A Lagrangian relaxation approach to an electricity system investment model with a high temporal resolution

OR Spectrum (2023) 45:1263–1294
https://doi.org/10.1007/s00291-023-00736-w

ORIGINAL ARTICLE

A Lagrangian relaxation approach to an electricity system investment model with a high temporal resolution

Caroline Granfeldt[1] · **Ann-Brith Strömberg**[1] · **Lisa Göransson**[2]

Received: 23 January 2023 / Accepted: 6 October 2023 / Published online: 9 November 2023

Abstract

The global production of electricity contributes significantly to the release of carbon dioxide emissions. Therefore, a transformation of the electricity system is of vital importance in order to restrict global warming. This paper proposes a modelling methodology for electricity systems with a large share of variable renewable electricity generation, such as wind and solar power. The model developed addresses the capacity expansion problem, i.e. identifying optimal long-term investments in the electricity system. Optimal investments are defined by minimum investment and production costs under electricity production constraints—having different spatial resolutions and technical detail—while meeting the electricity demand. Our model is able to capture a range of strategies to manage variations and to facilitate the integration of variable renewable electricity; it is very large due to the high temporal resolution required to capture the variations in wind and solar power production and the chronological time representation needed to model energy storage. Moreover, the model can be further extended—making it even larger—to capture a large geographical scope, accounting for the trade of electricity between regions with different conditions for wind and solar power. Models of this nature thus typically need to be solved using some decomposition method to reduce solution times. In this paper, we develop a decomposition method using so-called variable splitting and Lagrangian relaxation; the dual problem is solved by a deflected subgradient algorithm. Our decomposition regards the temporal resolution by defining 2-week periods throughout the year and relaxing the overlapping constraints. The method is tested and evaluated on some real-world cases containing regions with different energy mixes and conditions for wind power. Numerical results show shorter computation times as compared with the non-decomposed model and capacity investment options similar to the optimal solution provided by the latter model.

Keywords Variable renewable electricity · Variation management · Electricity system modelling · Long-term investment models · Cost optimization · Wind power integration · Solar power integration · Lagrangian relaxation · Variable splitting

Extended author information available on the last page of the article

Mathematics Subject Classification 90B90 · 90-08 · 90C06 · 90C08 · 90C59 · 90C90

1 Introduction

EU's roadmap 2050 establishes that the greenhouse gas emissions must decrease by some 85% until the year 2050 in order for global warming to be restricted to 2°C (COM 2011). The electricity system contributes significantly to the emissions of carbon dioxide, both in the EU and globally. A transformation of the electricity system is therefore needed, and electricity investment models can be used as a tool to make informed decisions regarding future electricity generation, storage, and transmission capacity. The mathematical optimization models describing the electricity system minimizes the investment and production costs of the system under electricity production constraints, while meeting the electricity demand. The existing systems mostly consist of thermal power (IEA 2020), and thus the traditional models are designed with this in mind. The characterization of such a system, dominated by dispatchable generation, includes the ability to regulate the electricity production to meet instant demand.

The vast majority of scenarios for the European electricity system complying with a 2°C warming target contain, however, large shares of variable renewable technologies, e.g. wind and solar power. These electricity generation technologies are characterized as being non-dispatchable, meaning that the electricity generation cannot be completely controlled but depend on weather and climatic conditions. Hence, models applied for investigating electricity systems with potentially large shares of wind and solar power need an ability to capture the variability in production and demand. Intuitively, variability can be accounted for by increasing the temporal resolution. Electricity investment models are, however, often computationally demanding due to their technological and/or spatial scope and level of detail.

To reduce computation time and computer memory requirements, previous work on electricity system investment models has typically focused on simplifying the time representation. Ringkjøb et al. (2018) reviewed 75 modelling tools used for analysing energy and electricity systems with large shares of variable renewable electricity (VRE); one remaining challenge identified regards how to represent short-term variability in long-term studies. A methodological review of strategies for integrating short-term variations is given by Collins et al. (2017); the authors discuss methods for improving the time representation in long-term electricity system investment models that employ traditional representations of time. Pfenninger et al. (2014) review several articles that discuss time representation for energy system models, which contain a substantial level of VRE.

Traditional time representation methods for electricity system investment models typically belong to a family of methods denoted as *time slices*. Integral time slices can for example be a single time slice per year or a small set of seasonal and daily time slices to represent the differences in demand dependent on season, weekday, or time of day. Time slicing methods that are based on an approximation of the joint probability distribution of the load and VRE generation are developed by, e.g. Wogrin et al.

(2014) and Lehtveer et al. (2017). Another time slicing method is the *representative days* method, suggested by Nahmmacher et al. (2016), which identifies a number of 24-hour segments based on load and VRE patterns over a day. Time reduction methods based on these principles have been implemented and shown promising results for long-term investment models; see, e.g. Mai et al. (2013); Gils (2016); Gerbaulet and Lorenz (2017), and Frew and Jacobson (2016); the methods have been compared and evaluated in Reichenberg et al. (2018).

The integral time slicing methods have, however, not worked when considering a larger geographical scope, including, e.g. regional trade. The reason for this is that approximating the joint probability distribution is challenging since, unlike variations in load, variations in VRE generation do not follow a common pattern across a wide geographical scope. Thus, the integral time slicing methods cannot properly account for wind and solar variations in models where a large geographical scope is considered. Furthermore, smoothing effects through trade is an important variation management strategy for VRE and thus, as is also concluded in Reichenberg et al. (2018), the integral time slicing method is not ideal for a multi-node electricity system model with large shares of VRE.

The representative days approach, on the other hand, can be employed in network models and therefore incorporate trade (see Frew et al. (2016)). This time representation can also handle short-term storage, but as the representative days typically consist of diurnal slices, it does not account for storage between days, which requires interconnected time steps. An alternative is to model over longer time periods, e.g. weeks, but this increases model complexity and thus computation time. Hence, simplification in the spatial or technological system detail dimensions might be necessary to compensate for the increased complexity.

This study aims to develop a methodology to decompose and solve long-term electricity system investment models by combining Lagrangian relaxation and variable splitting (Guignard and Kim 1987; Jörnsten and Näsberg 1986). A key objective is to examine how a high temporal resolution (i.e. 3 h) impacts the solution times using these decomposition methods, and—when possible—compare the solutions and solution times of the decomposed models with those of the non-decomposed model. The model investigated in this study considers only isolated regions and does not include the possibility to trade electricity with neighbouring regions (and thus over-invests in capacity to meet demand); the suggested solution method can be used for such models as well.

The paper outline is as follows. Section 2 defines the electricity system investment model, while Sect. 3 presents the mathematical methods to solve the model. Section 4 discusses how the model is implemented and solved using the introduced methods, and the results are evaluated in Sect. 5. Lastly, Sect. 6 concludes the main contributions of the paper.

2 Model

The problem consists of minimizing investment and operational costs while meeting the demand for electricity in a European electricity system. Europe is here divided into several regions, chosen according to country borders and, when existent, infrastructural bottlenecks within the countries.

Electricity is produced by different electricity generation technologies, e.g. coal, hydro-, wind, and solar power. The production capacity, measured in GW, determines an upper limit for the amount of electricity that can be produced at any instant. We assume an aggregated capacity within each region, instead of studying separate power plants. This reduces significantly the problem size, albeit losing some system detail. Previous work by Göransson (2014) shows, however, that the loss is marginal for the total system cost as well as for the average full load hours of each electricity production type, including wind power. Furthermore, we study a time period ranging from 2020 to 2050 and consider both existing production capacity as well as new investments. The total time period is divided into several investment periods to account for the life span of different power plants, i.e. the production capacity life span.

This work considers isolated regions, meaning that trade along an electricity grid is not possible. Hence, each region produces all electricity (measured in GWh/h) to meet its electricity demand.

Thermal cycling is included, as previous work has shown that it has a substantial impact on the cost-optimal electricity system composition (Göransson et al. 2017). Furthermore, to retain a linear model, thermal cycling is accounted for using a relaxed unit commitment approach as described by Weber (2005). The cost and technical limitations of cycling thermal generation is included through a cost for starting thermal generation capacity, a minimum down time, and a cost of operating it at part load. Started capacity can operate in the range between minimum load level and rated power. With a time resolution of 3 h, which we have used in our evaluation in Sect. 5, it is assumed to be technically possible to change the operational level within this range for all thermal generation. Thus, ramping limits are implicitly included for thermal power by limiting the possible start-up capacity in the constraints (3)–(5). However, for shorter time steps, explicit constraints similar to the constraints (10) may be necessary to include in order to properly capture the underlying unit commitment problem. For renewable electricity production, constraints regarding production and capacity limits from weather and terrain are included. Hydropower is constrained by ramp rate, i.e. the maximum rate at which the electricity output can increase or decrease and balance constraints for electricity storage. Lastly, the total system carbon dioxide emissions should be accounted for. This can be represented either by a hard constraint, or—as in our implementation—by including a cost (penalizing carbon-emitting technologies) in the objective.

2.1 Electricity system investment model

The problem is represented with a linear optimization model. The sets employed by the model are given in Table 1. The set $\mathcal{P}$ represents all the electricity production technologies, e.g. hydropower, wind power, and nuclear power. This set contains thermal power technologies, $\mathcal{P}_{\text{thermal}}$, such as nuclear power and waste incineration plants. It also contains renewable electricity generation technologies, and in this model, namely $\mathcal{P}_{\text{wind}}$, $\mathcal{P}_{\text{solar}}$, and $\mathcal{P}_{\text{hydro}}$. The set of renewables are then defined as $\mathcal{P}_{\text{ren}} := \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}} \cup \mathcal{P}_{\text{hydro}}$. The modelling years are given by $\mathcal{S}$, and the set of

time steps within a year is denoted $\mathcal{T}$. Since thermal cycling is considered in this model, $\mathcal{T}_{\text{start}}(p)$, is the set of hours in the start-up interval for technology $p \in \mathcal{P}$. The model uses the concept of *investment periods*, denoted $\mathcal{I}$, which are necessary in order to know during which year an investment in production capacity was made. This is relevant for two reasons: Firstly, the model covers several years. Hence, if an investment in capacity of technology type p is made during year s, it should not be possible to use that invested capacity prior to the year s. Secondly, since each production technology type $p \in \mathcal{P}$ has a specific life span U_p, it is crucial to know when the investment was made in order to estimate its remaining life. Therefore, the set $\mathcal{I}_{\text{active}}(s, p)$ is defined to contain the investment periods for technology type p that are still active at year $s \in \mathcal{S}$. Note that $\mathcal{S} \subset \mathcal{I}$.

A full nomenclature list is given in Appendix A. The mathematical constraints and objective for the problem are described below. All the decision variables in the model are nonnegative.

2.1.1 Meeting the demand

We begin by introducing the decision variables

$$x_{pist} = \text{ generated electricity [GWh/h] of technology type } p \in \mathcal{P} \text{ in year } s \in \mathcal{S} \text{ for investment period } i \in \mathcal{I}_{\text{active}}(s, p) \text{ and time step } t \in \mathcal{T}.$$

Letting d_{st} denote the demand at time step $t \in \mathcal{T}$ in year $s \in \mathcal{S}$, the constraints

$$\sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} x_{pist} \geq d_{st}, \qquad s \in \mathcal{S},\ t \in \mathcal{T}, \tag{1}$$

imply that the produced electricity meets the demand during all modelled time steps and years.

Table 1 The index sets used in the model

Symbol		Representation	Member
$\mathcal{P}$	$:= \mathcal{P}_{\text{thermal}} \cup \mathcal{P}_{\text{ren}}$	Electricity generation technologies	p
$\mathcal{P}_{\text{thermal}}$		Thermal power technologies	p
$\mathcal{P}_{\text{ren}}$	$:= \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}} \cup \mathcal{P}_{\text{hydro}}$	Renewable technologies	p
$\mathcal{P}_{\text{wind}}$		Wind technologies	p
$\mathcal{P}_{\text{solar}}$		Solar technologies	p
$\mathcal{P}_{\text{hydro}}$		Hydropower technologies	p
$\mathcal{I}$	$:= \{1960, 1970, \ldots, 2050\}$	Investment years, defining investment periods	i
$\mathcal{S}$	$:= \{2020, 2030, \ldots, 2050\}$	New capacity investment years; $\mathcal{S} \subset \mathcal{I}$	s
$\mathcal{I}_{\text{active}}(s, p)$	$:= \mathcal{I} \cap \{s - U_p, \ldots, s\}$	Investment periods for each technology type $p \in \mathcal{P}$ with life span U_p that is active at year $s \in \mathcal{S}$	i
$\mathcal{T}$	$:= \{1, \ldots, T\}$	Time steps within a year	t
$\mathcal{T}_{\text{start}}(p)$	$\subset \mathcal{T} \cup \{0\}$	Hours in the start-up interval for technology $p \in \mathcal{P}$	t

2.1.2 Generation limits

To represent the capacity installed in the system, we define the decision variables y_{pi} = installed capacity [GW] of technology type $p \in \mathcal{P}$ in investment period $i \in \mathcal{I}$. Letting b_{pi}^{gen} denote the existing production capacity of technology type $p \in \mathcal{P}$ in investment period $i \in \mathcal{I} \setminus \mathcal{S}$, the variables y_{pi} are fixed such that no new investments can be made prior to "now", as

$$y_{pi} = b_{pi}^{\text{gen}}, \qquad p \in \mathcal{P},\ i \in \mathcal{I} \setminus \mathcal{S}. \tag{2}$$

2.1.3 Thermal cycling

The property implying thermal cycling is that for thermal power plants, the capacity that has been taken out of operation has a minimum downtime that corresponds to the time it takes to start up the capacity before it can generate electricity once again. More importantly, the start-up cost of a unit is typically high, which property can be captured by thermal cycling constraints. To model this, we define two sets of decision variables:

z_{pist} = available hot capacity [GWh/h] of technology type $p \in \mathcal{P}_{\text{thermal}}$ in year $s \in \mathcal{S}$ for investment period $i \in \mathcal{I}_{\text{active}}(s,p)$ and time step $t \in \mathcal{T}$;

z^{+}_{pist} = started hot capacity [GWh/h] of technology type $p \in \mathcal{P}_{\text{thermal}}$ in year $s \in \mathcal{S}$ for investment period $i \in \mathcal{I}_{\text{active}}(s,p)$ from time step $t-1 \in \mathcal{T}$ to time step $t \in \mathcal{T}$.

The capacity that is currently up and running in a thermal power plant is referred to as *hot* or *available capacity*. The electricity generation should never exceed the available capacity. Likewise, it is required to generate a minimum level of electricity depending on the available capacity in order for it to stay hot. Letting ϕ_p denote the percentage corresponding to the minimum load level, this yields the constraints

$$\phi_p \cdot z_{pist} \le x_{pist} \le z_{pist}, \qquad i \in \mathcal{I}_{\text{active}}(s,p),\ p \in \mathcal{P}_{\text{thermal}},\ s \in \mathcal{S},\ t \in \mathcal{T}. \tag{3}$$

To connect the started hot capacity to the available capacity, the following constraints are used:

$$z^{+}_{pist} \ge \begin{cases} z_{pist} - z_{p,i,s,t-1}, & t \in \mathcal{T} \setminus \{1\}, \\ z_{pist} - z_{pisT}, & t = 1, \end{cases} \qquad i \in \mathcal{I}_{\text{active}}(s,p),\ p \in \mathcal{P}_{\text{thermal}},\ s \in \mathcal{S}. \tag{4}$$

The difference in available capacity between the time steps t and $t-1$ then corresponds to the started hot capacity. Here, we connect the first time step $t = 1$ to the very last time step $t = T$ to ensure continuity in the available capacity and thereby avoid boundary effects. Since it is costly to start new capacity the variable z^{+}_{pist} is penalized in the objective function. Hence, z^{+}_{pist} equals zero whenever $z_{pist} \le z_{p,i,s,t-1}$ holds. Moreover, $t_p \in \mathcal{T}_{\text{start}}(p)$, $p \in \mathcal{P}_{\text{thermal}}$, hours back in time, the started hot

capacity is limited by the available hot capacity z_{pist}. This yields, for $i \in \mathcal{I}_{\text{active}}(s,p)$, $p \in \mathcal{P}_{\text{thermal}}$, $s \in \mathcal{S}$, and $t \in \mathcal{T}$, the constraints

$$\sum_{j \in \mathcal{I}_{\text{active}}(s,p)} y_{pj} - z^{+}_{pist} \geq \begin{cases} z_{p,i,s,t-t_p}, & t_p \in \mathcal{T}_{\text{start}}(p) \setminus \{t, \ldots, T\}, \\ z_{p,i,s,T+t-t_p}, & t_p \in \mathcal{T}_{\text{start}}(p) \setminus \{0, \ldots, t-1\}, \end{cases} \tag{5}$$

which linearize the start-up constraints that are more intuitively modelled as integers. The idea is that capacity that has been taken out of operation has a minimum downtime before it can be started again. As a simple example, consider a case where the minimum downtime equals 3, implying that the model needs to consider the hot capacity up to three time steps back. The amount of start-up capacity that is available equals the total installed capacity minus the capacity that has not been hot in any of the previous three time steps (since any other capacity is either still hot or limited by its current downtime). This can, for some specific values of p, i, s, and t, be expressed mathematically as

$$z^{+}_{pist} \leq \sum_{j \in \mathcal{I}_{\text{active}}(s,p)} y_{pj} - \max\{z_{p,i,s,t-1},\, z_{p,i,s,t-2},\, z_{p,i,s,t-3}\}.$$

See also Weber (2005) for further details.

2.1.4 Renewables

For each technology $p \in \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}}$, the upper production limit of wind and solar power due to weather and climate is modelled using the profile $\theta_{pt} \in [0,1]$, $t \in \mathcal{T}$, and the total installed capacity that is still active in year s, using the inequalities

$$x_{pist} \leq \theta_{pt} \sum_{j \in \mathcal{I}_{\text{active}}(s,p)} y_{pj}, \qquad i \in \mathcal{I}_{\text{active}}(s,p),\ p \in \mathcal{P}_{\text{wind}} \cup \mathcal{P}_{\text{solar}},\ s \in \mathcal{S},\ t \in \mathcal{T}. \tag{6}$$

Not all areas are suitable for installation of wind farms, since wind speed and terrain vary across the region. Hence, due to reasons regarding land exploitation, there is an upper limit, W_p on the possible investments in wind capacity for wind type $p \in \mathcal{P}_{\text{wind}}$, which is modelled as

$$\sum_{i \in \mathcal{I}} y_{pi} \leq W_p, \qquad p \in \mathcal{P}_{\text{wind}}. \tag{7}$$

Define the decision variables

$$w_{st} = \text{hydropower storage [GWh] in year } s \in \mathcal{S} \text{ at time step } t \in \mathcal{T}.$$

Letting g_t denote the inflow into the reservoirs at time step $t \in \mathcal{T}$ (assumed constant over the years) and τ [h] denote the length of the time steps, the hydropower balance is then modelled as

$$w_{st} + g_t - \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} \tau \cdot x_{pist} \geq \begin{cases} w_{s,t+1}, & t \in \mathcal{T} \setminus \{T\}, \\ w_{s1}, & t = T, \end{cases} \qquad p \in \mathcal{P}_{\text{hydro}},\ s \in \mathcal{S}. \tag{8}$$

In this model presentation, $\tau = 1$ for simplicity but still necessary to include for the dimension analysis. The upper limit for the hydropower storage, denoted H, is modelled as

$$w_{st} \leq H, \qquad s \in \mathcal{S},\ t \in \mathcal{T}. \tag{9}$$

The production level of hydropower cannot change too quickly, which is modelled by ramping rate constraints. Letting $\delta^{\text{inc}} > 0$ and $\delta^{\text{dec}} > 0$ denote the shares corresponding to maximum change speeds, upper limits on the rate of increase and decrease, respectively, of the storage levels for $i \in \mathcal{I}_{\text{active}}(s,p)$, $p \in \mathcal{P}_{\text{hydro}}$, and $s \in \mathcal{S}$, are implied by the constraints

$$(1 + \delta^{\text{inc}}) x_{pist} \geq \begin{cases} x_{p,i,s,t+1}, & t \in \mathcal{T} \setminus \{T\}, \\ x_{pis1}, & t = T; \end{cases} \tag{10a}$$

$$(1 - \delta^{\text{dec}}) x_{pist} \leq \begin{cases} x_{p,i,s,t+1}, & t \in \mathcal{T} \setminus \{T\}, \\ x_{pis1}, & t = T. \end{cases} \tag{10b}$$

Lastly, no new investments in hydropower capacity are allowed. Thus,

$$y_{ps} = 0, \qquad p \in \mathcal{P}_{\text{hydro}},\ s \in \mathcal{S}. \tag{11}$$

2.1.5 Emissions

Emissions arise from running the power plants (e.g. fuel), but also from start-ups of plants since fuel is needed for this. Furthermore, there are extra emissions when not running on full capacity, due to reduced efficiency. Let e_{pi} denote the emissions released [CO_2/(GWh/h)] by technology type $p \in \mathcal{P}$ in investment period $i \in \mathcal{I}$. Let e^+_{pi} and $\tilde{e}_{pi}$ denote the emissions released from start-ups and from running on part-load, respectively, for technology type $p \in \mathcal{P}_{\text{thermal}}$ in investment period $i \in \mathcal{I}$. The total emissions for year $s \in \mathcal{S}$ and time step $t \in \mathcal{T}$ is then expressed as

$$e^{\text{tot}}_{st} := \sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} e_{pi} x_{pist} + \sum_{p \in \mathcal{P}_{\text{thermal}}} \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} \left(e^+_{pi} z^+_{pist} + \tilde{e}_{pi} (z_{pist} - x_{pist}) \right). \tag{12}$$

2.1.6 Objective

The objective is to minimize the total system costs. The sum (13a) considers the investment costs in electricity production technologies; c^{inv}_{ps} is the investment cost (including annuity costs) for technology type $p \in \mathcal{P}$ in year $s \in \mathcal{S}$, and c^{omf}_{p} is the fixed operation and maintenance costs for technology type $p \in \mathcal{P}$. The sum (13b) considers the costs of electricity production; c^{run}_{pi} are the running costs for tech-

nology type $p \in \mathcal{P}$ where the investment was made in period $i \in \mathcal{I}$. The sum (13c) describes the additional costs for thermal power technology types $p \in \mathcal{P}_{\text{thermal}}$; c^{+}_{ps} is the start-up cost for hot capacity in year $s \in \mathcal{S}$; $\tilde{c}_{ps}$ is the additional cost for running on part-load capacity in year $s \in \mathcal{S}$. The sum (13d) describes the costs for carbon dioxide emissions. Denoting the variables on vector form as $\boldsymbol{y}$, $\boldsymbol{x}$, z, z^{+}, and $\boldsymbol{e}^{\text{tot}}$, the objective function is defined as

$$C^{\text{tot}}(\boldsymbol{y}, \boldsymbol{x}, z, z^{+}, \boldsymbol{e}^{\text{tot}}) := \sum_{p \in \mathcal{P}} \sum_{s \in \mathcal{S}} (c^{\text{inv}}_{ps} + c^{\text{omf}}_{p}) y_{ps} \tag{13a}$$

$$+ \sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} c^{\text{run}}_{pi} x_{pist} \tag{13b}$$

$$+ \sum_{p \in \mathcal{P}_{\text{thermal}}} \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \left(c^{+}_{ps} z^{+}_{pist} + \tilde{c}_{ps} (z_{pist} - x_{pist}) \right) \tag{13c}$$

$$+ \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} c^{\text{CO}_2}_{s} e^{\text{tot}}_{st} . \tag{13d}$$

For the investment costs in (13a) only the investment periods that coincide with the years $s \in \mathcal{S}$ are considered, while no costs for prior investments are included. The costs in (13b)–(13c) cover all active investment periods.

3 Mathematical methodology

The above model is very large-scale, even for fairly small problem instances. In order to reduce computation times—or make the model even solvable for very large problem instances—special mathematical methods need to be utilized. Some decomposition methods relating to electrical energy applications are covered by Sagastizábal (2012); Göransson et al. (2021) develop a heuristic that divides an electricity system model into 26 subproblems consisting of 2-week segments. In this article, we adapt this heuristic approach to a mathematical method based on Lagrangian decomposition (Guignard 2003), for solving our model. Note that this entire section, besides Algorithm 2, uses a separate notation, which differs from that used in the rest of the article.

3.1 Lagrangian duality and Lagrangian relaxation

Many large optimization problems are structured such that they consist of several smaller problems, connected by some overlapping, typically complicating, constraints. Each separate problem is, however, often easily solved in comparison with the full problem. Lagrangian dual methods, such as Lagrangian relaxation, takes advantage of this problem structure; see Guignard (2003).

The idea of Lagrangian relaxation is to relax the overlapping constraints such that the remaining problem separates into several subproblems. Consider the convex optimization problem to

$$\begin{aligned} z^* := \text{minimum}_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x}, \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}^m, \\ & \boldsymbol{x} \in \boldsymbol{X}, \end{aligned} \tag{14}$$

where $\boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ such that each function g_i, $i = 1, \ldots, m$, is convex, $\boldsymbol{X} \subset \mathbb{R}^n$ is a convex set, and $m, n \in \mathbb{Z}_+$. We also assume that $\{\boldsymbol{x} \in \boldsymbol{X} \,|\, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}^m\} \neq \emptyset$ such that there exists a feasible solution. Here, $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}^m$ are the connecting constraints while the remaining constraints $\boldsymbol{x} \in \boldsymbol{X}$ are separable.

Define the Lagrangian function $L : \mathbb{R}^{m+n} \to \mathbb{R}$ such that $L(\boldsymbol{x}, \boldsymbol{\pi}) := \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\pi}^\top \boldsymbol{g}(\boldsymbol{x})$, where $\boldsymbol{\pi} \in \mathbb{R}^m$ are the Lagrangian multipliers. The Lagrangian dual problem is then defined as

$$h^* := \max_{\boldsymbol{\pi} \geq \boldsymbol{0}^m} \; h(\boldsymbol{\pi}), \tag{15}$$

where $h : \mathbb{R}^m \to \mathbb{R}$ denotes the concave Lagrangian dual function, defined as

$$h(\boldsymbol{\pi}) \; := \; \min_{\boldsymbol{x} \in \boldsymbol{X}} L(\boldsymbol{x}, \boldsymbol{\pi}) \; = \; \min_{\boldsymbol{x} \in \boldsymbol{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\pi}^\top \boldsymbol{g}(\boldsymbol{x}) \right\}. \tag{16}$$

For any $\boldsymbol{\pi} \geq \boldsymbol{0}^m$ the problem of minimizing the function L over its first argument $\boldsymbol{x}$ is referred to as the subproblem, which is separable due to the set $\boldsymbol{X}$ being a Cartesian product set.

By weak duality, $h(\boldsymbol{\pi}) \leq z^*$ holds whenever $\boldsymbol{\pi} \geq \boldsymbol{0}^m$. Hence, any feasible solution to the dual problem provides a lower bound on the optimal objective value of the original problem. Furthermore, any feasible solution $\bar{\boldsymbol{x}}$ to (14) provides an upper bound on z^*, since the inequality $z^* \leq \boldsymbol{c}^\top \bar{\boldsymbol{x}}$ holds. Moreover, since (14) is a convex optimization problem, strong duality implies that the equality $z^* = h^*$ holds.

For a fixed value of the Lagrangian multiplier $\boldsymbol{\pi}$, the remaining subproblems are typically easy to solve in comparison with the full problem. However, solving the dual problem requires the use of algorithms that (sequentially) update the values of the Lagrangian multipliers. Bundle methods are based on the idea to approximate the set of subgradients (i.e. subdifferential) of the objective function by gathering subgradients from previous iterations into a bundle. For comprehensive descriptions of bundle methods, see e.g. Lemaréchal et al. (1995) and Kiwiel (1990). The subgradient algorithm was developed by N.Z. Shor in 1962; see Shor (1991) for a full review of the early history of non-smooth optimization. This method is frequently applied to optimization problems, especially in the context of Lagrangian duality. Larsson et al. (1996) defined the conditional subgradient method, which generalizes the subgradient algorithm by projecting the subgradients themselves. An issue, however, with the classical subgradient algorithm is that it often stalls due to zigzagging (e.g. Larsson et al. (1996)). Thus, we instead consider the modified deflected subgradient (MDS) method by Belgacem and Amir (2018), which combines the modified gradient technique (MGT) by Camerini et al. (1975), and the average direction strategy (ADS) by Sherali and Ulular

(1989). In the MDS method, the step direction is a weighted average of a current subgradient and the previous step direction, which reduces the zigzagging behaviour. Consider the following definition.

Definition 1 A vector $\boldsymbol{\gamma} \in \mathbb{R}^m$ is a *subgradient* to the concave function h at $\bar{\boldsymbol{\pi}} \in \mathbb{R}^m$ if the inequality

$$h(\boldsymbol{\pi}) \leq h(\bar{\boldsymbol{\pi}}) + \boldsymbol{\gamma}^T(\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}) \tag{17}$$

holds for all $\boldsymbol{\pi} \in \mathbb{R}^m$. The set $\partial h(\bar{\boldsymbol{\pi}})$ of subgradients to h at $\bar{\boldsymbol{\pi}} \in \mathbb{R}^m$ is called the *subdifferential*.

Geometrically, a subgradient is the gradient of a supporting hyperplane to the hypograph of the function h at the point $(\bar{\boldsymbol{\pi}}, h(\bar{\boldsymbol{\pi}}))$ (see (Bazaraa et al. 2013, Sect. 3.2)). The MDS algorithm is provided in Algorithm 1. Here, we assume that $\boldsymbol{\pi} \in \Pi$, such that Π is the feasible set for the multipliers $\boldsymbol{\pi}$. In (15), $\Pi := \mathbb{R}^m_+$ but in our application additional dual constraints are implied (see the model (33) in Sect. 4). The operator Proj_Π refers to a Euclidean projection onto the set Π, which is modelled as a convex quadratic optimization problem; for our specific definition of the set Π this projection problem is efficiently solved using Lagrangian duality, the main computational complexity coming from calculating and comparing partial derivatives of the Lagrangian dual function. The step lengths α_k in Algorithm 1 are chosen according to some rule, which guarantees convergence. The Polyak step length rule (Polyak 1969) has seen much use in practice and is used in this work. It is defined as

$$\alpha_k := \frac{\theta_k \left(h^* - h(\boldsymbol{\pi}^k)\right)}{||\boldsymbol{\gamma}^k||^2}, \tag{18a}$$

under the conditions

$$0 < \epsilon_1 \leq \theta_k \leq 2 - \epsilon_2 < 2, \quad k = 0, 1, 2, \ldots. \tag{18b}$$

Here, θ_k acts as a scaling parameter for the step length, while the parameters ϵ_1 and ϵ_2 define positive limits for the scaling parameter.

The use of (18a) together with (18b) guarantees theoretical convergence to an optimal solution to the dual problem (15) (see Polyak (1969)). However, the dual optimal

1: Initiate $\boldsymbol{\pi}^1 \in \Pi$ and $h^1_{\text{best}} = h(\boldsymbol{\pi}^1)$, and define $\boldsymbol{d}^0 := \mathbf{0}^m$. Set $k := 1$.
2: Solve the subproblem: $\boldsymbol{x}(\boldsymbol{\pi}^k) \in \arg\min_{\boldsymbol{x} \in \boldsymbol{X}} L(\boldsymbol{x}, \boldsymbol{\pi}^k)$. $\boldsymbol{\gamma}^k := \boldsymbol{g}(\boldsymbol{x}(\boldsymbol{\pi}^k))$ is a subgradient to h at $\boldsymbol{\pi}^k$.
3: Calculate

$$\lambda_k := \max\left\{0, -\frac{(\boldsymbol{\gamma}^k)^\top \boldsymbol{d}^{k-1}}{\|\boldsymbol{\gamma}^k\| \, \|\boldsymbol{d}^{k-1}\|}\right\}, \qquad \Gamma^k_{\text{MGT}} := \max\left\{0, -\frac{(\boldsymbol{\gamma}^k)^\top \boldsymbol{d}^{k-1}}{(2-\lambda_k)\|\boldsymbol{d}^{k-1}\|^2}\right\}, \quad \Gamma^k_{\text{ADS}} := \frac{\|\boldsymbol{\gamma}^k\|}{\|\boldsymbol{d}^{k-1}\|},$$

$$\Gamma^k_{\text{MDS}} := (1-\lambda_k)\Gamma^k_{\text{MGT}} + \lambda_k \Gamma^k_{\text{ADS}}, \quad \text{and} \quad \boldsymbol{d}^k := \boldsymbol{\gamma}^k + \Gamma^k_{\text{MDS}} \boldsymbol{d}^{k-1}$$

4: For some $\alpha_k > 0$, the next iterate is calculated as $\boldsymbol{\pi}^{k+1} := \text{Proj}_\Pi\left\{\boldsymbol{\pi}^k + \alpha_k \boldsymbol{d}^k\right\}$
5: Update $h^{k+1}_{\text{best}} := \max\left\{h^k_{\text{best}}, h(\boldsymbol{\pi}^{k+1})\right\}$
6: If some termination criterion is fulfilled, then stop. Otherwise, let $k := k + 1$ and go to step 2

Algorithm 1 Modified deflected subgradient algorithm

value h^* is typically not known. If so, an upper bound $\bar{h} \geq h^*$ can be used instead in order to achieve finite convergence to an ϵ-optimal solution for some $\epsilon > \bar{h} - h^*$, where a point $\boldsymbol{\pi} \in \Pi$ is ϵ-optimal if $h(\boldsymbol{\pi}) \geq h^* - \epsilon, \epsilon > 0$; see (Polyak 1969, Theorem 4).

However, since the linear optimization model presented in Sect. 2 can be solved to optimality for our specific problem instances, by strong duality the dual optimal value h^* is available to us.

The value of the scaling parameter θ_k can be chosen in different ways. One method that in practice has been shown to give fast convergence to a close to optimal solution is presented in Caprara et al. (1999); the method is adaptive, such that the value of the parameter is updated every ℓ number of subgradient iterations:

$$\theta_{k+1} := \begin{cases} \frac{1}{2}\theta_k, & \text{if } \bar{h} - \underline{h} > 0.1 \cdot |\underline{h}|, \\ \frac{3}{2}\theta_k, & \text{if } \bar{h} - \underline{h} < 0.01 \cdot |\underline{h}|, \\ \theta_k, & \text{otherwise,} \end{cases} \qquad k = 1, 2, ..., \tag{19a}$$

where

$$\bar{h} := \max_{r=k-\ell+1,\ldots,k} h(\boldsymbol{\pi}^r) \qquad \text{and} \qquad \underline{h} := \min_{r=k-\ell+1,\ldots,k} h(\boldsymbol{\pi}^r). \tag{19b}$$

The values of ℓ and the numbers 0.1, and 0.01 should be modified depending on the results from the computations. Note that the updates (19) do not guarantee that the inequalities $0 < \theta_k < 2$ will hold for all iterations k.

3.2 Variable splitting

Relaxing over the time dimension is tricky since some variables are time-independent. For example, assume the year is divided into N time periods, such that $\mathcal{T}_n$ contains the time steps in time period $n \in \mathcal{N} := \{1, \ldots, N\}$. Define the variables $x_t \in X_t \subseteq \mathbb{R}_+$ and $y \geq 0$, representing the electricity generation in time step $t \in \mathcal{T}_n$, $n \in \mathcal{N}$, and the invested capacity, respectively. Then, consider the following simplified version of the model presented in Sect. 2:

$$\begin{aligned} \underset{y, x_t}{\text{minimize}} \quad & c^{\text{inv}} y + \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}_n} c_t^{\text{run}} x_t, \\ \text{subject to} \quad & x_t \leq y, && t \in \mathcal{T}_n,\ n \in \mathcal{N}, \\ & x_t \in X_t, && t \in \mathcal{T}_n,\ n \in \mathcal{N}, \\ & y \geq 0, \end{aligned} \tag{20}$$

where c^{inv} and c_t^{run}, $t \in \mathcal{T}_n$, $n \in \mathcal{N}$, are investment and run costs, respectively. Here, the time-independent variable y complicates the problem. Although the constraints $x_t \leq y$ can be relaxed, they represent an important property of the electricity system (one cannot produce more electricity than the installed capacity). Thus, they are very important for the model structure and should not be relaxed. Hence, we employ the concept of *variable splitting* (introduced by Jörnsten and Näsberg (1986)). The main idea is to introduce copies of certain primal variables (here, the variable y). Constraints are added to ensure consistency between the original variables and the copies, and then these consistency constraints are Lagrangian relaxed. This makes the problem separable, while the most important model structure is kept in each

subproblem; thus, the solution process can be parallelized. The variable splitting method was developed simultaneously by several research groups, thus also referred to as *Lagrangian decomposition* (Guignard and Kim 1987) and *variable layering* (Glover and Klingman 1988).

3.3 Generating primal solutions

In general, subgradient optimization methods often identify near-optimal dual solutions, but do not directly provide solutions to the primal problem. The conditional subgradient method constructs a sequence $\{\boldsymbol{x}(\boldsymbol{\pi}^k)\}$ of solutions to the Lagrangian subproblem, but these solutions are typically not feasible in the original primal problem since they do not have to satisfy the relaxed constraints. Thus, the sequence $\{\boldsymbol{x}(\boldsymbol{\pi}^k)\}$ does not converge to an optimal primal solution. Some method is therefore needed to remedy this. ADMM is a popular optimization algorithm, which relies on Lagrangian relaxation and the dual ascent method to converge to optimality and create primal feasible solutions (Boyd et al. 2011). In this work, we focus on using *ergodic sequences of subproblem solutions*.

As first presented by Larsson et al. (1999), ergodic sequences create approximations of primal solutions by averaging the subproblem solutions. The authors showed that the ergodic sequences in the limit produce optimal solutions to the original problem. An enhanced version in terms of convergence speed was introduced by Gustavsson et al. (2015). This version exploits more information from later subproblem solutions than from earlier ones. The ergodic sequence $\{\tilde{\boldsymbol{x}}_k\}$ is defined as

$$\tilde{\boldsymbol{x}}_k := \sum_{s=0}^{k-1} \mu_s^k \boldsymbol{x}(\boldsymbol{\pi}^s); \qquad \sum_{s=0}^{k-1} \mu_s^k = 1; \qquad \mu_s^k \geq 0,\ s = 0, ..., k-1, \tag{21}$$

where the convexity weights μ_s^k are chosen according to the s^n-rule:

$$\mu_s^k := \frac{(s+1)^n}{\sum_{r=0}^{k-1}(r+1)^n}, \qquad s = 0, ..., k-1,\ k = 1, 2, ...; \quad n \geq 0. \tag{22}$$

For $n > 0$, the s^n-rule yields a sequence in which later iterates are assigned higher weights than earlier ones. For increasing values of n, the weights are shifted towards later iterates. See also (Gustavsson et al. 2015, Def. 1).

It remains, however, to combine the ergodic sequences with the variable splitting approach. The idea used in this article is to calculate each subproblem's average investments over the subgradient iterations. This procedure converges in the limit for each subproblem, however, providing N (possibly) different solutions for the investments. Furthermore, these N solutions are most likely not feasible in the original full-scale model, since the subproblems possess different profiles for weather and demand. The heuristic in Algorithm 2 combines these solutions into a primal solution. The rationale behind the algorithm is that if enough many subproblems make a specific investment, that investment is likely necessary; a more expensive investment option requires a larger number of subproblems to be included in the solution, as compared to a cheaper investment option. Also, if exactly the same investment is

made in a large enough share of the subproblems, this investment is likely to be part of a near-optimal solution; it is thus used in the primal solution. Otherwise, we let the investment be the median of the nonzero subproblem solutions. However, our tests have shown that the heuristic tends to overestimate wind and solar power and underestimate intermediate production technology. Thus, the median is scaled by γ_p for some technologies p.

The presented algorithm does not consider the case when investments for the same technology are made by several subproblems but in different years. When this is the case, for each subproblem we find the maximum investment of that production technology over the years 2020–2050, which yields N different values in total. Then, the heuristic chooses the average of the nonzero options among these values. The heuristic does not guarantee primal feasibility, but in practice it seems to give near-feasible primal solutions. The values of the remaining variables (e.g. $\boldsymbol{x}$ and $\boldsymbol{z}$) are recomputed by letting the investment variables be fixed to their current values while solving the full model, which then simplifies to an easily solvable dispatch model. Note that the numbers used as limits here should be interpreted as a guideline and might require adjustments for different data instances.

4 Implementing Lagrangian relaxation and variable splitting

The idea is to separate the annual time steps of the full-scale model into M-week periods and solve them in parallel. Thus, some new sets and parameters are necessary:

$\mathcal{N} = \{1, \dots, N\}$ is the set of time periods within a year, with $M := \left\lfloor \frac{52}{N} \right\rfloor$;

$\mathcal{T}_n = \{(n-1)\Omega + 1, \dots, n\Omega\}$ is the set of time steps within each time period $n \in \mathcal{N}$, where $\Omega = \frac{24 \times 7 \times M}{\tau}$ is the length of a time period and τ is the time step length.

As an example, for $N = 26$ time periods and with the time step length $\tau = 1$, each period will contain $\Omega = 336$ time steps. Thus, $\mathcal{T}_1 = \{0 \cdot 336 + 1, \dots, 1 \cdot 336\} = \{1, \dots, 336\}$, $\mathcal{T}_2 = \{1 \cdot 336 + 1, \dots, 2 \cdot 336\} = \{337, \dots, 672\}$, $\dots$, $\mathcal{T}_{26} = \{25 \cdot 336 + 1, \dots, 26 \cdot 336\} = \{8401, \dots, 8736\}$.

Let Y_{psn} be the new investments generated by the ergodic sequences in subproblem $n \in \mathcal{N}$ for technology $p \in \mathcal{P}$, year $s \in \mathcal{S}$.
Define $\mathcal{R}_{ps} := \{n \in \mathcal{N} \mid Y_{psn} > 0\}$, $\alpha_{ps} := c_{ps}^{\text{inv}} / (1.5 \max_{r \in \mathcal{P}} \{c_{rs}^{\text{inv}}\})$, and $\lambda := 0.2|\mathcal{N}|$.
Let γ_p be a scaling factor to compensate for over- or underestimates.
for $p \in \mathcal{P}$, $s \in \mathcal{S}$ **do**
  if $|\mathcal{R}_{ps}| > \alpha_{ps}|\mathcal{N}|$ **then**
    Let a^{occ} be the value which occurs most, i.e. Λ_{ps}, times in $[Y_{psn}]_{n \in \mathcal{N}}$
    if $\Lambda_{ps} > \lambda$ **then**
      $\hat{Y}_{ps} := a^{\text{occ}}$
    else
      $\hat{Y}_{ps} := \gamma_p \operatorname{median}_{n \in \mathcal{R}_{ps}} \{Y_{psn}\}$
  else $\hat{Y}_{ps} := 0$

Algorithm 2 Heuristic generation of a primal investment solution $\left[\hat{Y}_{ps}\right]_{p \in \mathcal{P}, s \in \mathcal{S}}$

Let $y^{\text{split}}_{pin} := y_{pi}$, $n \in \mathcal{N}$, represent splitting variables for the investment variables, and consider again the objective function (13). Using the above notation, it is equivalently expressed as

$$C^{\text{tot}}(\boldsymbol{y}, \boldsymbol{x}, z, z^+, \boldsymbol{e}^{\text{tot}}) := \sum_{n \in \mathcal{N}} C^{\text{split}}_n(\boldsymbol{y}^{\text{split}}, \boldsymbol{x}, z, z^+, \boldsymbol{e}^{\text{tot}}), \tag{23}$$

where the system cost in each time period $n \in \mathcal{N}$ is defined as

$$C^{\text{split}}_n(\boldsymbol{y}^{\text{split}}, \boldsymbol{x}, z, z^+, \boldsymbol{e}^{\text{tot}}) := \sum_{p \in \mathcal{P}} \sum_{s \in \mathcal{S}} \left(\frac{c^{\text{inv}}_{ps} + c^{\text{omf}}_{p}}{N} y^{\text{split}}_{psn} + \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} \sum_{t \in \mathcal{T}_n} c^{\text{run}}_{pi} x_{pist} \right) \tag{24a}$$

$$+ \sum_{p \in \mathcal{P}_{\text{thermal}}} \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} \sum_{t \in \mathcal{T}_n} \left(c^+_{ps} z^+_{pist} + \tilde{c}_{ps}(z_{pist} - x_{pist}) \right) \tag{24b}$$

$$+ \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}_n} c^{\text{CO}_2}_{s} e^{\text{tot}}_{st} \tag{24c}$$

The investment variables in the model constraints should also be replaced by the splitting variables, and we add to the model the constraints

$$y^{\text{split}}_{psn} = y_{ps}, \qquad p \in \mathcal{P},\ s \in \mathcal{S},\ n \in \mathcal{N}. \tag{25}$$

These constraints are then Lagrangian relaxed using Lagrangian multipliers π^{one}_{psn}.

As for the other model variables, the constraints that cover multiple time steps need to be handled. More specifically, it is "the seams" that make the problem non-separable over time periods. Consider again the constraints (4); using the notation above, they are equivalently expressed as

$$z^+_{pist} \geq \begin{cases} z_{pist} - z_{p,i,s,t-1}, & t \in \mathcal{T}_n \setminus \{(n-1)\Omega + 1\},\ n \in \mathcal{N}, \\ z_{pist} - z_{p,i,s,t-1}, & t = (n-1)\Omega + 1,\ n \in \mathcal{N} \setminus \{1\}, \\ z_{pist} - z_{pisT}, & t = 1, \end{cases} \quad p \in \mathcal{P}_{\text{thermal}},\ i \in \mathcal{I}_{\text{active}}(s,p),\ s \in \mathcal{S}. \tag{26}$$

Here, it is the second and third sets of constraints—the seams—that need to be Lagrangian relaxed in order to make the model separable. For this, we denote the Lagrangian multipliers π^{two}_{pist}.

Moreover, for all $p \in \mathcal{P}_{\text{thermal}}$, $s \in \mathcal{S}$ and $i \in \mathcal{I}_{\text{active}}(s,p)$, the constraints (5) can be written as

$$\sum_{j \in \mathcal{I}_{\text{active}}(s,p)} y^{\text{split}}_{pjn} - z^+_{pist} \geq \begin{cases} z_{p,i,s,t-t_p}, & t_p \in \{m \in \mathcal{T}_{\text{start}}(p) : t - m \in \mathcal{T}_n\},\ t \in \mathcal{T}_n,\ n \in \mathcal{N}, \\ z_{p,i,s,t-t_p}, & t_p \in \{m \in \mathcal{T}_{\text{start}}(p) : t - m \in \mathcal{T}_{n-1}\},\ t \in \mathcal{T}_n,\ n \in \mathcal{N} \setminus \{1\}, \\ z_{p,i,s,N\Omega+t-t_p}, & t_p \in \{m \in \mathcal{T}_{\text{start}}(p) : t - m \leq 0\},\ t \in \mathcal{T}_n,\ n = 1. \end{cases} \tag{27}$$

Note that we have replaced the original investment variables by the splitting variables for investments in electricity generation. The first constraint corresponds to

time steps for z_{pist} within the M-week period, while the second constraint represents the time steps in the previous M-week period. The third constraint is the special case of first and last M-week period during the year. Here, we Lagrangian relax the second and third constraint using Lagrangian multipliers $\pi^{\text{three}}_{p,i,s,t,t_p,n}$.

Furthermore, the constraints (8) are replaced by

$$w_{st} + g_t - \sum_{i \in \mathcal{I}_{\text{active}}(s,p)} \tau \cdot x_{pist} \geq \begin{cases} w_{s,t+1}, & t \in \mathcal{T}_n \setminus \{n\Omega\},\ n \in \mathcal{N}, \\ w_{s,t+1}, & t = n\Omega,\ n \in \mathcal{N} \setminus \{N\}, \\ w_{s1}, & t = T, \end{cases} \qquad p \in \mathcal{P}_{\text{hydro}},\ s \in \mathcal{S}. \tag{28}$$

Once again, the second and third constraints should be relaxed with Lagrangian multipliers π^{four}_{pst}.

For all $p \in \mathcal{P}_{\text{hydro}}$, $i \in \mathcal{I}_{\text{active}}(s,p)$ and $s \in \mathcal{S}$, the constraints (10a) and (10b) can be written as

$$(1 + \delta^{\text{inc}}_r) x_{pist} \geq \begin{cases} x_{p,i,s,t+1}, & t \in \mathcal{T}_n \setminus \{n\Omega\},\ n \in \mathcal{N}, \\ x_{p,i,s,t+1}, & t = n\Omega,\ n \in \mathcal{N} \setminus \{N\}, \\ x_{pis1}, & t = T; \end{cases} \tag{29a}$$

$$(1 - \delta^{\text{dec}}_r) x_{pist} \leq \begin{cases} x_{p,i,s,t+1}, & t \in \mathcal{T}_n \setminus \{n\Omega\},\ n \in \mathcal{N}, \\ x_{p,i,s,t+1}, & t = n\Omega,\ n \in \mathcal{N} \setminus \{N\}, \\ x_{pis1}, & t = T. \end{cases} \tag{29b}$$

The second and third constraints in each of the expressions (29a) and (29b) should be relaxed, and we denote the corresponding Lagrangian multipliers by π^{five}_{pist} and π^{six}_{pist}.

To each subproblem $n \in \mathcal{N}$, the constraints (30), which limit the hot and start-up capacity, respectively, are added. These constraints are redundant in the original model but are here used to strengthen the dual formulation by making the subproblems tighter. They are formulated as

$$z_{pist} \leq \sum_{j \in \mathcal{I}_{\text{active}}(s,p)} y^{\text{split}}_{pjn}, \qquad i \in \mathcal{I}_{\text{active}}(s,p),\ p \in \mathcal{P}_{\text{thermal}},\ t \in \mathcal{T}, \tag{30a}$$

$$z^{+}_{pist} \leq \sum_{j \in \mathcal{I}_{\text{active}}(s,p)} y^{\text{split}}_{pjn}, \qquad i \in \mathcal{I}_{\text{active}}(s,p),\ p \in \mathcal{P}_{\text{thermal}},\ t \in \mathcal{T}. \tag{30b}$$

The new objective equals the sum of $C^{\text{tot}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{z}^{+}, \boldsymbol{e}^{\text{tot}})$—from (23)—and the Lagrangian penalty term $H^{\text{relax}}(\boldsymbol{y}, \boldsymbol{y}^{\text{split}}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{z}^{+}, \boldsymbol{\pi})$, where $\boldsymbol{\pi} := (\boldsymbol{\pi}^{\text{one}}, \boldsymbol{\pi}^{\text{two}}, \boldsymbol{\pi}^{\text{three}}, \boldsymbol{\pi}^{\text{four}}, \boldsymbol{\pi}^{\text{five}}, \boldsymbol{\pi}^{\text{six}})$, i.e. vector notation for all the Lagrangian multipliers, derived from the Lagrangian relaxation of the constraints as described above; the latter is expressed as

$$H^{\text{relax}}(\boldsymbol{y},\boldsymbol{y}^{\text{split}},\boldsymbol{x},\boldsymbol{w},\boldsymbol{z},\boldsymbol{z}^{+},\boldsymbol{\pi}) \;:=\; \sum_{n\in\mathcal{N}}\sum_{p\in\mathcal{P}}\sum_{s\in\mathcal{S}} \pi^{\text{one}}_{psn}\left(y_{ps}-y^{\text{split}}_{psn}\right) \tag{31a}$$

$$+\sum_{n\in\mathcal{N}\setminus\{1\}}\sum_{p\in\mathcal{P}_{\text{thermal}}}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_{\text{active}}(s,p)}\left(\pi^{\text{two}}_{p,i,s,(n-1)\Omega+1}\left(z_{p,i,s,(n-1)\Omega+1}-z_{p,i,s,(n-1)\Omega}-z^{+}_{p,i,s,(n-1)\Omega+1}\right)\right.$$
$$\left.+\sum_{t\in\mathcal{T}_n}\sum_{t_p\in\{m\in\mathcal{T}_{\text{start}}(p):\,t-m\in\mathcal{T}_{n-1}\}}\pi^{\text{three}}_{pist_pn}\left(z_{p,i,s,t-t_p}-\sum_{j\in\mathcal{I}_{\text{active}}(s,p)}y^{\text{split}}_{pjn}+z^{+}_{pist}\right)\right) \tag{31b}$$

$$+\sum_{p\in\mathcal{P}_{\text{thermal}}}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}_{\text{active}}(s,p)}\left(\pi^{\text{two}}_{pis1}\left(z_{pis1}-z_{pisT}-z^{+}_{pis1}\right)\right.$$
$$\left.+\sum_{t\in\mathcal{T}_1}\sum_{t_p\in\{m\in\mathcal{T}_{\text{start}}(p):\,m\geq t\}}\pi^{\text{three}}_{p,i,s,t,t_p,1}\left(z_{p,i,s,N\Omega+t-t_p}-\sum_{j\in\mathcal{I}_{\text{active}}(s,p)}y^{\text{split}}_{pj1}+z^{+}_{pist}\right)\right) \tag{31c}$$

$$+\sum_{n\in\mathcal{N}\setminus\{N\}}\sum_{p\in\mathcal{P}_{\text{hydro}}}\sum_{s\in\mathcal{S}}\left(\pi^{\text{four}}_{p,s,n\Omega}\left(w_{s,n\Omega+1}-w_{s,n\Omega}-g_{n\Omega}\right)\right.$$
$$+\sum_{i\in\mathcal{I}_{\text{active}}(s,p)}\left(\pi^{\text{four}}_{p,s,n\Omega}\,\tau\,x_{p,i,s,n\Omega}+\left(\pi^{\text{five}}_{p,i,s,n\Omega}-\pi^{\text{six}}_{p,i,s,n\Omega}\right)x_{p,i,s,n\Omega+1}\right. \tag{31d}$$
$$\left.\left.-\left(\pi^{\text{five}}_{p,i,s,n\Omega}(1+\delta^{\text{inc}}_{r})-\pi^{\text{six}}_{p,i,s,n\Omega}(1-\delta^{\text{dec}}_{r})\right)x_{p,i,s,n\Omega}\right)\right)$$

$$+\sum_{p\in\mathcal{P}_{\text{hydro}}}\sum_{s\in\mathcal{S}}\left(\pi^{\text{four}}_{psT}\left(w_{s1}-w_{sT}-g_{T}\right)\right.$$
$$+\sum_{i\in\mathcal{I}_{\text{active}}(s,p)}\left(\pi^{\text{four}}_{psT}\,\tau\,x_{pisT}+\pi^{\text{five}}_{pisT}\left(x_{pis1}-(1+\delta^{\text{inc}})x_{pisT}\right)\right. \tag{31e}$$
$$\left.\left.\left.+\,\pi^{\text{six}}_{pisT}\left((1-\delta^{\text{dec}})x_{pisT}-x_{pis1}\right)\right)\right).$$

The Lagrangian dual function, whose computation separates into $n\in\mathcal{N}$ subproblems, is defined as

$$\begin{aligned} h(\boldsymbol{\pi}) \;:=\; \underset{\boldsymbol{y},\boldsymbol{y}^{\text{split}},\boldsymbol{x},\boldsymbol{w},\boldsymbol{z},\boldsymbol{z}^{+},\boldsymbol{e}^{\text{tot}}}{\text{minimum}} \quad & \sum_{n\in\mathcal{N}} C^{\text{split}}_{n}(\boldsymbol{y}^{\text{split}},\boldsymbol{x},\boldsymbol{z},\boldsymbol{z}^{+},\boldsymbol{e}^{\text{tot}}) + H^{\text{relax}}(\boldsymbol{y},\boldsymbol{y}^{\text{split}},\boldsymbol{x},\boldsymbol{w},\boldsymbol{z},\boldsymbol{z}^{+},\boldsymbol{\pi}), \\ \text{subject to} \quad & (1),(2),(3),(6),(7),(9),(11),(12), \\ & (26),(27),(28),(29),(30) \end{aligned} \tag{32}$$

and the Lagrangian dual problem is defined as

$$\underset{\boldsymbol{\pi}}{\text{maximum}} \qquad h(\boldsymbol{\pi}), \tag{33a}$$

$$\text{subject to} \qquad \sum_{n \in \mathcal{N}} \pi^{\text{one}}_{psn} \geq 0, \quad p \in \mathcal{P},\, s \in \mathcal{S}, \tag{33b}$$

$$\pi^{\text{one}}_{psn} \leq N^{-1}\left(c^{\text{inv}}_{ps} + c^{\text{omf}}_{p}\right), \qquad p \in \mathcal{P},\, s \in \mathcal{S},\, n \in \mathcal{N}, \tag{33c}$$

$$\boldsymbol{\pi}^{\text{two}}, \boldsymbol{\pi}^{\text{three}}, \boldsymbol{\pi}^{\text{four}}, \boldsymbol{\pi}^{\text{five}}, \boldsymbol{\pi}^{\text{six}} \geq \mathbf{0}. \tag{33d}$$

Although the variables $\boldsymbol{y}$ are included in the objective of the subproblem in (32), their values will take the value 0 (∞) for any values of the dual variables $\boldsymbol{\pi}^{\text{one}}$ that can (cannot) be optimal in the Lagrangian dual defined by (33a) and (33d). The constraints (33b)–(33c) restrict the dual space to yield only finite (i.e. 0) values of $\boldsymbol{y}$ in the solution to (32); see also (Guignard 1993, Le. 2.2) for a theoretical motivation. Note that the constraints (33b)–(33d) define the set Π in Algorithm 1. A detailed theoretical motivation for this formulation, as well an algorithm for updating these dual variables in the subgradient algorithm, are found in Granfeldt (2021).

5 Implementation, data, and results

The implementation of the presented model and method is done in Julia (Bezanson et al. 2017) and JuMP (Dunning et al. 2017), using the Gurobi Optimizer version 9.1.1 (Gurobi Optimization, LLC 2021) on a system with Intel Xeon CPU E5-2650 v4 processor (2.20 GHz) with 24 cores and 256 GB RAM.

5.1 Data

The investment and variable costs, year 2050, for the electricity generation technologies considered in the model are presented in Table 2. The investment costs and fixed operation and maintenance costs are based on the World Energy Outlook (IEA 2016). The costs for wind and solar power are, however, based on data from the Danish Energy Agency (Energistyrelsen 2016). The models use annualized investment costs, assuming 5% interest rate. Technology learning for thermal generation is included as gradual improvements in the efficiencies of these technologies, which is reflected in a reduced variable cost for later years in the model. Also the cost of carbon dioxide emissions varies between years as it is assumed to become more expensive in forthcoming years. Moreover, the variable costs do not include costs from thermal cycling generation. Instead, the start-up and part-load costs are included explicitly as part of the thermal cycling constraints in the model. These costs, including the minimum load level, are based on the report by Jordan and Venkataraman (2012). The cycling properties of nuclear power are based on the paper by Persson et al. (2012), which considers a start-up time of 20 h and a minimum load level of 70%. Biogas is assumed to be produced through gasification of solid biomass

with 70% conversion efficiency. The cost of the gasifier equipment is included in the form of 20 €/MWh added to the fuel cost, rather than being incorporated in the investment cost of biogas technologies; this is due to biogas being storable, such that the gasifier equipment may attain a much larger number of full-load hours than the power plant consuming the biogas. The total cost of gasification equipment is from Thunman et al. (2019); 8,000 full-load hours are assumed.

Wind and solar power and sites are ordered in classes. Offshore wind sites are represented by one class while onshore wind sites are organized into several classes corresponding to different wind conditions, where each class is represented as one generation technology. For solar power, the classes represent land-based solar PV or rooftop solar PV. Wind and solar supply profiles and available capacities for different resource classes are based on a code set presented by Mattsson et al. (2021), which combines data with high spatial resolution and high temporal resolution. Table 3 provides the full load hours (FLH) and maximum capacities (Cap) for the onshore classes, as well as the offshore wind.

5.2 Results from the computations

Three regions with different conditions for wind power and with various existing energy mixes are evaluated: Hungary (HU), Ireland (IE), and southern Sweden (SE3), the latter being defined as the corresponding bidding area on the Nord Pool power exchange market (Nord Pool AS 2020). Table 4 presents the existing

Table 2 Costs and some technical properties for the electricity generation technologies

Technology $\in \mathcal{P}$	Investment costs [M€/MW]	Variable costs [€/MWh]	Fixed O &M costs [k€/MW,yr]	Lifetime [yr]	Minimum load level [share]	Start-up time [h]	Start-up cost [€/MW]
Coal ST	1.98	23.2	42.9	40	0.3	12	250
Natural gas CCGT	0.90	58.6	16.5	30	0.3	6	45
Natural gas GT	0.45	89.6	15.0	30	0.3	0	35
Biomass ST	1.98	86.6	51.9	40	0.3	12	240
Biomass CCGT	0.90	106.9	16.5	30	0.3	6	45
Biomass GT	0.45	164.1	15.0	30	0.3	0	50
Hydropower	2.39	1.0	54.4	500	0	0	0
Nuclear	3.98	24.5	149.0	60	0.7	24	660
Solar PV, land	0.42	1.1	6.50	40	0	0	0
Solar PV, rooftop	0.59	1.1	10.8	40	0	0	0
Onshore wind (A1–B5)	0.96	1.1	12.6	30	0	0	0
Offshore wind	1.53	1.1	36.0	30	0	0	0

All costs are given for the year 2050. The onshore wind classes all have the same costs and properties. The following abbreviations are used: *ST* steam turbine, *CCGT* combined cycle gas turbine, *GT* gas turbine, and *PV* photovoltaics

energy mixes in these regions. As we can see, the installed production capacity in HU consists mostly of nuclear and natural gas (gas turbines and combined-cycle gas turbines). The IE region has most of its production capacity in wind and natural gas, and also some coal. The production capacity in SE3 has historically been largely dominated by nuclear power, followed by some hydropower and wind power. Some of the nuclear power capacity used for this instance has, however, been recently decommissioned in Sweden. In this evaluation, trade is disregarded as only a single region is studied in each case. For each studied case, 2-week periods are used which thus yields 26 subproblems. In all our instances, we have used a 3 h time step throughout the year. As is evident by Table 5, the full model instance size is much larger than that of a single subproblem.

Figure 1 illustrates results from 50 deflected subgradient iterations of the decomposed model for the three instances. All costs are normalized, so that the optimal objective value for each instance equals 1. The dual objective value is the value of the Lagrangian objective function, while the primal objective value is calculated by solving the full model with fixed investments according to the solution provided by the heuristic in Algorithm 2. The scaling parameter θ_k in (19a) for the step length α_k in (18a) in Algorithm 1, is updated every $\ell = 5$ for all three regions.

For the HU-case, the best primal objective is found after only a few iterations. For the IE-region, however, the primal objective decreases drastically at first before increasing again, and then finally stabilizing around the same value as the upper bound (UBD) concluded during the first few iterations. For the SE3-region, it is necessary to greatly penalize the "seams" of the constraints relating to the hydropower reservoir storage level in order to avoid boundary effects. Thus, the large values of

Table 3 Full-load hours (FLH) and maximum capacity limits (Cap) for wind and solar power classes in the examined regions

Technology	Wind class	HU		IE		SE3	
$\in \mathcal{P}_{\text{wind}}$		FLH [h]	Cap [MW]	FLH [h]	Cap [MW]	FLH [h]	Cap [MW]
Onshore	A1	1387	5970	Ø	Ø	1288	150
Onshore	A2	1831	22510	1965	7	1895	34180
Onshore	A3	2409	5580	2715	300	2389	32750
Onshore	A4	3181	6.7	3419	2750	3254	6120
Onshore	A5	Ø	Ø	4243	24800	3999	1320
Onshore	B1	Ø	Ø	Ø	Ø	1264	380
Onshore	B2	Ø	Ø	Ø	Ø	1868	5970
Onshore	B3	Ø	Ø	Ø	Ø	2515	6140
Onshore	B4	Ø	Ø	Ø	Ø	3097	680
Onshore	B5	Ø	Ø	5661	0.8	4141	10
Offshore	–	Ø	Ø	4955	8340	4402	67770
Solar PV	A1	544	–	382	–	401	–
Solar PV	B1	Ø	–	Ø	–	349	–
Solar PV	R1	546	–	390	–	407	–

Absence of available sites is indicated by Ø

Table 4 Existing capacity in the three different regions, measured in GW

	HU	IE	SE3
Coal ST	0.20	0.86	–
Natural gas CCGT	1.00	3.22	0.08
Natural gas GT	1.00	3.22	0.08
Biomass ST	0.31	–	1.15
Hydropower	–	0.25	2.59
Nuclear	1.89	–	9.49
Solar PV	0.62	2.24	0.51
Wind onshore	0.33	4.13	2.45
Wind offshore	–	0.02	0.15

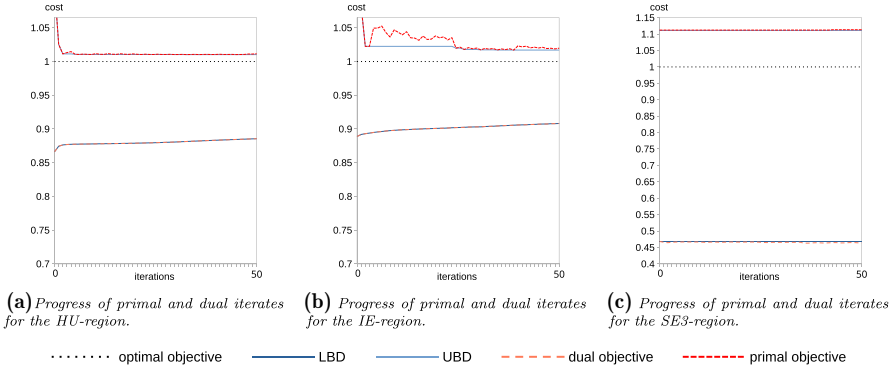
the dual multipliers $\boldsymbol{\pi}^{\text{four}}$ need to be compensated by a smaller Polyak step length in the subgradient algorithm, which results in the behaviour in Fig. 1c where neither the primal nor dual objective values make much progress. However, if the dual multipliers and the step length parameter are given similar values as in the other two instances, then the duality gap decreases as the number of iterations increase; see also Fig. 3 in Appendix B. However, due to the resulting boundary effects, this produces unreasonable feasible solutions with large primal objective values.

To conclude, the lower bound improves for the instances HU and IE (and SE3 dependent on parameter settings), implying that the subgradient algorithm performs well, albeit very slowly. However, we see that our subgradient algorithm is very sensitive to parameter settings and requires some calibration before producing good result. This is especially the case for the SE3-region, which is likely related to the large share of hydropower in Sweden as compared to the other examined regions. Indeed, our hypothesis is that the results for the HU-region are better due to the lack of hydropower (see Table 4), and thus those storage constraints and corresponding multipliers are not present in the model. Moreover, dependent on the initial values of the dual multipliers, the starting point for the dual objective value greatly varies. Hence, combined with the slow dual convergence, the size of the duality gap at our early termination is dependent on the initial subgradient algorithm settings.

Based on the results presented in Fig. 1, we stop the subgradient algorithm after three iterations as the potential savings in total system costs from running more iterations are not enough to compensate for the increased computation time. Figure 2 presents the results in terms of new production capacity for the three different cases,

Table 5 The number of variables and constraints for the full model and for a single subproblem in the SE3-region instance with a 3 h time step

	# Variables	# Constraints
Full model	4 321 715	3 972 177
Subproblem	172 497	180 273



(a) *Progress of primal and dual iterates for the HU-region.*

(b) *Progress of primal and dual iterates for the IE-region.*

(c) *Progress of primal and dual iterates for the SE3-region.*

Fig. 1 Dual and primal objective values—the latter calculated by solving the model (1)–(13) with fixed investments from Algorithm 2—versus subgradient iteration numbers. LBD and UBD denote, respectively, the lower and upper bound on the optimal value. All values are normalized by the optimal objective value for the respective instance

where the solutions provided by the decomposed model are compared with the optimal solutions provided by solving the full-scale model.

For the HU-region in Fig. 2a, the duality gap at termination is 13.3%. The feasible solution generated by Algorithm 2 is very similar to the optimal solution, besides some investments in natural gas and a slight overestimation of wind power capacity. The scaling parameter in Algorithm 2 was set to increase the investments in biomass production capacity, while decreasing the investments in natural gas. The difference in total system cost for this case is 1.11%. For the IE-region in Fig. 2b, the duality gap after three iterations is 12.7%. The installed capacity in wind power is underestimated, but this is mainly the result of the scaling parameter in Algorithm 2 being set too harsh to reduce wind power investments. As for the HU-region, the scaling parameter was moreover set to increase the investments in biomass production capacity, while decreasing the investments in natural gas. Thus, some biomass CCGT (combined cycle gas turbines) replaces natural gas GT (gas turbines) in the heuristic solution. The total system cost difference is 1.97% for this case. Lastly, as expected based on the results in Fig. 1c, the subgradient algorithm for the SE3-region terminated with the very large duality gap 57.7%. The feasible solution provided by the heuristic suggests CCGT running on natural gas for intermediate peak power, as compared to biomass in the optimal solution. Moreover, solar power is overestimated, and some offshore wind power is replaced by onshore wind power. As for the feasibility heuristic, the scaling parameter in Algorithm 2 was set to decrease investments in both solar and wind power to counteract the overestimations provided by the subgradient algorithm. The difference in total system cost is 10.75%, which is much worse if compared to the other tested instances. It should be noted however that

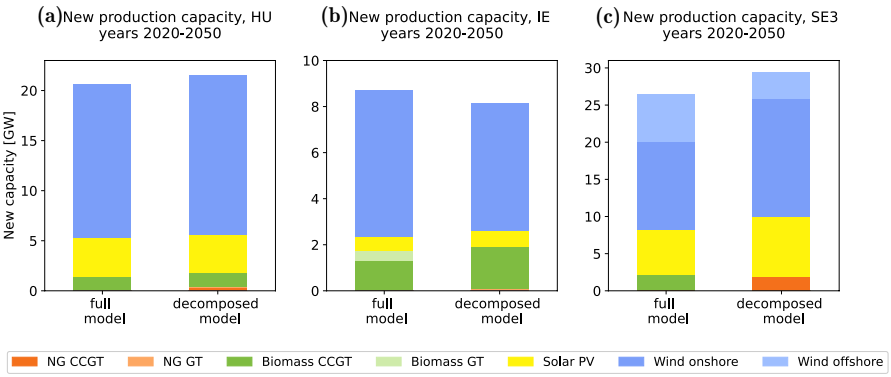


Fig. 2 Investments in new capacity for the full and the decomposed model in three different regions. The increase in total system costs for the decomposed model compared to the full model is (**a**) 1.11%, **b** 1.97%, and **c** 10.75%

if natural gas CCGT is manually replaced by biomass CCGT in the solution, the total system cost difference is reduced to 2.33%.

The decomposed model is implemented so that all computations are performed sequentially. The computation process can, however, be parallelized, and therefore the numbers presented correspond to a simulated parallel solve process. The argument for a sequential implementation over a parallel implementation on a single multi-core computer is that the decomposed model has the potential to be solved in parallel using a computer cluster (thus, 26 identical computers). Each subproblem is then solved on a computer where no parallelization is necessary and thus full processor capacity, and all CPU cores can be used to solve a single subproblem. Table 6 compares computation times for the different instances and models. The *build time* is the time it takes to build the model in Julia, i.e. convert data from files to create sets and parameters and declare the model with its variables, constraints, and objective function. For the decomposed model, we use the maximum build time among the 26 subproblems. The *solve time* is the time it takes to solve the model. For the decomposed model, the maximum solve time among the 26 subproblems is saved each iteration and finally summarized to provide the total solution time. The *algorithm time* is the computation time within the subgradient algorithm, including calculating new values of the Lagrangian multipliers. The *total time* is the sum of the computation times. The *average solve time for a subproblem* and the total *number of subgradient iterations* are also included.

As is evident from Table 6, in all instances the computing times for the decomposed model using the subgradient algorithm are lower, as compared to the full model. The total computing times for the regions HU, IE, and SE3 decrease by

Table 6 Computation times (all measured in seconds) for the different models and instances. For the decomposed model, the build and solve times are given presuming that all 26 subproblems are solved in parallel

Region	Model	Build time	Solve time	Algorithm time	Total time	Average subproblem solve time	# Subgradient iterations
HU	Full model	329	153	–	482	–	–
	Decomposed model	136	97	205	438	23	3
IE	Full model	150	350	–	500	–	–
	Decomposed model	135	88	207	431	20	3
SE3	Full model	155	1031	–	1186	–	–
	Decomposed model	138	109	201	448	27	3

9.1%, 13.8%, and 62.2%, respectively. This assumes, however, a parallel approach since a sequential implementation would take much longer time to solve. Moreover, the computation time for the decomposed model is highly dependent on the number of subgradient iterations. It should also be noted that the programming implementation of the models leaves room for improvement, and therefore the numbers presented here besides the solution times merely hint at the time complexity. Furthermore, for instances of this size it is likely unnecessary to implement the decomposition method since the full model computing times are quite short. Nevertheless, if several regions with options to trade electricity are included in the instances, the full model computing times are known to increase dramatically. Further, in Göransson et al. (2021), where we evaluated different spatial scope sizes, we showed that this choice of a 2-week period decomposition became increasingly beneficial in terms of computing times—as compared to the full model—as more regions were included in the problem instances. Since increasing the temporal resolution (from 3 h) will scale similarly as increasing the spatial scope, we argue that our proposed method likely has the potential to solve models with both a high temporal resolution and a rich spatial scope within reasonable solution times.

6 Discussion, conclusions, and future work

The model presented in this paper is a single-node electricity system model that can be used as a tool to analyse long-term investments in an electricity system containing a large share of variable renewable electricity generation. The model includes a fine discretization of time, using 3-hour time steps throughout the year, which allows for wind and solar power variations to be captured. In particular, the high temporal resolution captures the value of variable electricity production with respect to curtailment. The model is decomposed into M-week periods (where we in our tests let $M = 2$) using variable splitting and Lagrangian relaxation, which allows for the capture of wind events that last several days. However, although not evaluated in our tests, the proposed decomposition method can be used with different values of M to represent electricity systems that are dominated by variability on other time scales. Although, this will most likely affect the computation times as well as the accuracy of the results. The dual problem is solved by a deflected subgradient algorithm, and ergodic sequences are utilized to combine primal solutions from different iterations. A heuristic finally merges the 26 different solutions—one for each subproblem—into a single solution. The decomposition enables a separate solution of each subproblem which reduces memory requirements drastically. When utilizing computers with several CPUs or computer clusters, the subproblems can be solved in parallel to reduce computation times compared to solving the non-decomposed model. Furthermore—although neither included in our modelling nor evaluated in our tests—the suggested method has the potential to include trade between regions as a variation management strategy. The inclusion of trade in electricity system models using a fine temporal resolution typically leads to computationally very demanding problems, but with the proposed decomposition method it is likely that such models could be solved within reasonable computing times; further computational experiments are, however, needed in order to examine this presumed property. The choice of excluding trade from our model naturally leads to excess production capacity in the solutions to the problem instances. On the other hand, the issue of a restricted grid capacity is indeed faced during the current times. Moreover, we expect that batteries and hydrogen storage as well as demand flexibility will be key components in future electricity system. The mathematical implementation of these strategies follows a similar structure as Constraint (8) (demand flexibility is for example implemented using this structure in Göransson et al. (2014)) and can thus be fitted into the suggested structure. It is however unclear if our method will be able to properly dimension seasonal storage due to the 2-week period decomposition.

When evaluated for three different cases containing regions with varying conditions for variable renewable electricity sources, the method provides capacity investments similar to the optimal solution for investments provided by the

non-decomposed model. The heuristic used to combine the subproblem solutions can, however, be further developed as it is sensitive to parameter settings and requires some fine-tuning to provide good results. Furthermore, the decomposed model makes investments in electricity generation technology with the same production pattern (i.e. base, peak, and intermediate power plants) as the non-decomposed model, but occasionally choose natural gas over biomass as fuel. This is likely due to them having very similar costs using the assumed costs of carbon dioxide emissions. When these costs are increased, the decomposed model provides solutions containing biomass instead of natural gas.

The most important extensions of our work are to further develop the model to include and evaluate several regions and electricity trade with neighbouring regions, as well as technology options for batteries and demand flexibility such as hydrogen storage. However, if electricity trade or wind power are not of relevance for the investigated instances, other time representation methods such as *representative days* or *time slicing* are likely more efficient approaches than our proposed method.

A Nomenclature

See Tables 7, 8

Table 7 The variables used in the model

Symbol	Restriction	Explanation	Unit
y_{pi}	≥ 0	Investments in capacity (both new and old) for technology $p \in \mathcal{P}$ during investment period $i \in \mathcal{I}$	GW
y_{pin}^{split}	≥ 0	Investments in capacity (both new and old) for technology $p \in \mathcal{P}$ during investment period $i \in \mathcal{I}$ for subproblem $n \in \mathcal{N}$	GW
x_{pist}	≥ 0	Generated electricity of technology type $p \in \mathcal{P}$ using technology from investment period $i \in \mathcal{I}_{\text{active}}(s,p)$ in year $s \in \mathcal{S}$ and time step $t \in \mathcal{T}$	GWh/h
w_{st}	≥ 0	Stored hydropower for year $s \in \mathcal{S}$ at time step $t \in \mathcal{T}$	GWh
z_{pist}	≥ 0	Available hot capacity of technology type $p \in \mathcal{P}$ using technology from investment period $i \in \mathcal{I}_{\text{active}}(s,p)$ in year $s \in \mathcal{S}$ at time step $t \in \mathcal{T}$	GWh/h
z_{pist}^{+}	≥ 0	Increase in hot capacity from time step $t-1 \in \mathcal{T}$ to $t \in \mathcal{T}$ for technology type $p \in \mathcal{P}$ using technology from investment period $i \in \mathcal{I}_{\text{active}}(s,p)$ in year $s \in \mathcal{S}$	GWh/h
e_{st}^{tot}	≥ 0	Auxiliary definition variable for the total system emissions at year $s \in \mathcal{S}$ in time step $t \in \mathcal{T}$	Tonnes CO_2
$(\boldsymbol{y}, \boldsymbol{y}^{\text{split}}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{z}^{+}, \boldsymbol{e}^{\text{tot}})$		Vector notations for all model variables	

Table 8 The parameters used in the model

Symbol	Representation	Unit
b_{pi}^{gen}	Existing electricity generation capacity of technology $p \in \mathcal{P}$ in investment period $i \in \mathcal{I} \setminus \mathcal{S}$	GW
c_{ps}^{inv}	Investment cost for technology type $p \in \mathcal{P}$ during year $s \in \mathcal{S}$. Includes an annuity factor	k€/GW
c_{p}^{omf}	Fixed operation and maintenance costs for technology type $p \in \mathcal{P}$	k€/GW
c_{pi}^{run}	Run cost for technology type $p \in \mathcal{P}$ using technology from investment period $i \in \mathcal{I}$	k€/(GWh/h)
c_{ps}^{+}	Upstart cost for technology type $p \in \mathcal{P}$ in year $s \in \mathcal{S}$	k€/(GWh/h)
$\tilde{c}_{ps}$	Part load cost for technology type $p \in \mathcal{P}$ using technology at year $s \in \mathcal{S}$	k€/(GWh/h)
$c_{s}^{\text{CO}_2}$	Costs for emissions at year $s \in \mathcal{S}$	k€/tonnes CO_2
d_{st}	Demand at year $s \in \mathcal{S}$ and time step $t \in \mathcal{T}$	GWh/h
e_{pi}	Emissions per produced GWh of technology $p \in \mathcal{P}$ using technology from investment period $i \in \mathcal{I}$	Tonnes CO_2/(GWh/h)
$\tilde{e}_{pi}$	Extra emissions when running on part-load for technology type $p \in \mathcal{P}_{\text{thermal}}$ using technology from investment period $i \in \mathcal{I}$	Tonnes CO_2/(GWh/h)
e_{pi}^{+}	Upstart emissions for technology type $p \in \mathcal{P}_{\text{thermal}}$ using technology from investment period $i \in \mathcal{I}$	Tonnes CO_2/(GWh/h)
θ_{pt}	Weather profile for renewable technologies $p \in \mathcal{P}_{\text{ren}}$ at time step $t \in \mathcal{T}$	Share
g_t	Inflow into hydro power from rain, ground, etc. during time step $t \in \mathcal{T}$	GWh
δ^{inc}	Maximum ramping rate for hydropower water level increase	Share
δ^{dec}	Maximum ramping rate for hydropower water level decrease	Share
W_p	Maximum capacity of wind, i.e. land availability for wind technology $p \in \mathcal{P}_{\text{wind}}$	GW
ϕ_p	Minimum load level for technology $p \in \mathcal{P}_{\text{thermal}}$	Share
U_p	Life span of technology type $p \in \mathcal{P}$	Years
H	Upper limit for hydropower storage in region	GWh
T	Total number of time steps in the model	

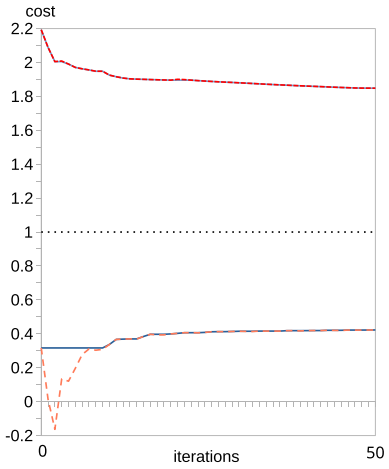


Fig. 3 Dual and primal objective values—the latter calculated by solving the model (1)–(13) with fixed investments from Algorithm 2—versus subgradient iteration numbers for the SE3-region using other values on the dual multipliers and step length. LBD and UBD denote, respectively, the lower and upper bound on the optimal value. All values are normalized by the optimal objective value for the instance

B Additional results

Figure 3 shows the result from 50 deflected subgradient iterations of the decomposed model for the SE3-region. Here, the values on the dual multipliers and the scale parameter in (18a) are similar to the ones used when solving the HU and IE problem instances. We see that the duality gap is very large, but progress is made both on the upper and lower bound as the number of iterations increases.

Acknowledgements We thank Professor Tom Brown, TU Berlin, for valuable input to method development. The authors would like to thank the anonymous referees for their valuable suggestions and feedback which contributed to an improved quality of the results of the paper. The research leading to the results presented in this article was supported by the Swedish Energy Agency (project number 39907-1).

Funding Open access funding provided by Chalmers University of Technology.

Data availability The input data applied in this work is given in Sect. 5.1 together with relevant references. The datasets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Bazaraa M, Sherali H, Shetty C (2013) Nonlinear programming: theory and algorithms. Wiley, https://books.google.se/books?id=nDYz-NIpIuEC

Belgacem R, Amir A (2018) A new modified deflected subgradient method. J King Saud Univ Sci 30(4):561–567. https://doi.org/10.1016/j.jksus.2017.08.004

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: a fresh approach to numerical computing. SIAM Rev 59(1):65–98. https://doi.org/10.1137/141000671

Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn 3(1):122. https://doi.org/10.1561/2200000016

Camerini PM, Fratta L, Maffioli F (1975) On improving relaxation methods by modified gradient techniques. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 26–34. https://doi.org/10.1007/BFb0120697

Caprara A, Fischetti M, Toth P (1999) A heuristic method for the set covering problem. Oper Res 47(5):730–743. https://doi.org/10.1287/opre.47.5.730

Collins S, Deane JP, Poncelet K, Panos E, Pietzcker RC, Delarue E, Gallachóir ÓBP (2017) Integrating short term variations of the power system into integrated energy system models: a methodological review. Renew Sustain Energy Rev 76:839–856. https://doi.org/10.1016/j.rser.2017.03.090

COM(2011) 885 (2011) Energy Roadmap 2050. European Commission

Dunning I, Huchette J, Lubin M (2017) JuMP: a modeling language for mathematical optimization. SIAM Rev 59(2):295–320. https://doi.org/10.1137/15M1020575

Energistyrelsen (2016) Technology data—energy plants for electricity and district heating generation. Report, Energistyrelsen, Copenhagen V, Denmark. https://ens.dk/sites/ens.dk/files/Statistik/technology_data_catalogue_for_el_and_dh_-_0009.pdf

Frew BA, Jacobson MZ (2016) Temporal and spatial tradeoffs in power system modeling with assumptions about storage: an application of the POWER model. Energy 117:198–213. https://doi.org/10.1016/j.energy.2016.10.074

Frew BA, Becker S, Dvorak MJ, Andresen GB, Jacobson MZ (2016) Flexibility mechanisms and pathways to a highly renewable US electricity future. Energy 101:65–78. https://doi.org/10.1016/j.energy.2016.01.079

Gerbaulet C, Lorenz C (2017) dynELMOD: a dynamic investment and dispatch model for the future European electricity market. Tech. rep., Deutsches Institut für Wirtschaftsforschung (DIW), Berlin, https://www.econstor.eu/bitstream/10419/161634/1/888201575.pdf

Gils HC (2016) Energy system model REMix. Report, Deutsches Zentrum für Luft- und Raumfahrt (DLR). www.dlr.de/tt/system

Glover F, Klingman D (1988) Layering strategies for creating exploitable structure in linear and integer programs. Math Program 40(1):165–181. https://doi.org/10.1007/BF01580728

Göransson L (2014) The impact of wind power variability on the least-cost dispatch of units in the electricity generation system. PhD thesis, Chalmers University of Technology

Göransson L, Goop J, Odenberger M, Johnsson F (2017) Impact of thermal plant cycling on the cost-optimal composition of a regional electricity generation system. Appl Energy 197:230–240. https://doi.org/10.1016/j.apenergy.2017.04.018

Göransson L, Granfeldt C, Strömberg AB (2021) Management of wind power variations in electricity system investment models. A parallel computing strategy. Oper Res Forum 2(2):25. https://doi.org/10.1007/s43069-021-00065-0

Granfeldt C (2021) Mathematical modelling and methodology for cost optimization of variable renewable electricity integration. Licentiate thesis, Chalmers University of Technology, Gothenburg, Sweden

Göransson L, Goop J, Unger T, Odenberger M, Johnsson F (2014) Linkages between demand-side management and congestion in the European electricity transmission system. Energy 69:860–872. https://doi.org/10.1016/j.energy.2014.03.083

Guignard M (1993) Solving makespan minimization problems with Lagrangean decomposition. Discret Appl Math 42(1):17–29. https://doi.org/10.1016/0166-218X(93)90176-O

Guignard M (2003) Lagrangean relaxation. Top 11(2):151–200. https://doi.org/10.1007/BF02579036

Guignard M, Kim S (1987) Lagrangean decomposition: a model yielding stronger Lagrangean bounds. Math Program 39(2):215–228. https://doi.org/10.1007/BF02592954

Gurobi Optimization, LLC (2021) Gurobi optimizer reference manual. https://www.gurobi.com

Gustavsson E, Patriksson M, Strömberg AB (2015) Primal convergence from dual subgradient methods for convex optimization. Math Program 150(2):365–390. https://doi.org/10.1007/s10107-014-0772-2

IEA (2016) World energy outlook 2016. Report, International Energy Agency. https://www.iea.org/reports/world-energy-outlook-2016

IEA (2020) Total energy supply (TES) by source, World 1990-2018. International Energy Agency

Jordan G, Venkataraman S (2012) Analysis of cycling costs in western wind and solar integration study. Report, National Renewable Energy Laboratory, Golden, Colorado, US. https://doi.org/10.2172/1047933

Jörnsten K, Näsberg M (1986) A new Lagrangian relaxation approach to the generalized assignment problem. Eur J Oper Res 27(3):313–323. https://doi.org/10.1016/0377-2217(86)90328-0

Kiwiel K (1990) Proximity control in bundle methods for convex nondifferentiable minimization. Math Program 46:105–122. https://doi.org/10.1007/BF01585731

Larsson T, Patriksson M, Strömberg A-B (1996) Conditional subgradient optimization–theory and applications. Eur J Oper Res 88(2):382–403. https://doi.org/10.1016/0377-2217(94)00200-2

Larsson T, Patriksson M, Strömberg AB (1999) Ergodic, primal convergence in dual subgradient schemes for convex programming. Math Program 86(2):283–312. https://doi.org/10.1007/s101070050090

Lehtveer M, Mattsson N, Hedenus F (2017) Using resource based slicing to capture the intermittency of variable renewables in energy system models. Energy Strategy Rev. https://doi.org/10.1016/j.esr.2017.09.008

Lemaréchal C, Nemirovskii A, Nesterov Y (1995) New variants of bundle methods. Math Program 69(1):111–147. https://doi.org/10.1007/BF01585555

Mai T, Drury E, Eurek K, Bodington N, Lopez A, Perry A (2013) Resource planning model: an integrated resource planning and dispatch tool for regional electric systems. Report, National Renewable Energy Laboratory. https://doi.org/10.2172/1067943

Mattsson N, Verendel V, Hedenus F, Reichenberg L (2021) An autopilot for energy models–automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions. Energy Strategy Rev 33:100606. https://doi.org/10.1016/j.esr.2020.100606

Nahmmacher P, Schmid E, Hirth L, Knopf B (2016) Carpe diem: a novel approach to select representative days for long-term power system modeling. Energy. https://doi.org/10.1016/j.energy.2016.06.081

Nord Pool AS (2020) Day-ahead overview Nord Pool. https://www.nordpoolgroup.com/en/maps/. Accessed 23 June 2022

Persson J, Andgren K, Henriksson H, Loberg J, Malm C, Pettersson L, Sandström J, Sigfrids T (2012) Additional costs for load-following nuclear power plants—experiences from Swedish, Finnish, German, and French nuclear power plants. Report 12:71, Elforsk, Stockholm, Sweden

Pfenninger S, Hawkes A, Keirstead J (2014) Energy systems modeling for twenty-first century energy challenges. Renew Sustain Energy Rev 33:74–86. https://doi.org/10.1016/j.rser.2014.02.003

Polyak BT (1969) Minimization of unsmooth functionals. USSR Comput Math Math Phys 9(3):14–29. https://doi.org/10.1016/0041-5553(69)90061-5

Reichenberg L, Siddiqui AS, Wogrin S (2018) Policy implications of downscaling the time dimension in power system planning models to represent variability in renewable output. Energy 159:870–877. https://doi.org/10.1016/j.energy.2018.06.160

Ringkjøb HC, Haugan PM, Solbrekke IM (2018) A review of modelling tools for energy and electricity systems with large shares of variable renewables. Renew Sustain Energy Rev 96:440–459. https://doi.org/10.1016/j.rser.2018.08.002

Sagastizábal C (2012) Divide to conquer: decomposition methods for energy optimization. Math Program 134(1):187–222. https://doi.org/10.1007/s10107-012-0570-7

Sherali HD, Ulular O (1989) A primal-dual conjugate subgradient algorithm for specially structured linear and convex programming problems. Appl Math Optim 20(1):193–221. https://doi.org/10.1007/BF01447654

Shor NZ (1991) The development of numerical methods for nonsmooth optimization in the USSR. In: Lenstra JK, Rinnoy Kan AHG, Schrijver A (eds) History of mathematical programming: a collection of personal reminiscences. North-Holland, Amsterdam, The Netherlands, pp 135–139

Thunman H, Gustavsson C, Larsson A, Gunnarsson I, Tengberg F (2019) Economic assessment of advanced biofuel production via gasification using cost data from the GoBiGas plant. Energy Sci Eng 7:217–229. https://doi.org/10.1002/ese3.271

Weber C (2005) Uncertainty in the electric power industry: methods and models for decision support. Springer, New York

Wogrin S, Duenas P, Delgadillo A, Reneses J (2014) A new approach to model load levels in electric power systems with high renewable penetration. IEEE Trans Power Syst 29(5):2210–2218. https://doi.org/10.1109/TPWRS.2014.2300697

Authors and Affiliations

Caroline Granfeldt[1] · Ann-Brith Strömberg[1] · Lisa Göransson[2]

✉ Caroline Granfeldt
cargranf@chalmers.se

[1] Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

[2] Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, Sweden