\mathrm{m}^3 \tag{2.1}$$

   where

   V_{gas} : volume of the high pressure gas network $[\mathrm{m}^3]$
   l_{gas} : length of the high pressure gas network $[\mathrm{m}]$
   d_{DN} : average nominal diameter of the high pressure gas pipes $[\mathrm{m}]$.

3. According to [38], the high pressure network's operating pressure p_{op} can vary from $1.6 \cdot 10^6$ to $10^7\,\mathrm{Pa}$, depending on the pipe type. An average pressure difference allowed per pipe of $\Delta p = 5 \cdot 10^6\,\mathrm{Pa}$ is assumed, varying between $p_1 = 5 \cdot 10^6\,\mathrm{Pa}$ and $p_2 = 10^7\,\mathrm{Pa}$.
4. A temperature of $273.15\,\mathrm{K}$ is assumed for all calculations.
5. The annual gas consumption in Germany in 2030 is estimated to be $779\,\mathrm{TWh_{th}}$, as in the reference case of [39].

As the gases are under high pressure, one cannot assume the gas to be an ideal gas. The real gas law applies:

$$pV = mZR_iT \tag{2.2}$$

where

p : pressure [Pa]
V : volume [m^3]
m : mass [kg]
Z : compressibility [$-$]
R_i : specific gas constant of gas i [$\mathrm{Jkg^{-1}K^{-1}}$]
T : temperature [K].

With the gas-specific values in Table 2.5, the gas mass Δm that can be stored follows:

$$\begin{aligned} \Delta m_{\mathrm{CH_4}} &= \left(\frac{p_2}{Z_{2,\mathrm{CH_4}}} - \frac{p_1}{Z_{1,\mathrm{CH_4}}}\right) \cdot \frac{0.95 V_{gas}}{R_{\mathrm{CH_4}} \cdot T} \approx 2.866 \cdot 10^9 \,\mathrm{kg} \\ \Delta m_{\mathrm{H_2}} &= \left(\frac{p_2}{Z_{2,\mathrm{H_2}}} - \frac{p_1}{Z_{1,\mathrm{H_2}}}\right) \cdot \frac{0.05 V_{gas}}{R_{\mathrm{H_2}} \cdot T} \approx 1.342 \cdot 10^7 \,\mathrm{kg} \end{aligned} \tag{2.3}$$

where

i : Index for type of gas [-]
Δm_i : Storable mass of gas i in gas network [kg]
p_2 : Pressure at point 2 in gas network [Pa]
p_1 : Pressure at point 1 in gas network [Pa]
$Z_{2,i}$: Compressibility of gas i at p_2 [$-$]
$Z_{1,i}$: Compressibility of gas i at p_1 [$-$]
V_{gas} : Volume of the high pressure gas network [m^3]
R_i : Specific gas constant of gas i [$\mathrm{Jkg^{-1}K^{-1}}$].

Table 2.5: Gas-specific values of methane ($\mathrm{CH_4}$) and hydrogen ($\mathrm{H_2}$).

	Formula symbol	$i = \mathrm{CH_4}$	$i = \mathrm{H_2}$	Source
Compressibility of gas i at p_2	$Z_{1,i}$	0.92	1.03	[40]
Compressibility of gas i at p_1	$Z_{2,i}$	0.84	1.06	[40]
Specific gas constant	R_i in $\frac{\mathrm{J}}{\mathrm{kgK}}$	518.4	4124.2	[41]
Lower heating value	$H_{U,i}$ in $\frac{\mathrm{MJ}}{\mathrm{kg}}$	50.01	119.97	[42]

Using the lower heating values H_U of the gases, the energy that can be stored in the network is:

$$\begin{aligned} E_{\mathrm{CH_4},C1} &= \Delta m_{\mathrm{CH_4}} \cdot H_{U,\mathrm{CH_4}} = 39.813 \,\mathrm{TWh_{th}} \\ E_{\mathrm{H_2},C1} &= \Delta m_{\mathrm{H_2}} \cdot H_{U,\mathrm{H_2}} = 0.447 \,\mathrm{TWh_{th}} \end{aligned} \tag{2.4}$$

$$E_{C1} = E_{\mathrm{CH_4},C1} + E_{\mathrm{H_2,C1}} = 40.261 \,\mathrm{TWh_{th}} \tag{2.5}$$

where

i : Index for type of gas [-]
$E_{i,C1}$: Energy storage capacity of gas i in the gas network based on the pressure difference (estimation case 1) $[\mathrm{TWh_{th}}]$
Δm_i : Storable mass of gas i in gas network based on the pressure difference (estimation case 1) $[\mathrm{kg}]$
$H_{U,i}$: Lower heating value of gas i $[\mathrm{MJ/kg}]$
E_{C1} : Energy storage capacity of methane-hydrogen mixture in the gas network based on the pressure difference (estimation case 1) $[\mathrm{TWh_{th}}]$.

The high-pressure gas grid can contain a maximum of 79.6 $\mathrm{TWh_{th}}$ at $p = 10^7$ Pa while the annual gas consumption in Germany in 2030 is assumed to be 779 $\mathrm{TWh_{th}}$. Using a very rough estimation, the contents of the gas network at the maximum pressure would have to be replenished about 9.8 times per year to transport the demanded gas to the consumer. Hence theoretically the calculated storage capacity by pressure difference could be used 9.8 times per year, resulting in:

$$E_{ann,C1} \approx E_{C1} \cdot \frac{779\,\mathrm{TWh_{th}}}{79.6\,\mathrm{TWh_{th}}} = 394.012\,\mathrm{TWh_{th}} \tag{2.6}$$

where

E_{C1} : Energy storage capacity of methane-hydrogen mixture in the gas network based on the pressure difference (estimation case 1) $[\mathrm{TWh_{th}}]$
$E_{ann,C1}$: Estimated annual energy storage capacity of methane-hydrogen mixture in the gas network based on the pressure difference (estimation case 1) $[\mathrm{TWh_{th}}]$.

This estimation exceeds the 130 $\mathrm{TWh_{th}}$ stated by i.a. the BNetzA by a factor of 3 [34]. As the input values in this estimation included rough simplifications regarding the pipe diameter and the permitted pressure differences, it makes sense that less gas can realistically be stored within the grid. The 130 $\mathrm{TWh_{th}}$ can be viewed as within scope based on this estimation calculation. However, a more precise calculation should be conducted before this value is integrated in a scientific energy model.

Case 2 - Overall hydrogen storage potential within the gas network

This estimation calculates the overall storage energy capacity of hydrogen, adopting some assumptions of case 1:

1. A hydrogen share of 5 vol%.
2. A temperature of 273.15 K.
3. A yearly natural gas demand in 2030 of 779 $\mathrm{TWh_{th}}$.
4. The gas network is operated at the pressure $p1 = 5 \cdot 10^6$ Pa.

Retracing the calculation of Götz et al. [36], the volumetric gas demand in 2030 is estimated. The yearly natural gas demand in 2030 assumed by the *Deutsche Energie-Agentur* is 1.17% higher than the demand in 2015 [39]. With this ratio and the volumetric gas demand in 2015 of $77 \cdot 10^9$ $\mathrm{m^3}$ [43], the volumetric

gas demand in 2030 follows as:

$$V_{dem,2030} \approx V_{dem,2015} \cdot \frac{E_{dem,2030}}{E_{dem,2015}} = 77 \cdot 10^9\,\mathrm{m}^3 \cdot \frac{779\,\mathrm{TWh_{th}}}{770\,\mathrm{TWh_{th}}} = 77.9 \cdot 10^9\,\mathrm{m}^3 \tag{2.7}$$

where

$V_{dem,a}$: Gas demand in year a by volume $[\mathrm{m}^3]$
$E_{dem,a}$: Gas demand in year a by energy $[\mathrm{TWh}]$.

The volumetric hydrogen demand at 5 vol% hydrogen is:

$$V_{\mathrm{H_2,2030}} = 0.05 \cdot V_{dem,2015} = 3.895 \cdot 10^9\,\mathrm{m}^3 \tag{2.8}$$

where

i : Index for type of gas [-]
$V_{i,a}$: Demand of gas i in year a by volume $[\mathrm{m}^3]$.

At $p1 = 5 \cdot 10^6$ Pa, the hydrogen in [TWh] in this volume follows:

$$E_{H_2,C2} = \frac{p1}{Z_{1,H_2}} \cdot \frac{V_{H_2,2030}}{R_{H_2} \cdot T} \cdot H_{U,H2} \approx 16.784\,\mathrm{TWh} \tag{2.9}$$

where

i : Index for type of gas [-]
$E_{i,C2}$: Energy storage capacity of gas i in the gas network based on 5 vol% rule (estimation case 2) $[\mathrm{TWh_{th}}]$
p_1 : Pressure at point 1 in gas network $[\mathrm{Pa}]$
$Z_{1,i}$: Compressibility of gas i at p_1 $[-]$
$V_{i,a}$: Demand of gas i in year a by volume $[\mathrm{m}^3]$
R_i : Specific gas constant of gas i $[\frac{\mathrm{J}}{\mathrm{kgK}}]$
T : Temperature $[\mathrm{K}]$.

The estimation of Götz et al. could be reproduced with a solution deviating by 1 $\mathrm{TWh_{th}}$ / 11.9% from the one in the paper [36].
As will be discussed in Section 2.2, at the time of this work, only electrolyzers are considered to be feeding into the gas network within the model applied. Therefore, solely the potential hydrogen storage within the gas grid needs to be integrated. For this, the storage capacity of Götz et al. of 15 $\mathrm{TWh_{th}}$ hydrogen is followed because this number was comprehensible within this estimation. For further works, the gas storage potential should be taken into account due to the pressure difference. It is advised to carry out more precise calculations before doing so.

2.1.3 The Power-to-gas expansion planned in the German network development plan

To reach the Federal Government's goal to reduce GHG emissions by 95 % in 2050 in relation to 1990, the gas sector needs to be decarbonized [6]. Following the concept of sector coupling, P2G is a key technology to accomplish this goal, as it can produce carbon-neutral gas, given that the power used originates from renewable sources[3]. P2G can furthermore contribute to more gas import independence and security of supply [45].
A limit of P2G is its economics: Every transformation step causes losses. Electrolysis and methanation today both have an efficiency of 80 % (alkaline or PEM electrolysis / catalytic methanation) [5]. These result in a combined Power-to-Methane efficiency and a Power-to-Hydrogen-to-Power efficiency of both 64 % (assuming a storage efficiency of 1) [5]. See Table 2.11 for an overview of P2G efficiencies. Different studies behold hydrogen and methane produced with P2G from RES as not competitive with natural gas and grey hydrogen (produced from fossil fuels) today, while projecting that it could become competitive in the German market within ten years [46], [47].

Table 2.11: P2G efficiencies [5].

	Formula symbol	Value
Electrolysis efficiency	$\eta_{P\rightarrow H_2}$	0.80
Methanation efficiency	$\eta_{H_2\rightarrow CH_4}$	0.80
Power-to-hydrogen-to-methane efficiency	$\eta_{P\rightarrow H_2\rightarrow CH_4}$	0.64
Power-to-hydrogen-to-hydrogen efficiency	$\eta_{P\rightarrow H_2\rightarrow P}$	0.64
Hydrogen storage efficiency	η_{H_2stor}	1.00

This work is conducted assuming demand for P2G and hydrogen as stated in the *2035 Network Development Plan, Scenario C 2035, Version 2021 Draft* from January 2020 (*NEP 2035 C*) [6]. During this thesis, a new version of the *2035 Network Development Plan* was published by the BNetzA. As this version is not yet implemented in the used model *eTraGo*, this thesis is based on the numbers from the draft version from January 2020 for the gas side of the model. The *NEP 2035 C* scenario includes the highest share of sector coupling in the system of all presented scenarios in the plan, which are listed in Table 2.12. It assumes a capacity of 7.5 GW of Power-to-Hydrogen and 0.5 GW of Power-to-Methane facilities by 2035. 3 GW of the Power-to-Hydrogen facilities are assumed to be power-network-oriented to avoid congestions in the transmission network. This is a favourable assumption, as e.g. the *German National Hydrogen Strategy* from June 2020 only plans to establish up to 5 GW of electrolysis generation capacity in 2030 [48].

[3] A major issue in a scenario to reduce GHG emissions by 95 % is finding a source for the CO_2 required for methanation. Most of the industrial processes in which CO_2 nowadays can be separated will have to be carbon-neutral and not emit GHGs by then. Capturing industrial CO_2 exhaust for methanation would not be in line with carbon-neutrality, as the CO_2 would, in the end, be emitted to the atmosphere.
One possibility to get around this problem is to extract CO_2 via direct air capture. This energy-intensive technology is not yet widespread, as it has a price of today more than 200 €/t CO_2. Some authors assume that the price of CO_2 from direct air capture could be lowered down to 23 €/t CO_2 in 2050, under the assumption of a low electricity price and the usage of waste heat for the process [44].

Table 2.12: Power demand and P2G capacities in the *NEP 2035 C* [6].

	Value
Power demand	634 TWh_{el}
thereof: P2G	30 TWh_{el}
Heat	35 TWh_{el}
Transport	52 TWh_{el}
P2G capacities	
H_2	7.5 GW
CH_4	0.5 GW

2.2 The energy system model *eTraGo* and its need for complexity reduction

This segement introduces energy system models (ESMs) in general and the model applied in this thesis, *eTraGo*, in particular. After 2.2.1 gives an overview of the existing types of ESMs, 2.2.2 explains how the power network model of *eTraGo* works. The techno-economical optimisation on which the model is based is introduced in 2.2.3. Section 2.2.4 then explains how the gas sector is included in the formerly power-only model. Details on the scenario applied and on the modelling are provided in 2.2.5. The segment concludes by discussing the validity of *eTraGo* (2.2.6) and explaining its need for complexity reduction (2.2.7).

2.2.1 Overview of energy system models

The increasing relevance of climate and energy policies, as discussed in Section 2.1, together with economic interests in the energy field and the need for energy security lead to the development of many different ESMs. They are used to describe how energy today and/or in the future will be extracted, converted, transported and utilized [7].
Depending on their modelling goal, ESMs can be set up in a variety of ways. Pfenninger et al. distinguished them based on the model's focus (see Table 2.13) [7]. While energy system optimisation and simulation take into account physical constraints such as line loading, power system and market models often model the entire power grid as a copper plate and do not take into account line capacities or transport losses.

Table 2.13: Distinction of energy system model groups [7].

	Model type	Focus
1	Energy systems optimisation models	Normative scenarios
2	Energy systems simulation models	Predictions and Forecasts
3	Power systems and electricity market models	Operational and business decisions

ESMs generally are based on optimisation methods that look for an optimal solution that is determined in the target function. This can e.g. be the minimization of system costs. Next to the target function, set

constraints represent technological and economical aspects of the system to be modelled, such as line capacities or capital and operational expenditures (CAPEX/OPEX). Methods to model energy systems include linear programming, mixed-integer programming, agent-based simulation or computable general equilibrium. In a review of 75 modelling tools, Ringkjob et al. [49] found linear programming to be the most often applied.

This thesis applies and extends *eTraGo*, a linear programming based energy system optimisation model. It models a power network as a graph of nodes connected by lines with line limits. This takes real power flow limitations into account. *eTraGo* optimises for total system costs. For better readability, this thesis will continue to use the broader term ESM when talking about *eTraGo*.

2.2.2 The power network model *eTraGo*

eTraGo was originally built for the research project *open_eGo* (open electricity Grid optimisation) to serve as an open data network model for the technical and economic optimisation of the transmission network expansion in Germany on the extra-high-voltage (EHV) and high voltage (HV) level (380, 220 and 110 kV) [2]. It uses the open-data-network-model developed within the project, which was published on the *OpenEnergyDatabase*[4]. In the follow-up project eGo^n (energy Grid optimisation of n flexibilities), this power-only model is expanded by other sectors to allow for modelling sector coupling in Germany [50]. Figure 2.3 gives an overview on the power-only functionality of *eTraGo*.

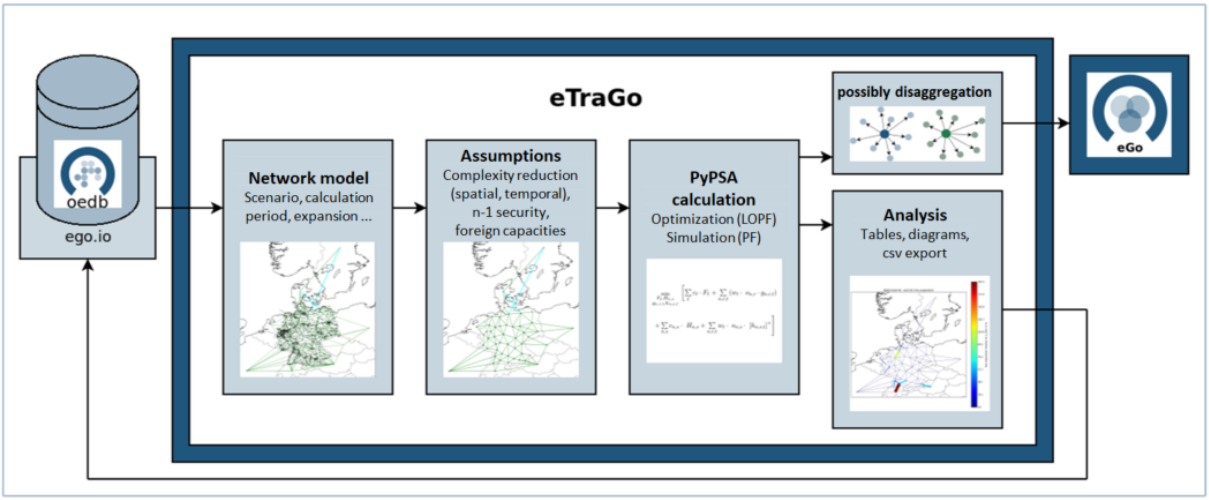

Figure 2.3: Scheme of *eTraGo*, taken and translated from [2].

The goal of *eTraGo* is to optimise a given energy system by minimizing the total system costs. The steps applied by the power-only version of this model to achieve this can be broken down into:

1. Import network from the *OpenEnergyDatabase* database.
2. Cluster the EHV network.
3. Apply spatial clustering using a k-means algorithm (algorithm explained in 2.4.1).
4. Apply temporal clustering, e.g. snapshot skipping or hierarchical clustering using methods from the Python tool *tsam* (algorithm explained in 2.4.2).

[4] https://openenergy-platform.org/

5. Perform a linear optimal power flow (LOPF) calculation using the Python tool *PyPSA* (Python for Power System Analysis) over a set number of loops. Doing this, the output network of loop n is set as the input network for loop $n + 1$.

After the LOPF calculation, the results from the optimisation can be evaluated with further functions, or after disaggregation can be passed on as an input to the cross-network calculation tool *eGo* which allows for subsequent optimisation of the distribution grids.

The following chapter sums up the techno-economical optimisation used within *eTraGo* and the spatial and temporal complexity reduction methods that were already included in the original power-only version of *eTraGo*. Detailed documentation of all data and methods used is published in the final report of *open_eGo* [2].

2.2.3 Techno-economical optimisation

eTraGo uses the LOPF method from *PyPSA* to perform the techno-economical optimisation on the network. The tool *PyPSA* is given certain components that it then uses for its optimisation. The important components for this thesis are briefly defined in Table 2.14.

Table 2.14: Selection of *PyPSA* components used within this work [8].

Component	Definition
Network	Container of all other network components
Bus	Fundamental nodes to which all other components attach
Carrier	Energy carriers (e.g. wind, solar, gas, etc.)
Load	A consumer of energy
Generator	Generator whose feed-in can be flexible subject to minimum loading or minimum down and up times, or variable according to a given time series of power availability
Storage Unit	A device which can shift energy from one point of time into the future, subject to efficiency losses. If the nominal capacity is extendable, it gets optimised in the LOPF according to the investment costs
Store	A more fundamental storage object with no restrictions on charging or discharging power

Target of the LOPF is to minimize the yearly total system costs, consisting of OPEX and CAPEX of the generation, storage and transmission units [8]. Within this work the focus lies on the expansion of the transmission lines and energy storages, the target function of the LOPF hence is:

$$\min_{F_\ell, G_{n,r}, H_{n,s}, g_{n,r,t}, h_{n,s,t}} \Bigg[\overbrace{\sum_\ell c_\ell \cdot F_\ell}^{\text{grid expansion}} + \overbrace{\sum_{n,r} c_{n,r} \cdot G_{n,r}}^{\text{generation CAPEX}} + \overbrace{\sum_{n,r,t} w_t \cdot o_{n,r} \cdot g_{n,r,t}}^{\text{generation OPEX}} + \underbrace{\sum_{n,s} c_{n,s} \cdot H_{n,s}}_{\text{storage expansion}} + \underbrace{\sum_{n,r,t} w_t \cdot o_{n,s} \cdot [h_{n,s,t}]^+}_{\text{storage OPEX}} \Bigg] \tag{2.10}$$

where

ℓ	: Index for line/transformer	w_t	: Weighting of time step t
n	: Index for bus	$o_{n,r}$	: OPEX of generator r at bus n
r	: Index for generator/load technology	$g_{n,r,t}$	: Dispatch of generator r at bus n at time step t
s	: Index for storage		
t	: Index for time step	$c_{n,s}$	: Annualised CAPEX of storage s at bus n
c_ℓ	: Annualised capital expenditures of line/transformer ℓ	$H_{n,s}$	: Rated capacity of storage s at bus n
F_ℓ	: Branch capacity of line/transformer ℓ	$o_{n,s}$	: OPEX of storage dispatch s at bus n
$c_{n,r}$	: Annualised CAPEX of generator r at bus n	$[h_{n,s,t}]^+$	: Positive part of storage dispatch n, s, t
$G_{n,r}$	: Generator capacity of generator r at bus n.		

Electrotechnical conditions such as Kirchhoff's laws and maximum generator dispatch are set in constraints that have to hold when optimising the target function. For the full description of *PyPSA*'s target function and constraints please refer to the *PyPSA* documentation in [8]. The constraints added to integrate the gas sector will be described in the following section.

2.2.4 Including the gas sector into *eTraGo*

The research project ego^n aims to extend *eTraGo* and couple it with the sectors gas, heat and mobility. This thesis is conducted within ego^n to extend the model with the German gas network. Following the *NEP 2035 C*, the focus of modelling sector coupling lies on the effects on the power grid, as shown in Figure 2.4. A simplified model for the gas sector is used, as shown in Figure 2.4:

- No calculation of the gas flow in the gas grid is executed.
- To account for local restrictions for the feed-in of hydrogen into the gas grid, local capacities are derived for the nodes of the model (see *I.1 Setting hydrogen feed-in restrictions*).
- To account for seasonal fluctuations in gas withdrawal which influence the size of possible hydrogen injections, seasonal feed-in constraints are set (see *II.2 Setting seasonal feed-in constraints*).
- The hydrogen demand that has to be covered is derived from an annual hydrogen production volume given in the *NEP 2035 C* (see *II.2 Setting a fixed annual hydrogen demand*).
- Hydrogen can be stored within the gas grid with a storage capacity of 15 $\mathrm{TWh_{th}}$ (see *Gas storage*).
- The hydrogen demand can be covered by the gas stored within the gas grid or by the hydrogen production of the electrolyzers (see II.3 *Covering hydrogen demand with production and storage*).
- It is assumed that the large methane import capacities, e.g. via pipelines, still exist in 2035. Therefore it is not explicitly analyzed whether the supply of conventional gas, e.g. by residential gas boilers or gas power plants, can be covered. Instead, only the hydrogen demand produced from electrolysis is modelled (see *II.3 Covering hydrogen demand with production and storage*).

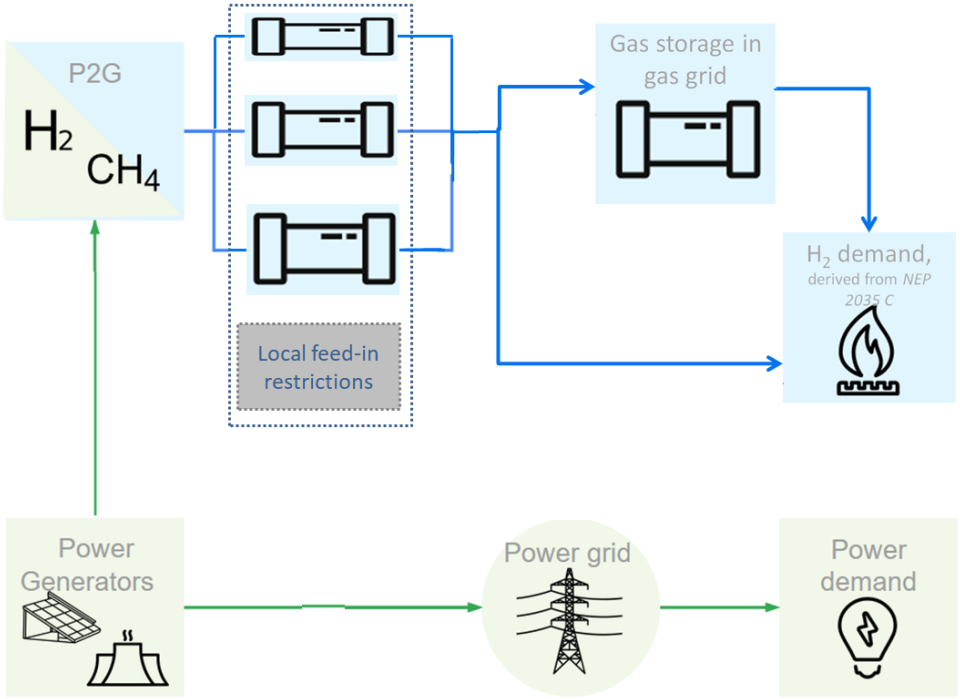

Figure 2.4: Sector coupling of the power and gas sector as integrated in *eTraGo* within this work.

Gas network

In the simplified sector coupling, only the impact of the electrolyzer load on the power system is implemented. This means that electrolyzers can be built by the model to transform power into hydrogen with an efficiency $\eta_{P\to \mathrm{H}_2}$. This hydrogen then is used to fuel a set hydrogen demand or can be stored within the gas network, which serves as a gas storage. Stored gas can fuel the set hydrogen demand at a later time. In the future, *eTraGo* is planned to include the gas grid with its volume flow and pipe capacities next to an option to re-transform produced gas into power using gas turbines.

For the sector coupling of the electrolyzers to the power model, the following values can be optimised:

I.1 **Setting hydrogen feed-in restrictions:** The maximum hydrogen feed-in capacity at each bus n is determined based on the locally available gas network capacities and the 5 vol% H_2 limit [38], [51]. A map of the maximum installable electrolyzer load is shown in Figure 2.5.

Electrolyzer capacity expansion: The electrolyzer capacity can be optimised within the minimum and maximum installable capacities (2.11). The minimum installable electrolyzer capacity is set to be zero (2.12). The maximum installable electrolyzer capacity at each bus is determined to be dependent on the maximum hydrogen feed-in (2.13).

$$\tilde{L}_{n,P\to \mathrm{H}_2} \leq L_{n,P\to \mathrm{H}_2} \leq \bar{L}_{n,P\to \mathrm{H}_2} \tag{2.11}$$

$$\tilde{L}_{n,P\to \mathrm{H}_2} = 0\,\mathrm{MW_{el}} \tag{2.12}$$

$$\bar{L}_{n,P\to \mathrm{H}_2} = \frac{\bar{N}_{n,\mathrm{H}_2}}{\eta_{P\to \mathrm{H}_2}} \tag{2.13}$$

where

$\tilde{L}_{n,P\to\mathrm{H_2}}$: Minimum installable electrolyzer capacity at bus n [$\mathrm{MW_{el}}$]
$L_{n,P\to\mathrm{H_2}}$: Electrolyzer capacity at bus n [$\mathrm{MW_{el}}$]
$\bar{L}_{n,P\to\mathrm{H_2}}$: Maximum installable electrolyzer capacity at bus n [$\mathrm{MW_{el}}$]
$\bar{N}_{n,\mathrm{H_2}}$: Maximum hydrogen feed-in capacity in gas network at bus n [$\mathrm{MW_{th}}$]
$\eta_{P\to\mathrm{H_2}}$: Electrolyzer efficiency [$\mathrm{MW_{th}/MW_{el}}$].

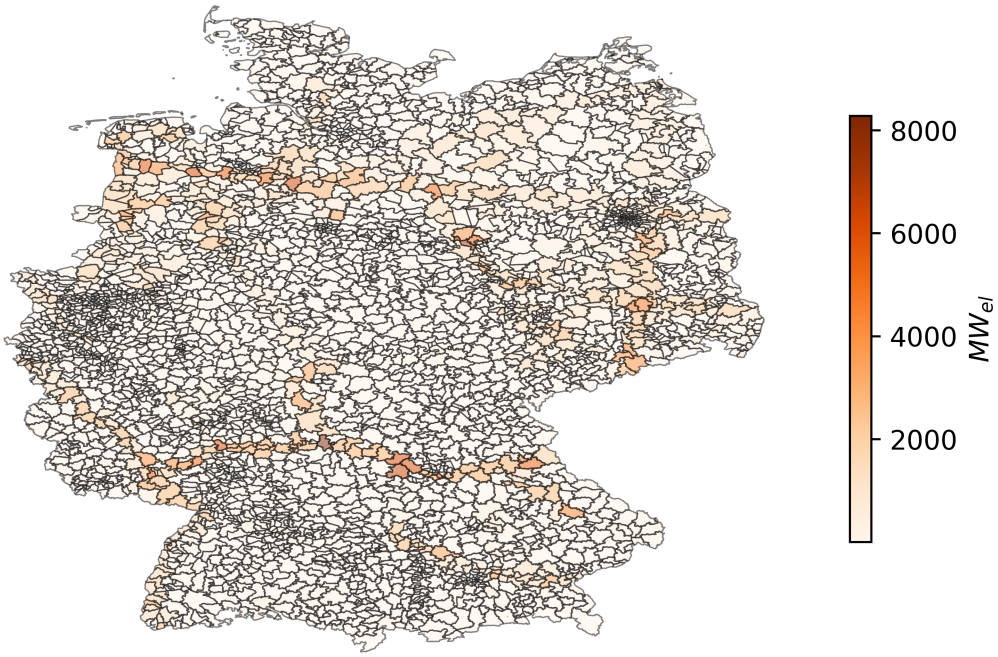

Figure 2.5: Spatial depiction of maximum installed electrolyzer capacity in *eTraGo* [3].

I.2 **Optimising the electrolyzer dispatch:** The electrolyzer dispatch $g_{P\to\mathrm{H_2},n,t}$ at bus n shall be minimised. As a constraint, it cannot exceed the installed electrolyzer capacity at the bus (2.14). The produced hydrogen $g_{\mathrm{H_2},n,t,th}$ cannot exceed the maximum feed-in capacity at bus n at time step t (2.15). Notice that hydrogen within the model is handled by its thermal energy content in $\mathrm{MW_{th}}$, and not by volume or mass.

$$g_{P\to\mathrm{H_2},\mathrm{n},\mathrm{t}} \leq L_{n,P\to\mathrm{H_2}} \tag{2.14}$$

$$g_{\mathrm{H_2},n,t,th} = g_{P\to\mathrm{H_2},n,t} \cdot \eta_{P\to\mathrm{H_2}} \leq \bar{N}_{n,t,\mathrm{H_2}} \tag{2.15}$$

where

$g_{P\to \mathrm{H}_2,n,t}$: Electrolyzer dispatch at bus n at time step t [$\mathrm{MW_{el}}$]
$L_{n,P\to \mathrm{H}_2}$: Installed electrolyzer capacity at bus n [$\mathrm{MW_{el}}$]
$g_{H_2,n,t,th}$: Hydrogen produced at bus n at time step t [$\mathrm{MW_{th}}$]
$\eta_{P\to \mathrm{H}_2}$: Electrolyzer efficiency [$\mathrm{MW_{th}/MW_{el}}$]
$\bar{N}_{n,t,\mathrm{H}_2}$: Maximum hydrogen feed-in capacity in gas network at bus n at time step t [$\mathrm{MW_{th}}$].

To model the sector coupling, additional constraints are necessary:

II.1 **Setting seasonal feed-in constraints:** The maximum electrolyzer load time series over all buses, $l_{P\to \mathrm{H}_2,\mathrm{t}}$ [$\mathrm{MW_{el}}$] is set as an input for the model, see Figure 2.6.

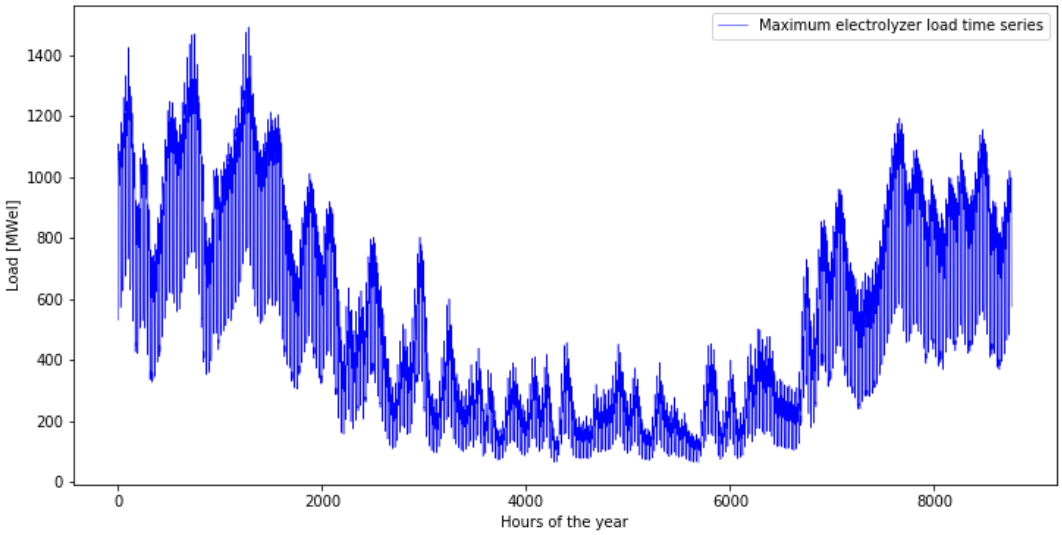

Figure 2.6: Maximum electrolyzer load time series $\bar{l}_{P\to \mathrm{H}_2,t}$ [$\mathrm{MW_{el}}$] over 8760 hours.

This time series shows that from late spring until early autumn (hour 2880 until 6550), when less gas is consumed for heating, significantly less hydrogen can be fed into the network than in winter. The electrolyzer load over all buses during time step t must not exceed this maximum time series:

$$\sum_{n} g_{P\to \mathrm{H}_2,n,t} =\leq \bar{l}_{P\to \mathrm{H}_2,t} \tag{2.16}$$

where

$g_{P\to \mathrm{H}_2,n,t}$: Electrolyzer dispatch at bus n at time step t [$\mathrm{MW_{el}}$]
$l_{P\to \mathrm{H}_2,t}$: Maximum electrolyzer load time series over all buses at time step t [$\mathrm{MW_{el}}$].

II.2 **Setting a fixed annual hydrogen demand:** In order to fulfill the *NEP 2035 C*'s grid-oriented expansion of electrolyzers of 3 $\mathrm{MW_{el}}$ at 1500 full load hours (FLH), the annual production of

hydrogen from electrolyzers must be:

$$\begin{aligned} E_{P\rightarrow \mathrm{H_2,th}} &= L_{P\rightarrow \mathrm{H_2},NEP} \cdot FLH_{P\rightarrow \mathrm{H_2,NEP}} \cdot \eta_{P\rightarrow \mathrm{H_2}} \\ &= 3 \cdot 10^3\, MW_{el} \cdot 1500\,\mathrm{h} \cdot 0.8 \frac{\mathrm{MW_{th}}}{\mathrm{MW_{el}}} = 3.6 \cdot 10^6\,\mathrm{MWh_{th}} \end{aligned} \tag{2.17}$$

where

$E_{P\rightarrow \mathrm{H_2},th}$: Annual hydrogen production [$\mathrm{MWh_{th}}$]
$L_{P\rightarrow \mathrm{H_2},NEP}$: Total electrolyzer built capacity in *NEP 2035 C* [$\mathrm{MW_{el}}$]
$FLH_{P\rightarrow \mathrm{H_2},NEP}$: Electrolyzer FLH in *NEP 2035 C* scenario [h]
$\eta_{P\rightarrow \mathrm{H_2}}$: Electrolyzer efficiency [$\mathrm{MW_{th}/MW_{el}}$].

From this annual production, the hydrogen load $d_{\mathrm{H_2},n,t,th}$ at bus n during time step t is derived and set using the relative grid-oriented electrolyzer load $d_{P\rightarrow \mathrm{H_2,n,t,pu}}$:

$$d_{\mathrm{H_2},n,t,th} = d_{P\rightarrow \mathrm{H_2},n,t,pu} \cdot \frac{E_{P\rightarrow \mathrm{H_2},th}}{8760\,\mathrm{h}} \tag{2.18}$$

where

$d_{\mathrm{H_2},n,t,th}$: Hydrogen load at bus n at time step t [$\mathrm{MW_{th}}$]
$d_{P\rightarrow \mathrm{H_2},n,t,pu}$: Per unit grid-oriented electrolyzer load at bus n at time step t [$-$]
$E_{P\rightarrow \mathrm{H_2},th}$: Annual hydrogen production [$\mathrm{MWh_{th}}$].

II.3 **Covering hydrogen demand with production and storage:** All $\mathrm{H_2}$ produced is fed into the bus $m_{\mathrm{H_2}}$, see Figure 2.7. Each bus n is connected with $m_{\mathrm{H_2}}$ via a link. Within *eTraGo*, each of these links depicts a buildable electrolyzer. The hydrogen fed into $m_{\mathrm{H_2}}$ at time step t is used to supply the hydrogen demand from (2.18). The difference between hydrogen demand and production is balanced by the hydrogen storage:

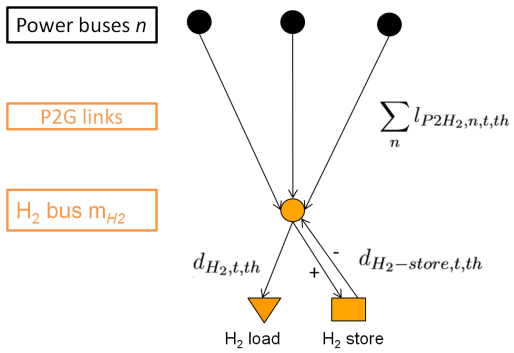

Figure 2.7: Schematic hydrogen flow in *eTraGo*.

$$\underbrace{g_{P\rightarrow \mathrm{H_2,t,th}}}_{\mathrm{H_2\ production\ at}\ t} = \sum_n l_{P\rightarrow \mathrm{H_2,n,t,th}} = \underbrace{d_{\mathrm{H_2,t,th}}}_{\mathrm{H_2\ demand\ at}\ t} + \underbrace{d_{H_2stor,t,th}}_{\mathrm{H_2\ storage/discharge\ at}\ t} \tag{2.19}$$

where

$g_{P\rightarrow \mathrm{H_2},t,th}$: Total hydrogen production by electrolysis at time step t [$\mathrm{MW_{th}}$]
$l_{P\rightarrow \mathrm{H_2},n,t,th}$: Hydrogen production by electrolysis at bus n at time step t [$\mathrm{MW_{th}}$]
$d_{\mathrm{H_2},t,th}$: Hydrogen demand at time step t [$\mathrm{MW_{th}}$]
$d_{\mathrm{H_2}stor,t,th}$: Hydrogen storage or discharge at time step t [$\mathrm{MW_{th}}$].

II.4 **Setting the electrolyzer CAPEX:** The CAPEX of electrolyzer expansion for the optimisation function 2.10 is set according to Brown et al. [5]:

$$c_{P\rightarrow \mathrm{H_2}} = 350000 €/\mathrm{MW_{el}} \tag{2.20}$$

where

$c_{P\rightarrow \mathrm{H_2}}$: Electrolyzer CAPEX [€/$\mathrm{MW_{el}}$].

Gas storage

Storage within the gas grid needs to be integrated for the sector-coupled version of *eTraGo*. Following the estimations in Section 2.1.2, a hydrogen storage capacity within the gas grid of 15 $\mathrm{TWh_{th}}$ is included in *eTraGo* (equation 2.21) [36]. Hydrogen production by the electrolyzers as set by *NEP 2035 C* is 3.6 $\mathrm{TWh_{th}}$, so the set storage capacity allows *eTraGo* to store the entire annual hydrogen production from electrolysis of the *NEP 2035 C*. The state of charge at the beginning and end of the year is set as equal so that the optimisation does not drain the storage for minimising costs (equation 2.22).

$$E_{\mathrm{H_2}stor} = 15 \cdot 10^6\,\mathrm{MWh_{th}} \tag{2.21}$$

2.22).

$$E_{\mathrm{H_2}stor,\mathrm{o}} = E_{\mathrm{H_2}stor,8_{759}} \tag{2.22}$$

where

$E_{\mathrm{H_2}stor}$: Hydrogen storage energy capacity [$\mathrm{MWh_{th}}$].
$E_{\mathrm{H_2}stor,t}$: Hydrogen storage SOC at time step t [$\mathrm{MWh_{th}}$].

2.2.5 *NEP 2035* scenario construction and modelling details

During the conduction of this thesis, *eTraGo* was only partially updated to the input values of the *NEP 2035 C*. While the values for the gas system, such as the P2G capacities, were set as described in 2.1.3, the values for the power-side of the model were still based on the Network Development Plan 2025, scenario B2-2035 (*NEP 2025 B*) [10], which had been implemented within *open_eGo* [2]. The installed generator capacities in neighbouring countries were adopted from the *European Network of Transmission System Operators for Electricity* (ENTSO-E)'s *Scenario Outlook and adequacy forecast 2014-2030* dataset ([52], cited in [2]). Table 2.26 presents key parameters of *NEP 2025 B*. See Tables B.1 and B.3 in Appendix B for installed capacities and CAPEX by generation type. Hereinafter, the abbreviation *NEP 2035* will be used for the merged scenarios of *NEP 2025 B* for the power side and *NEP 2035 C* for the gas side of the model.

The reference weather year used for the input time series of generation from RES, electric and electrolyzer load is set as 2011. This year also serves as the reference year for prices, so that the costs optimised by the model are given as 2011 real. To convert the costs into real prices of other years, an

inflation factor must be applied.

Table 2.26: Key parameters of *NEP 2025 B* [2].

	Germany	Germany and neighbouring countries
Share of renewable energy generators by installed capacity [%]	75	67
Annual net electricity consumption [TWh]	506	1569
Annual peak load [TWh]	87	253

2.2.6 Validating the power and gas model

While forecasts or predictions of future energy systems cannot be validated against data from the future, the reliability of the data obtained with *eTraGo* can be checked against other ESMs and regarding the validity of its input data sets. A validation against historic data will not be performed within this section, as it would go beyond the scope of this thesis.

The validity of the power-only version of *eTraGo* was discussed in the final report of *open_eGo* [2]. When validating the border-crossing link capacities of *eTraGo* against data from ENTSO-E and the *Agency for the Cooperation of Energy Regulators*, the results were determined to be sufficiently similar. However, the report stated that it was impossible to conclude with certainty that the model is valid. During the *open_eGo* project, no comparable grid model existed with which *eTraGo* could have been compared. A validity check hence would have only been possible with the real grid data of the network operators. In most cases, this is private data and therefore hard to obtain. In an exemplary comparison with data sets provided by one network operator, *open_eGo* data proved to have a high correspondence [2]. In a further step to validate *eTraGo*, Peters et al. [53] compared network characteristics and load flow results for a limited region in north-west Germany, testing the *eTraGo* power model against the proprietary model for the EHV and the HV grid of the regional distribution system operator (DSO) *Avacon Netz GmbH*. They found the deviations between the two models in the EHV-level to be acceptable and showed that *eTraGo* had better results than a simple aggregation of loads and generation in that region. However, there may occur significant deviations between the *etrago* model and the DSO model on the HV level, for example at the detection of line congestions [53].

A validation of the gas model of *eTraGo* was not yet conducted when writing this thesis, as the gas network was not fully integrated within the model. The plausibility of the storage capacity of hydrogen within the gas network that got implemented within *eTraGo* is discussed in Section 2.1.2. Regarding the quality of the used gas data set for setting the local hydrogen feed-in restrictions (Section 2.2.4), the data documentation of said data notes that it is limited in its quality of infrastructure data, as transmission system operators only provide limited details of the technical features of their gas grid publicly [38].

2.2.7 Achieving computability by reducing complexity

Many ESM application cases today are too complex to be solved directly by computational power and need aggregation methods, which can be seen in the number of recent publications about methods of complexity reduction [12]. Heuristics to reduce the complexity may only degrade a model's accuracy by an acceptable amount.

The model considered here, *eTraGo*, contains a spatial resolution of more than 3000 nodes in Germany and an hourly temporal resolution of 8760 hours per year next to many constraints. The compute time needed to solve the model exceeds practical limits. In an exemplary calculation, the current version of *eTraGo* was run without temporal complexity reduction and a low amount of spatial complexity reduction to 3000 nodes over only 24 hours, which resulted in a computing time of 72.66 minutes (see Appendix B for information about the server *torch* this calculation was carried out on).

Computing time in this thesis is defined as wall-clock time from the execution of the python command to start *eTraGo* until the end of the spatial disaggregation, see code in Appendix A.1.

Within a spatially and temporally resolved model, complexity can be reduced in these two areas. *eTraGo* already contains several options to reduce complexity:

- spatial clustering to n nodes using a k-means algorithm
- temporal complexity reduction applying hierarchical clustering while linking the states of charge following Kotzur et al.'s method [4] (see 2.6.3).

In a study on the power-only version of *eTraGo*, Raventos et al. found that the reduction in spatial complexity leads to a good reduction in computing time without losing too much accuracy [54]. When looking at a sector-coupled model however, the accurate depiction of the spatial distribution of flexibility options is relevant. Figure 2.8 depicts the spatial allocation of the electrolyzer expansion in three exemplary runs of the version of *eTraGo* described earlier (2.2). Spatial complexity reduction to fewer nodes leads to fewer line congestions (displayed as line loading in per unit values of line capacity), as many nodes were clustered together. This results in a more centralized expansion of electrolyzers. Spatial complexity reduction with the included k-means algorithm, therefore, can only be used to a limited extent for the sector-coupled system.

Since spatial complexity reduction reaches its limits with this sector-coupled model, this work is dedicated to applying and analysing temporal clustering on *eTraGo* to reach shorter computing times.

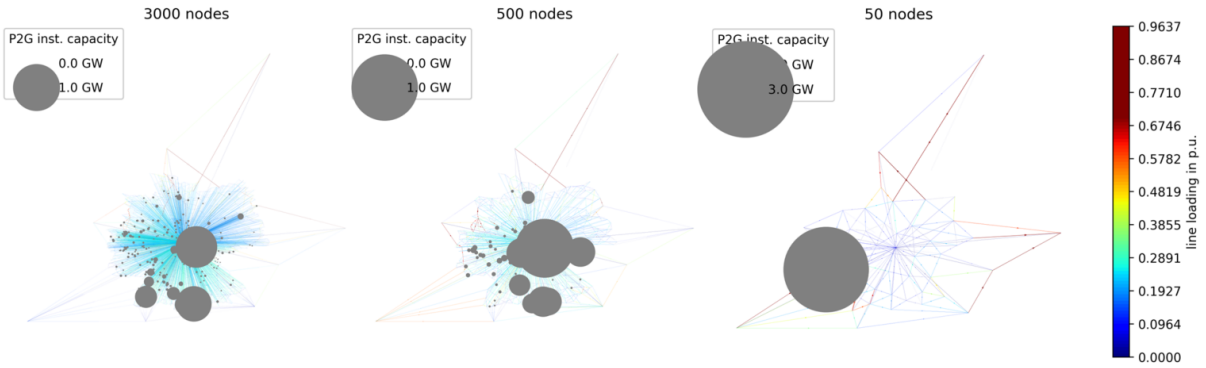

Figure 2.8: Spatial distribution of the electrolyzer expansion in Germany with a spatial complexity reduction to 3000 (left), 500 (centre) and 50 nodes (right), model run over first three hours of the year.

2.3 Overview on methods to reduce temporal complexity

In recent years, a variety of time series aggregation methods (TSAMs) to reduce temporal complexity have been proposed in scientific literature. Such methods can be distinguished in four categories, see Table 2.28. Time-based approaches include **downsampling**, which simply lowers the time-based reso-

Table 2.28: Major time series aggregation methods.

Type of merging	Variation in resolution	Typical periods
Time-based	**Snapshot Skipping / Downsampling** Only takes every n-th snapshot into account.	**Time Slices** Creates a set of representative slices e.g. for season or day of the week
Feature-based	**Segmentation** Clusters similar adjacent time steps together	**Cluster Typical Periods** Clusters similar periods together

lution by representing consecutive time steps by one longer time step. When downsampling e.g. to every 6^{th} hour, this approach skips five hours and thus loses a lot of accuracy of the input data set [12].
The other time-based approach, **time slices**, retains more accuracy, as it uses a certain set of time steps which represent important aspects of temporal variability and thereby preserves information on typical periods. A time slices approach would mean e.g. to use a time slice for a work- and a weekend-day for every month, resulting in 24 time slices total for one year [12].
Time series can also be aggregated concerning their features by applying clustering methods. These methods exclude redundant information while maintaining a lot of the original data set accuracy. They generally have a smaller impact on the model result than time-based methods [7]. These methods can be divided into segmentation and clustering methods. In **segmentation**, the data set is reduced by lowering the resolution. A day of 24 hours can be segmented to only 12 segments, which are created by merging the most similar hours, i.e. in regards to demand and supply, together. The proximity of time segments is determined using clustering algorithms, which will be described in Section 2.5.2.
The **cluster typical periods** method creates typical periods by merging similar data points into clusters. These clusters are then represented by one value only, hence reducing the input data while covering the variability and the statistical structure of the original time series rather accurately [12].
In the more recent literature, feature-based TSAMs are favoured due to their better representation of the original data [12]. Therefore, only feature-based clustering algorithms are considered for the complexity reduction in this thesis.

2.4 Clustering algorithms

Clustering algorithms have the goal to cluster individual data points together. When applied in TSAM, this clustering is used to decrease the amount of data points that need to be taken into consideration by a model while not losing too much of the original data sets information.
This thesis will focus on two of the most used clustering algorithms, partitional and hierarchical clustering. This section first gives a broad overview of these algorithms and then describes important clustering

metrics, as the choice of representative, distance metric, clustering method, usage of extreme periods and whether they link periods.

2.4.1 Partitional clustering: k-means and k-medoids

This clustering divides a data set $\{x_t\}_{t=1,\dots,T}$ into k pairwise disjoint clusters $\{T_t \subset \{1,\dots,T\}\}_{c=1,\dots,k}$, where a cluster comprises data points that are close to each other. A commonly used type of partitional clustering is the **k-means** clustering, in which all data points of one cluster are represented by the cluster's centroid $\{\mu_c\}_{c=1,\dots,k}$. The optimisational algorithm of this method minimizes the sum of the squared distances between each k-centroid μ_c and the data points of its cluster $\{x_t\}_{t\in T_t}$ [55]:

$$\min_{\{\mu_c\}} \sum_{c=1}^{k} \sum_{t\in T_t} \|x_t - \mu_c\|^2 \tag{2.23}$$

where

k : Set number of clusters $[-]$
c : Index for cluster $[-]$
t : Index for time step $[-]$
T_t : Amount of time steps t $[-]$
μ_c : k-centroid of cluster c []
x_t : Data point at t $[-]$.

See Figure 2.9 for the application of a k-means algorithm with three clusters on an exemplary data set. The borders between the clusters are displayed with Voronoi lines.

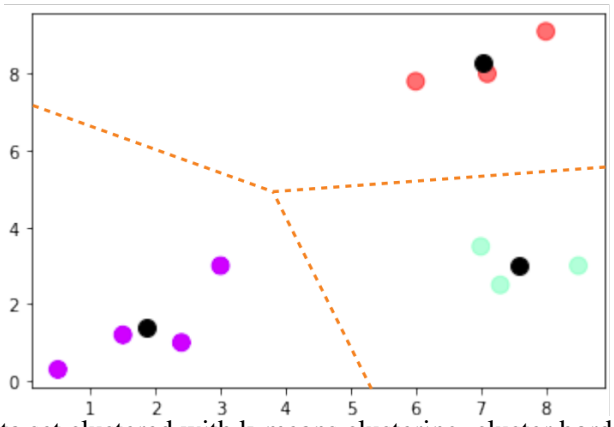

Figure 2.9: Exemplary data set clustered with k-means clustering, cluster borders depicted with Voronoi lines.

The steps of the algorithm are [56]:

1. Select an initial set of centroids μ_c.
2. Assign each datapoint x_t to its closest centroid μ_c and thereby create a set of clusters.
3. Compute the mean value of all data points within each cluster $\{x_t\}_{t\in T_t}$ and specify this as the new centroids μ_c.
4. If the difference between the previous and the new centroid values is smaller than a fixed tolerance finish, else return to step 2.

2.4.2 Hierarchical clustering

In hierarchical clustering, data points are clustered together one after another, which results in a dendrogram, a nested family of partitions, whose appearance is similar to a tree (see Figure 2.10).
In hierarchical clustering the following steps are generally taken:

1. Set each sample as its own cluster at the start.
2. Calculate the distance between all clusters, resulting in a proximity matrix.
3. Find the smallest non-zero value in the matrix and merge the two associated clusters into one. Set the mean value of the two clusters to be the value of the new cluster.
4. Repeat starting from Step 2 until all clusters are merged into one or until the number of clusters is smaller than a set value k [57].

Figure 2.10 depicts the application of hierarchical clustering on the same data set as previously used with k-means, shown by location (left) and by a dendogram (right) .

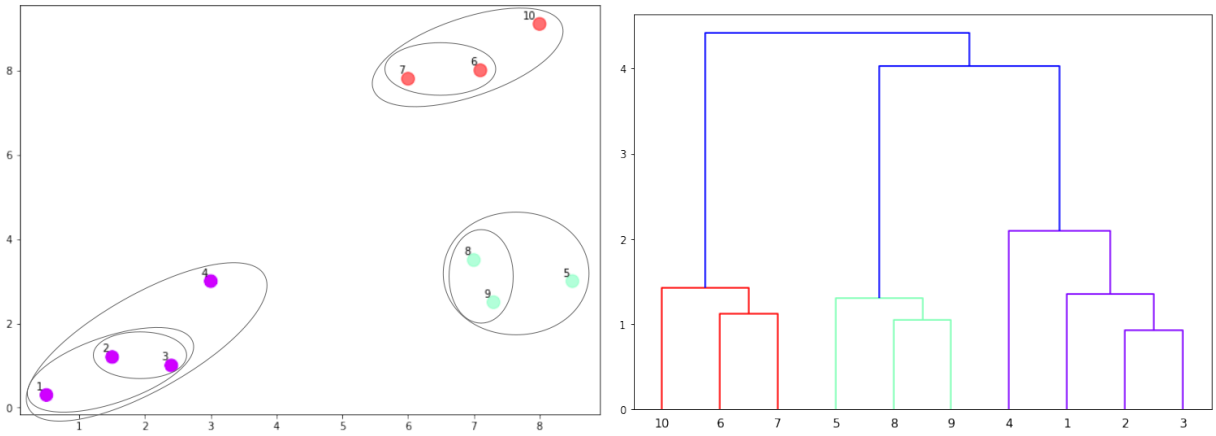

Figure 2.10: Exemplary data set clustered with hierarchical clustering, depicted in a diagram and as a dendogram.

2.4.3 Cluster metrics

This section describes common methods to determine cluster representatives and distance metrics for partitional clustering and hierarchical clustering.

Determine cluster representatives

Aggregation methods cluster multiple data points together which are then represented by a single value. The most common methods to decide on this representative value in time series aggregation are:

- the **centroid** method. It takes the mean value of the cluster as the representative, creating a fictional day. This results in the information of all elements being taken into account equally.
- the **medoid** representation. Instead of a fictional day, it uses the nearest neighbouring point to the centroid. It chooses the cluster element with the smallest Euclidean distance to the cluster's centroid. When a cluster's medoid instead of its centroid is chosen as its representative, this alteration of k-means is called **k-medoid** clustering.

Both the medoid and centroid method have a smoothing effect and cut off-peak periods. This leads to an underestimation of the system costs and models an energy system that cannot meet peak demands [58]. A method to counteract this issue is to include **extreme periods** into the methods described above. This allows integrating more variability within the set of clustered days. Adding extreme periods can be done by i.a. [58]:

- Appending: Adding some extreme periods to the previously calculated representatives.
- Adding cluster centres: Set the extreme periods as additional cluster centres and reassign data points that are closer to the new cluster centre than to their original cluster centre to the new clusters.
- Replace representative period: Set the extreme period, i.e. the data point in a cluster with the biggest distance to the centroid, as the representative of its cluster.

Distance metrics for partitional clustering

In the second step of the k-means clustering, the closest neighbour is located using a distance metric. The choice of the metric influences the shape of the clusters, as two points can be closer together according to one metric, but further apart according to another method. The paragraphs below describe two commonly used metrics to calculate the distance between two n-dimensional vectors $x = \{x_1, x_2, ..., x_n\}$ and $y = \{y_1, y_2, ..., y_n\}$ in clustering.

The **Manhattan** or city-block distance is a linear metric and described in equation 2.24. It describes the distance one would need to walk to get from x to y in a city made up of rectangular blocks [57]. The **Euclidean** distance, in contrast, is a non-linear metric and takes the square root of the square of distances (equation 2.25) [57]. It measures the distance in an Euclidean space. Hence, it weighs longer distances more than smaller ones.

$$\mathrm{d_M}(x, y) = \sum_{i=1}^{n} |x_i - y_i| \tag{2.24}$$

$$\mathrm{d_E}(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{2.25}$$

where:

$d_{\mathrm{M}}(x, y)$: Manhattan distance of two vectors x and y $[-]$
$d_{\mathrm{E}}(x, y)$: Euclidean distance of two vectors x and y $[-]$
x : n-dimensional vector $\{x_1, x_2, ..., x_n\}$ $[-]$
y : n-dimensional vector $\{y_1, y_2, ..., y_n\}$ $[-]$
x_i : Vector x at i $[-]$
y_i : Vector y at i $[-]$.

Singh et al. applied k-means with different distance metrics on a data set and found that the Manhattan distance lead to most distortion, the Euclidean distance to less distortion and an overall better result [59]. In applications of k-means and k-medoids clustering, the Euclidean is used more often than the Manhattan distance.

Distance metrics for hierarchical clustering

In the second step of the hierarchical clustering described above, a proximity matrix is compiled using the calculated distance between the clusters. These distances are calculated using a linkage method. The linkage method used in this work is Ward's linkage, which calculates the dissimilarity between all pairs of clusters i, j by using the Euclidean distance (from equation 2.25) [60]:

$$D_{Ward}(i,j) = d_E(\{i\},\{j\}) = ||\mathbf{x}_i - \mathbf{x}_j||^2 \tag{2.26}$$

where:

$D_{Ward}(i,j)$: Dissimilarity between pair of clusters i, j $[-]$
$d_{\mathrm{E}}(x,y)$: Euclidean distance of two vectors x and y $[-]$
i, j : Pair of clusters $[-]$
x_i : Vector x at i $[-]$
x_j : Vector x at j $[-]$.

Other linkage methods include the Single-Linkage, Average-Linkage or Complete-Linkage. They will not be further described here as they will not be used in this thesis.

2.5 Comparison of time series aggregation methods for energy system models with seasonal storage

The deployment of RES in an energy system makes the usage of seasonal storage relevant, as described in Section 2.1. Models designed to be decision-making aids for systems with these kinds of storage, like a system where the gas network can be used as seasonal storage, therefore need to model and correctly predict the demand and expansion of such storage options. As a consequence, clustering approaches for these ESMs need to make a good prediction of storage possible.
In an application of clustering typical days on a model without seasonal storage options, Green et al. found 10 typical days to be the best trade-off between computing time and model accuracy [61]. However, these methods do not allow the exchange of energy between periods. Table 2.34 shows a comparison of recent papers where one or more TSAMs were applied on a model which included some kind of storage option. Most of these publications were found within the review paper of Hoffmann et al [12]. The storage options included in the models used in these papers included battery, thermal, (pumped) hydro and hydrogen storage. For hydrogen storage, caverns [54], pressure vessels [4], or the gas grid [62] were included in the model setup.
To depict storage behaviour, most methods in Table 2.34 use a way to connect periods with the TSAM. This was either done by segmentation or using multiple time grids, as further explained in the following sections.

Table 2.34: Comparison of applied time series aggregation methods in recent scientific publications.

Year	Author	Applied ESM	Clustering Method	Representative	Extreme Periods	Connecting Periods	Applicable?
2020	Raventos and Bartels [54]	Model with battery and hydrogen storage	Hierarchical	Medoid	No	Yes, clustered adjacent periods (days)	Yes
2019	Baumgärtner et al. [62]	Model with battery and hydrogen storage	k-means	Centroid	No	By multiple time grids	Not stated
2018	Kotzur et al. [4]	Model with thermal, battery and hydrogen storage	k-medoids	Medoid	No	By multiple time grids	Yes
2018	Pineda et al. [63]	Model with intra- and interday storage	Hierarchical	Medoid	No	By clustering adjacent periods (hours within a day)	Yes
2017	Pfenninger [7]	Model with pumped hydro and battery storage	Hierarchical, k-means	Medoid, centroid	Yes	By clustering adjacent periods	Yes
2016	Nahmmacher et al. [56]	Model with intraday storage options	Hierarchical	Medoid, centroid	No	No	Maybe, if representative weeks instead of representative days are used

2.5.1 Segmentation / Clustering adjacent periods

Pineda et al. developed a method to cluster only neighbouring periods together while applying hierarchical clustering [63]. Within the method, the authors only allowed adjacent periods within a day to be clustered together, calling these clusters segments. The amount of segments per day hence could range from 1 to 24. See Figure 2.11 for an exemplary application of this segmentation method with 6 or 12 segments per day on the solar input time series of *eTraGo*.

Raventos et al. modified Pineda et al.'s approach by clustering adjacent days instead of hours together [54]. Within this method, the number of segments could variate between 1 and 365, repeating a daily pattern for adjacent days. They applied this method, which they called chronological, in comparison with a non-chronological hierarchical TSAM and a spatial k-means clustering method on a 152 node power transmission grid in Northern Germany with high generation from wind turbines. Within that work, the spatial clustering was found to be much more efficient in reducing the computing time than the TSAMs. In comparing the chronological and non-chronological hierarchical time clustering, they concluded the chronological to perform better for models which include large storage planning.

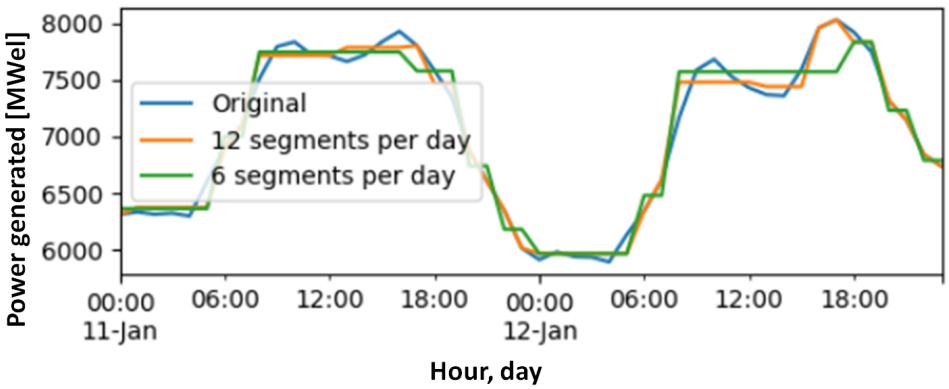

Figure 2.11: Exemplary segmentation of the *eTraGo* solar generation input time series, depicting the curves of the original data next to segmented data with 6 and 12 segments per day.

2.5.2 Linking typical periods with multiple time grids

Kotzur et al. applied both k-means and hierarchical clustering to aggregate typical days in an island system with thermal, battery and hydrogen storage and found that the typical period aggregation generally performed poorly in systems with a high share of RES and overestimated the system costs [58]. This happened as the model did not take into account the sequence of typical periods and therefore could not properly depict long term storage. The authors, therefore, developed a method to split the time layer up in two, where one depicts the intra-day state-of-charge (SOC), the other the inter-day SOC of the storage (see Figure 2.12). This way, they were able to use a typical days clustering approach, while still being able to account for the storage expansion needed [4].

Baumgärtner et al. applied Kotzur et al.'s method on their multi-node model with battery and hydrogen storage and found that they thus obtained feasible solutions for these storage options [62].

Pfenninger tested 42 different TSAMs on a model with pumped hydro and battery storage and found that a method clustering to typical days while linking storage states resulted in less computing tractability,

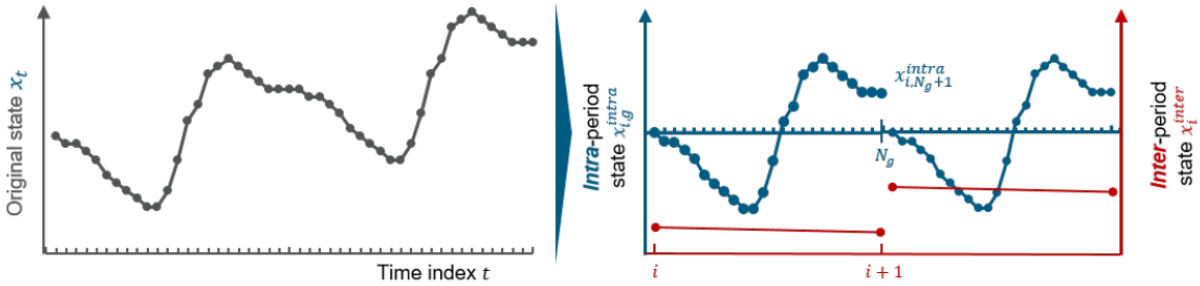

Figure 2.12: Scheme of separating the original state of charge (SOC) into two SOC (intra- and inter-period SOC) on two different time layers [4].

but allowed for accurate enough results at a lower time resolution [64].
Nahmmacher et al. applied clustering to representative days without a type of linkage, on a system with few intraday storage options [56]. The authors did not see this method as applicable on systems with larger storage in its current setup, but considered the method to work better for storage in case representative weeks were used.

2.5.3 Choice of clustering metrics

Regarding the decisions on clustering metrics, the publications compared decided on different paths. Nahmmacher et al. tested both k-medoids and k-means and found that their clustered data set resulted in a load duration curve closer to the original one when they used medoids instead of centroids as representatives [56].
In their comparison paper, Buchholz et al. found that taking the centroid or the medoid of a cluster as its representative often leads to underestimations of the production and the need for storage capacities, resulting in an underestimation of total system costs [65]. This could be remedied by including extreme periods to the clustered data. Pfenninger did so in his *Min/max solar and wind day* approach [7]. He extracted up to 4 days with minimum and maximum generation from intermittent RES from their data set, clustered the rest of the days into 10 clusters with a k-means algorithm and then merged the 10 chosen representative days with the extracted extreme solar and wind days. This method was found to have a lower error in the levelized costs of electricity than other methods applied.
Within all papers included in this comparison, the authors used Euclidean as their distance metric.

2.6 Applying complexity reduction methods

Within the literature review above, two main pathways emerged to allow modelling of longer-term storage when applying a TSAM, using a segmentation method or using clustering to few typical days and linking the periods with a second time layer. Each one variant of these methods will be tested and analysed in *eTraGo*. While a method linking periods following Kotzur et al. [58] already is implemented in the model, a segmentation method needs to be newly integrated. The following sections describe how the reference case to compare the results is set, and how the methods tested are applied in *eTraGo*. The amended model files used to run the model with the TSAMs described are provided on the CD/ the cloud folder of this work.

2.6.1 Setting the reference case

A reference case is set to evaluate the results of the complexity reduction methods with. This case has to be modelled for an entire year to depict seasonal patterns of the sector-coupled model. Since the server capacity for the optimisation is limited, the spatial resolution needs to be severely restricted to run the model repeatedly over 8760 hours during the time limits of this thesis. Therefore the reference case is clustered to 50 nodes only, applying the build-in k-means clustering method. To compensate for the fluctuations in computing time, 3 runs of the reference case were carried out and the mean values of the results were used for further analyzes.
As the scenario, *NEP 2035* is applied. The run outputs show the structure and costs of the German energy system in 2035.

Figure 2.13 depicts the shares of the various energy generators [MWh/MWh] in the average reference case[5]. When comparing these shares with the ones displayed in the final report of *open_eGo*, which applied the same scenario in the power-only model, the generator shares are mostly similar. Deviating numbers occur for production from wind energy, where the reference case is 10 percentage points short the generation in the final report, and run-of-river, with a 5 percentage points higher generation [2]. The high shares of nuclear power stem from imports from neighbouring countries which are included in the model.

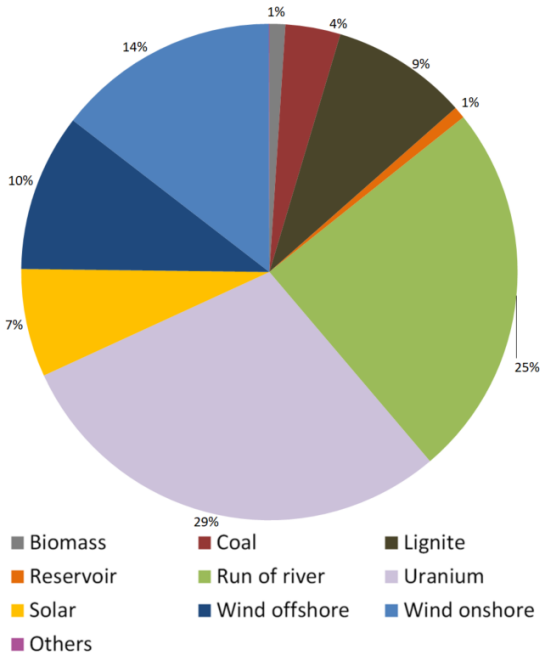

Figure 2.13: Shares of energy generators in reference case for NEP 2035 C Scenario.

The curtailment of intermittent RES in the reference case is shown in Table 2.37. Likely due to the spatial resolution to only 50 nodes, only up to 0.03% of the yearly power generation per generator type is curtailed, summing up to 0.106 $\mathrm{TWh_{el}}$. This value is noticeably lower than the 6-12 $\mathrm{TWh_{el}}$ of curtailed power estimated by other sources for the years 2030 to 2050 [21], [23]. As this value in the reference case is likely to be largely underestimating the curtailment in 2035, a comparison of curtailed energy between the model runs with applied TSAMs and the reference case is not seen to have an added value.

2.6.2 Applying segmentation

As done by [63] and [54] previously, a segmentation method that applies hierarchical clustering for adjacent periods of time is implemented in *eTraGo*. In difference to the previously listed approaches which clustered adjacent days or adjacent hours within a day, this work clusters adjacent hours within the whole year together, with a possible number of clusters between 1 and 8760. This is done to create a better depiction of seasonal patterns, as it allows for hours of different days to be clustered together.

[5] Shares from gas and waste power plants were too small to be shown in this graph.

Table 2.36: Curtailment of energy from intermittent RES in reference case, scenario *NEP 2035*.

Generator type	Total power curtailed [$\mathrm{MWh_{el}}$]	Share of curtailed power from overall production of generator type [%]
Solar	$6.121 \cdot 10^3$	0.01%
Onshore wind	$6.867 \cdot 10^4$	0.03%
Offshore wind	$3.161 \cdot 10^4$	0.02%
Sum	$1.064 \cdot 10^5$	

Following the most used metric methods, the distance is computed with the Euclidean metric, and Ward's linkage is applied. The segmentation method can then be described in the style of [63]:

1. The initial number of clusters n is set to the total number of hours N.
2. The centroid $\bar{x}_i$ of each cluster i is calculated with

$$\overline{\mathbf{x}}_i = \frac{1}{|i|} \sum_{i \in i} \mathbf{x}_i. \tag{2.27}$$

where

$\overline{\mathbf{x}}_i$: Centroid of cluster i $[-]$
i : Index for cluster $[-]$.

3. The dissimilarity between all adjacent pairs of clusters i, j is computed following Ward's method (eq. 2.26).
4. In the dissimilarity matrix, the two closest adjacent clusters (i', j') are found with $(i', j') \in \operatorname{argmin} D_{Ward}(i, j)$ s.t. $j \in \mathcal{A}(i)$, $\mathcal{A}(i)$ being the set of clusters adjacent to cluster i. They are merged.
5. Update $n \longleftarrow n - 1$
6. If $n > N'$ go to step 2. Otherwise, continue with step 7.
7. The clusters' centroids $\overline{x_i}$ are determined as the set of representative hours.
8. The value of the weight parameter ω_t is set to equal the number of hours belonging to each cluster.

Within *eTraGo*, segmentation is included using the segmentation method of the Python package *tsam* which is based on the agglomerative clustering method of the package *scikit learn* [66]. The altered code in the *eTraGo* file *snapshot.py* to include this method is shown in Appendix A.2.1.

The time series to be clustered are the input time series of *eTraGo*. In the power-only model, these included time series for the potential generation from onshore and offshore wind, from solar and for the electricity load. When adding the gas sector, the electrolyzer load time series is added to this input time series. See Figure 2.14 for the unclustered input time series. As the patterns of on- and offshore wind have a quite random pattern, for the analysis of the pattern depiction by the cluster methods only one of the wind patterns (onshore) will be included.

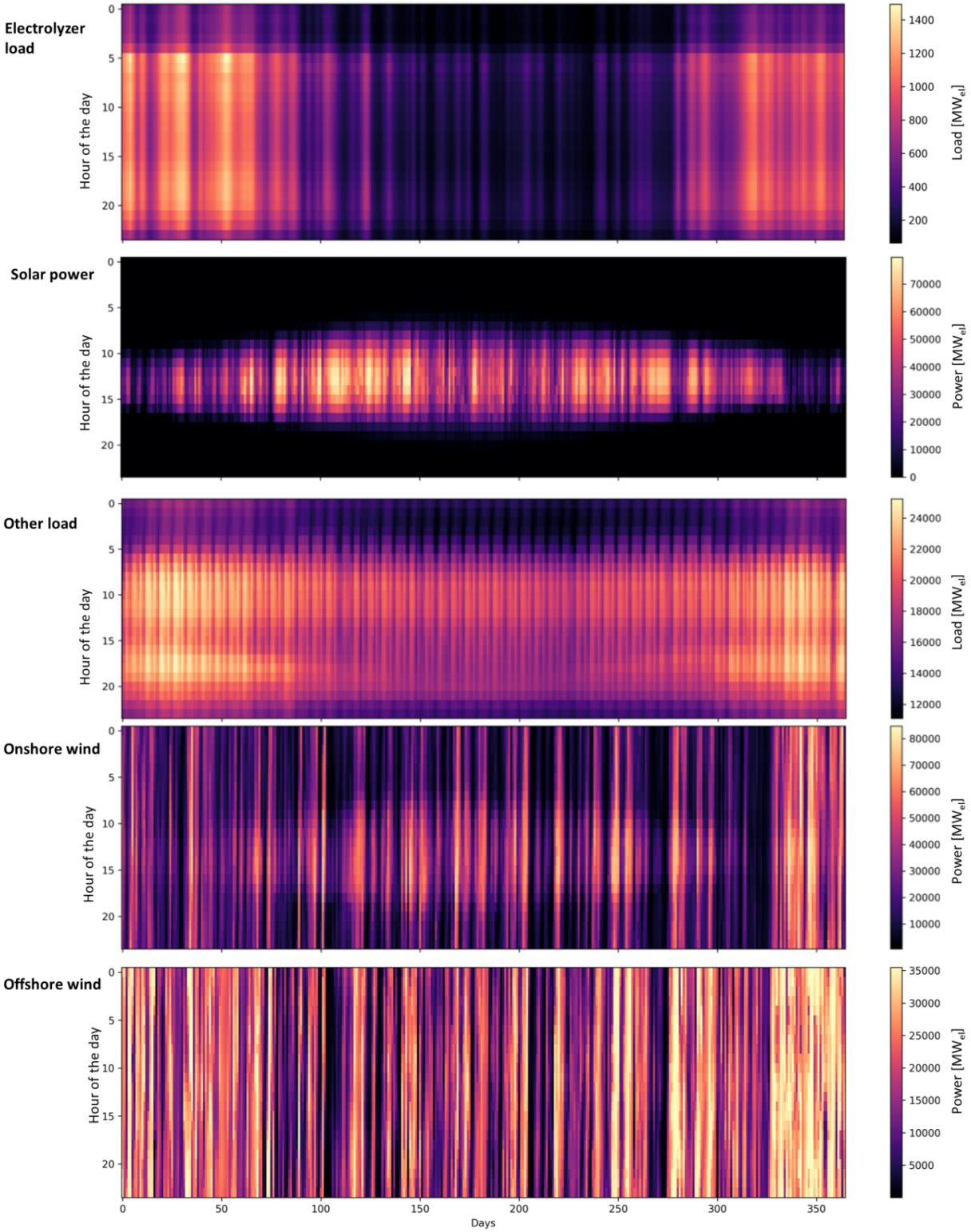

Figure 2.14: Original input time series of electrolyzer load, solar power, electric load, onshore wind, offshore wind (from top to bottom).

When segmentation is applied, the model only optimises for the number of segments instead of for the total number of hours, thereby reducing the number of optimisation variables. As a result, the input time series and also the output time series only contain values for the cluster centres, but not for all hours of the year.

2.6.3 Applying linkage of typical periods

The linkage of typical periods was included in *eTraGo* in a previous master thesis within the project *open_eGo* [67]. This method is applied by setting the condition $'\text{storage\_constraints}' : '\text{soc\_constraints}'$ in the execution file *appl.py*.
For this method, the variable SOC^{intra} is added to the model for each storage at every time step. The value of SOC^{intra} is calculated following equation 2.28. Its limits are set in equation 2.29, upwards by the nominal power of the storage multiplied with the maximum amount of hours that this power can be delivered and downwards by zero.

$$SOC^{intra}_{s,t+1} = SOC^{intra}_{s,t} \left(1 - \eta^{self}_s\right) + \eta^{char}_s P^{char}_{s,t} - \frac{P^{dis}_{s,t}}{\eta^{dis}_s} \tag{2.28}$$

$$0 \leq SOC^{intra}_{s,t} \leq t_{\max} P^{nom}_s \tag{2.29}$$

where

$SOC^{intra}_{s,t}/SOC^{intra}_{s,t+1}$: State of charge of storage s at time step t/$(t+1)$ [MWh]
η^{self}_s : Self discharge efficiency of storage s $[-]$
η^{char}_s : Efficiency of storage charging s $[-]$
η^{dis}_s : Efficiency of storage discharging s $[-]$
$P^{char}_{s,t}$: Charging capacity of storage s at time step t [MW]
$P^{dis}_{s,t}$: Discharging capacity of storage s at time step t [MW]
$t_{\max,s}$: Maximum amount of hours storage s can be discharged at full capacity [h]
P^{nom}_s : Nominal power of storage s [MW].

Büttner found that when she applied this method to link typical periods and clustered to 30 representative days, the deviations in grid expansion lay above 10%, which she considered too high to reliably determine the need for network expansion [67].
Within this work, the linkage of typical periods is applied upon a normal hierarchical clustering algorithm which clusters to a set number of typical days. With this TSAM, the model only optimises for the typical days which are the cluster centres of the algorithm, thereby reducing the number of optimisation variables.

2.6.4 Applying the temporal complexity reduction on the gas sector solely and on the coupled sectors

eTraGo has a set of time series as its input values. In the power-only model, these included time series for the generation from onshore and offshore wind, from solar and for the electric load. When adding the gas sector, the electrolyzer load time series was added to this input time series.

To be able to estimate to what extent the coupling of the sectors affects the quality of the clustering and the model results, the two selected methods are applied in two different ways:

1. in a first step only on the gas sector, which means on the time series of the electrolyzer load
2. in a second step on the coupled energy and gas sector, which means on all input time series shown in Figure 2.14.

Appendix A.2.2 shows how a TSAM can be applied on the gas sector by itself or on both sectors together in the *eTraGo* code.

2.6.5 Traceability of the results

This research focuses on the computing time reduction of the LOPF of *eTraGo*. The computing time is highly dependent on the soft- and hardware that it is run on, so that different systems will lead to different total computing times when optimising the same problems, while the relative change in computing time achievable by the methods compared in this thesis should be similar on different machines. For transparency of the results presented in Section 3, the technical framework conditions for the optimisation runs that are compared in this work are briefly listed in Appendix B.
This Appendix also lists how to install and run *eTraGo* to reproduce the results and which specific version of the model was used to obtain them. To make the results obtained traceable, the exact run files of the model which include the clustering settings as well as the complete output data of the model runs conducted for this thesis are provided on the CD / the cloud folder of this work.

2.7 Result indicators

This section describes the indicators used to compare and evaluate the model's in- and output values with different applied TSAMs.

2.7.1 Comparing the deviation of the input time series

The accuracy of the performed clustering is measured against the reference input data. Following Nahmmacher et al. [56], the normalized root mean square error (nRMSE) is used to compare the input and output duration curves. The duration curve is created by sorting the values of the time series in descending order. The nRMSE is calculated using the error between the respective t-largest value of the original and the clustered time series and dividing this value by the mean of the original time series (equation 2.30). The normalization is performed to compare the values of the different input time series.

$$\mathrm{nRMSE}\left(x_t\right) = \frac{\sqrt{\frac{\sum_{t=1}^{T_t}\left(x_{ref,t} - x_{app,t}\right)^2}{T_t}}}{\overline{|x_{ref}|}} \cdot 100\% \tag{2.30}$$

where

$nRMSE$: Normalized root mean square error [%]
$x_{ref,t}$: Reference value at time step t [−]
$x_{app,t}$: Approximated value at time step t [−]
$\overline{|x_{ref}|}$: Average of absolute reference values of all time steps t [−]
T_t : Number of time steps t [−].

2.7.2 Comparing the model outputs

Like previously done by Pineda et al. [63], the total system costs of the different model runs were compared using the average objective error (AOE):

$$AOE = \left| \frac{x_{\text{ref}} - x_{\text{app}}}{x_{\text{ref}}} \right| \cdot 100\% \tag{2.31}$$

where

AOE : Average objective error [%]
x_{ref} : Reference value [−]
x_{app} : Approximated value [−].

When just the gas input time series is used for clustering, the output time series only includes this time series. To compare the nRMSE of these approaches with the ones using all input time series, the hours of the year set as cluster centres and their weighing is determined. With this information, the hours of the intermittent RES and electric load time series which were fed into *eTraGo* when applying the clustering only over the gas sector can be traced back. This information can then be used to calculate the nRMSE for all *eTraGo* input time series.
To compare the reduction in computing time, a similar measure to the AOE is defined:

$$ATR = \frac{\overline{t_{\text{ref}}} - \overline{t_{\text{TSAM}}}}{\overline{t_{\text{ref}}}} \cdot 100\% \tag{2.32}$$

where

ATR : Average time reduction [%]
$\overline{t_{ref}}$: Average computing time over reference case runs [−]
$\overline{t_{TSAM}}$: Average computing time over runs of one TSAM applied [−].

Computing time thoughout this thesis is defined as wall-clock time from the execution of the python command to start *eTraGo* until the end of the spatial disaggregation, see code in Appendix A.1.

Chapter 3

Results and analysis

In Section 2.5, two applicable time series aggregation methods (TSAMs) to reduce temporal complexity were found when the SOC of the storage between time periods must be taken into account: segmentation, which clusters only adjacent time periods, or linkage, which has periods linked together with a second time grid. For each of these methods, one variant is applied in *eTraGo* to address the research question *(3) How can these complexity reduction methods be used across sectors for both power and gas? eTraGo* is run using an hourly granularity, so that the input time series to be clustered have periods with the length of one hour.

This section analyzes the results of this TSAM application. The analysis provides answers to:

- How much do the clustered input time series differ from the reference case's ones?
- How much deviate the runs with the TSAMs from the reference output data sets?
- What are the gains in computing time?

The four different TSAMs used in this analysis are abbreviated as:

- **Seg0**: The segmentation method as described in 2.6.2 is applied on the input time series of the electrolyzer load only. The clustered periods are then used as the input for a run of *eTraGo*, whereby the cluster centres, which were created by considering the electrolyzer load time series only, are applied to all input time series (intermittent RES, electric load, electrolyzer load).
- **Seg1**: The segmentation method of *Seg0* is applied on the input time series of all intermittent RES (solar, onshore wind, offshore wind), the electric load input time series and the input time series of the electrolyzer load.
- **Lin0**: The method *linkage of typical periods* as described in 2.6.3 is applied on the input time series of the electrolyzer load only. Like in *Seg0*, the typical days created by considering the electrolyzer load time series only in the model run are applied to all input time series (intermittent RES, electric load, electrolyzer load).
- **Lin1**: The linkage method of *Lin0* is applied on the input time series of all intermittent RES (solar, onshore wind, offshore wind), the electric load and the input time series of the electrolyzer load.
- The reference case run is abbreviated as **Ref**.

Test computations are performed for these four TSAM approaches as well as for the reference case (see summary in Table 3.1). As the segmentation method is newly introduced to the model, runs with a broad selection of segments per year (seg/y) are carried out to find a suitable amount of segments. The linkage method had already been used in the model before. In previous studies, hierarchical clustering of 50

typical days lead to AOEs $\leq 5\%$ of the annual system costs [54], so that fewer runs are launched within this method, spreading out the clustering to five runs with few typical days per year (tday/y) (10 to 50) and two runs with many tday/y (100 and 200). The goal of these tests is to track the computing time used and capture the clustered input time series as well as the model outputs for later analysis. Apart from the temporal complexity reduction, all other inputs to the runs are identical (see Appendix B for how to reproduce the results), with a spatial clustering to 50 nodes using a k-means algorithm.

Table 3.1: Summary of time series aggregation method (TSAM) configurations used in tests.

TSAM	Time span / periods	Aggregation	Aggregation applied on time series
Ref	8760 hours / 1 year	—	—
Seg0	100, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000 and 8000 segm/y	Segmentation with hierarchical clustering	electrolyzer load
Seg1	100, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000 and 8000 segm/y	Segmentation with hierarchical clustering	electrolyzer load, solar, onshore wind, offshore wind, electric load
Lin0	5, 10, 20, 30, 40, 50, 100 and 200 tday/y	Hierarchical clustering with period linkage	electrolyzer load
Lin1	5, 10, 20, 30, 40, 50, 100 and 200 tday/y	Hierarchical clustering with period linkage	electrolyzer load, solar, onshore wind, offshore wind, electric load

Computing times on the server used for this thesis depended not only on the complexity of the model but also on the load caused by the concurrent execution of other jobs on the same server. To reduce the effects of these jobs on the numbers presented here, each computation was run three times and their computing times averaged. Results of the runs, including computing times, annual system costs (as defined in Section 2.2.3) and built electrolyzer capacity are provided in Appendix C.1. Since the individual executions of the same job yield almost identical simulation results (see Appendix C.1 for details), the analysis in this chapter is mostly based on the values of the first run (marked with the index 1, e.g. $Seg1_1$), unless stated otherwise (average values are marked e.g. as $\overline{Seg0}$). Within this analysis, an acceptable deviation (in nRMSE) of the input data considered is set at 5%, while the error allowed for the output data is allowed slightly higher at 10% (in AOE). The results are based on the *NEP 2035* scenario within *eTraGo*, so all results describe the energy model in 2035 based on the assumptions listed in section 2.2.5.

To have targetable error limits, the acceptable errors for nRMSE of the input data sets are set at 5%, for the AOE of the output data at 10%.

3.1 Reference scenario results

The average over the three runs of the reference case resulted in:

- annual system costs of $8.574 \cdot 10^9$ €
- an expansion of the electrolyzer capacity to $3471\,\mathrm{MW_{el}}$
- while requiring a computing time of 286 minutes

The hydrogen load, production, and storage charge are shown in Figure 3.1. Note that the curves are smoothed, using a moving average over 100 hours, as otherwise, the flicker of the curves would make reading the diagram difficult. It can be seen that the hydrogen production curve closely follows the simulated hydrogen load. The hydrogen charge and discharge to the gas grid balances the difference between the load and production. In summer, hydrogen is stored into the grid, while in winter the hydrogen discharge often exceeds the charge. The total hydrogen storage SOC in 2035 begins (hour 1) and ends (hour 8760) at 65.03 TWh, which is set as a constraint to avoid draining the storage (not shown in Figure 3.1).

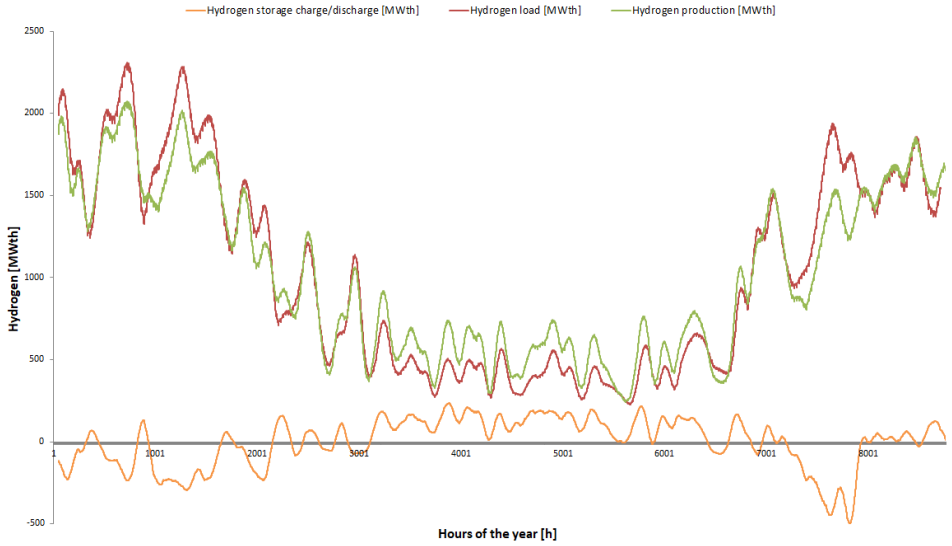

Figure 3.1: Hydrogen load, production from electrolyzers and storage (charging: positive values, discharging: negative values) as a moving average over 100 hours in 2035 [$\mathrm{MW_{th}}$] of Ref_1.

Figure 3.2 shows the stacked power generation from intermittent RES in relation to the hydrogen storage charge/discharge to/from the gas grid in the scenario year 2035, for the exemplary months January and June. As shown before in Figure 3.1, hydrogen is mostly discharged from the gas grid in winter (see January graph) and charged into it in summer (see July graph). A slight correlation can be seen that lower generation from RES results in lower storage charge, and higher production results in higher charge.
Next to the hydrogen storage output time series, the spatial expansion of electrolyzers could be an interesting measure for the accuracy of the different TSAMs. However, the spatial expansion in the three executed runs for the reference case already varied greatly (see Figure 3.3). While electrolyzers are only built at two nodes, no clear local preference for their location can be seen in the different runs. This is probably caused by the model being spatially clustered to 50 nodes. The distribution of the electrolyzers in the runs conducted in this thesis hence is not a meaningful measure for the compliance of the runs with TSAMs to the reference case outputs. This can be examined with a selection of spatial expansion of electrolyzers shown in Appendix C.3. Other theses in the project eGo^n will investigate the spatial expansion of electrolyzers with different model settings at a higher spatial resolution more closely [3].

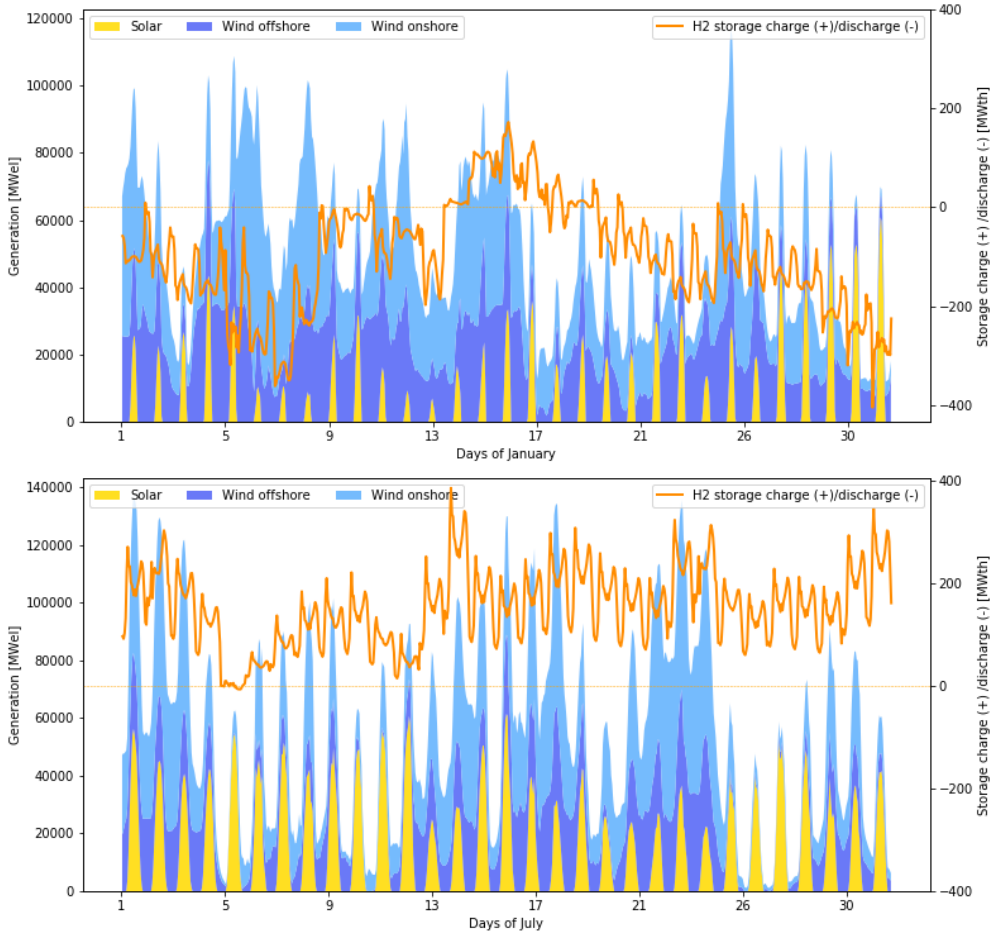

Figure 3.2: Stacked intermittent RES generator dispatch and hydrogen storage charge/discharge curve in two exemplary months (January, July 2035) of Ref_1.

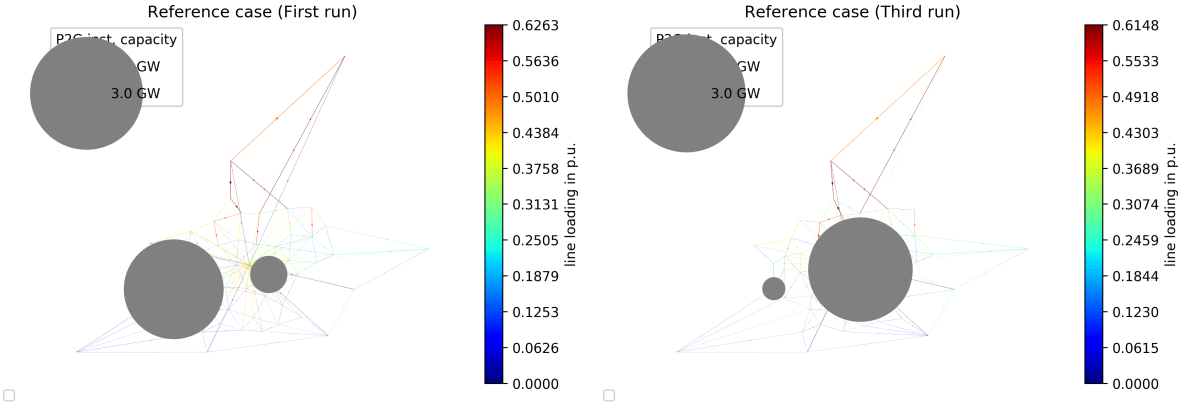

Figure 3.3: Spatial expansion of electrolyzers [GW_{el}] and line loading [per unit of line capacity] in Ref_1 and Ref_3 in 2035.

3.2 Accuracy of the model inputs with applied temporal complexity reduction

In the first step of this analysis, the focus lies on how well the applied TSAMs represent the input data sets of the reference case. This is done by graphical comparison and by calculating the duration curve errors of the clustered input data sets in relation to the input data sets of the reference case. The distribution of segment lengths and typical days are interpreted to show seasonal patterns in the clustering algorithms.

3.2.1 Changes in input patterns

Figure 3.5 depicts the input patterns of the reference case, *Seg0* and *Seg1* clustered to 1000 segments; and *Lin0* and *Lin1* clustered to 10 typical days. As should be expected, the representations of the electrolyzer time series in *Seg0* and *Lin0*, the TSAMs only applied on said time series, are more accurate. It is interesting to look at the behavior of the TSAMs applied to all input time series, *Seg1* and *Lin1*. While both these clustering algorithms lead to wind, solar and electric load patterns that resemble the reference case time series, *Lin1* overpredicts the number of days with much solar power production and *Seg1* misses more of the peak solar hours. In the segmentation methods, a smoothing of the patterns can be seen, as would be expected when clustering adjacent hours without applying an extreme period method. The linkage methods result in more repetitive input time series, while e.g. artefacts can be seen in the *Seg1* onshore wind graph due to the strong aggregation.
The difference between the approximated input time series and the reference case time series can further be shown with input duration curves. These order the time series not by time, but by value in descending order. Figure 3.4 shows the electrolyzer load duration curves of various *Seg0* and *Seg1* runs compared to the reference case one. Also within these duration curves, *Seg0* and *Lin0* show a good approximation of the reference case data already at 500 segm/y or 5 typical days. However, when other time series are also taken into account in the TSAMs, *Seg1* seems to depict the reference case electrolyzer load duration curve better than *Lin1* in general.

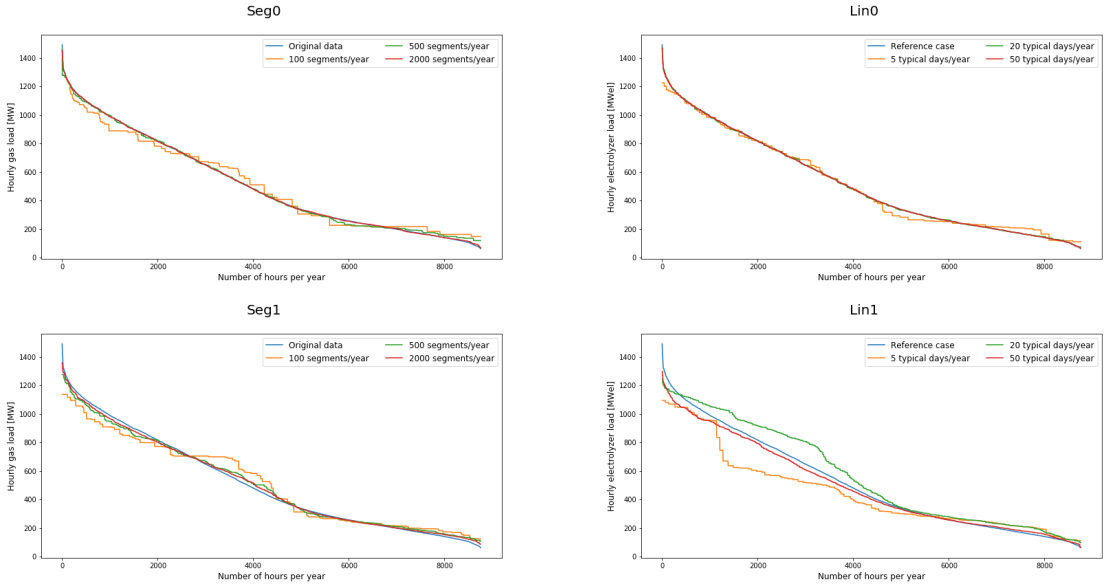

Figure 3.4: Electrolyzer load duration curves of $Seg0_1$ (top left) and $Seg1_1$ (bottom left) with different numbers of segments compared to the reference case, $Lin0_1$ (top right) and $Lin1_1$ (bottom right) with sifferent numbers of tday/y compared to the reference case.

To calculate the duration curve errors (DCEs), the nRMSE (equation 2.30) is used. See Figure 3.6 for the error values of the different agglomerated time series in reference to the reference case time series. Detailed nRMSE values are listed in Appendix C.2.

As previously discussed, when the clustering algorithms are only applied to the gas sector (*Seg0* and *Lin0*), the nRMSE of the electrolyzer load curve is lower than if more inputs are clustered (*Seg1* and *Lin1*). For all other time series, the methods applied on all time series (*Seg1* and *Lin1*) produce a lower nRMSE.

When setting the boundary for an acceptable DCE at 5% nRMSE, *Seg1* achieves this at 1000 segments or more per year in all time series. The linkage approach *Lin1* gets to nRMSE of $\leq 5\%$ in electric load at 10 tday/y, but requires 100 typical days to achieve an nRMSE of $\leq 5\%$ in electrolyzer load and 200 typical days in wind and solar time series. Overall, the trend in nRMSE is clear within the segmentation method, while more typical days in the linkage method not always lead to better results.

3.2.2 Variation of segment lengths / location of typical days over the year

Seg0 and *Seg1* vary in their choice of segments, as shown in Figure 3.5. To look at this more closely, Figure 3.7 compares the average segment length per week number set in these two approaches.

In the segmentation with only 1000 segm/y, *Seg0* chooses long segments in the summer, weeks 28 until 37, with a maximum weekly average segment duration of 43 h. In the winter weeks, this segmentation has relatively short weekly average segment lengths. In *Seg1*, the segmentation with 1000 segments/year has a maximum average segment duration of 16 h, which is substantially lower. Throughout the year, the average length of segments does not prefer one season but has zones of low mean segment duration both in spring (weeks 11-22) and autumn (weeks 35-43). The segmentation runs with more segm/y show a similar trend as the segmentation runs with 1000 segm/y. However, due to the larger number of segments, this trend is weaker and the average segment lengths are closer together.

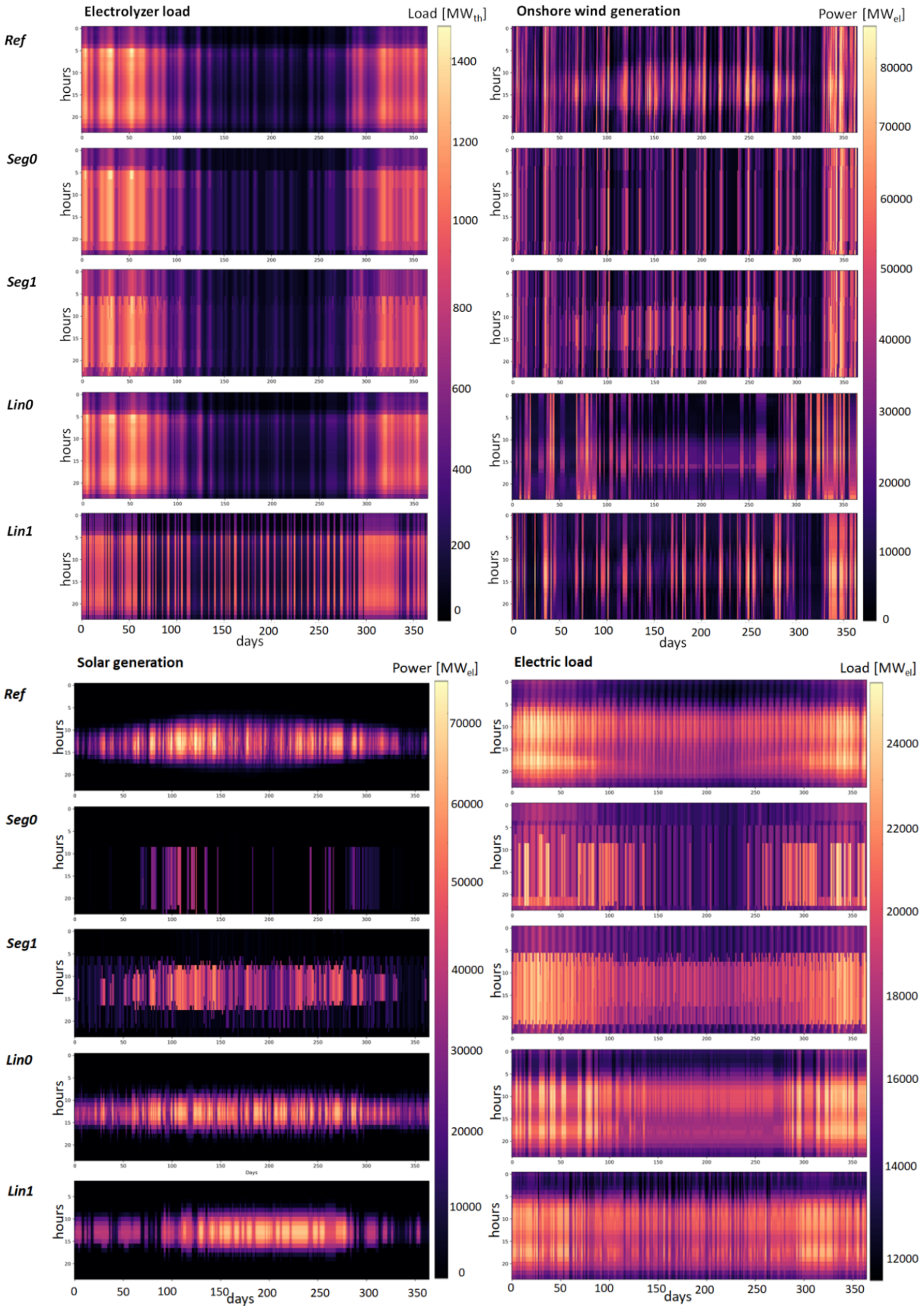

Figure 3.5: Input time series for *eTraGo*'s LOPF. Electrolyzer load, onshore wind power, solar power, electric load for each hour of the year. From top to bottom: Reference case, $Seg0_1$ and $Seg1_1$ (both applying 1000 segments/year), $Lin0_1$ and $Lin1_1$ (both with 10 typical days/year).

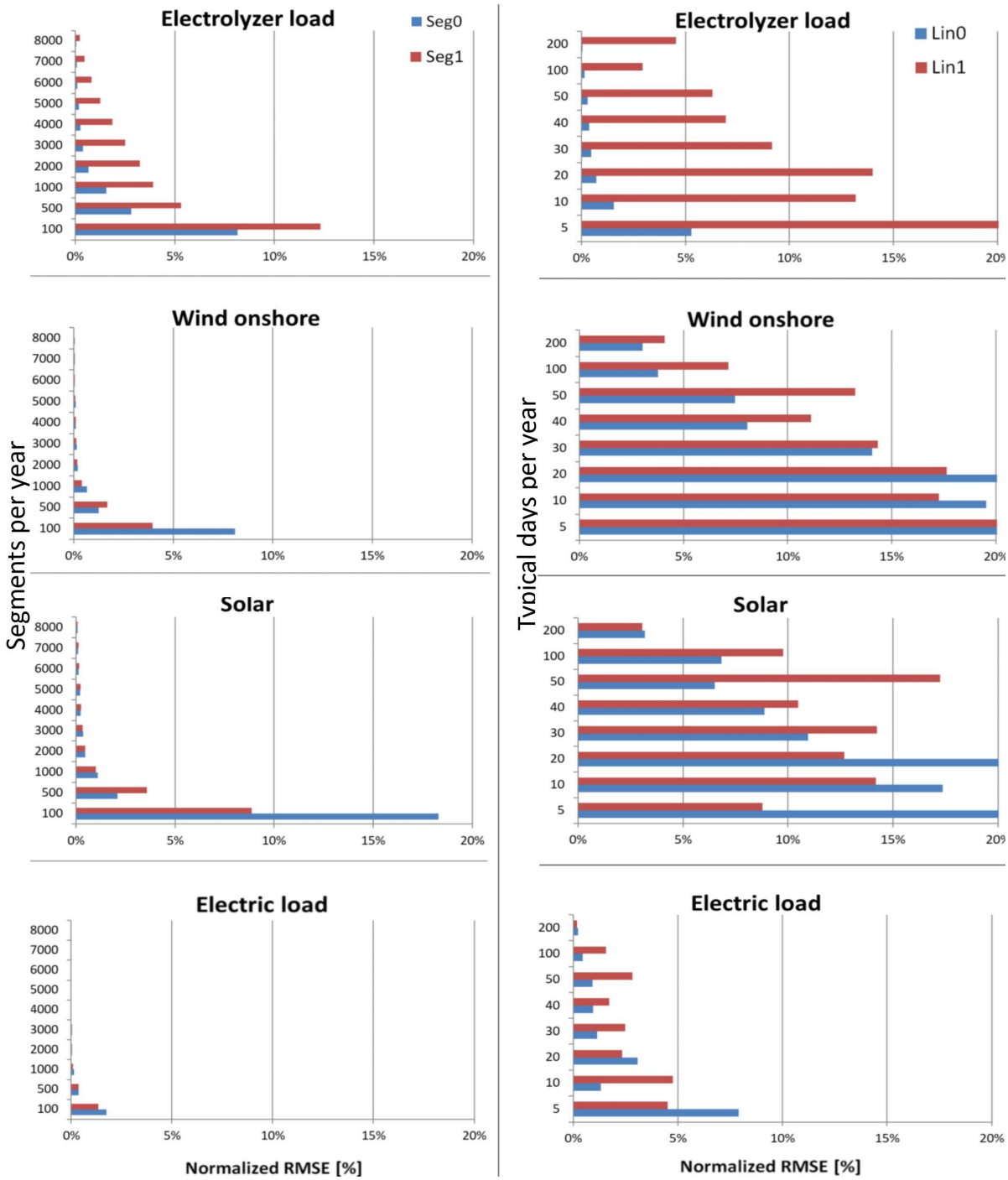

Figure 3.6: nRMSE values of the agglomerated input time series. Left: $Seg0_1$ and $Seg1_1$, Right: $Lin0_1$ and $Lin1_1$ (values cut off at 20%).

When looking at the distribution of typical days in *Lin0* and *Lin1* (Figure 3.8), a similar behaviour can be seen. The linkage method applied on the electrolyzer load only picks more typical days in winter (see in particular 200 tday/y, grey line) and fewer typical days in summer. When applying the linkage method on all input time series, however (*Lin1*), the number of typical days picked per month is distributed more regularly.

This seasonal pattern can be explained by looking at the reference case time series in Figure 3.5. The

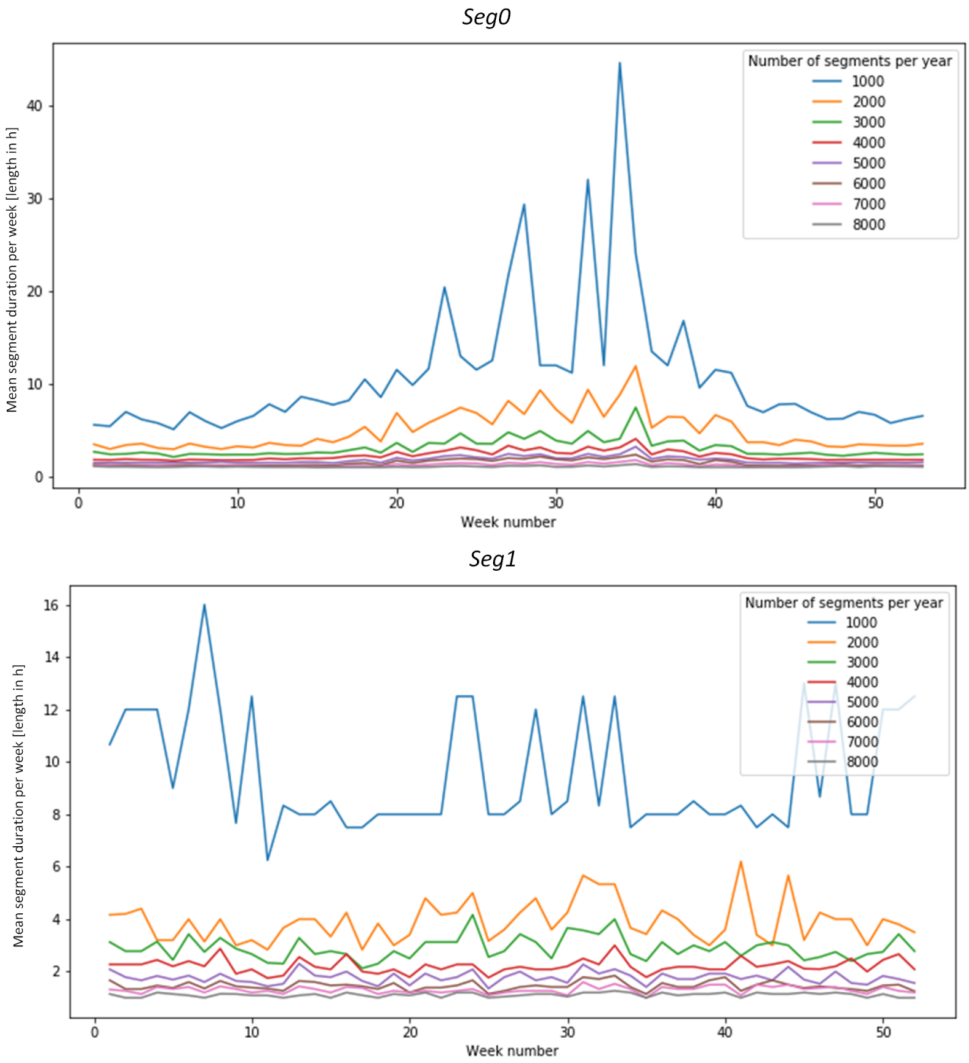

Figure 3.7: Weekly average of the segment length in hours for $Seg0_1$ (top) and $Seg1_1$ (bottom), comparing number of segments ranging from 1000 to 8000 per year.

electrolyzer load has a strong seasonal pattern with a larger load in winter than in summer. When applying a TSAM on this time series only, the winter periods get clustered more finely than the summer to generate a good representation of this load. When however this time series is clustered together with the electric load and the intermittent generation from RES, the daily and seasonal patterns of all time series are taken into account. Onshore wind has a seasonal pattern with a larger generation in December. Solar generation and electric load above all exhibit a daily cycle that is superimposed onto a seasonal cycle. This overlay of seasonal patterns then leads to a more regular distribution of the segment lengths or typical days.

3.3 Changes to the simulation output and computing time reduction

After regarding the changes in input patterns due to the application of TSAMs, this section focuses on analyzing the model outputs of the test computations with TSAMs in relation to the reference case. It looks at how accurate the runs with the TSAMs comply to the reference output data sets and which gains

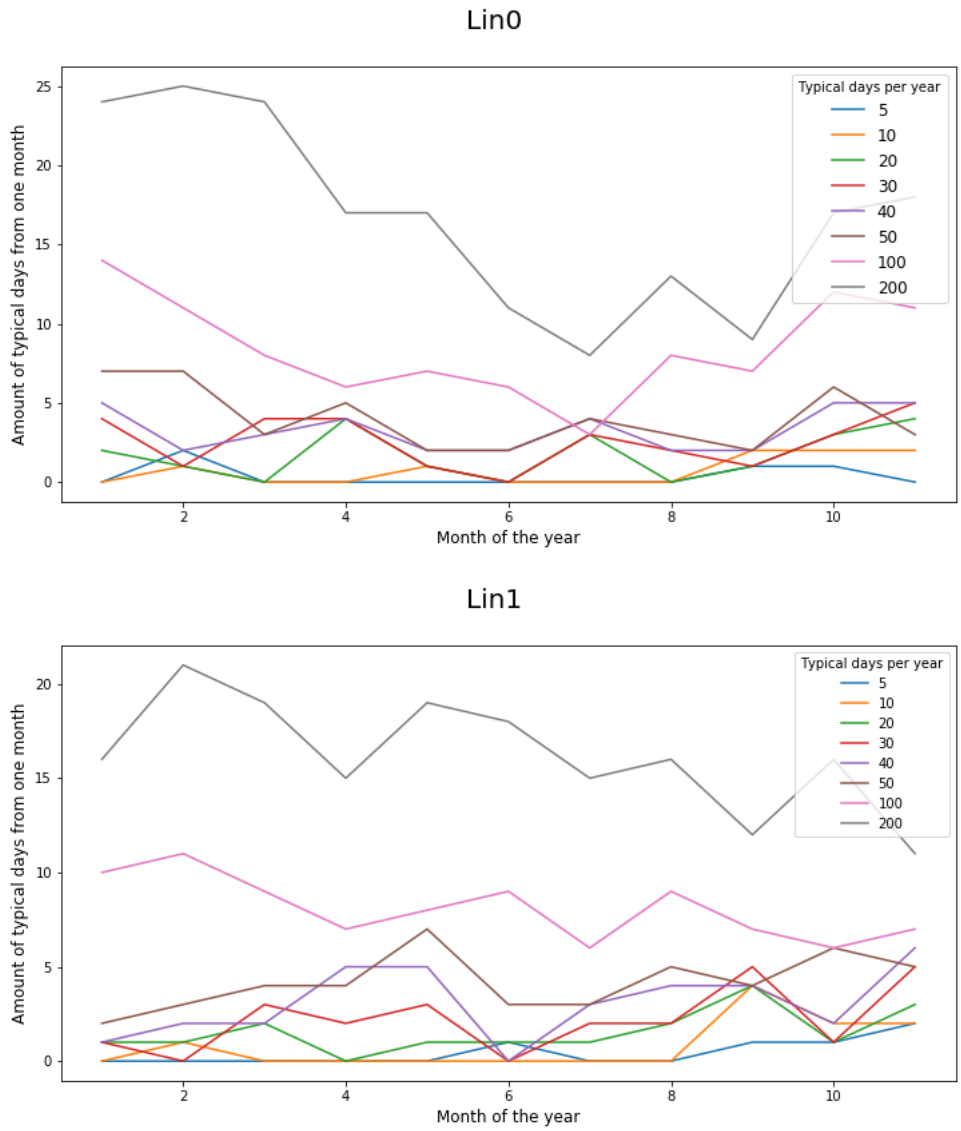

Figure 3.8: Distribution of typical days in $Lin0_1$ and $Lin1_1$ over the months of the year, comparison of tday/y ranging between 5 and 200.

in computing time they result in.

Beginning with the changes in computing time, Figure 3.9 plots the mean average speedup (in ATR) over the each three runs executed (see average absolute computing time in Figure 3.10, detailed computing time of all runs in Appendix C.3).

Both the segmentation, as well as the linkage method, can shorten the time needed to simulate the model and show a negative correlation between the number of segments or typical days and the ATR achieved. The ATR of the segmentation methods starts at 90% (100 segm/y) and decreases until 15% (8000 segm/y), while both *Seg0* and *Seg1* show a similar trend. For the linkage methods, the ATR starts at 40% (*Lin0*) and 30% (*Lin1*) (5 tday/y) and decreases, until it goes negative at 100 tday/y. A negative ATR means that the linkage methods took longer to solve than the reference case. This is presumably caused by the additional overhead for the clustering, which could not be compensated for by a faster LOPF calculation.

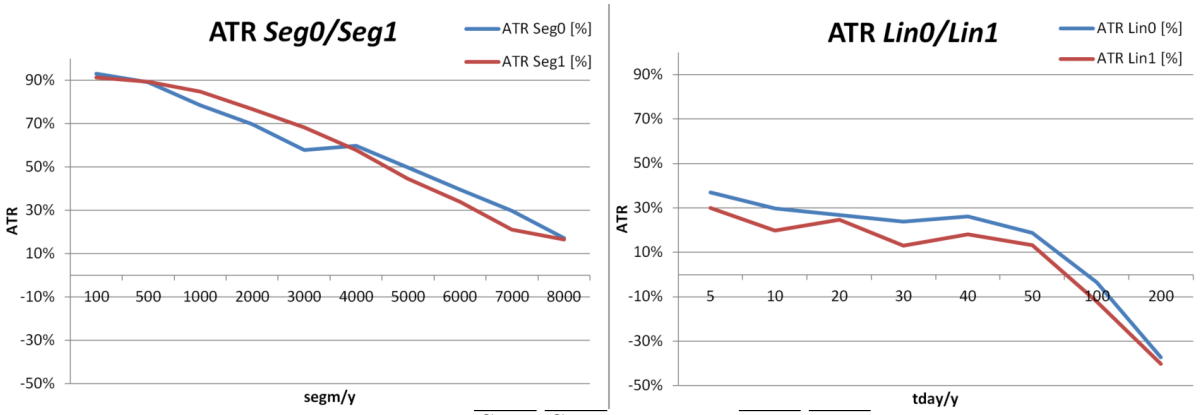

Figure 3.9: Average time reduction of $\overline{Seg0}/\overline{Seg1}$ (left) and $\overline{Lin0}/\overline{Lin1}$ (right) in comparison to the reference case.

When looking at the reduction in computing time alone, the segmentation method seems to be advantageous over the linkage method in *eTraGo*. Looking at the changes in computing results will give us further information about the applicability of both methods.

3.3.1 Trading off computing time and simulation accuracy

In comparing the results of the runs with the agglomerated input time series, one crucial optimisation variable to observe is the annual system costs (see equation 2.10). Figure 3.10 shows the average objective error (AOE) of the annual system costs of the runs with clustering applied and the computing time of the runs on the left. On the right, the AOE of the electrolyzer expansion is shown compared to the AOE of the annual system costs.

Beginning with *Seg0* and *Seg1*, an overall trend to more time needed for the run with an increase in segments can be seen. The AOE of the annual system costs in *Seg0* reaches its minimum at 100 segm/y, then rises, hits a local minimum at 2000 segm/y and rises again. Thus, the error interestingly mostly grows as the resolution of the time series increases. The AOE-graph of *Seg1* only shows a global minimum at 2000 segm/y.

The other interesting variable for the sector-linked model is the deviation in the electrolyzer expansion. *Seg0* exhibits an AOE of around 50%, independent from the number of segments. *Seg1* showed a slightly higher AOE in electrolyzer expansion of 55% when starting with 100 segm/y, which then slowly decreased to 50% at 8000 segm/y.

When looking at similar graphs for *Lin0* and *Lin1*, a different trend shows up. Within this method, the computing time does not differ much when applying different amounts of typical days. While the AOE of the annual system costs decreases initially when more typical days are applied, it afterwards increases again with a further increase in typical days. Similar to *Seg1*, the AOE of the electrolyzer expansion starts at 65% (*Lin0*), 60% (*Lin1*) at 5 tday/y and then decreases 50% at 200 tday/y. However, *Lin0*, which applied linkage only on the electrolyzer time series, has a stronger decrease in electrolyzer expansion AOE initially and oscillates around 50% when applied on 20 to 200 typical days.

Within the methods applied, only *Seg0* and *Seg1* with 2000 segm/y are able to achieve an AOE of $\leq 5\%$ in annual system costs. All methods at best get an AOE in electrolyzer expansion of 50% and always underestimate the electrolyzer expansion (detailed electrolyzer expansion values in Appendix C.3).

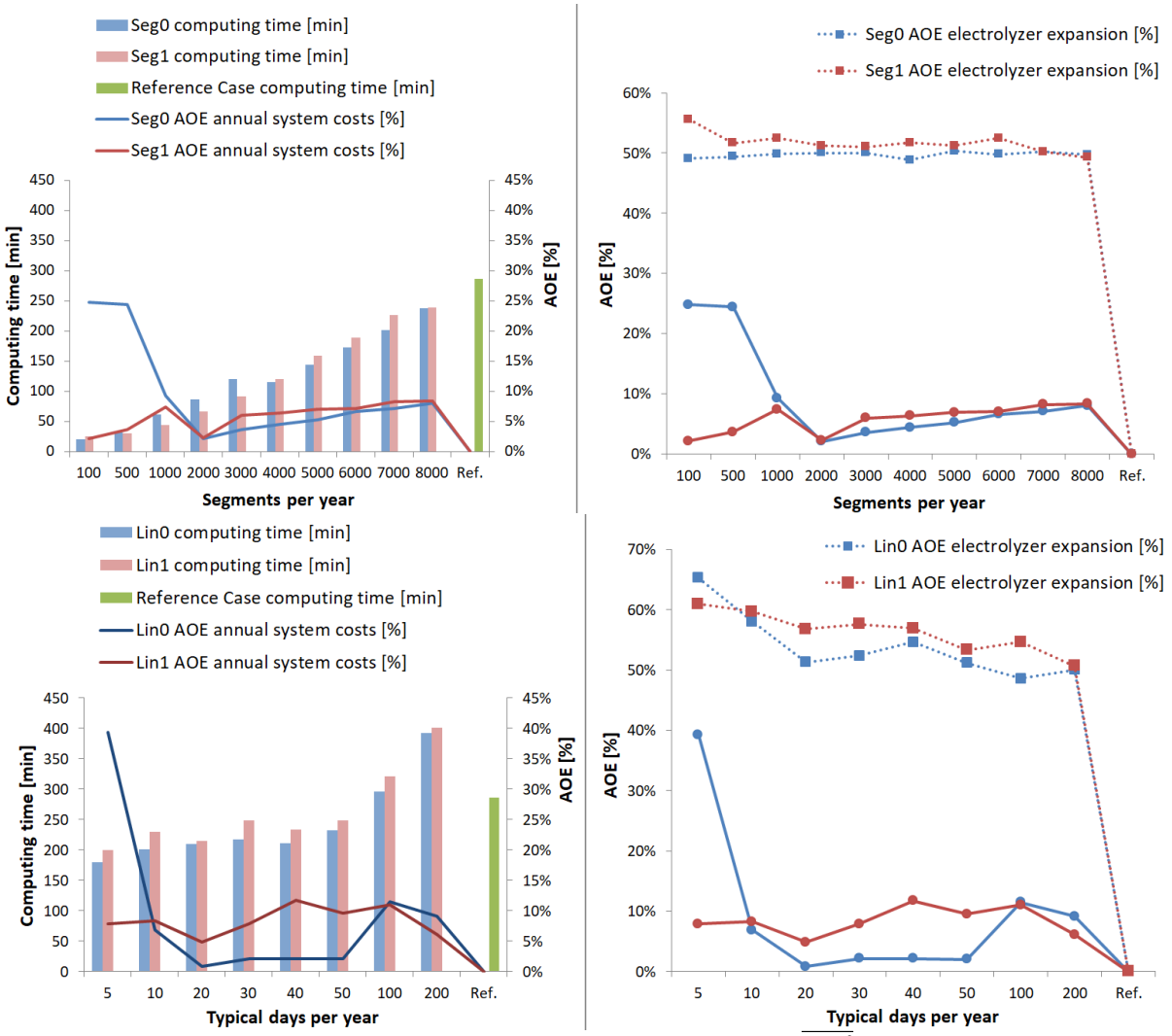

Figure 3.10: Computing time for runs with TSAM and reference case ($\overline{Ref}$) [min] vs average objective error of annual system costs [%] (left), comparison AOE of annual system costs vs AOE of electrolyzer expansion [both in %] (right).Top: $\overline{Seg0}$ and $\overline{Seg1}$, bottom: $\overline{Lin0}$ and $\overline{Lin1}$.

3.3.2 Comparison of generation dispatch

Comparing annual system costs of different model runs has limits as a comparative value. In theory, two model results with completely different expansion of generation units could coincidentally amount to the same costs. As a more detailed comparison, this section looks at another result of the runs, the shares of generation dispatch over the year. Figure 3.11 compares the relative amounts of power from different generation units when applying the compared TSAMs.

The TSAMs used over all input time series (*Seg1/Lin1*) resulted in results closer to the reference case than the ones applied on the gas time series only (*Seg0/Lin0*). The *Seg1* method underrepresented photovoltaic generation and overrepresented lignite generation until a segment number of 2000 per year. The *Lin1* method underrepresented wind and overrepresented coal generation at a small number of tday/y, while at high amounts of tday/y slightly overestimated wind generation.

In general, both the segmentation as well as the linkage method resulted in a relative generation dispatch similar to the reference case at 2000 or more segm/y (*Seg1*) and 30 or more tday/y (Lin1).

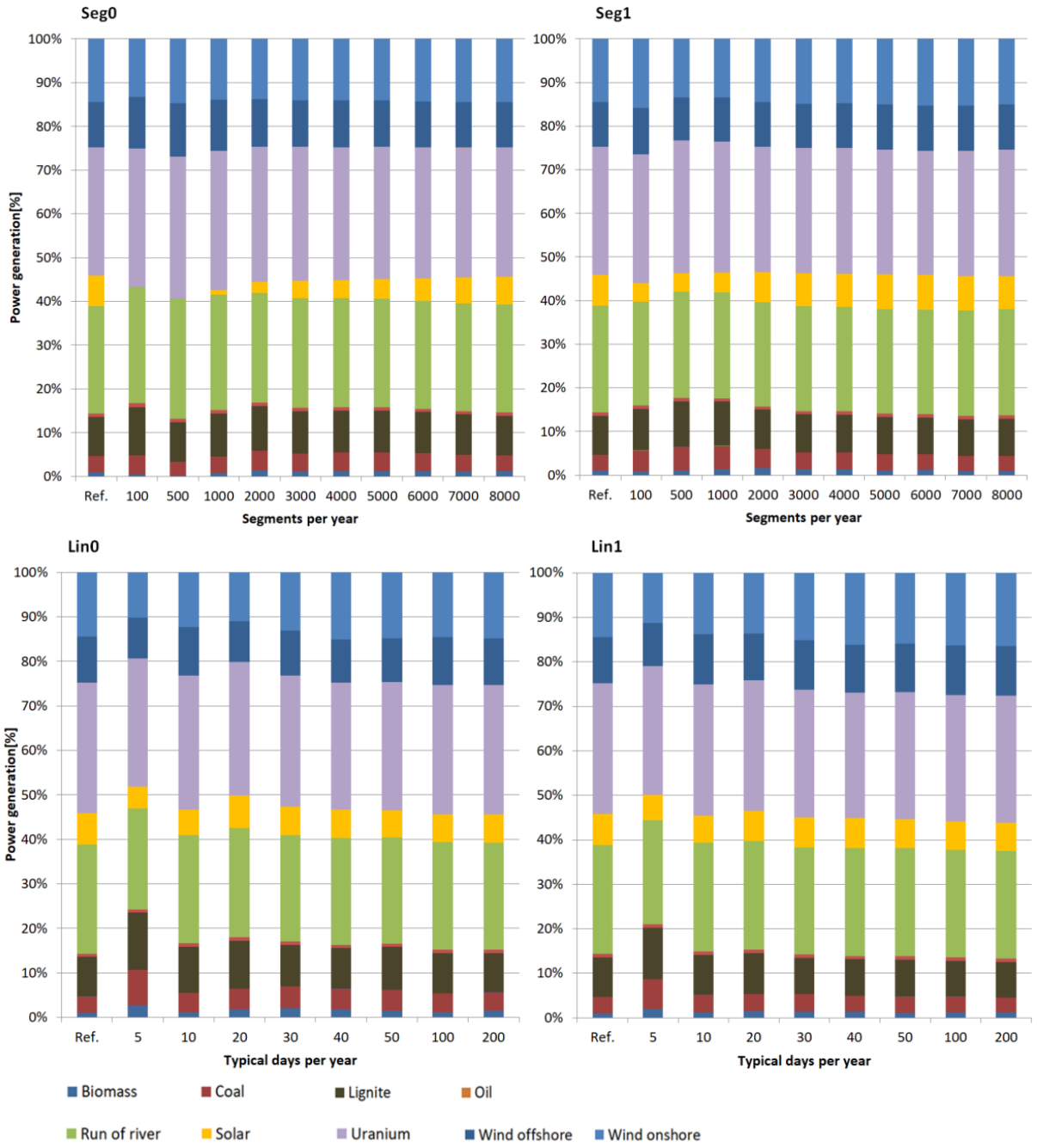

Figure 3.11: Relative generation dispatch in $Seg0_1$ (upper left), $Seg1_1$ (upper right), $Lin0_1$(bottom left) and $Lin1_1$ (bottom right) runs compared to the reference case (Ref_1).

3.3.3 Comparison of hydrogen storage and electrolyzer dispatch

Within the coupled gas and power systems in *eTraGo*, the focus for selecting TSAMs lay on properly representing the inter-period SOC of the gas storage. To analyse how well the selected methods models the storage use, Figure 3.12 looks at the hydrogen storage charge and discharge in all methods in an exemplary month.

TSAMs reduce the computing time by only optimising the model on the cluster centres. The output data hence only include values for these centres. To create comparable continuous output graphs like in Figure 3.12, the output data needs to be desegregated again. This is done by inserting the output value of

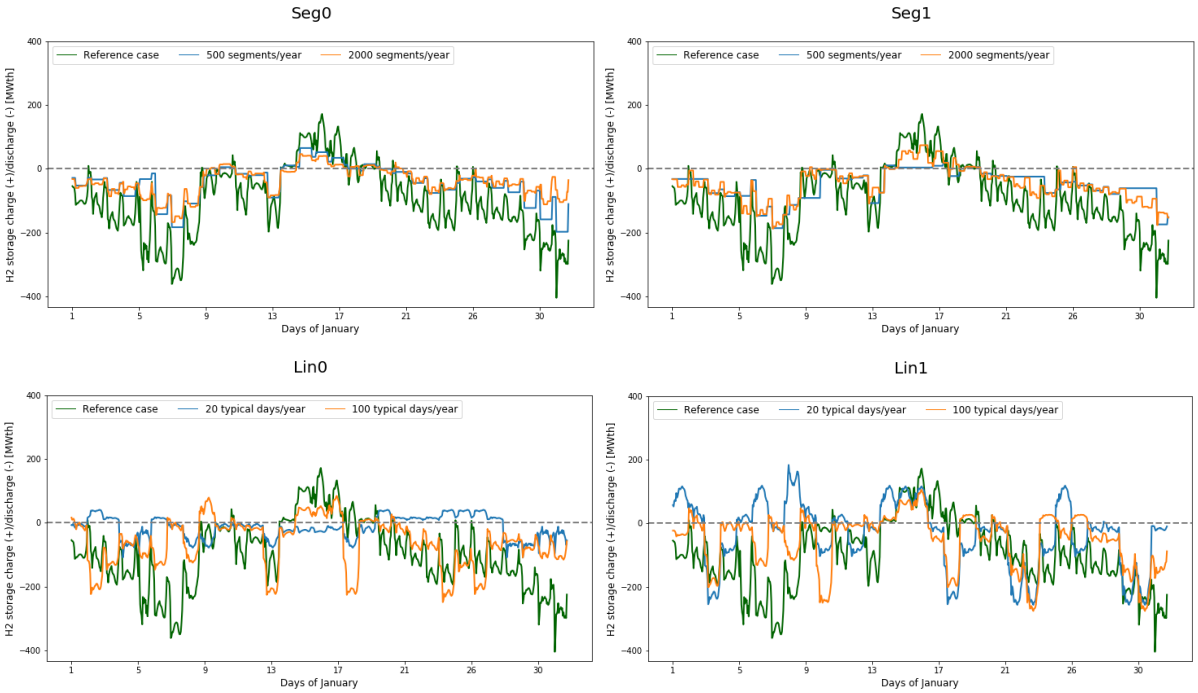

Figure 3.12: Hydrogen storage charge and discharge [$\mathrm{MW_{th}}$] in the reference case (green) compared to *Seg0* (top left) / *Seg1* (top right) with 500 (blue), 4000 (orange) segm/y, *Lin0* (bottom left) / *Lin1* (bottom right) with 20 (blue), 100 (orange) tday/y.

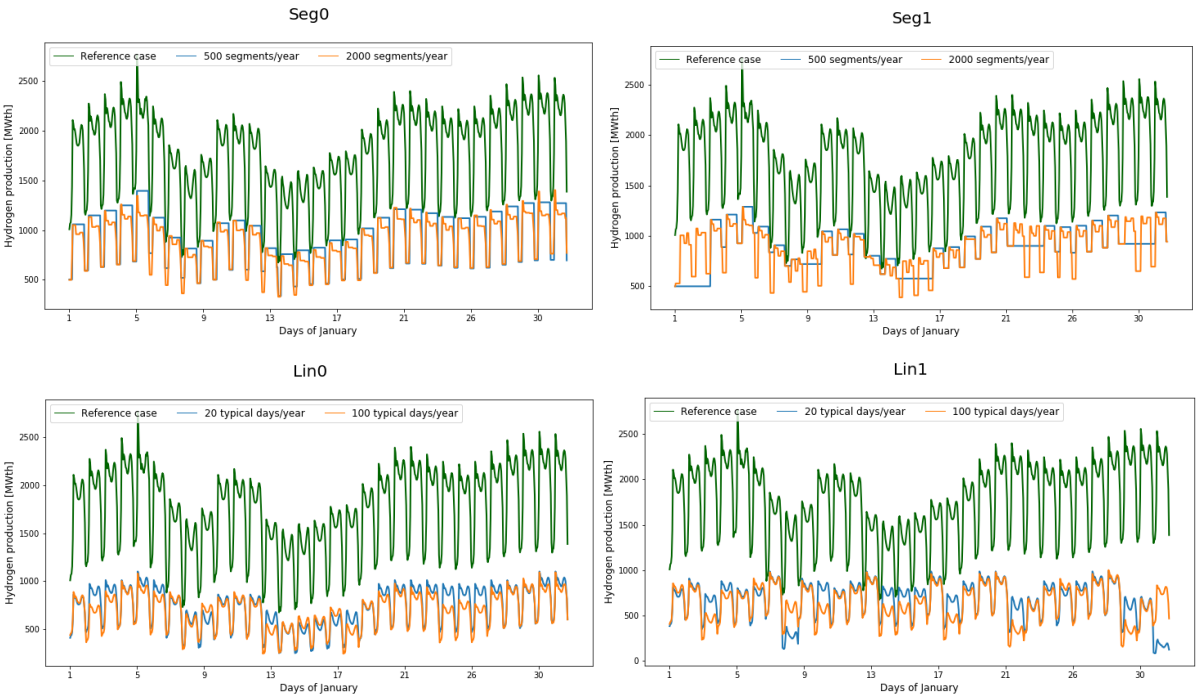

Figure 3.13: Hydrogen production [$\mathrm{MW_{th}}$] in the reference case (green) compared to *Seg0* (top left) / *Seg1* (top right) with 500 (blue), 4000 (orange) segm/y, *Lin0* (bottom left) / *Lin1* (bottom right) with 20 (blue), 100 (orange) tday/y.

each cluster centre also for the hours which are included in the respective cluster.

In Figure 3.12, the exemplary hydrogen charge diagrams of *Seg0* and *Seg1*, smoothing effect can be seen

that the method had on the overall model, as well as the angularity of the curve, as adjacent hours have the same value. The output time series in the runs with applied segmentation do not include the peaks that the reference case shows and overall they underpredict the hydrogen storage.
In the exemplary hydrogen charge diagrams of *Lin0* and *Lin1*, one can see typical days repeating within the time series, while the graphs from the runs with applied linkage do not follow the reference case curve as closely as the ones with applied segmentation did. Similar to the *Seg0* and *Seg1* curves, the curves of *Lin0* and *Lin1* do not display the peaks that the reference case included; a smoothing effect and underprediction are also present here.
Continuing the comparison of hydrogen output data, Figure 3.13 shows exemplary hydrogen production by electrolyzers for the runs with applied TSAMs in comparison to the reference output data. Similar to the hydrogen storage data, all TSAMs underrepresented the hydrogen production. As the electrolyzer expansion is based on the maximum value of hydrogen production within the output data, the underestimation of the electrolyzer expansion with AOEs of 50% and more (Figure 3.10 in Section 3.3.1) is likely due to the smoothing effect of the TSAMs.

3.3.4 Applicability of the clustering methods tested to reduce temporal complexity in *eTraGo*

The four methods picked in Section 2 are applied and tested on a branch of *eTraGo* which included sector coupling of the gas and power sector as described in Section 2.2.4. Returning to the questions asked at the beginning of this chapter, it is found:

- The segmentation method is better equipped to represent the reference case input data set than the linkage method and results in nRMSE over all time series considered of $\leq 5\%$ when applied with 1000 segments or more per year.
- All four methods considered result in AOE of the annual system costs of $< 10\%$ when applied with more than 1000 segments or 10 tday/y. Regarding the relative generation dispatch, *Seg1* closely approximates *Ref* when applied with > 2000 segm/y, *Lin1* with > 30 tday/y. Regarding the output data sets of hydrogen production and storage, however, all methods heavily underrepresented the reference outputs.
- Looking at gains in computing time, the segmentation methods are superior to the linkage methods (Section 3.3).

3.4 Critical appraisal of the methods

Throughout the different runs conducted, the computing time varied due to executing jobs concurrently on the same server. The run times of all runs with TSAMs are shown in Figure 3.14 with their respective trend lines. While differences in run times in a repetition of a run are 79 minutes at maximum ($Seg0_1$ and $Seg0_2$ with 2000 segm/y, with a deviation of 119%), the trend lines show a correlation between the amount of segments/typical days per year and the run time needed.
Regarding the methods applied, the nRMSE of the clustered input values with the linkage methods exceeded those of the segmentation method when applied on all time series by up to 450% ($Seg1_1$ at 100 segm/y, $Lin1_1$ at 5 tday/y, at a computing time of 26.22 and 219.01 minutes, respectively). This could imply that, assuming that the linkage algorithm has been implemented correctly, it is not applicable

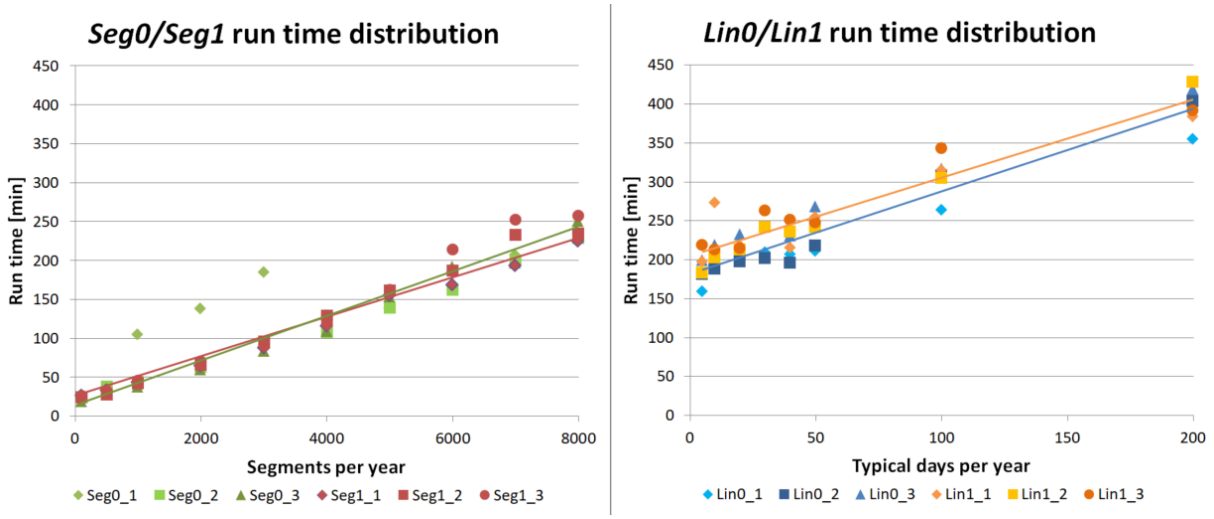

Figure 3.14: Distribution of run times [min] of each three runs of *Seg0/Seg1* (left) and *Lin0/lin1* (right) with linear trend lines.

to the data or has to be tuned further to suit the data. As the AOE of the annual system costs of both methods did not differ much, this may also imply that the nRMSE is not a measure that predicts the accuracy of the model results well.

Both the spatial expansion of the electrolyzers as well as the curtailment of power from intermittent renewables could not be analyzed, as these results are very noisy. This likely is a result of spatially clustering the data to only 50 nodes within Germany. Other results could also have been affected by this spatial clustering, since with this few nodes, the model may have degraded to a set of large interconnected copper plates, which removed a lot of line capacity constraints.

Chapter 4

Resume and outlook

Looking back on the research questions stated in Chapter 1, some answers can now be given. Concerning *(1) Which methods for temporal complexity reduction over time can be used to efficiently reduce the computing time of the model?*, Section 2.2.3 found that both time-based and feature-based methods can be applied. They either vary the resolution or determine typical periods to reduce the input variables of the model (see Table 2.34). Recently published papers prefer feature-based time series aggregation methods (TSAMs), as these tend to more accurately represent the original data (Section 2.3).
For question *(2) To what extent are these methods applicable when the storage capacity of the gas network is taken into account? What adjustments are necessary for that?*, Section 2.5 comprised a literature review. In recent studies on ESMs with long-term storage, two categories of TSAMs were found to be useful: Segmentation methods which cluster adjacent periods, and linkage methods which apply a clustering algorithm with a constraint for tracking the intertemporal state-of-charge.
Regarding question *(3) How can these complexity reduction methods be used across sectors for both power and gas?*, each one variant of segmentation and linkage was applied and tested within *eTraGo* using the scenario *NEP 2035*, which projects a share of renewable energy generators of 75% and Power-to-gas (P2G) capacities of 8 GW in Germany in 2035 (2.1.3, 2.2.5). For segmentation, an innovative method of hierarchical clustering of consecutive hours with variable segment lengths was introduced in *eTraGo*. It was compared to an existing method that links hierarchical clustering of typical days with a constraint for the intertemporal state-of-charge.
Chapter 3, Results and Analysis, compared the in- and outputs of test runs conducted applying the implemented TSAMs with a set reference case without temporal complexity reduction. The segmentation method applied could represent the reference case input data sets within an error margin of 5% while reducing the computing time by 85% (values at 1000 segments per year (segm/y)). The linkage method was not able to reach this error margin for the input data sets while also reducing the computing time. Both methods achieved an error in the annual system costs of $\leq$ 10% as well as a close approximation of the relative generation dispatch when used with $\geq$ 2000 segm/y or $\geq$ 30 typical days per year (tday/y). However, both methods underrepresented the hydrogen production and storage with an error in electrolyzer expansion of $\geq$ 50%. This is likely due to peak shaving, a common problem of clustering algorithms [58].
The segmentation and linkage method were tested in two scenarios: One only applied the TSAM on the input time series of the gas sector, the other applied it on all input time series of the power and gas sectors.

When applied on the gas sector only, both methods preferred to represent the hours in the winter with more resolution, using shorter segment lengths or more typical days. This seasonal behaviour vanished out when applied to both sectors, as it was superimposed with other seasonal and daily patterns in the power sector.

Different aspects of the methods applied in this thesis could be improved upon in further works. The reference case with a spatial clustering to 50 nodes leads to fewer line congestions, curtailment of energy from intermittent renewable sources and decentralized expansion of electrolyzers than expected. To test the methods on a more realistic reference case, a smaller area of Germany could be chosen or the runs could be conducted with less spatial clustering for the whole German system on a server with more computing power.
To receive a smaller error in electrolyzer expansion, both methods could be set up with extreme period methods to decrease the smoothing effects of the TSAMs (as described in Section 2.4.3).
An interesting thing to look at would be amending the weighting of the sectors in the sector-coupled model. Within this thesis, the gas and power input time series were weighted equally. However, if some input time series were weighted higher than others, their seasonal effects would translate stronger to the choice of clusters. This could improve the outcome of the model runs with TSAMs for specific parts of the model, e.g. the electrolyzer usage.
When including the gas flow in *eTraGo*, the gas storage potential due to pressure difference should be included within the model hydrogen storage potential. A more precise calculation on the storage potential of this case than in this thesis would be useful for this.

Following the trend of the past years, the need for reducing complexity in ESMs will likely continue to increase. Within the model *eTraGo*, this need in the future will be due to increasing the length of the simulation by adding more weather years, adding more sectors like heat and mobility as well as a more defined version of the gas sector, also including the gas flow and gas reconversion. This raised the number of input variables, necessitating the application of suitable methods to reduce complexity even more in order to ensure the computability of the model. While this thesis focused on temporal complexity reduction, methods to reduce the spatial complexity while allowing for a better representation of e.g. the electrolyzer capacity expansion will be necessary as well.

Appendix A

Python code

A.1 Calculation of computing time in *appl.py*

Within the file execution file of *eTraGo*, *appl.py*, the commands x, y and z were added to calculate the computing time of a run:

```
def run_etrago(args, json_path):
    x = time.time()

    etrago = Etrago(args, json_path)

    # import network from database
    etrago.build_network_from_db()

    # adjust network, e.g. set (n-1)-security factor
    etrago.adjust_network()

    # ehv network clustering
    etrago.ehv_clustering()

    # k-mean clustering
    etrago.kmean_clustering()

    # PtG Modelling Input Parameters
    P_grid_oriented_installations_H2 = 3000.0
    """[MWel] NEP_2035_v_2021_Szenariorahmen_2035_Entwurf , p.52, Scenario C,
    netzdienliche Power-to-Hydrogen Anlagen"""
    n_Full_load_hours_H2 = 1500.0
    """ NEP_2035_v_2021_Szenariorahmen_2035_Entwurf , p.52,
    Scenario C, netzdienliche Power-to-Hydrogen Anlagen"""
    e_store_gas_grid = 15.0 * 10**6
    """ [MWh] Kapazität Speicher Wasserstoff im Gasnetz: 15 TWh
    Speicherkapazität, Götz et al (2011). Speicherung von regenerativ erzeugter
    elektrischer Energie in der Erdgasinfrastruktur. gwf-Gas, April.
    gwf-Gas Erdgas. 152. 200-210."""
```

```
    #add ptg installations
    ptg_addition(etrago.network,
                 etrago.args['network_clustering_kmeans']['n_clusters'],
                P_grid_oriented_installations_H2 = P_grid_oriented_installations_H2,
                n_Full_load_hours_H2 = n_Full_load_hours_H2,
                e_store_gas_grid = e_store_gas_grid)

    # skip snapshots
    etrago.skip_snapshots()

    # snapshot clustering
    etrago.snapshot_clustering()

    # start linear optimal powerflow calculations
    etrago.lopf()

    # TODO: check if should be combined with etrago.lopf()
    etrago.pf_post_lopf()

    # spatial disaggregation
    etrago.disaggregation()

    # calculate central etrago results
    etrago.calc_results()
    print(etrago.results)

    #Calculate runtime
    y = time.time()
    z = (y - x) / 60
    print("Runtime etrago [min]:", round(z, 2))

    return etrago
```

A.2 Changes in *snapshot.py*

A.2.1 For implementing the segmentation method

```
def tsam_cluster(timeseries_df,
                 typical_periods=10,
                 how='daily',
                 extremePeriodMethod = 'None',
                 segmentation = False,
                 segment_no = 1000):
    """
    Parameters
    ----------
```

```
    timeseries_df : pd.DataFrame
        DataFrame with timeseries to cluster
    segmentation: Is given by the run-function, can be True or False
    segment_no: Only is used when segmentation is true, no of segments per year

    Returns
    -------
    timeseries : pd.DataFrame
        Clustered timeseries, only used if clustering is applied
    """

    aggregation = tsam.TimeSeriesAggregation(
        timeseries_df,
        noTypicalPeriods = typical_periods,
        extremePeriodMethod = extremePeriodMethod,
        addPeakMin = ['residual_load'],
        addPeakMax = ['residual_load'],
        rescaleClusterPeriods = False,
        hoursPerPeriod = 8760,
        clusterMethod = 'hierarchical',
        segmentation = segmentation,
        noSegments = segment_no)
    print('Snapshot clustering to ' + str(typical_periods) + period +
          ' using extreme period method: ' + extremePeriodMethod +
          ', Segmentation ' + str(segmentation))

    timeseries_creator = aggregation.createTypicalPeriods()
    timeseries = timeseries_creator.copy()

    #If Segmentation is True, insert 'Dates' and 'SegmentNo' column in
        #timeseries output
    if segmentation is True:
        weights=timeseries.index.get_level_values(2)
        dates_df= timeseries_df.index.get_level_values(0)
        dates=[]
        segmentno=[]
        wcount=0
        count=0
        for weight in weights:
            dates.append(dates_df[wcount])
            wcount = wcount + weight
            segmentno.append(count)
            count = count +1
        timeseries.insert(0, "dates", dates, True)
        timeseries.insert(1, "SegmentNo", segmentno, True)
        timeseries.insert(2, "SegmentDuration", weights, True)
        timeseries.set_index(['dates', 'SegmentNo', 'SegmentDuration'],
            inplace=True)
```

```
        if 'Unnamed: 0' in timeseries.columns:
            del timeseries['Unnamed: 0']
        if 'Segment Step' in timeseries.columns:
            del timeseries['Segment Step']
        print(timeseries)
    timeseries.to_csv('results/timeseries_' + str(segment_no))
```

A.2.2 For applying TSAMs on the electrolyzer load time series only

```
def prepare_pypsa_timeseries(network, normed=False):
    """
    """

    if normed:
        normed_loads = network.loads_t.p_set / network.loads_t.p_set.max()
        normed_loads.columns = 'L' + normed_loads.columns
        normed_renewables = network.generators_t.p_max_pu
        normed_renewables.columns = 'G' + normed_renewables.columns

        df = pd.concat([normed_renewables,
                        normed_loads], axis=1)
    else:
        loads = network.loads_t.p_set.copy()
        load_gas = loads['Gas_Load']
        load_gas.column = 'L_Gas'
        renewables = network.generators_t.p_max_pu.mul(
                network.generators.p_nom[
                network.generators_t.p_max_pu.columns], axis = 1).copy()
        renewables.columns = 'G' + renewables.columns
        residual_load=pd.DataFrame()
        residual_load['residual_load']=loads.sum(axis=1)-renewables.sum(axis=1)
        df = pd.concat([load_gas], axis=1)
        #df = pd.concat([renewables, loads, residual_load], axis=1)
```

Appendix B

How to reproduce the results

This appendix lists how the model *eTraGo* can be downloaded and run to reproduce the results from this thesis in Section B.1. Section B.2 lists the applied solver settings, the dimensions of the server used for the optimisation as well as further input data of the scenario applied to show how the computing times and the reference case outputs were generated.

B.1 Installing and running *eTraGo*

eTraGo is a Python tool to be run in a Linux environment. A detailed description of how it can be downloaded and installed can be found on its website *readthedocs*[1]. For the optimisation runs within this thesis, a version of an *eTraGo* branch[2] was used, in which the sector coupling of power and gas as described in Section 2.2.4 was introduced. The modified *eTraGo* files to run the selected TSAMs (as described in Appendix A.2.1, A.2.2), are included in the cloud folder of this thesis.

eTraGo is developed as part of publicly funded open-source projects, so all packages, tools and data sets mentioned in this work are publicly available and open-source. For optimizations and load flow calculations, *eTraGo* resorts to *PyPSA*[3], a tool developed at the Frankfurt Institute for Advanced Studies [8]. The tool *tsam*[4], developed at the Jülich Research Centre, is used for the time series aggregation methods of *eTraGo* [12].

B.2 Server and scenario settings applied within this thesis

The solver applied was *Gurobi*, a licensed mathematical optimization solver, using a student's license[5]. The options that *Gurobi* was applied with were: Barrier Method with a Barrier convergence tolerance of 10^{-5} and a feasibility tolerance of 10^{-2}, which is the standard conditions in *eTraGo* when applying *Gurobi*. For more information on these options please refer to the *Gurobi* documentation [68].

[1] https://etrago.readthedocs.io/en/latest/installation.html

[2] https://github.com/openego/eTraGo/commit/293865da9d6fa860d794aab0b3d766502f60bb7e

[3] https://github.com/PyPSA/PyPSA

[4] https://github.com/FZJ-IEK3-VSA/tsam

[5] Available on: https://www.gurobi.com

The solver ran on the server *torch* of the DLR with 48 Intel® Xeon® Gold 6128 central processing units at 3.40 GHz and 3 TB random-access memory.
The model was run applying its *NEP 2035* scenario. Next to the parameters laid out in 2.1.3 and 2.2.5, the installed power generation and CAPEX of generator units set within this scenario are shown in Tables B.1 and B.3.

Table B.1: Installed power generation capacity in the *NEP 2035* scenario [2].

Capacity installed of generator type	Germany [GW_{el}]	Germany and neighbouring countries[GW_{el}]
Natural gas	0.0	57.5
Lignite	9.1	25.7
Waste	11.0	27.3
Oil	35.7	91.2
Nuclear	0.8	5.6
Biomass	1.6	1.6
Coal	1.0	1.0
Run of the river	59.2	209.9
Storage	90.3	154.9
Pumped storage	16.4	42.8
Photovoltaics	60.6	114.3
Wind Onshore	8.3	36.0
Wind Offshore	5.8	70.7
Geothermy	181.5	418.8
Other renewable generation	240.7	628.6

Table B.3: CAPEX of generator units in NEP 2035 scenario in *eTraGo* [9].

Generator type	CAPEX[€/kW_{el}]
Natural gas	1322
Lignite	2718
Waste	1800
Oil	400
Nuclear	6000
Biomass	2141
Coal	1800
Run of the river	3000
Storage	2000
Pumped storage	2000
Photovoltaics	555
Wind Onshore	1154
Wind Offshore	2396
Geothermy	3216
Other renewable generation	3216

Appendix C

Further results

C.1 Test computation outputs

Table C.1: Reference case test computation outputs.

Round number	Annual system costs [€]	Computing time [min]	Electrolyzer capacity [MW_{el}]
1	$8.57 \cdot 10^9$	320.86	3475.66
2	$8.57 \cdot 10^9$	261.11	3449.75
3	$8.58 \cdot 10^9$	276.70	3489.98

Table C.2: $Seg0$ test computation outputs.

Round number	segm/y	Annual system costs [€]	Computing time [min]	Electrolyzer capacity [MW_{el}]
1	100	$6.38 \cdot 10^9$	18.60	1790.89
	500	$6.46 \cdot 10^9$	27.20	1743.15
	1000	$7.78 \cdot 10^9$	104.71	1761.41
	2000	$8.76 \cdot 10^9$	137.63	1752.32
	3000	$8.27 \cdot 10^9$	185.17	1754.09
	4000	$8.20 \cdot 10^9$	129.21	1799.39
	5000	$8.12 \cdot 10^9$	138.89	1722.00
	6000	$8.00 \cdot 10^9$	166.74	1728.09
	7000	$7.96 \cdot 10^9$	206.72	1728.46
	8000	$7.87 \cdot 10^9$	233.84	1728.87
2	100	$6.58 \cdot 10^9$	23.60	1756.99
	500	$6.47 \cdot 10^9$	37.89	1743.16
	1000	$7.78 \cdot 10^9$	43.57	1728.59
	2000	$8.76 \cdot 10^9$	62.85	1725.22
	3000	$8.26 \cdot 10^9$	93.46	1721.00
	4000	$8.19 \cdot 10^9$	107.35	1800.99
	5000	$8.12 \cdot 10^9$	138.84	1728.71
	6000	$8.00 \cdot 10^9$	161.69	1728.09

Table C.2: $Seg0$ test computation outputs.

Round number	segm/y	Annual system costs [€]	Computing time [min]	Electrolyzer capacity [MW_{el}]
	7000	$7.96 \cdot 10^9$	199.52	1728.46
	8000	$7.88 \cdot 10^9$	228.12	1728.87
3	100	$6.38 \cdot 10^9$	18.73	1757.00
	500	$6.50 \cdot 10^9$	28.63	1781.40
	1000	$7.77 \cdot 10^9$	37.24	1736.93
	2000	$8.75 \cdot 10^9$	59.57	1725.22
	3000	$8.26 \cdot 10^9$	83.47	1727.50
	4000	$8.18 \cdot 10^9$	109.40	1730.44
	5000	$8.13 \cdot 10^9$	153.10	1722.00
	6000	$8.01 \cdot 10^9$	191.41	1775.75
	7000	$7.97 \cdot 10^9$	197.00	1728.46
	8000	$7.90 \cdot 10^9$	250.24	1776.84

Table C.3: $Seg1$ test computation outputs.

Round number	segm/y	Annual system costs [€]	Computing time [min]	Electrolyzer capacity [MW_{el}]
1	100	$8.23 \cdot 10^9$	26.22	1491.51
	500	$8.90 \cdot 10^9$	33.33	1610.72
	1000	$9.21 \cdot 10^9$	43.70	1609.91
	2000	$8.76 \cdot 10^9$	64.84	1608.30
	3000	$8.05 \cdot 10^9$	87.80	1661.96
	4000	$8.02 \cdot 10^9$	116.08	1608.45
	5000	$7.98 \cdot 10^9$	154.80	1738.71
	6000	$7.96 \cdot 10^9$	168.49	1608.23
	7000	$7.86 \cdot 10^9$	193.48	1729.01
	8000	$7.85 \cdot 10^9$	225.08	1728.86
2	100	$8.68 \cdot 10^9$	23.70	1508.02
	500	$8.87 \cdot 10^9$	27.78	1760.60
	1000	$9.19 \cdot 10^9$	41.93	1610.33
	2000	$8.77 \cdot 10^9$	65.83	1734.39
	3000	$8.07 \cdot 10^9$	95.68	1758.19
	4000	$8.04 \cdot 10^9$	128.28	1736.33
	5000	$7.97 \cdot 10^9$	160.78	1608.16
	6000	$7.96 \cdot 10^9$	186.52	1608.20
	7000	$7.87 \cdot 10^9$	232.84	1729.03
	8000	$7.86 \cdot 10^9$	234.48	1775.68
3	100	$8.46 \cdot 10^9$	26.02	1627.90
	500	$8.91 \cdot 10^9$	31.38	1661.88

Table C.3: $Seg1$ test computation outputs.

Round number	segm/y	Annual system costs [€]	Computing time [min]	Electrolyzer capacity [MW_{el}]
	1000	$9.23 \cdot 10^9$	45.36	1730.99
	2000	$8.78 \cdot 10^9$	70.37	1735.17
	3000	$8.06 \cdot 10^9$	90.02	1680.21
	4000	$8.02 \cdot 10^9$	118.18	1682.14
	5000	$7.98 \cdot 10^9$	161.55	1734.80
	6000	$7.98 \cdot 10^9$	213.74	1734.77
	7000	$7.87 \cdot 10^9$	251.95	1729.02
	8000	$7.86 \cdot 10^9$	257.00	1775.68

Table C.4: $Lin0$ test computation outputs.

Round number	tday/y	Annual system costs [€]	Computing time [min]	Electrolyzer capacity [MW_{el}]
1	5	1.21E+10	199.61	1203.34252
	10	$7.98 \cdot 10^9$	219.28	1409.803709
	20	$8.51 \cdot 10^9$	232.58	1719.367763
	30	$8.40 \cdot 10^9$	242.4	1708.085511
	40	$8.40 \cdot 10^9$	230.71	1581.567141
	50	$8.42 \cdot 10^9$	267.99	1767.285057
	100	$7.58 \cdot 10^9$	316.61	1706.937993
	200	$7.79 \cdot 10^9$	418.94	1722.385634
2	5	$1.19 \cdot 10^{10}$	159.29	1209.81
	10	$8.01 \cdot 10^9$	194.35	1456.95
	20	$8.51 \cdot 10^9$	198.64	1720.14
	30	$8.39 \cdot 10^9$	209.42	1614.09
	40	$8.39 \cdot 10^9$	206.69	1581.56
	50	$8.40 \cdot 10^9$	211.32	1697.04
	100	$7.61 \cdot 10^9$	263.58	1789.06
	200	$7.81 \cdot 10^9$	355.35	1763.68
3	5	$1.19 \cdot 10^{10}$	181.21	1209.79
	10	$7.99 \cdot 10^9$	188.53	1470.69
	20	$8.51 \cdot 10^9$	197.59	1678.63
	30	$8.40 \cdot 10^9$	202.12	1708.61
	40	$8.40 \cdot 10^9$	195.52	1581.57
	50	$8.40 \cdot 10^9$	217.91	1708.73
	100	$7.60 \cdot 10^9$	308.13	1798.67
	200	$7.78 \cdot 10^9$	403.50	1722.39

Table C.5: $Lin1$ test computation outputs.

Round number	tday/y	Annual system costs [€]	Computing time [min]	Electrolyzer capacity [MW_{el}]
1	5	$9.36 \cdot 10^9$	219.01	1206.69
	10	$8.14 \cdot 10^9$	212.66	1274.26
	20	$8.75 \cdot 10^9$	214.95	1540.91
	30	$7.69 \cdot 10^9$	262.95	1479.83
	40	$7.71 \cdot 10^9$	251.39	1551.01
	50	$7.49 \cdot 10^9$	248.05	1537.10
	100	$7.61 \cdot 10^9$	342.96	1580.97
	200	$8.09 \cdot 10^9$	391.49	1599.42
2	5	$9.27 \cdot 10^9$	197.98	1461.16
	10	$8.30 \cdot 10^9$	273.02	1440.60
	20	$7.97 \cdot 10^9$	218.14	1556.38
	30	$8.17 \cdot 10^9$	241.21	1435.30
	40	$7.50 \cdot 10^9$	215.75	1447.45
	50	$8.52 \cdot 10^9$	254.79	1809.89
	100	$7.70 \cdot 10^9$	314.83	1573.34
	200	$8.07 \cdot 10^9$	384.13	1805.46
3	5	$9.16 \cdot 10^9$	183.22	1421.2985842
	10	$7.17 \cdot 10^9$	202.99	1502.510735
	20	$8.11 \cdot 10^9$	212.72	1425.1767183
	30	$7.85 \cdot 10^9$	241.70	1522.141939
	40	$7.51 \cdot 10^9$	235.50	1510.802896
	50	$7.27 \cdot 10^9$	241.45	1538.101988
	100	$7.59 \cdot 10^9$	305.23	1592.072878
	200	$8.01 \cdot 10^9$	428.06	1755.35

C.2 nRMSE between agglomerated and reference time series

Table C.6: nRMSE values between input time series with segmentation applied and the reference case.

Method	segm/y	Electrolyzer load	Onshore wind	Solar	Electric load
$Seg0_1$	100	0.0814	0.0811	0.1830	0.0175
	500	0.0280	0.0125	0.0209	0.0036
	1000	0.0156	0.0066	0.0111	0.0015
	2000	0.0066	0.0021	0.0048	0.0004
	3000	0.0039	0.0016	0.0036	0.0003
	4000	0.0024	0.0011	0.0024	0.0002
	5000	0.0016	0.0010	0.0022	0.0002
	6000	0.0010	0.0007	0.0013	0.0001
	7000	0.0007	0.0005	0.0011	0.0001

Table C.6: nRMSE values between input time series with segmentation applied and the reference case.

Method	segm/y	Electrolyzer load	Onshore wind	Solar	Electric load
	8000	0.0003	0.0005	0.0008	0.0001
$Seg1_1$	100	0.1232	0.0397	0.0888	0.0134
	500	0.0532	0.0168	0.0359	0.0036
	1000	0.0390	0.0041	0.0101	0.0008
	2000	0.0324	0.0018	0.0046	0.0004
	3000	0.0250	0.0013	0.0033	0.0003
	4000	0.0185	0.0010	0.0027	0.0002
	5000	0.0123	0.0008	0.0023	0.0002
	6000	0.0080	0.0007	0.0016	0.0001
	7000	0.0046	0.0005	0.0013	0.0001
	8000	0.0021	0.0003	0.0009	0.0001

Table C.7: nRMSE values between input time series with linkage method applied and the reference case.

Method	tday/y	Electrolyzer load	Onshore wind	Solar	Electric load
$Lin0_1$	5	0.0528	0.3588	0.2721	0.0790
	10	0.0156	0.1953	0.1737	0.0132
	20	0.0072	0.3779	0.2689	0.0309
	30	0.0049	0.1405	0.1096	0.0116
	40	0.0038	0.0806	0.0888	0.0095
	50	0.0029	0.0746	0.0653	0.0093
	100	0.0015	0.0376	0.0684	0.0046
	200	0.0006	0.0302	0.0317	0.0022
$Lin1_1$	5	0.2169	0.2198	0.0879	0.0451
	10	0.1318	0.1726	0.1418	0.0476
	20	0.1400	0.1764	0.1268	0.0234
	30	0.0916	0.1432	0.1423	0.0248
	40	0.0693	0.1113	0.1048	0.0172
	50	0.0630	0.1325	0.1726	0.0282
	100	0.0293	0.0714	0.0976	0.0156
	200	0.0455	0.0408	0.0306	0.0019

C.3 Spatial electrolyzer expansion

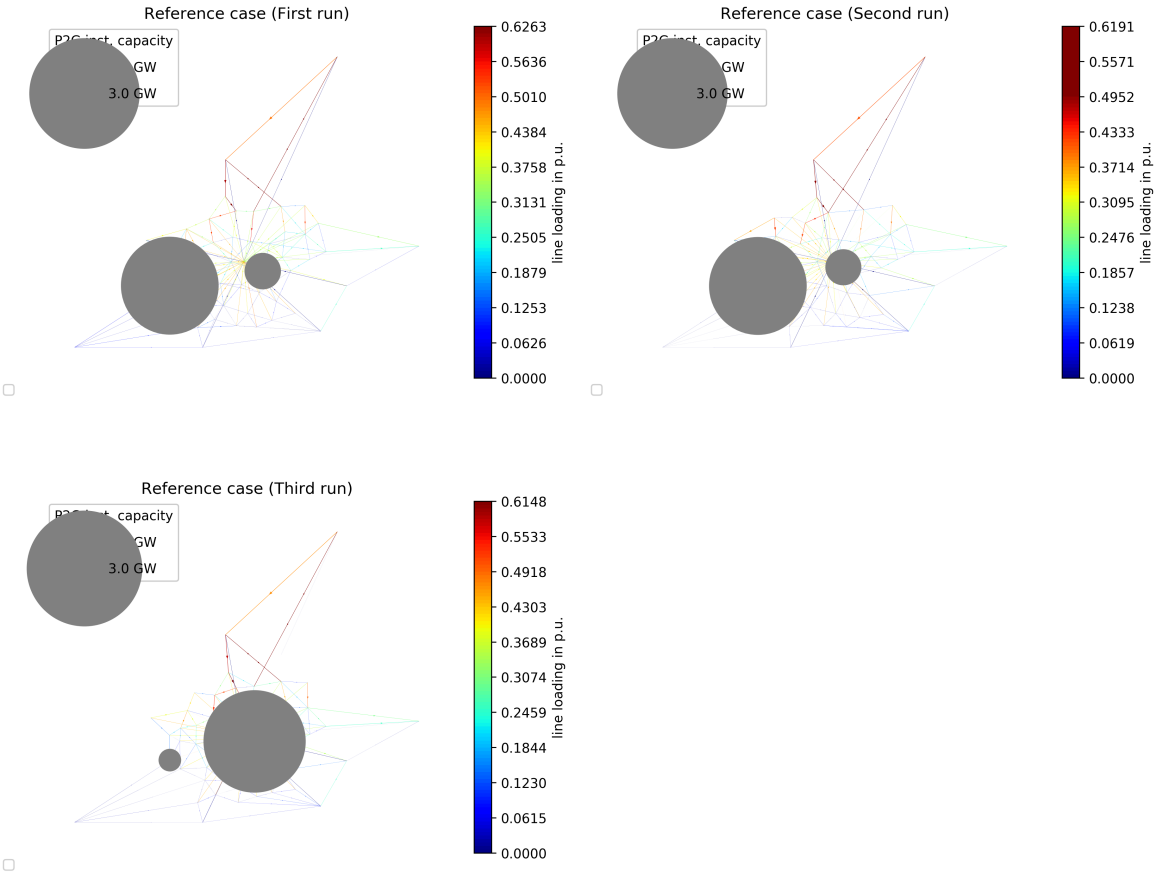

Figure C.1: Spatial electrolyzer expansion [GW_{el}] of the reference case runs and line loading [pu].

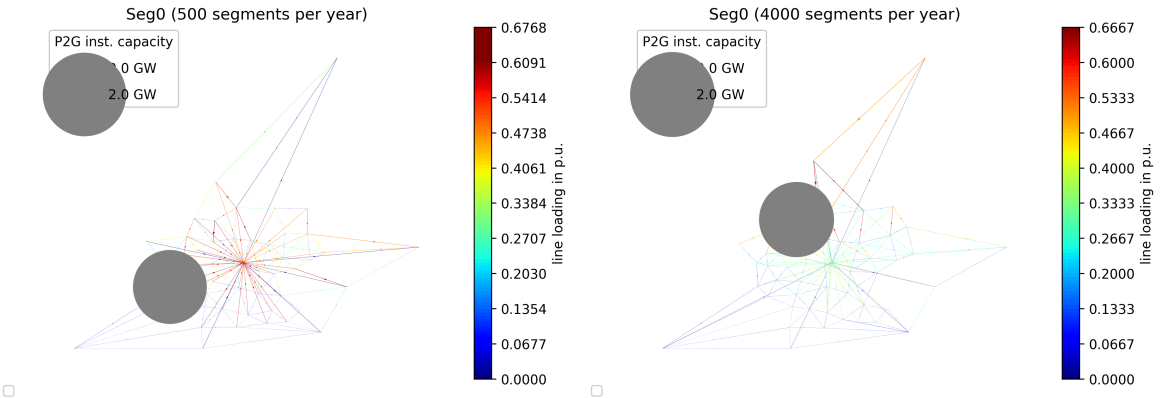

Figure C.2: Spatial electrolyzer expansion [GW_{el}] in $Seg0_1$, clustered to 500 (left), 4000 (right) segm/y and line loading [pu].

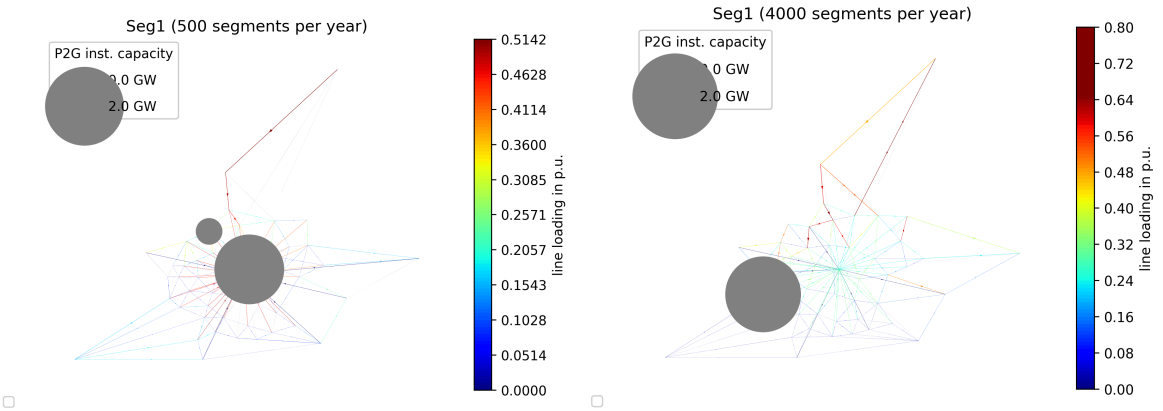

Figure C.3: Spatial electrolyzer expansion [GW_{el}] in $Seg1_1$, clustered to 500 (left), 4000 (right) segm/y and line loading [pu].

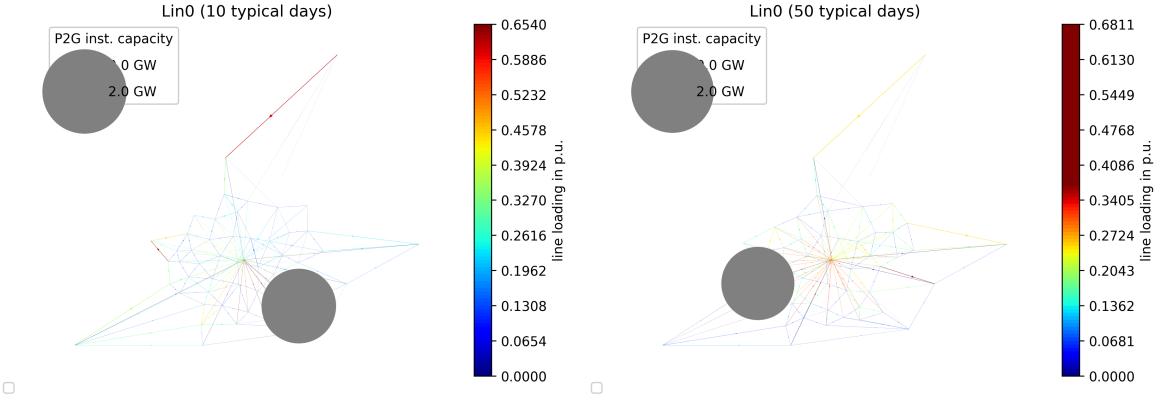

Figure C.4: Spatial electrolyzer expansion [GW_{el}] $Lin0_1$, clustered to 10 (left), 50 (right) tday/y and line loading [pu].

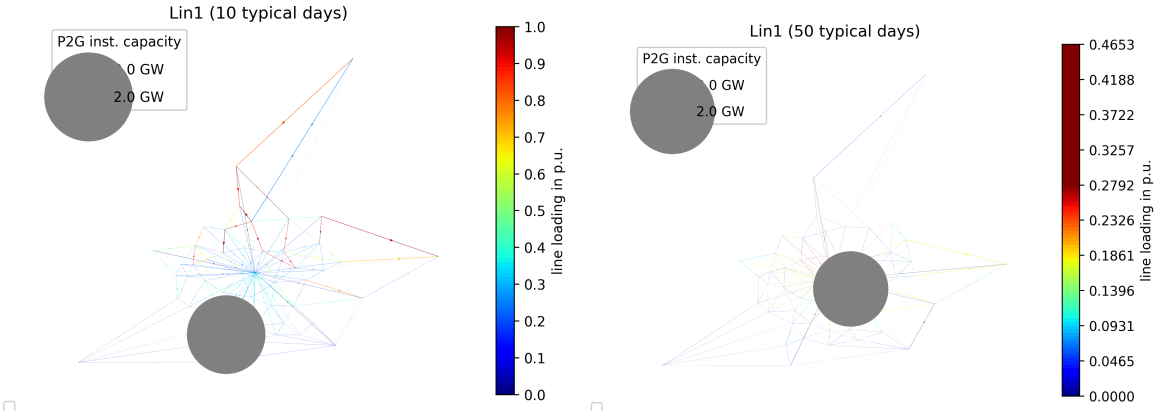

Figure C.5: Spatial electrolyzer expansion [GW_{el}] $Lin1_1$, clustered to 10 (left), 50 (right) tday/y and line loading [pu].

Bibliography

[1] Agentur für erneuerbare Energien, "Grafiken: Sektorenkopplung," Sep. 2020. [Online]. Available: https://www.unendlich-viel-energie.de/mediathek/grafiken/sektorenkopplung2019

[2] U. P. Müller, B. Schachler, W.-D. Bunke, J. Bartels, M. Glauer, C. Büttner, S. Günther, E. Kötter, I. Cußmann, L. Hülk, M. Scharf, T. Mossakowski, and J. Wendiggensen, "open_ego - Netzebenenübergreifendes Planungsinstrument - zur Bestimmung des optimalen Netz- und Speicherausbaus in Deutschland - integriert in einer OpenEnergyPlatform," Flensburg, Apr. 2019.

[3] N. Neuland, "Räumlich und zeitlich aufgelöste Potentialanalyse und Optimierung der Allokation von Power-to-Gas-Anlagen in Deutschland [unpublished]," Aachen, Jan. 2021.

[4] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten, "Time series aggregation for energy system design: Modeling seasonal storage," *Applied Energy*, vol. 213, pp. 123–135, Mar. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261918300242

[5] T. Brown, D. Schlachtberger, A. Kies, S. Schramm, and M. Greiner, "Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system," *Energy*, vol. 160, pp. 720–739, Oct. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S036054421831288X

[6] 50Hertz Transmission GmbH, Amprion GmbH, TenneT TSO GmbH, and TransnetBW GmbH, "Netzentwicklungsplan 2035, Version 2021 – Erster Entwurf der Übertragungsnetzbetreiber," 2020. [Online]. Available: https://www.netzentwicklungsplan.de/de/netzentwicklungsplaene/netzentwicklungsplan-2035-2021

[7] S. Pfenninger, A. Hawkes, and J. Keirstead, "Energy systems modeling for twenty-first century energy challenges," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 74–86, May 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364032114000872

[8] T. Brown, J. Hörsch, and D. Schlachtberger, "PyPSA: Python for Power System Analysis," *Journal of Open Research Software*, vol. 6, no. 4, 2018. [Online]. Available: https://doi.org/10.5334/jors.188

[9] A. Schröder, F. Kunz, J. Meiss, R. Mendelevitch, and C. Von Hirschhausen, "Current and prospective costs of electricity generation until 2050. Data Documentation," 2013. [Online]. Available: https://www.diw.de/documents/publikationen/73/diw_01.c.424566.de/diw_datadoc_2013-068.pdf

[10] 50Hertz Transmission GmbH, Amprion GmbH, TenneT TSO GmbH, and TransnetBW GmbH, "2015 Netzentwicklungsplan Strom 2025, Version 2015 – Erster Entwurf der Übertragungsnetzbetreiber," 2014. [Online]. Available: https://www.netzentwicklungsplan.de/sites/default/files/paragraphs-files/NEP_2025_1_Entwurf_Teil1_0.pdf

[11] United Nations, "The Paris Agreement," 2015. [Online]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

[12] M. Hoffmann, L. Kotzur, D. Stolten, and M. Robinius, "A Review on Time Series Aggregation Methods for Energy System Models," *Energies*, vol. 13, no. 3, p. 641, Feb. 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/3/641

[13] WMO, *WMO statement on the state of the global climate in 2019*, 2020, OCLC: 1153984970. [Online]. Available: https://library.wmo.int/doc_num.php?explnum_id=10211

[14] Umweltbundesamt, "Erneuerbare Energien in Deutschland. Daten zur Entwicklung im Jahr 2019," Dessau-Roßlau, Mar. 2020. [Online]. Available: https://www.umweltbundesamt.de/publikationen/erneuerbare-energien-in-deutschland-2019

[15] Bundesministerium für Wirtschaft und Energie, "Integrierter Nationaler Energie- und Klimaplan," Berlin, 2020. [Online]. Available: https://www.bundesregierung.de/breg-de/themen/energiewende/energiewende-1758720

[16] BMUB, "Climate Action Plan 2050. Principles and goals of the German government's climate policy," Berlin, 2016. [Online]. Available: https://www.bmu.de/en/publication/climate-action-plan-2050/

[17] M. Robinius, A. Otto, P. Heuser, L. Welder, K. Syranidis, D. Ryberg, T. Grube, P. Markewitz, R. Peters, and D. Stolten, "Linking the Power and Transport Sectors—Part 1: The Principle of Sector Coupling," *Energies*, vol. 10, no. 7, p. 956, Jul. 2017. [Online]. Available: http://www.mdpi.com/1996-1073/10/7/956

[18] Bundesministerium für Wirtschaft und Energie, "Grünbuch Energieeffizienz – Diskussionspapier des Bundesministeriums für Wirtschaft und Energie," Berlin, 2017. [Online]. Available: https://www.bmwi.de/Redaktion/DE/Artikel/Energie/gruenbuch-energieeffizienz.html

[19] Deutscher Verein des Gas- und Wasserfaches, *Sektorenkopplung – Synergien sinnvoll nutzen*, 2020. [Online]. Available: https://www.dvgw.de/themen/energiewende/sektorenkopplung/

[20] Bundesnetzagentur, "Quartalsbericht Netz- und Systemsicherheit - Gesamtes Jahr 2019," Berlin, 2020. [Online]. Available: https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Berichte/2020/Quartalszahlen_Gesamtjahr_2019.pdf?__blob=publicationFile&v=9

[21] Fraunhofer-Institut für solare Energiesysteme ISE, "Energiesystem Deutschland 2050. Sektor- und Energieträgerübergreifende, modellbasierte, ganzheitliche Untersuchung zur langfristigen Reduktion energie- bedingter CO 2 -Emissionen durch Energieeffizienz und den Einsatz Erneuerbarer Energien," 2013. [Online]. Available: https://www.ise.fraunhofer.de/de/veroeffentlichungen/studien/studie-energiesystem-deutschland-2050.html

[22] Forschungsstelle für Energiewirtschaft e.V., "Merit Order der Energiespeicherung im Jahr 2030 – Executive Summary," 2016. [Online]. Available: https://www.ffe.de/themen-und/414-merit-order-der-energiespeicherung-im-jahr-2030

[23] Fraunhofer IWES/IBP, "Wärmewende 2030. Schlüsseltechnologien zur Erreichung der mittel- und langfristigen Klimaschutzziele im Gebäudesektor. Study on behalf of Agora Energiewende," 2017. [Online]. Available: https://www.agora-energiewende.de/veroeffentlichungen/waermewende-2030-1/

[24] S. McDonagh, S. Ahmed, C. Desmond, and J. D. Murphy, "Hydrogen from offshore wind: Investor perspective on the profitability of a hybrid system including for curtailment," *Applied Energy*, vol. 265, p. 114732, May 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261920302440

[25] A. Moser, M. Zdrallek, H. Krause, and F. Graf, "Nutzen von Smart-Grid-Konzepten unter Berücksichtigung der Power-to-Gas-Technologie. DVGW-Förderkennzeichen G3-03-12," Bonn, 2014. [Online]. Available: https://www.dvgw.de/medien/dvgw/forschung/berichte/g3_03_12.pdf

[26] M. Sterner and I. Stadler, *Energiespeicher: Bedarf, Technologien, Integration*. Berlin: Springer Vieweg, 2014.

[27] Deutscher Verein des Gas- und Wasserfaches, *Technische Regel – Arbeitsblatt DVGW G 262 (A). Nutzung von Gasen aus regenerativen Quellen in der öffentlichen Gasversorgung*, Sep. 2011.

[28] Bundesnetzagentur and Bundeskartellamt, "Monitoringbericht 2019," Bonn, 2019. [Online]. Available: https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/DatenaustauschundMonitoring/Monitoring/Monitoringberichte/Monitoring_Berichte_node.html

[29] Bundesverband der Energie- und Wasserwirtschaft, *Erdgasaufkommen und -verbrauch in Deutschland*, 2019. [Online]. Available: https://www.bdew.de/service/daten-und-grafiken/erdgasaufkommen-und-verbrauch-deutschland/

[30] R. Baldick, "Flexibility and Availability: Can the Natural Gas Supply Support These Needs? [In My View]," *IEEE Power and Energy Magazine*, vol. 12, no. 6, pp. 104–101, Nov. 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6928577/

[31] H. Neumann, "Faktencheck: Ist das Gasnetz als Speicher geeignet?" Aug. 2018. [Online]. Available: https://www.topagrar.com/energie/news/faktencheck-ist-das-gasnetz-als-speicher-geeignet-9835668.html

[32] VNG AG, "Erdgas kann mehr. Wir auch. Gasinfrastruktur." [Online]. Available: https://vng.de/de/erdgas-kann-mehr-wir-auch/gasinfrastruktur

[33] T. Wencker, "Dekarbonisierungsstrategie im Erdgasnetz," Magdeburg, Oct. 2019. [Online]. Available: https://asue.de/sites/default/files/asue/themen/bio-erdgas/2019/vortraege/ASUE_Thomas-Wencker_Dekarbonisierungsstrategie-im-Erdgasnetz_191009.pdf

[34] G. Volk, "Die Herausforderung an die Bundesnetzagentur die Energiewende zu meistern," Berlin, Dec. 2012. [Online]. Available: file:///C:/Users/viol_is/AppData/Local/Temp/VortragVolkBNA.pdf

[35] N. Hartmann, L. Eltrop, N. Bauer, J. Salzer, S. Schwarz, and M. Schmidt, "Stromspeicherpotenziale für Deutschland," Jul. 2012. [Online]. Available: http://www.zfes.uni-stuttgart.de/deutsch/downloads/20120727_Final_Stromspeicherpotenziale_fuer_Deutschland-.pdf

[36] M. Götz, F. Ortloff, S. Bajohr, and F. Graf, "Speicherung von regenerativ erzeugter elektrischer Energie in der Erdgasinfrastruktur," *gwf-Gas*, vol. 04/2011, Apr. 2011.

[37] Bundesverband der Energie- und Wasserwirtschaft, "Gasnetzentwicklung in Deutschland," 2020. [Online]. Available: https://www.bdew.de/media/documents/Erdgasnetzlaengen_D_Entw_10J_o_online_jaehrlich_Ki_13032020.pdf

[38] F. Kunz, M. Kendziorski, W.-P. Schill, J. Weibezahn, J. Zepter, C. von Hirschhausen, P. Hauser, M. Zech, D. Möst, S. Heidari, B. Felten, and C. Weber, "Data Documentation 92. Electricity, Heat, and Gas Sector Data for Modeling the German System," 2017. [Online]. Available: https://www.diw.de/documents/publikationen/73/diw_01.c.574130.de/diw_datadoc_2017-092.pdf

[39] Deutsche Energie-Agentur GmbH (dena), "dena-Leitstudie Integrierte EnergiewendeImpulse für die Gestaltung des Energiesystems bis 2050," Berlin, Jul. 2018. [Online]. Available: https://www.dena.de/fileadmin/dena/Dokumente/Pdf/9261_dena-Leitstudie_Integrierte_Energiewende_lang.pdf

[40] V. Tietze and D. Stolten, "Thermodynamics of Pressurized Gas Storage," in *Hydrogen Science and Engineering : Materials, Processes, Systems and Technology*, P. D. D. Stolten and D. B. Emonts, Eds. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, Apr. 2016, pp. 601–628. [Online]. Available: http://doi.wiley.com/10.1002/9783527674268.ch25

[41] K. Langeheinecke, P. Jany, and G. Thieleke, *Thermodynamik für Ingenieure: ein Lehr- und Arbeitsbuch für das Studium ; mit 57 Tabellen*, 7th ed., ser. Studium. Wiesbaden: Vieweg + Teubner, 2008, oCLC: 229449902.

[42] A. Schmidt, "Combustion Processes," in *Technical Thermodynamics for Engineers*. Cham: Springer International Publishing, 2019, pp. 655–731. [Online]. Available: http://link.springer.com/10.1007/978-3-030-20397-9_23

[43] statista, "Erdgasverbrauch in Deutschland in den Jahren von 1980 bis 2019 (in Milliarden Kubikmeter)," 2020. [Online]. Available: https://de.statista.com/statistik/daten/studie/41033/umfrage/deutschland---erdgasverbrauch-in-milliarden-kubikmeter/

[44] M. Fasihi, O. Efimova, and C. Breyer, "Techno-economic assessment of CO2 direct air capture plants," *Journal of Cleaner Production*, vol. 224, pp. 957–980, Jul. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0959652619307772

[45] J. Haumaier, P. Hauser, H. Hobbie, and D. Möst, "Grünes Gas für die Gaswirtschaft – Regionale Power-to-Gas-Potentiale aus Onshore-Windenergie in Deutschland," *Zeitschrift*

für Energiewirtschaft, vol. 44, no. 2, pp. 61–83, Jun. 2020. [Online]. Available: http://link.springer.com/10.1007/s12398-020-00274-w

[46] M. Götz, J. Lefebvre, F. Mörs, A. McDaniel Koch, F. Graf, S. Bajohr, R. Reimert, and T. Kolb, "Renewable Power-to-Gas: A technological and economic review," *Renewable Energy*, vol. 85, pp. 1371–1390, Jan. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148115301610

[47] G. Glenk and S. Reichelstein, "Economics of converting renewable power to hydrogen," *Nature Energy*, vol. 4, no. 3, pp. 216–222, Mar. 2019. [Online]. Available: http://www.nature.com/articles/s41560-019-0326-1

[48] Bundesministerium für Wirtschaft und Energie, "The National Hydrogen Strategy," Jun. 2020. [Online]. Available: https://www.bmbf.de/de/nationale-wasserstoffstrategie-9916.html

[49] H.-K. Ringkjøb, P. M. Haugan, and I. M. Solbrekke, "A review of modelling tools for energy and electricity systems with large shares of variable renewables," *Renewable and Sustainable Energy Reviews*, vol. 96, pp. 440–459, Nov. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364032118305690

[50] I. Cußmann, "Projektvorstellung_egon," Flensburg, Sep. 2020. [Online]. Available: https://ego-n.org/presentations/Projektvorstellung_eGon.pdf

[51] J. Hüttenrauch, G. Müller-Syring, H. Krause, W. Fichtner, C. Nolden, P. Hauser, T. Müller, D. Möst, P. Härtel, N. Gerhardt, M. Sterner, M. Thema, P. Markewitz, B. Gillessen, H. Heinrichs, K. Münch, T. Bongers, I. Marjanovic, J. Müller-Kirchenbauer, S. Hotopp, F. Merten, M. Buddeke, P. Schaube, A. Taubitz, B. Dresen, D. Möllenbrink, H.-J. Wagner, W. Köppel, K. Bär, C. Degünther, and J. Benthin, "Integration fluktuierender erneuerbarer energien durch konvergente nutzung von strom- und gasnetzen - konvergenz strom- und gasnetze (konstgas) : Abschlussbericht," DBI - Gastechnologisches Inst., Freiberg, Tech. Rep., 2017. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:swb:90-766306

[52] European Network of Transmission System Operators for Electricity (ENSTO-E), "Scenario Outlook & Adequacy Forecast 2014-2030," ENTSO-E, Brussels, Tech. Rep., 2018. [Online]. Available: https://docs.entsoe.eu/dataset/scenario-outlook-adequacy-forecast-2014-2030

[53] D. Peters, W. Heitkoetter, R. Völker, A. Möller, T. Gross, B. Petters, F. Schuldt, and K. v. Maydell, "Validation of an open source high voltage grid model for AC load flow calculations in a delimited region," *IET Generation, Transmission & Distribution*, May 2020. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-gtd.2020.0107

[54] O. Raventós and J. Bartels, "Evaluation of Temporal Complexity Reduction Techniques Applied to Storage Expansion Planning in Power System Models," *Energies*, vol. 13, no. 4, p. 988, Feb. 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/4/988

[55] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume*

1: Statistics. Berkeley, Calif.: University of California Press, 1967, pp. 281–297. [Online]. Available: https://projecteuclid.org/euclid.bsmsp/1200512992

[56] P. Nahmmacher, E. Schmid, L. Hirth, and B. Knopf, "Carpe diem: A novel approach to select representative days for long-term power system modeling," *Energy*, vol. 112, pp. 430–442, Oct. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0360544216308556

[57] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*, ser. Prentice Hall advanced reference series. Englewood Cliffs, N.J: Prentice Hall, 1988.

[58] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten, "Impact of different time series aggregation methods on optimal energy system design," *Renewable Energy*, vol. 117, pp. 474–487, Mar. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148117309783

[59] A. Singh, A. Yadav, and A. Rana, "K-means with Three different Distance Metrics," *International Journal of Computer Applications*, vol. 67, no. 10, pp. 13–17, Apr. 2013. [Online]. Available: http://research.ijcaonline.org/volume67/number10/pxc3886785.pdf

[60] J. H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, Mar. 1963. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845

[61] R. Green, I. Staffell, and N. Vasilakos, "Divide and conquer? k-means clustering of demand data allows rapid and accurate simulations of the british electricity system," *IEEE Transactions on Engineering Management*, vol. 61, pp. 251–260, 05 2014.

[62] N. Baumgärtner, F. Temme, B. Bahl, M. Hennen, D. Hollermann, and A. Bardow, "RiSES4 Rigorous Synthesis of Energy Supply Systems with Seasonal Storage by Relaxation and Time-Series Aggregation to Typical Periods," Wroclaw, Poland, Jun. 2019.

[63] S. Pineda and J. M. Morales, "Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7162–7170, Nov. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8369128/

[64] S. Pfenninger, "Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability," *Applied Energy*, vol. 197, pp. 1–13, Jul. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261917302775

[65] S. Buchholz, M. Gamst, and D. Pisinger, "A comparative study of time aggregation techniques in relation to power capacity expansion modeling," *TOP*, vol. 27, no. 3, pp. 353–405, Oct. 2019. [Online]. Available: http://link.springer.com/10.1007/s11750-019-00519-z

[66] tsam Developer Team, "The Segmentation Functionality." [Online]. Available: https://tsam.readthedocs.io/en/latest/segmentationDoc.html

[67] C. Büttner, "Technisch-ökonomische Optimierung des Übertragungsnetzausbaus in Deutschland unter Berücksichtigung von (n-1)-Ausfallanalysen," 2020.

[68] (2020) Gurobi Optimization. Documentation. [Online]. Available: https://www.gurobi.com/documentation/