Optimal Price Zones of Electricity Markets: A Mixed-Integer Multilevel Model and Global Solution Approaches

Veronika Grimm, Thomas Kleinert, Frauke Liers, Martin Schmidt, Gregor Zöttl

ABSTRACT. Mathematical modeling of market design issues in liberalized electricity markets often leads to mixed-integer nonlinear multilevel optimization problems for which no general-purpose solvers exist and which are intractable in general. In this work, we consider the problem of splitting a market area into a given number of price zones such that the resulting market design yields welfare-optimal outcomes. This problem leads to a challenging multilevel model that contains a graph-partitioning problem with multi-commodity flow connectivity constraints and nonlinearities due to proper economic modeling. Furthermore, it has highly symmetric solutions. We develop different problem-tailored solution approaches. In particular, we present an extended KKT transformation approach as well as a generalized Benders approach that both yield globally optimal solutions. These methods, enhanced with techniques such as symmetry breaking and primal heuristics, are evaluated in detail on academic as well as on realistic instances. It turns out that our approaches lead to effective solution methods for the difficult optimization tasks presented here, where the problem-specific generalized Benders approach performs considerably better than the methods based on KKT transformation.

1. Introduction

Bilevel and, in general, multilevel optimization problems become increasingly important in applied optimization. Especially in fields like economics, where proper models often need to consider different objectives of different agents, optimization tasks naturally lead to multilevel problems that are extremely hard to solve. Even in their easiest instantiation, i.e., bilevel problems with linear lower and upper level, they are strongly NP hard; cf. Dempe et al. [18], Deng [19], and Garey and Johnson [28]. Moreover, Vicente et al. [62] have shown that even the verification of local optimality is NP hard in general. Besides these theoretical properties, the inherent violation of constraint qualifications adds to their hardness; cf. Ye and Zhu [63]. Obviously, many bilevel models from practice are not linear and, even worse, do not only contain continuous variables. In these situations one is faced with mixed-integer nonlinear bi- or multilevel problems. When confronted with such a model, one typically has to develop problem-tailored solution approaches since—at least to the best of our knowledge—no general-purpose solvers exist.

In this paper we consider an application from electricity market design that is of special economic and political importance: We analyze zonal pricing in electricity markets with redispatch as it is implemented, e.g., in Europe, Australia, or Latin America. In those regimes all intra-zonal network constraints are ignored at the spot market, which implies distorted investment incentives for generation capacity, leading to inefficiencies. In case the number of price zones cannot be adjusted, a partial remedy

consists in implementing an optimal configuration of price zones, such that congestion issues are reflected most appropriately in view of generation investment and production incentives. In other words, given an electricity market area with a corresponding electricity transmission network, a regulatory authority partitions the market area into "optimally configured" price zones. The corresponding zonal configuration specifies the set of physical constraints that have to be respected upon day-ahead spot-market trading of electricity. Due to the incomplete consideration of network constraints at the spot market, possibly infeasible outcomes have to be redispatched ex post, which gives rise to additional costs. The objective of the regulator (upon the optimal configuration of price zones) is to maximize overall social welfare, i.e., net welfare from spot-market trading minus investment costs and costs incurred at the redispatch level.

Rigorous mathematical modeling of this economic setting yields a mixed-integer nonlinear trilevel optimization problem. On the one hand, it combines various ingredients from different fields of optimization like its overall multilevel structure, a graph partitioning problem, a classical direct current (DC) power flow approximation, and, e.g., a multi-commodity flow problem as an auxiliary technique for modeling connectivity of the resulting zones. On the other hand, it suffers from integer programming (IP) symmetry and its large model size, even for comparably small networks. Thus, it is intractable in general and no off-the-shelf solution techniques exist to tackle the problem.

Our contribution is the following. First, we present a clear-cut model formulation. Since it is required to exploit the specific structure of the problem, the presentation of the complex model is as brief as possible but as detailed as required. Based on this model, we develop different solution techniques such as a KKT reformulation approach using a problem-tailored technique for reducing the number of levels. In addition, we present a problem-specific generalized Benders approach. We furthermore discuss the used techniques for breaking IP symmetry, present primal heuristics, and discuss how to algorithmically exploit structural variants of the overall objective function of the multilevel problem. Lastly, we evaluate the developed techniques on a set of academic instances from the literature as well as on further realistic instances modeling the German electricity market. We show that our methods lead to effective solution approaches.

The paper is structured as follows. In Sect. 2 we present the mixed-integer multilevel model for the specification of welfare-optimal price zones in electricity markets. Afterward, in Sect. 3 we discuss two different solution approaches for solving this model. Section 4 then presents some enhanced solution techniques and Sect. 5 contains a detailed computational study. The paper closes with some concluding remarks in Sect. 6.

Since the relevant parts of the literature are widespread, we discuss the relevant literature in the corresponding sections.

2. The Mixed-Integer Multilevel Market Model

Electricity markets are frequently organized as a system of interconnected price zones, where price formation at the spot market only takes into account inter-zonal network constraints but disregards all other (intra-zonal) network constraints; e.g., in Europe, Australia, or Latin-America. Feasibility of the final dispatch is typically guaranteed by so-called redispatch operations ex post; cf., e.g., Holmberg and Lazarczyk [37], or The European Commission [59] and ENTSO-E [24]. Whenever network constraints are not reflected in spot-market prices, this potentially distorts investment incentives for the efficient location of new generation capacity; cf., e.g., Grimm et al. [33]. As the principle decision to use price zones is often not questioned, a proper configuration of price zones (that most appropriately reflects congestion issues for a given number of zones) is of high interest and relevance. In fact, many recent academic contributions discuss and
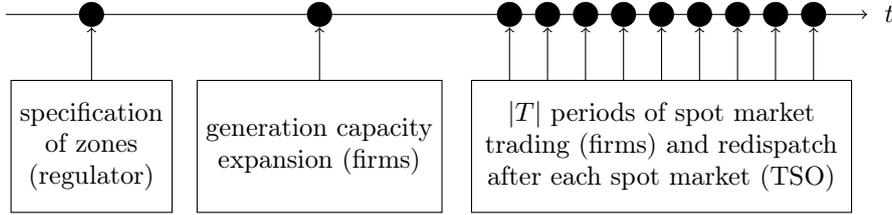
FIGURE 1. Timing of the multilevel game

analyze the impact of zonal configuration; cf., e.g., Bjørndal and Jørnsten [8, 9], Breuer and Moser [12], Egerer et al. [23], Stoft [58], and Trepper et al. [60]. However, none of those contributions proposes models or even techniques which allow to determine optimal zone configurations, as we do in this article.

In this section we first briefly sketch the structure of our model before we introduce the details at all levels. We consider a trilevel problem in which the regulator decides about a specification of price zones anticipating an energy-only market (within price zones) with cost-based redispatch. Figure 1 illustrates the economic setup we have in mind. Later in Fig. 2 and 3 we illustrate how the situation is captured in a trilevel model. The specification of zones is done once by the regulator, followed by generation capacity investment by private firms and multiple periods of spot-market trading and, in the case of congestion, redispatch by the transmission system operator (TSO). This game can be translated into a trilevel model as follows; see Grimm et al. [33] for a more detailed discussion. At the first level, the regulator specifies the price zones, anticipating the outcomes at all subsequent levels. The objective of the regulator is to maximize social welfare. At the second level, we model investment decisions of private firms in generation capacity as well as trading at a sequence of $|T|$ spot markets with fluctuating demand. In contrast to the regulator who is driven by welfare concerns, firms take investment and supply decisions in order to maximize profits. We assume that spot-market rules yield no price signals within price zones and assume perfectly competitive spot markets. The redispatch of the $|T|$ spot-market results is modeled in the third level and is anticipated by firms when they decide on generation investment, production, and demand at level two. Redispatch is operated by the regulated TSO. We assume that the TSO's objective is to maximize welfare (due to regulation of his activities). Redispatch occurs whenever traded quantities turn out to be infeasible subject to transmission constraints that have been ignored at the spot market. Note that consideration of redispatch in a separate third level is possible since, once generation capacities as well as price zones have been chosen, redispatch in time period t does not have any impact on supply decisions at any later point in time.

2.1. **Basic Economic and Technical Setup.** In this section we present the basic notation that is used throughout the paper.

We consider an electricity transmission network $\mathcal{G} = (N, L)$ with a set of nodes N and a set of transmission lines $L \subseteq N \times N$. Different lines l are characterized by their capacity $\bar{f}_l$ and their susceptance B_l. Throughout the paper we make use of the standard δ-notation, i.e., the sets of in- and outgoing lines of a node set $N' \subseteq N$ are denoted by $\delta_{N'}^{\text{in}}$ and $\delta_{N'}^{\text{out}}$, respectively. More formally, we have

$$\delta_{N'}^{\text{in}} := \{l \in L \colon l = (n, m) \text{ with } n \notin N', m \in N'\},$$
$$\delta_{N'}^{\text{out}} := \{l \in L \colon l = (n, m) \text{ with } n \in N', m \notin N'\}.$$

Price zones Z_i are modeled as parts of a partition $N = Z_1 \cup \cdots \cup Z_k$ of the node set, i.e., $i \in [k] := \{1, \ldots, k\}$.

At every node $n \in N$ we introduce a set of consumers C_n (with $0 \le |C_n| < \infty$) that are located at that node. We further assume a given set of scenarios $T = \{t_1, \ldots, t_{|T|}\}$

with scenario durations $\tau_\ell = t_{\ell+1} - t_\ell$ for $\ell = 1, \ldots, |T| - 1$. Elastic demand of a consumer $c \in C_n$ at node $n \in N$ in scenario $t \in T$ is modeled by a continuous and strictly decreasing function $p_{t,c} = p_{t,c}(d_{t,c})$, where $d_{t,c}$ denotes demand and $p_{t,c}(d_{t,c})$ is the resulting market price. Note that the gross consumer surplus

$$\int_0^{d_{t,c}} p_{t,c}(\omega)\,\mathrm{d}\omega$$

is a strictly concave function in this case. For what follows it is important (both for theory and practice) that the models under consideration can be solved efficiently. Thus, we restrict ourselves to the case of strictly decreasing linear demand functions $p_{t,c} = p_{t,c}(d_{t,c})$ that yield concave-quadratic gross consumer surplus.

For a given network node $n \in N$, G_n^{all} (with $0 \le |G_n^{\mathrm{all}}| < \infty$) denotes a finite set of existing technologies and candidate technologies that firms can invest in. We use the set G_n^{ex} for already existing generation technologies and the set G_n^{new} for candidate generation technologies. Thus, $G_n^{\mathrm{all}} = G_n^{\mathrm{new}} \cup G_n^{\mathrm{ex}}$ holds. For all existing generators $g \in G_n^{\mathrm{ex}}$, their capacity $\bar{q}_g^{\mathrm{ex}}$ is given. In contrast, new generation capacity $\bar{q}_g^{\mathrm{new}}$ can be installed for candidate generators. The corresponding investment costs are denoted by $c_g^{\mathrm{inv}} \in \mathbb{R}_{\ge 0}$ for $g \in G_n^{\mathrm{new}}$. Variable costs for production $q_{t,g}$ are denoted by $c_g^{\mathrm{var}} \in \mathbb{R}_{\ge 0}$ for all $g \in G_n^{\mathrm{all}}$ and $n \in N$.

In what follows, a superindex "spot" indicates quantities contracted upon spot-market trading and a superindex "red" denotes quantities actually produced and consumed after redispatch.

2.2. **First-Level Problem: Specification of Price Zones.** At the first level, the regulator decides on a specification of price zones as to maximize welfare, which is given as the difference of gross consumer surplus from all markets and investment and generation costs (after redispatch) of the producers:

$$\psi_1 := \sum_{t \in T} \sum_{n \in N} \sum_{c \in C_n} \int_0^{d_{t,c}^{\mathrm{red}}} p_{t,c}(\omega)\,\mathrm{d}\omega - \sum_{n \in N} \left(\sum_{g \in G_n^{\mathrm{new}}} c_g^{\mathrm{inv}} \bar{q}_g^{\mathrm{new}} + \sum_{t \in T} \sum_{g \in G_n^{\mathrm{all}}} c_g^{\mathrm{var}} q_{t,g}^{\mathrm{red}} \right).$$

The specification of price zones is modeled as follows. A price zone Z_i is part of a partition of the node set N. The number of parts of the partition is set to $k \in \mathbb{N}$, i.e., $N = Z_1 \cup \cdots \cup Z_k$, where k is given as input. For every node n the binary variables $x_{n,i} \in \{0,1\}, i \in [k]$, indicate to which zone i it belongs, i.e., $x_{n,i} = 1$ if node n belongs to zone i and $x_{n,i} = 0$ if not. Obviously, every node has to be located in one zone, i.e., we have the SOS-1 type constraint

$$\sum_{i \in [k]} x_{n,i} = 1 \quad \text{for all } n \in N. \tag{1}$$

For the application considered here, we are only interested in connected partitions that we model using a multi-commodity flow problem. Every commodity models a zone and every zone needs a sink to which every other node of that zone must be able to send a certain amount of flow in order to ensure connectivity. To be specific, we introduce another binary variable $z_{n,i} \in \{0,1\}$ with $z_{n,i} = 1$ if and only if node n is the (artificial) sink of zone i. The constraints

$$\sum_{n \in N} z_{n,i} = 1 \qquad\qquad \text{for all } i \in [k], \tag{2a}$$

$$z_{n,i} \le x_{n,i} \qquad\qquad \text{for all } n \in N,\ i \in [k] \tag{2b}$$

ensure that every zone has exactly one sink. Finally, we need two more constraints for modeling connectivity. To this end, we define the bi-directed graph $\mathcal{G}' := (N, A)$ with A consisting of lines $a^1(l) = (n, m)$ and $a^2(l) = (m, n)$ for all $l = (n, m) \in L$. Now

let $u^i = (u^i_a)_{a \in A} \geq 0$ be the vector of flows of commodity $i \in [k]$ and consider the conditions

$$\sum_{a \in \delta^{\text{out}}_n} u^i_a \leq M x_{n,i} \qquad \text{for all } n \in N, \ i \in [k], \qquad (3a)$$

$$\sum_{a \in \delta^{\text{out}}_n} u^i_a - \sum_{a \in \delta^{\text{in}}_n} u^i_a \geq x_{n,i} - M z_{n,i} \qquad \text{for all } n \in N, \ i \in [k], \qquad (3b)$$

where M is a sufficiently large number. Constraints (3a) model that flow of commodity i can only flow out of node n if this node belongs to zone i and Constraints (3b) state the following: First, if node n belongs to zone i and is not the sink of that zone, then n supplies at least one unit of flow of commodity i. Second, this supplied flow has to be transported to the sink of that zone because it is the only "demand" node in that zone.

Due to the nature of the application, we have modeled an exact k-cut problem with connectivity constraints. For the spot-market model (cf. Sect. 2.3) we additionally need indicator variables for inter-zonal lines, i.e., for the cut edges obtained from the graph partition. To this end, we introduce binary variables $y_l \in \{0, 1\}$ with $y_l = 1$ if and only if line l is an inter-zonal link, i.e., if it connects nodes of different zones. This is modeled by the constraints

$$x_{n,i} - x_{m,i} \leq y_l \qquad \text{for all } l = (n,m) \in L, \qquad (4a)$$

$$x_{m,i} - x_{n,i} \leq y_l \qquad \text{for all } l = (n,m) \in L, \qquad (4b)$$

$$x_{n,i} + x_{m,i} + y_l \leq 2 \qquad \text{for all } l = (n,m) \in L. \qquad (4c)$$

Altogether, we obtain the following first-level problem:

$$\max \quad \psi_1 \quad \text{s.t.} \quad (1)\text{–}(4). \qquad (5)$$

In the following, we assume that social welfare ψ_1 is non-negative, which is reasonable for well-functioning markets. Model (5) consists of a concave-quadratic objective function and a set of mixed-integer linear constraints that contain both a graph partitioning as well as a multi-commodity flow problem as substructures, i.e., we are faced with a mixed-integer quadratic program (MIQP).

Graph partitioning problems have been studied extensively in the literature—often specifically for the case $k = 2$. Many beautiful structural insights could be obtained. Out of the many references, we refer to the book Deza and Laurent [20] as well as to Barahona and Mahjoub [6], Boros and Hammer [11] for the case $k = 2$, and to Chopra and Rao [15, 16] for general values of k.

Effective branch-and-bound approaches based on linear relaxations have been developed for $k = 2$ in Barahona et al. [5] and Liers et al. [43], as well as for the k-equipartition problem by Mitchell [50]. Furthermore, positive semidefinite relaxations yield very strong bounds and lead to effective global solution approaches, see Rendl et al. [55] for $k = 2$. For general values of k, we refer to Anjos et al. [3] and to Lisser and Rendl [44] for the the k-equipartition problem. In contrast, graph partitioning with connectivity constraints has been considered much less. A related problem is the maximum-weight connected subgraph problem that has several applications, for example in designing fiber-optic networks discussed in Lee and Dooly [42], in wildlife corridor design considered in Dilkina and Gomes [21], and in forest planning studied in Carvajal et al. [13]. According to our knowledge, multi-level graph partitioning problems have not yet been studied in the literature.

2.3. **Second-Level Problem: Generation Investment and Spot-Market Behavior.** At the second level we model the behavior of firms with respect to generation investment and spot-market trading. The wholesale electricity market is assumed to be perfectly competitive, i.e., no firm can directly affect prices by strategic investment or supply decisions. It has been shown that a perfectly competitive environment

yields welfare maximizing investment and production decisions in our context; see, e.g., Grimm et al. [34]. We are aware of the issue that the assumption of perfect competition may not be adequate for power systems in general. However, this assumption is necessary in order to keep the multilevel problem tractable—both theoretically and computationally. For a discussion about this, see Grimm et al. [33].

When making investment and supply decisions, firms only consider physical constraints for which they receive price signals. If the market is not divided into zones, firms receive no signals concerning network capacities and thus will not account for them. If the market is divided into two or more zones, firms consider those physical constraints that are expressed in price differences due to market splitting: Between any pair of zones, electricity flow cannot exceed the maximum capacity of the respective inter-zone network links and congestion implies price differences across zones. This is modeled by the following zonal version of Kirchhoff's first law: total generation in zone i plus total inflow of zone i equals total demand in zone i plus total outflow of that zone. More formally, using the binary variables of the first level, we obtain

$$d_{t,n}^{\mathrm{spot}} = \sum_{c \in C_n} d_{t,c}^{\mathrm{spot}}, \quad q_{t,n}^{\mathrm{spot}} = \sum_{g \in G_n^{\mathrm{all}}} q_{t,g}^{\mathrm{spot}} \qquad \text{for all } n \in N,\ t \in T, \quad (6a)$$

$$D_t^i = \sum_{n \in N} x_{n,i} d_{t,n}^{\mathrm{spot}}, \quad Q_t^i = \sum_{n \in N} x_{n,i} q_{t,n}^{\mathrm{spot}} \qquad \text{for all } i \in [k],\ t \in T, \quad (6b)$$

$$F_{i,t}^{\mathrm{in}} = \sum_{l=(n,m) \in L} (1 - x_{n,i}) x_{m,i} f_{t,l}^{\mathrm{spot}} \qquad \text{for all } i \in [k],\ t \in T, \quad (6c)$$

$$F_{i,t}^{\mathrm{out}} = \sum_{l=(n,m) \in L} x_{n,i} (1 - x_{m,i}) f_{t,l}^{\mathrm{spot}} \qquad \text{for all } i \in [k],\ t \in T, \quad (6d)$$

$$D_t^i + F_{i,t}^{\mathrm{out}} = Q_t^i + F_{i,t}^{\mathrm{in}} \qquad \text{for all } i \in [k],\ t \in T. \quad (6e)$$

Market splitting based flow restrictions are modeled by flow capacity constraints on inter-zonal lines, i.e.,

$$-\bar{f}_l - (1 - y_l) M \le f_{t,l}^{\mathrm{spot}} \le \bar{f}_l + (1 - y_l) M \quad \text{for all } l \in L,\ t \in T. \quad (7)$$

Again, M is a sufficiently large number. In addition, we have lower and/or upper bounds on demand, power generation, and capacity:

$$0 \le d_{t,c}^{\mathrm{spot}} \qquad \text{for all } t \in T,\ n \in N,\ c \in C_n, \quad (8a)$$

$$0 \le q_{t,g}^{\mathrm{spot}} \le \tau \bar{q}_g^{\mathrm{new}} \qquad \text{for all } t \in T,\ n \in N,\ g \in G_n^{\mathrm{new}}, \quad (8b)$$

$$0 \le q_{t,g}^{\mathrm{spot}} \le \tau \bar{q}_g^{\mathrm{ex}} \qquad \text{for all } t \in T,\ n \in N,\ g \in G_n^{\mathrm{ex}}. \quad (8c)$$

To summarize, at level two we consider welfare-maximizing generation investment and supply decisions, i.e.,

$$\psi_2 := \sum_{t \in T} \sum_{n \in N} \sum_{c \in C_n} \int_0^{d_{t,c}^{\mathrm{spot}}} p_{t,c}(\omega)\, \mathrm{d}\omega - \sum_{n \in N} \left(\sum_{g \in G_n^{\mathrm{new}}} c_g^{\mathrm{inv}} \bar{q}_g^{\mathrm{new}} + \sum_{t \in T} \sum_{g \in G_n^{\mathrm{all}}} c_g^{\mathrm{var}} q_{t,g}^{\mathrm{spot}} \right),$$

where supply is constrained by generation capacities installed and transmission constraints across zones. Thus, the second-level problem reads

$$\max \quad \psi_2 \quad \text{s.t.} \quad (6)\text{--}(8). \quad (9)$$

In the case where every zone consists of exactly one network node, (6e) coincides with Kirchhoff's first law and ensures power balance at every network node:

$$\sum_{c \in C_n} d_{t,c}^{\mathrm{spot}} + \sum_{l \in \delta_n^{\mathrm{out}}} f_{t,l}^{\mathrm{spot}} = \sum_{g \in G_n^{\mathrm{all}}} q_{t,g}^{\mathrm{spot}} + \sum_{l \in \delta_n^{\mathrm{in}}} f_{t,l}^{\mathrm{spot}} \quad \text{for all } n \in N,\ t \in T. \quad (10)$$

The spot-market model is a concave-quadratic maximization problem over mixed-integer nonlinear constraints. All binary variables are variables of the first level and all nonlinearities in the constraints appear in the zonal version of Kirchhoff's first law (6).
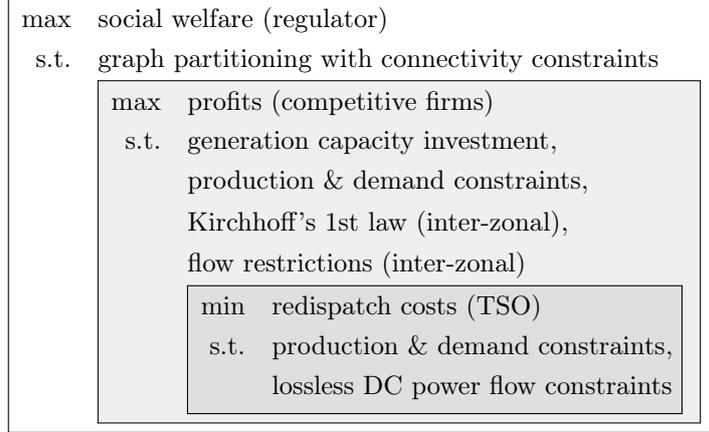
```
max   social welfare (regulator)
s.t.  graph partitioning with connectivity constraints

      max   profits (competitive firms)
      s.t.  generation capacity investment,
            production & demand constraints,
            Kirchhoff's 1st law (inter-zonal),
            flow restrictions (inter-zonal)

            min   redispatch costs (TSO)
            s.t.  production & demand constraints,
                  lossless DC power flow constraints
```

FIGURE 2. Structure of the trilevel market model

These nonlinearities are of the type xw with $x \in \{0,1\}$ and $w \in \mathbb{R}$. Assuming that the continuous variable w is bounded below and above, i.e., $w \in [\underline{w}, \bar{w}]$, we linearize this product by introducing the new variable $v \in \mathbb{R}$ and by adding the constraints $\underline{w}x \leq v \leq \bar{w}x$ and $\underline{w}(1-x) \leq w - v \leq \bar{w}(1-x)$. Thus, we again end up with a concave-quadratic maximization problem over a set of mixed-integer linear constraints, where all discrete variables are first-level variables.

2.4. **Third-Level Problem: Optimal Redispatch.** At the third level, the TSO simultaneously decides on cost-based redispatch for all $|T|$ spot markets. Reallocation of spot-market outcomes is realized in a way that ensures feasibility with respect to transmission constraints at lowest costs. The latter are given by

$$\psi_3 := \sum_{t \in T} \sum_{n \in N} \sum_{c \in C_n} \int_{d_{t,c}^{\mathrm{red}}}^{d_{t,c}^{\mathrm{spot}}} p_{t,c}(\omega)\,\mathrm{d}\omega + \sum_{t \in T} \sum_{n \in N} \sum_{g \in G_n^{\mathrm{all}}} c_g^{\mathrm{var}}(q_{t,g}^{\mathrm{red}} - q_{t,g}^{\mathrm{spot}})$$

and the redispatch decision has to account for all physical transmission constraints. Beside the already stated constraints, this includes Kirchhoff's second law, which determines the voltage angles $\theta_{t,n}$, $t \in T$, $n \in N$, in the network:

$$f_{t,l}^{\mathrm{red}} = B_l(\theta_{t,n} - \theta_{t,m}) \quad \text{for all } l = (n,m) \in L,\ t \in T. \tag{11}$$

In order to obtain unique physical solutions, we have to fix the voltage angle at an arbitrary reference node $\hat{n} \in N$ in every time period:

$$\theta_{t,\hat{n}} = 0 \quad \text{for all } t \in T. \tag{12}$$

Furthermore, all transmission flows are limited by lower and upper bounds, i.e.,

$$-\bar{f}_l \leq f_{t,l}^{\mathrm{red}} \leq \bar{f}_l \quad \text{for all } l \in L,\ t \in T. \tag{13}$$

The problem at the redispatch level then reads

$$\min \quad \psi_3 \quad \text{s.t.} \quad (8), (10)\text{--}(13),$$

where we replaced $d_{t,n}^{\mathrm{spot}}, q_{t,g}^{\mathrm{spot}}$ in (8) and $d_{t,n}^{\mathrm{spot}}, q_{t,g}^{\mathrm{spot}}, f_{t,l}^{\mathrm{spot}}$ in (10) by the redispatch variables $d_{t,n}^{\mathrm{red}}, q_{t,g}^{\mathrm{red}}, f_{t,l}^{\mathrm{red}}$. Finally, we note that the redispatch model is a convex-quadratic minimization problem over linear constraints.

The entire trilevel market model is sketched in economic and technical terms in Fig. 2. More formally, the trilevel model is given in Fig. 3 (left), where $W_i \in \mathbb{R}^{n_i}$ and $X_i \in \{0,1\}^{m_i}$ are the continuous and discrete variables of level $i = 1, 2, 3$, and Ω_i is the corresponding discrete-continuous feasible set. The dependencies between the three levels is as follows; cf. Fig. 3 (right). The first level depends on the continuous variables of the second and third level (spot-market and redispatch results), the second level

$$\begin{aligned}
\max \quad & \psi_1(W_2, W_3) \\
\text{s.t.} \quad & (W_1, X_1) \in \Omega_1 \\
& \quad \begin{aligned}
\max \quad & \psi_2(W_2) \\
\text{s.t.} \quad & (W_2, X_1) \in \Omega_2 \\
& \quad \begin{aligned}
\min \quad & \psi_3(W_2, W_3) \\
\text{s.t.} \quad & (W_2, W_3) \in \Omega_3
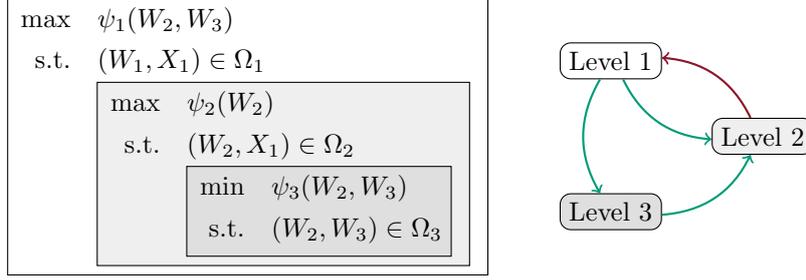\end{aligned}
\end{aligned}
\end{aligned}$$

FIGURE 3. Mathematical structure of the trilevel market model (left) and dependencies of the three market-model levels (right). Green arcs denote dependencies on continuous variables, red arcs denote dependencies on discrete variables.

depends on the discrete variables of the first level (zone specification) and the third level depends on the continuous variables of the second level (spot-market results).

2.5. **The Integrated Planner Model.** It is helpful to additionally consider a related optimization problem that is less complex than the full trilevel problem. It models an integrated planner. In a version slightly adapted from Jenabi et al. [38], a fictitious integrated generation and transmission company (IGTC) simultaneously determines capacity expansion and production at the spot markets such that social welfare ψ_{IGTC} is maximized. The latter is defined as the difference of gross consumer surplus and generation capacity investment costs as well as variable costs of production:

$$\psi_{\mathrm{IGTC}} := \sum_{t \in T} \sum_{n \in N} \sum_{c \in C_n} \int_0^{d_{t,c}} p_{t,c}(\omega) \, \mathrm{d}\omega - \sum_{n \in N} \left(\sum_{g \in G_n^{\mathrm{new}}} c_g^{\mathrm{inv}} \bar{q}_g^{\mathrm{new}} + \sum_{t \in T} \sum_{g \in G_n^{\mathrm{all}}} c_g^{\mathrm{var}} q_{t,g} \right).$$

An integrated planner has to account for the full physical network. Kirchhoff's first law, i.e., flow conservation at each node, reads

$$\sum_{c \in C_n} d_{t,c} + \sum_{l \in \delta_n^{\mathrm{out}}} f_{t,l} = \sum_{g \in G_n^{\mathrm{all}}} q_{t,g} + \sum_{l \in \delta_n^{\mathrm{in}}} f_{t,l} \quad \text{for all } t \in T, n \in N. \tag{14}$$

As before, Kirchhoff's second law determines the voltage angles θ_n:

$$f_{t,l} = B_l \left(\theta_{t,n} - \theta_{t,m} \right) \quad \text{for all } l = (n, m) \in L, \ t \in T. \tag{15}$$

We again fix the voltage angle at a reference node $\hat{n} \in N$ in every time period:

$$\theta_{t,\hat{n}} = 0 \quad \text{for all } t \in T. \tag{16}$$

In addition, we have lower and/or upper bounds on demand, power generation, and capacity, i.e.,

$$0 \le d_{t,c} \qquad \text{for all } t \in T, \ n \in N, \ c \in C_n, \tag{17a}$$

$$0 \le q_{t,g} \le \tau \bar{q}_g^{\mathrm{new}} \qquad \text{for all } t \in T, \ n \in N, \ g \in G_n^{\mathrm{new}}, \tag{17b}$$

$$0 \le q_{t,g} \le \tau \bar{q}_g^{\mathrm{ex}} \qquad \text{for all } t \in T, \ n \in N, \ g \in G_n^{\mathrm{ex}}. \tag{17c}$$

In summary, the integrated planner model is a continuous maximization problem with linear constraints and a concave-quadratic objective function:

$$\begin{aligned}
\max \quad & \psi_{\mathrm{IGTC}} & \text{(18a)} \\
\text{s.t.} \quad & \text{Kirchhoff's first law: (14),} & \text{(18b)} \\
& \text{Kirchhoff's second law: (15), (16),} & \text{(18c)} \\
& \text{Bounds: (17).} & \text{(18d)}
\end{aligned}$$

This convex problem can be solved efficiently. Furthermore, it is easy to see that the social welfare ψ_{IGTC} is non-negative since the trivial solution is feasible.

Finally note that the relation of the physical parts (14), (15) of the IGTC model and the zonal spot-market model is given by interpreting every node as a single zone.

3. Solution Approaches

As described in the last section, our electricity market model is a trilevel optimization model including mixed-integer aspects for modeling the exact k-cut problem. We now develop problem-specific solution techniques that are required to solve instances of relevant size.

In this section, we discuss two different approaches. The first approach, cf. Sect. 3.1, exploits problem-tailored reformulations of the trilevel model in combination with general transformation techniques based on first-order optimality conditions. These techniques yield a single-level but large and numerically challenging MIQP. The advantage however is that it can be solved using standard solvers. Second, in Sect. 3.2 we present a tailored variant of generalized Benders decomposition, where the decomposition explicitly exploits the relation between discrete and continuous aspects of the overall model.

Enhanced techniques that can be used within the MIQP and/or the Benders approach like, e.g., symmetry breaking or primal heuristics, are topic of Sect. 4.

3.1. **An MIQP Approach.** We first show how reformulation techniques that exploit the problem structure can be used to reformulate the trilevel market model as an equivalent mixed-integer bilevel model with concave-quadratic objectives at both levels. Afterward, we employ a standard KKT transformation approach for the lower level of the bilevel problem; cf., e.g., Dempe et al. [18]. This leads to a single-level mixed-integer concave-quadratic maximization problem that can be solved by general-purpose MIQP solvers. The main observation for the reformulation of the trilevel to an equivalent bilevel problem is the following:

Proposition 1. *Let ψ_1, ψ_2, and ψ_3 be the objective functions of the trilevel market model. Then, $\psi_1 = \psi_2 - \psi_3$ holds.*

This proposition is also used in Grimm et al. [33] in a slightly modified setting. It reveals that the first- and third-level problem have affine equivalent objective functions and thus have identical optimization directions. Hence, we can equivalently replace the original trilevel model (cf. Fig. 3) by the following bilevel model:

$$
\begin{aligned}
\max \quad & \psi_1(W_2, W_3) \\
\text{s.t.} \quad & (W_1, X_1) \in \Omega_1, \ (W_2, W_3) \in \Omega_3, \\
& W_2 \in \arg\max \left\{ \psi_2(W_2) \colon (W_2, X_1) \in \Omega_2 \right\}.
\end{aligned}
$$

The upper level maximizes the original first-level objective subject to the constraints of the original first and third level. This new upper level is again an MIQP containing an exact k-cut problem with multi-commodity flow connectivity constraints as well as a lossless DC flow problem as substructures. The new lower-level problem is the original second level. All discrete variables appearing in the lower level stem from the upper level, i.e., there are no discrete lower-level variables.

We now replace this bilevel problem by an equivalent single-level MIQP using the standard KKT transformation approach for the lower level. We note that the lower level model is a concave-quadratic maximization problem subject to constraints that are linear in the continuous second-level variables. Thus, for given discrete upper-level variables the KKT conditions are both necessary and sufficient. Hence, we can replace the lower level of our bilevel problem by its KKT conditions. These are given by dual

feasibility

$$p_{t,c}(d_{t,c}^{\mathrm{spot}}) + \sum_{i\in[k]} \epsilon_{i,t}x_{n,i} + \kappa_{t,c}^{-} = 0 \qquad \text{for all } t\in T,\ n\in N,\ c\in C_n, \qquad (19a)$$

$$-c_g^{\mathrm{var}} - \sum_{i\in[k]} \epsilon_{i,t}x_{n,i} + \pi_{t,g}^{-} - \pi_{t,g}^{+} = 0 \qquad \text{for all } t\in T,\ n\in N,\ g\in G_n^{\mathrm{new}}, \qquad (19b)$$

$$-c_g^{\mathrm{var}} - \sum_{i\in[k]} \epsilon_{i,t}x_{n,i} + \nu_{t,g}^{-} - \nu_{t,g}^{+} = 0 \qquad \text{for all } t\in T,\ n\in N,\ g\in G_n^{\mathrm{ex}}, \qquad (19c)$$

$$-c_g^{\mathrm{inv}} + \sum_{t\in T} \tau_t \pi_{t,g}^{+} = 0 \qquad \text{for all } n\in N,\ g\in G_n^{\mathrm{new}}, \qquad (19d)$$

$$\eta_{t,l}^{-} - \eta_{t,l}^{+} - \sum_{i\in[k]} \epsilon_{i,t}(x_{m,i} - x_{n,i}) = 0 \qquad \text{for all } t\in T,\ l=(n,m)\in L, \qquad (19e)$$

primal feasibility (6)–(8), non-negativity of dual variables of inequality constraints

$$\eta_{t,l}^{-}, \eta_{t,l}^{+} \ge 0 \qquad \text{for all } l\in L,\ t\in T,$$

$$\pi_{t,g}^{-}, \pi_{t,g}^{+} \ge 0 \qquad \text{for all } t\in T,\ n\in N,\ g\in G_n^{\mathrm{new}},$$

$$\nu_{t,g}^{-}, \nu_{t,g}^{+} \ge 0 \qquad \text{for all } t\in T,\ n\in N,\ g\in G_n^{\mathrm{ex}},$$

$$\kappa_{t,c}^{-} \ge 0 \qquad \text{for all } t\in T,\ n\in N,\ c\in C_n,$$

and KKT complementarity conditions

$$\eta_{t,l}^{-}(-\bar{f}_l - (1-y_l)M - f_{t,l}^{\mathrm{spot}}) = 0 \qquad \text{for all } t\in T,\ l\in L, \qquad (20a)$$

$$\eta_{t,l}^{+}(f_{t,l}^{\mathrm{spot}} - \bar{f}_l - (1-y_l)M) = 0 \qquad \text{for all } t\in T,\ l\in L, \qquad (20b)$$

$$\kappa_{t,c}^{-}d_{t,c}^{\mathrm{spot}} = 0 \qquad \text{for all } t\in T,\ n\in N,\ c\in C_n, \qquad (20c)$$

$$\pi_{t,g}^{-}q_{t,g}^{\mathrm{spot}} = \pi_{t,g}^{+}(q_{t,g}^{\mathrm{spot}} - \tau_t\bar{q}_g^{\mathrm{new}}) = 0 \qquad \text{for all } t\in T,\ n\in N,\ g\in G_n^{\mathrm{new}}, \qquad (20d)$$

$$\nu_{t,g}^{-}q_{t,g}^{\mathrm{spot}} = \nu_{t,g}^{+}(q_{t,g}^{\mathrm{spot}} - \tau_t\bar{q}_g^{\mathrm{ex}}) = 0 \qquad \text{for all } t\in T,\ n\in N,\ g\in G_n^{\mathrm{ex}}. \qquad (20e)$$

Note that the dual variables of the primal auxiliary constraints (6a)–(6d) are already eliminated in the dual feasibility conditions (19).

The KKT conditions, except for the complementarity conditions (20), are linear in the primal and dual lower level variables. Using a standard linearization trick, we now get rid of the nonlinearities caused by the KKT complementarity conditions; cf. Fortuny-Amat and McCarl [26]. To this end, for a dual variable $\xi \ge 0$ let $z(\xi) \in \{0,1\}$ be a binary variable that is zero, if $\xi = 0$ and 1 otherwise. Furthermore, let the constant $\tilde{M}$ be a sufficiently large number. Now, the nonlinear conditions (20) can be replaced by the mixed-integer linear constraints

$$\eta_{t,l}^{-} \le \tilde{M}z(\eta_{t,l}^{-}), \quad \bar{f}_l + (1-y_l)M + f_{t,l}^{\mathrm{spot}} \le (1 - z(\eta_{t,l}^{-}))\tilde{M},$$

$$\eta_{t,l}^{+} \le \tilde{M}z(\eta_{t,l}^{+}), \quad \bar{f}_l + (1-y_l)M - f_{t,l}^{\mathrm{spot}} \le (1 - z(\eta_{t,l}^{+}))\tilde{M}$$

for all $t\in T, l\in L$,

$$\kappa_{t,c}^{-} \le \tilde{M}z(\kappa_{t,c}^{-}), \quad d_{t,c}^{\mathrm{spot}} \le (1 - z(\kappa_{t,c}^{-}))\tilde{M}$$

for all $t\in T, n\in N, c\in C_n$,

$$\pi_{t,g}^{-} \le \tilde{M}z(\pi_{t,g}^{-}), \quad q_{t,g}^{\mathrm{spot}} \le (1 - z(\pi_{t,g}^{-}))\tilde{M}$$

$$\pi_{t,g}^{+} \le \tilde{M}z(\pi_{t,g}^{+}), \quad \tau\bar{q}_g^{\mathrm{new}} - q_{t,g}^{\mathrm{spot}} \le (1 - z(\pi_{t,g}^{+}))\tilde{M}$$

for all $t\in T, n\in N, g\in G_n^{\mathrm{new}}$, and

$$\nu_{t,g}^{-} \le \tilde{M}z(\nu_{t,g}^{-}), \quad q_{t,g}^{\mathrm{spot}} \le (1 - z(\nu_{t,g}^{-}))\tilde{M},$$

$$\nu_{t,g}^{+} \le \tilde{M}z(\nu_{t,g}^{+}), \quad \tau\bar{q}_g^{\mathrm{ex}} - q_{t,g}^{\mathrm{spot}} \le (1 - z(\nu_{t,g}^{+}))\tilde{M}$$

for $t\in T, n\in N, g\in G_n^{\mathrm{ex}}$.

Finally, after applying all reformulations and linearizations we end up with a single-level mixed-integer quadratic problem that is equivalent to the trilevel market model of Sect. 2.

3.2. **A Generalized Benders Decomposition Approach.** In general, the intrinsic difficulty of multilevel problems stem from the interdependence of the different levels, which makes the overall problem extremely hard to solve. Fortunately, after fixing the discrete (i.e., graph partitioning) variables of the first level, the considered trilevel problem exhibits a one-way dependence of the remaining two levels: Whereas the constraints of each lower-level problem depend on variables of the level above, the other way around does not appear. In fact, no first- or second-level constraint contains any respective lower-level variables. A similar observation holds for the objective functions. The second-level objective depends on second-level variables and the third-level objective depends on second- and third-level variables. Only the first-level objective interconnects all three levels; cf. Fig. 3.

We now exploit this weak coupling by iteratively computing a connected graph partition as a feasible solution of the first level. After fixing this partition, we solve the second level and then fix its solution in the third level. Finally, we use the solutions of the second and third level in order to compute the objective value of the first-level, i.e., of the overall trilevel problem. In principle, these steps need to be performed for every feasible partition.

In each iteration of this decomposition approach, a connected graph partition is determined and two convex-quadratic problems are successively solved. To prove correctness of such a decomposition algorithm, the optimal solution in the second level necessarily has to be unique as otherwise the third-level problem's solution would depend on an ambiguous second-level solution. Fortunately, uniqueness of the second level solution can be shown under certain assumptions that are satisfied in our case; see Grimm et al. [34] for a proof.

For the algorithmic details on globally optimizing the trilevel problem, we now present a problem-specific generalized Benders decomposition approach. These methods are extensions of the variable partitioning approach by Benders [7] for solving large linear programs as proposed in Geoffrion [29]; cf. Bonnans et al. [10] for an introduction. For briefly summarizing the generalized Benders approach, let us consider an optimization problem of the form

$$\max_{x,y} \quad f(x,y) \quad \text{s.t.} \quad g(x,y) \geq 0, \ (x,y) \in X \times Y,$$

where X and Y are assumed to be convex sets. This problem can be rewritten as the bilevel problem

$$\max_x \quad v(x) \tag{21a}$$

$$\text{s.t.} \quad v(x) = \max_y \{f(x,y) \colon g(x,y) \geq 0, \ y \in Y\}, \tag{21b}$$

$$x \in X \cap \{x \colon \exists y \in Y \text{ with } g(x,y) \geq 0\}. \tag{21c}$$

The main idea is that it is often easier to decompose Problem (21) and to optimize over the sets (21c) and Y separately than to solve the entire problem directly. We now assume that Problem (21b) is easy to solve, and we assume further that $v(x)$ is convex. The idea is to solve a relaxed version of (21) that ignores all but a few constraints of (21). If its solution also satisfies all ignored constraints, it is a solution of (21). Otherwise, violated constraints are added to the relaxed problem, which is then solved again. Let F denote the corresponding set of feasibility cuts, i.e., supporting hyperplanes of the feasible set (21c). Furthermore, we denote by O a set of optimality cuts, i.e., elements of the subdifferential of $v(x)$ for some given x. We

now consider the relaxation

$$\max_{x \in X} \quad \tau$$
$$\text{s.t.} \quad \tau \le a^\top x + b \quad \text{for all } (a,b) \in O, \tag{22}$$
$$d \le c^\top x \quad \text{for all } (c,d) \in F$$

of Problem (21). Problem (22) is called the master problem. A solution of (21) is determined by alternately solving master problems and subproblems of the form (21b). First, we solve the master problem and obtain a solution $\hat{x} \in X$. If (21b) is infeasible for fixed $\hat{x}$, a feasibility cut is added to F in (22). If (21b) has an optimal solution for fixed $\hat{x}$, an optimality cut is added to O in (22). The difficulty is the generation of the cut sets F and O. The general idea relies on knowledge of the dual of the subproblem. However, for the application considered here, the cuts are computed by exploiting problem-specific knowledge without appealing to duality. Assuming we can implement the required oracles for F and O, the overall procedure can be performed as displayed in Alg. 1.

Algorithm 1: Generalized Benders decomposition framework

---
1 Set $F \leftarrow \emptyset$, $O \leftarrow \emptyset$, $\Theta \leftarrow -\infty$, $\phi \leftarrow \infty$.
2 **while** $\Theta < \phi$ **do**
3   Solve master problem (22).
4   **if** (22) is infeasible **then** return *infeasible*
5   Let $\hat{x}$ be an optimal solution of (22) and set ϕ to its optimal value.
6   Solve subproblem (21b) with fixed $\hat{x}$.
7   **if** (21b) is *feasible* **then**
8   Let $\hat{y}$ be the optimal solution of (21b) and γ its optimal value.
9   **if** $\gamma > \Theta$ **then** set $\Theta \leftarrow \gamma$, $(x^*, y^*) \leftarrow (\hat{x}, \hat{y})$
10   Compute a subgradient s of v at $\hat{x}$ and add $(s, -s^\top \hat{x})$ to O.
11   **else**
12   Compute (c,d) so that $c^\top x \ge d$ separates $\hat{x}$ from (21c) and add (c,d) to F.
13 **return** (x^*, y^*)

Geoffrion [29] showed that this approach converges under suitable assumptions. However, Sahinidis and Grossmann [56] demonstrated that a naive application of the generalized Benders decomposition to nonconvex problems may not even lead to local optima since for general nonconvex problems the standard construction of optimality cuts only gives validity over the set (21c). In order to obtain global optima, validity over the whole set X is needed. We will show that such globally valid cuts can be obtained in our case.

Applying the basic ideas of the described decomposition approach to the trilevel problem we obtain the master problem

$$\max \quad \tau \quad \text{s.t.} \quad \tau \le a^\top x + b \quad \text{for all } (a,b) \in O, \tag{23a}$$
$$\text{graph partition with connectivity: (1)–(3).} \tag{23b}$$

We later show that no additional feasibility cuts F are needed. In our context, the subproblem consists of the original second- and third-level problem. However, after fixing the master problem's solution, this subproblem decomposes into two convex QPs that can be solved successively: (1) The zonal spot-market model with fixed zones (given as the result of the master problem) and (2) the redispatch model for fixed capacity investment, production, and demand decisions (given as the result of the zonal spot-market model).

In order to construct suitable optimality cuts that are valid for the whole feasible domain of (23b), we need some more notation. In what follows, let $\hat{x}$ be a connected partition and let $\psi_2(\hat{x})$ be the corresponding optimal zonal spot-market welfare, i.e., the optimal value of the second level for fixed first-level variables. Let $\psi_3(\hat{x})$ be the optimal value of the third level for fixed spot-market solutions in dependence of the fixed graph partition. Thus, the overall objective value of the trilevel market model for a fixed graph partition $\hat{x}$ is given by $\psi_1(\hat{x}) = \psi_2(\hat{x}) - \psi_3(\hat{x})$. We denote the optimal value of the integrated planner problem by ψ_{IGTC}^*.

We now show that

$$\tau \leq \psi_2(\hat{x}) - \psi_3(\hat{x}) + \psi_{\mathrm{IGTC}}^* \left(\sum_{i \in [k]} \sum_{n \in N : \hat{x}_{n,i} = 0} x_{n,i} + \sum_{i \in [k]} \sum_{n \in N : \hat{x}_{n,i} = 1} (1 - x_{n,i}) \right) \quad (24)$$

is a valid optimality cut. We first show that the optimal solution ψ_{IGTC}^* of the integrated planner (18) yields an upper bound for the optimal objective value ψ_1^* of the trilevel problem, i.e.,

$$\psi_1^* \leq \psi_{\mathrm{IGTC}}^*. \quad (25)$$

To this end, let $X^* = (x^*, q^*, d^*, f^*, (\bar{q}^{\mathrm{new}})^*, \theta^*)$ be part of an optimal solution of the trilevel problem, where q^*, d^*, f^*, and θ^* denote quantities after redispatch. Thus, we omitted all spot-market variable values in X^* except for the optimal capacity investment values $(\bar{q}^{\mathrm{new}})^*$. Then, $X = (q^*, d^*, f^*, (\bar{q}^{\mathrm{new}})^*, \theta^*)$ is feasible for the integrated planner problem (18) and, hence,

$$\psi_1^* = \psi_1(X^*) = \psi_{\mathrm{IGTC}}(X) \leq \psi_{\mathrm{IGTC}}(X_{\mathrm{IGTC}}^*) = \psi_{\mathrm{IGTC}}^*$$

holds, where X_{IGTC}^* is an optimal solution for the integrated planner problem (18).

The right-hand side of the cut (24) equals $\psi_1(\hat{x})$ for $x = \hat{x}$, otherwise it has a value of at least $2\psi_{\mathrm{IGTC}}^*$, as we assume $\psi_1(\hat{x}) \geq 0$ for all $\hat{x}$. Thus, by optimality of the master problem, the cut (24) ensures that whenever the master problem is solved it determines a partition that has not been examined yet (if one still exists). In order to bound the master problem in the first iteration, we initialize the set of optimality cuts O with the cut $\tau \leq \psi_{\mathrm{IGTC}}^*$.

The second-level problems as well as the third-level problems are always feasible. Furthermore, the master problem is feasible as we assume that the network is connected. Thus, feasibility cuts are not needed. An outline of the algorithm is given in Alg. 2. We show next that it is correct and that it terminates within a finite number of iterations.

Algorithm 2: Generalized Benders decomposition for the trilevel problem

Input: The trilevel problem
Output: A solution $(x^*, (q^{\mathrm{spot}})^*, (d^{\mathrm{spot}})^*, (\bar{q}^{\mathrm{new}})^*, q^*, d^*, f^*, \theta^*)$ for the trilevel problem.

1 Set $O \leftarrow \{(0, \psi_{\mathrm{IGTC}}^*)\}$, $\Theta \leftarrow 0$, $\phi \leftarrow \infty$.
2 **while** $\Theta < \phi$ **do**
3   Solve (23). Let $\hat{x}$ be its optimal solution, set ϕ to its optimal value.
4   Solve the second-level problem with fixed $\hat{x}$. Let $q^{\mathrm{spot}}, d^{\mathrm{spot}}$, and $\bar{q}^{\mathrm{new}}$ be part of its optimal solution and let $\psi_2(\hat{x})$ be its optimal value.
5   Solve the third-level problem with fixed $q^{\mathrm{spot}}, d^{\mathrm{spot}}$, and $\bar{q}^{\mathrm{new}}$. Let (q, d, f, θ) be the optimal solution and let $\psi_3(\hat{x})$ be its optimal value.
6   **if** $\psi_2(\hat{x}) - \psi_3(\hat{x}) > \Theta$ **then**
7   Set $\Theta \leftarrow \psi_2(\hat{x}) - \psi_3(\hat{x})$ and $(x^*, (q^{\mathrm{spot}})^*, (d^{\mathrm{spot}})^*, (\bar{q}^{\mathrm{new}})^*, q^*, d^*, f^*, \theta^*) \leftarrow (\hat{x}, q^{\mathrm{spot}}, d^{\mathrm{spot}}, \bar{q}^{\mathrm{new}}, q, d, f, \theta)$.
8   Add cut (24) to O.
9 **return** $(x^*, (q^{\mathrm{spot}})^*, (d^{\mathrm{spot}})^*, (\bar{q}^{\mathrm{new}})^*, q^*, d^*, f^*, \theta^*)$.

Theorem 1. *Assume that the social welfare $\psi_1(\hat{x})$ is non-negative for all $\hat{x}$, that the second-level problem's solutions $q(\hat{x}), d(\hat{x})$, and $\bar{q}^{new}(\hat{x})$ are unique for given $\hat{x}$, and that the network is connected. Then, Alg. 2 terminates within a finite number of iterations and returns a globally optimal solution for the trilevel problem.*

Proof. As argued above, every optimality cut (24) is valid over the whole feasible region (23b).

We now consider an iteration of Alg. 2 and assume that there exists at least one feasible solution of the master problem, i.e., a connected graph partition, that the algorithm has not considered in the preceding iterations. Suppose now that Alg. 2 computes a solution $\hat{x}$ that has already been computed before. Thus, the set O contains the corresponding optimality cut and from (24) it follows that the master problem's objective value ϕ has to satisfy $\phi \leq \psi_1(\hat{x})$. Moreover, by optimality of the master problem's solution $\hat{x}$, the latter is fulfilled with equality, i.e., $\phi = \psi_1(\hat{x})$. By assumption, there is another master problem solution x' that has not been considered so far. However, this yields $\tau \leq \psi_1(\hat{x}) + 2\psi_{\text{IGTC}}^*$ with a right-hand side larger than ϕ, which is, again by optimality, a contradiction. Thus, the algorithm does not consider a connected graph partition more than once.

Since the number of connected graph partitions is finite, it remains to prove that the algorithm terminates with an optimal solution if no connected graph partition x' exists that has not been considered so far. Let x^* be the solution for which $\Theta = \psi_2(x^*) - \psi_3(x^*)$ is maximal and let ϕ^* be the objective value of the corresponding master problem. Since no more connected graph partitions exist that have not been considered so far, an additional iteration would yield a master problem's objective value of $\phi = \Theta = \psi_2(x^*) - \psi_3(x^*)$ and hence, the termination criterion is satisfied. $\square$

As the number of iterations of this algorithm scales with the number of feasible connected graph partitions, the number of iterations grows prohibitively large, already for medium-sized instances. In the next section, we introduce algorithmic improvements that enhance the performance of the introduced solution approaches.

4. Enhanced Solution Techniques

In the last section we described two different solution strategies for the mixed-integer multilevel problem of Sect. 2. These approaches can, in principle, be used to solve practical instances. However, their computational performance might be unnecessarily weak. To strengthen their capabilities we discuss some enhanced solution techniques in this section.

4.1. **Breaking IP Symmetry.** The first-level problem uses binary variables to specify whether a node is contained in a specific zone or not. This leads to symmetric integer solutions that typically lead to a large number of branch-and-bound nodes in our MIQP approach or a large number of iterations in our Benders approach. It is therefore desirable to exploit techniques for breaking this symmetry.

Since the early 2000s, different approaches have been introduced to cope with integer symmetries. Among the first, Sherali and Smith [57] proposed simple symmetry breaking constraints. Another technique is to perturb the objective function as proposed by Ghoniem and Sherali [30]. More complex approaches exploit the structure of the branch-and-bound tree. Important contributions are isomorphism pruning developed in Margot [45, 46], orbital branching suggested by Ostrowski et al. [52], as well as orbital fixing by Kaibel et al. [39]. A recent overview is given in Margot [47] and, together with a computational comparison of different methods, in Pfetsch and Rehn [53]. For the ease of implementation and because they can be used both in our MIQP and the Benders approach we focus on symmetry breaking constraints in the following.

Several formulations of graph partitioning problems exist. One of them uses $|L|$ binary variables defined on the edges L of the graph. For each edge, the corresponding variable states whether the end points of that edge are in different zones or not. The alternative formulation that we use in the first-level problem requires $k|N|$ binary variables for a graph with $|N|$ nodes and k zones. For each node $n \in N$ and each zone $i \in [k]$, the node variable states whether the node belongs to the corresponding zone or not. Whereas symmetry does not play a major role in the first formulation, the one that we use suffers from many symmetric solutions. Indeed, given a solution of the graph partition problem as a matrix $Z = [x_{n,i}]_{n \in N}^{i \in [k]} \in \{0,1\}^{|N| \times k}$ it is easy to see that any permutation of the zones, i.e., of the columns of Z, results in another feasible solution. Clearly, this fact leads to many unnecessary branchings in standard IP solvers.

Nevertheless, the node formulation cannot easily be avoided in our context. In fact, if all we had to know is whether a line is an inter-zonal line or not, a graph partition model with edge variables would suffice. Instead, we (i) need node variables in our multi-commodity flow formulation (2), (3) for modeling connectedness of all partitions and (ii) for the formulation of the zonal spot-market model (6). This is why we chose the formulation using node variables together with methods to break the resulting IP symmetry.

Another way to view IP symmetry is by group theory. To be more specific, the symmetric group $\mathcal{S}_k$ of order k acts on solutions Z by permuting the columns of Z. In our context, it is easy to see that $\mathcal{S}_k$ acts in such a way that the corresponding second-level problem (9) is invariant along every orbit of the group. Thus, when considering, e.g., the Benders approach discussed in Sect. 3.2, it is obvious that the algorithm will perform many redundant iterations because many second-level problems are parameterized by symmetric first-level solutions. Therefore, the effect of symmetry breaking can be highly significant (even on small network instances).

In the rest of this section we review all types of symmetry breaking constraints that we later compare in Sect. 5 w.r.t. their computational behavior. One central idea of symmetry breaking constraints is to make use of a lexicographic ordering of the representatives of the orbits and to cut off all points from each orbit except the one that is the maximal point w.r.t. the ordering. From now on we assume that the node set is given as $N = \{1, \ldots, |N|\}$. A solution $Z \in \{0,1\}^{|N| \times k}$ of a general graph partitioning problem is called maximal w.r.t. to a lexicographic ordering if and only if the columns of Z are in non-increasing lexicographical order, i.e., if they satisfy

$$\sum_{n=1}^{|N|} 2^{|N|-n} x_{n,i} \geq \sum_{n=1}^{|N|} 2^{|N|-n} x_{n,i+1} \quad \text{for all } i \in [k-1]. \tag{26}$$

In fact, these inequalities single out the lexicographically maximal representative from each orbit from the set of all 0/1-matrices with exactly one 1-entry per row. Hence, we can add these inequalities to the first-level problem in order to break the IP symmetry. However, although the number of additional inequalities is small compared to the total number of constraints of the problem, Constraints (26) are likely to result in numerical problems for larger networks due to the order of magnitude of the coefficients $2^{|N|-n}$.

In what follows we make use of the fact that a lexicographical ordering always yields a zero upper triangle of Z, i.e., we have $x_{n,i} = 0$ for all $i > n$. A different approach than (26) to ensure lexicographical order has been proposed by Méndez-Díaz and Zabala in [48, 49]. They used the inequalities

$$x_{n,i} - \sum_{\ell=1}^{n-1} x_{\ell,i-1} \leq 0 \quad \text{for all } 2 \leq i \leq n \tag{27}$$

to cut off all points of each orbit except the lexicographical maximal. This formulation is expected to be numerically more stable than (26). However, we have to add $\mathcal{O}(k|N|)$ inequalities instead of only $\mathcal{O}(k)$.

A strengthening of these inequalities is given by so-called column inequalities

$$\sum_{\ell=i}^{k} x_{n,\ell} - \sum_{\ell=1}^{n-1} x_{\ell,i-1} \leq 0 \quad \text{for all } 2 \leq i \leq n, \tag{28}$$

cf. Kaibel and Pfetsch [40]. It is known that the column inequalities (28) are not strong enough to obtain complete descriptions of partitioning orbitopes. To this end, one needs so-called shifted column inequalities that build a substantially richer class of exponentially many (in k) inequalities. Their description is quite technical and we refer the interested reader to Kaibel and Pfetsch [40] for the details. Since the exponential number of shifted column inequalities is not prohibitively large in our application, we also statically add all constraints to the models as we do for (26), (27), and (28).

4.2. **Primal Heuristics.** It is folklore knowledge that highly symmetric (mixed-)integer problems like ours suffer from the fact that primal feasible solutions of good quality are typically found quite late in the branch-and-bound process because solutions of the LP relaxations typically contain many fractional entries. This is why we now discuss primal heuristics in order to determine good feasible solutions early in the process.

For the maximum k-cut problem, randomized greedy heuristics such as the GRASP heuristic introduced in Festa et al. [25] as well as variable-neighborhood search methods as proposed in Hansen and Mladenović [36] are known to generate good solutions quickly. These methods can easily be extended to partition the graph into more than two parts. Good graph partitions can be extracted from approximation algorithms based on positive semidefinite optimization; cf. Goemans and Williamson [31] for maximum 2-cut and Frieze and Jerrum [27] for the generalization to maximum k-cut. Recently, several effective solution algorithms and implementations have been developed for graph partitioning and graph clustering problems in Bader et al. [4], where the focus lies on partitioning very large unweighted graphs.

In the following, we will present two primal heuristics for the trilevel problem we study here.

4.2.1. *A Minimum k-Cut Approximation Heuristic.* The key idea of the heuristic is simple. If one neglects the economic data of producers and consumers, bottlenecks in the network mainly depend on the capacity of the transport lines. Since infeasible spot-market outcomes typically tend to violate these capacity constraints it is reasonable to consider zonings that stem from minimum k-cuts in the capacitated network.

The algorithmic realization of this idea is straight forward and based on Chap. 4 of Vazirani [61]: First, we compute a Gomory–Hu tree T of the network graph $\mathcal{G}$ and then delete the $k-1$ lightest edges (w.r.t. their capacity) of T. These Gomory–Hu tree edges correspond to cuts in G. In Vazirani [61] it is shown that this yields a k-cut in the original graph G and that this k-cut approximates a minimum k-cut with approximation factor $2 - 2/k$.

Applied to a connected graph, this procedure yields a connected graph partition that can be directly translated into an assignment of the first-level variables $x_{n,i}$ and y_l. Given these values it is then straight forward to construct feasible values of all other variables of the first level. However, in the MIQP approach, it is reasonable to fix all other variables as well—or, at least, the remaining discrete variables that occur due to the linearization of the KKT complementarity conditions. This can be done by solving the second-level problem for fixed first-level variables (that are outcome of the heuristic) and by finally solving the third-level problem for fixed second-level variables.

Both completion steps require solving a QP. By doing so, we have used the key idea of the Benders approach of Sect. 3.2: Fixing the first-level discrete decisions yields decoupled second- and third-level problems that can be solved successively. Thus, we could also use the presented heuristic in the first iteration of the Benders approach; cf. Sect. 3.2.

4.2.2. *A Relaxation-Based Rounding Heuristic.* We now present a relaxation-based rounding heuristic. Since there is no suitable relaxation involved in the Benders decomposition, this heuristic can only be used in the MIQP approach.

The relaxation-based rounding heuristic is formally stated in Alg. 3. Let $\hat{x}_{n,i} \in [0,1]$ for $n \in N, i \in [k]$, be part of a relaxation solution of the single-level MIQP of Sect. 3.1. We interpret these relaxation solutions as the probability that node n should be part of zone i. The idea is now as follows. First, we sort the vector $\hat{x} \in [0,1]^{k|N|}$ in descending order (Line 1). Moreover, we encode the information to which pair of node and zone each entry of the sorted vector belongs using the mappings ν and ζ, which map every entry to the corresponding node and to the corresponding zone, respectively. For instance, assume the entry $\hat{x}_{n,i}$ has index $\alpha \in \{1, \ldots, k|N|\}$ after sorting. Then, $\nu(\alpha) = n$ and $\zeta(\alpha) = i$ holds. Next, we assign to every zone the node with the highest probability of being assigned to that zone (while-loop). The set M used in the algorithm collects all nodes that have already been associated to a zone. Afterward, we again iterate over the sorted vector of relaxation solutions and assign every node that is not yet assigned to a zone if this assignment does not violate connectivity (for-loop). It is possible that in early iterations of Alg. 3 a node

---

Algorithm 3: A relaxation-based rounding heuristic

Input: A vector $\hat{x} = (\hat{x}_{n,i})_{n \in N}^{i \in [k]} \in [0,1]^{k|N|}$.

1 Sort the vector $\hat{x}$ in descending order, yielding the vector $(x_\ell)_{\ell \in [k|N|]}$ with $\nu(\ell) = n \in N$ and $\zeta(\ell) = i \in [k]$.

2 Set $Z_i = \emptyset$ for all $i \in [k]$, $\ell = 1$, and $M = \emptyset$.

3 **while** $\exists i \in [k]$ with $Z_i = \emptyset$ and $\nu(\ell) \notin M$ **do**

4   **if** $Z_{\zeta(\ell)} = \emptyset$ **then** set $Z_{\zeta(\ell)} = \{\nu(\ell)\}$ and $M \leftarrow M \cup \{\nu(\ell)\}$.

5   Set $\ell \leftarrow \ell + 1$.

6 **for** $\ell = 1, \ldots, k|N|$ with $\nu(\ell) \notin M$ **do**

7   **if** $\nu(\ell) \in \delta(Z_{\zeta(\ell)})$ **then** set $Z_{\zeta(\ell)} \leftarrow Z_{\zeta(\ell)} \cup \{\nu(\ell)\}$ and $M \leftarrow M \cup \{\nu(\ell)\}$

---

cannot be assigned to a favorable zone because it is not yet connected to any other node of that zone. Thus, we subsequently apply local improvement steps in which we iteratively check for each node whether it should be assigned to a different zone than to the one to which it is currently assigned to. The method is given in Alg. 4. More formally, suppose a node n is assigned to zone i. If the value of the relaxation $\hat{x}_{n,i}$ is smaller than $\hat{x}_{n,j}$ for some other zone $j \neq i$ (cf. Line 5), then n is moved from zone i to j if this leaves all partitions non-empty and if the resulting node partition is still connected (cf. Line 6). As it is the case for the Gomory–Hu based minimum k-cut heuristic, the rounding heuristic (with or without the 1-opt improvements of Alg. 4) yields a connected graph partition that can be used to compute the remaining variables as explained in Sect. 4.2.1.

4.3. **Algorithmic Improvements for the Benders Decomposition Approach.** In case the first-level problem contains objective function terms that directly depend on first-level variables, additional optimality cuts can be added to reduce the overall number of iterations of the Benders approach of Sect. 3.2. The description of the first-level problem in Sect. 2.2 does not have any such first-level terms in the objective function ψ_1. However, in practice, such non-zero contributions may be present if, e.g.,

---

Algorithm 4: A 1-opt improvement heuristic

Input: A vector $\hat{x} = (\hat{x}_{n,i})_{n \in N, i \in [k]} \in [0,1]^{k|N|}$ and a connected graph
 partition $N = \cup_{i=1}^k Z_i$.
Output: A connected graph partition $N = \cup_{i=1}^k Z_i$.

1 improved = true
2 **while** improved **do**
3   improved = false
4   **for** $i = 1, \ldots, k$ with $|Z_i| > 1$ **do**
5   **if** $\exists n \in Z_i, j \in [k], i \neq j$, with $\hat{x}_{n,j} > \hat{x}_{n,i}$ **then**
6   **if** $Z_i \setminus \{n\}$ and $Z_j \cup \{n\}$ are connected **then**
7   Set $Z_i \to Z_i \setminus \{n\}$, $Z_j \to Z_j \cup \{n\}$, and improved $\to$ true

---

one wants to penalize undesired shapes of the resulting zones. This can be realized by so-called acceptance costs for every pair of nodes that may, e.g., be proportional to the geographical proximity of each pair of nodes because nodes that are geographically "close" are more desirable to be contained in one zone than those that are "far apart" from each other.

To this end, we impose additional first-level variables x_i^{acc} that represent the acceptance costs for each zone $i \in [k]$. They are determined by the largest acceptance costs data $c_{n,m}^{\mathrm{acc}} \geq 0$ for any two nodes n and m within the zone i. Expressed in quadratic constraints, this leads to

$$x_i^{\mathrm{acc}} \geq c_{n,m}^{\mathrm{acc}} x_{n,i} x_{m,i} \quad \text{for all } i \in [k], \ n, m \in N, \ n < m. \tag{29}$$

For each binary quadratic expression of the form $x_{n,i} x_{m,i}$, we add an auxiliary binary variable $x_{n,m,i}$ together with the linearization constraints

$$x_{n,m,i} \geq x_{n,i} + x_{m,i} - 1, \quad x_{n,m,i} \leq x_{n,i}, \quad x_{n,m,i} \leq x_{m,i}, \tag{30}$$

which enables us to linearize (29) as

$$x_i^{\mathrm{acc}} \geq c_{n,m}^{\mathrm{acc}} x_{n,m,i} \quad \text{for all } i \in [k], \ n, m \in N, \ n < m. \tag{31}$$

The objective function of the first-level problem then is rephrased to

$$\psi_1 \leftarrow \psi_1 - \psi_{\mathrm{acc}}, \quad \psi_{\mathrm{acc}} := \sum_{i \in [k]} x_i^{\mathrm{acc}}$$

to prioritize graph partitions with low acceptance costs. We note that $\mathcal{O}(k|N|^2)$ constraints of the form (30) and (31) are added, which makes the first-level problem more difficult. However, this setting enhances the decomposition approach as additional optimality cuts can be added in the first level. The goal is to cut off partitions with high acceptance costs that cannot improve the overall objective function value. This can be modeled with constraints of the form

$$\psi_{\mathrm{IGTC}}^* - \psi_{\mathrm{acc}}(x) \geq \psi_1^{\mathrm{inc}},$$

where ψ_1^{inc} is the best known objective value of the trilevel problem and $\psi_{\mathrm{acc}}(x)$ denotes the acceptance costs for a given connected partition x. As no acceptance costs are present in the integrated planner model, its optimal value ψ_{IGTC}^* is again a valid upper bound for ψ_1; cf. (25).

In the following, we informally describe how acceptance costs are embedded into the Benders decomposition approach of Sect. 3.2. The adopted master problem in the

generalized Benders approach reads

$$\max \quad \tau - \psi_{\mathrm{acc}}$$
$$\text{s.t.} \quad \tau \leq a^\top x + b \quad \text{for all } (a, b) \in O,$$
$$\text{first-level constraints: (1)–(3),}$$
$$\text{(linearized) acceptance costs: (30), (31).}$$

Optimality cuts (24) can be added to the master problem—they remain valid for the whole feasible set of the master problem without adjustments.

Taking these modifications into account, one can easily extend the generalized Benders approach in Alg. 2 to the case of acceptance costs. We remark that it is needed to restrict τ from above by ψ_{IGTC}^*. Thus, we initialize O again with the cut $\tau \leq \psi_{\mathrm{IGTC}}^*$. The decomposition algorithms works as follows. We alternately solve master- and subproblems. As long as there exist yet unexplored partitions with acceptance costs lower than $\psi_{\mathrm{IGTC}}^* - \Theta$, the master problem provides a new partition x' with minimum acceptance cost $\psi_{\mathrm{acc}}(x')$. The reason is that for x' the right-hand side of every optimality cut in set O is equal to or greater than ψ_{IGTC}^*, as x' has not yet been investigated. Hence, the initial cut $\tau \leq \psi_{\mathrm{IGTC}}^*$ becomes active and we obtain the master problem's objective value of $\phi = \psi_{\mathrm{IGTC}}^* - \psi_{\mathrm{acc}}(x') > \Theta$.

On the other hand, if acceptance costs exceed $\psi_{\mathrm{IGTC}}^* - \Theta$ for every yet unexplored x', the objective value of the master problem falls below Θ when choosing such a solution. Hence, an already investigated solution is chosen and an optimality cut becomes active. In this case, a partition x^* is determined that yields the master objective value $\phi = \Theta$. Thus, the stopping criterion is reached and a solution is returned. Since we compute partitions in non-decreasing order with respect to acceptance costs, it is correct to stop when $\psi_{\mathrm{acc}}(x')$ exceeds $\psi_{\mathrm{IGTC}}^* - \psi$ for the first time since no more connected partition with lower acceptance costs exist.

The actual number of iterations clearly depends on the choice of the parameters $c_{n,m}^{\mathrm{acc}}$ and is crucial for the efficiency of the algorithm; a detailed discussion of the choice of the parameters $c_{n,m}^{\mathrm{acc}}$ is given in the thesis [41] by Kleinert.

5. Computational Study

In this section we present an extensive computational study of the presented global solution approaches of Sect. 3 as well as of the enhanced techniques discussed in Sect. 4. To this end, we first describe the computational setup and the test instances in Sect. 5.1. Afterward, we present the results for the MIQP approach in Sect. 5.2 and finally discuss the results of the generalized Benders approach in Sect. 5.3, where we also compare both approaches.

5.1. **Test Instances and Computational Setup.** We use a set of test instances that divides into purely academic instances as well as realistic instances that model the German electricity market. All test networks together with their graph sizes and the size of the corresponding scenario sets are given in Table 1. The first three networks Grimm-et-al-2015-3, Chao-Peck-1998, and Grimm-et-al-2016-6 are test instances from the literature on energy economics; the references are given in Table 1 as well. The DE networks all are based on the policy report [32] by Grimm et al. The DE-28 network is exactly the network that is used in the report. It models each of the 16 federal states of Germany as single nodes and also includes 12 additional nodes for the neighboring countries. More information about the network and the scenario data of this network is given in [32]. We note that the policy report uses a scenario set of size 8760 (which corresponds to an hourly discretized year) that we reduced to a scenario set of size 52 (which corresponds to a weekly discretized year). The other DE networks DE-09, DE-12, DE-16, and DE-23 have been constructed by aggregating nodes and lines of the original DE-28 network. All models that we have to solve within our solution

TABLE 1. Test networks with number of nodes ($|N|$), number of lines ($|L|$), and the number of scenarios ($|T|$)

Network name	$|N|$	$|L|$	$|T|$	Reference
Grimm-et-al-2015-3	3	3	4	Grimm et al. [33]
Chao-Peck-1998	6	6	4	Chao and Peck [14]
Grimm-et-al-2016-6	6	6	52	Grimm et al. [33]
DE-09	9	19	52	—
DE-12	12	23	52	—
DE-16	16	27	52	—
DE-23	23	39	52	—
DE-28	28	39	52	Grimm et al. [32]

approaches are MIQPs, MIPs, or convex QPs and we use Gurobi 6.5.2 [35] for solving the instances. All models as well as the primal heuristics and the generalized Benders approach have been implemented in Python 2.7.6 using the graph library NetworkX [51] contained in Anaconda 2.7 [2]. All computational experiments have been executed on a compute cluster; cf. [54] for the details about installed hardware. The time limit for all computations is set to 2 h.

Throughout this section we use log-scaled performance profiles as proposed by Dolan and Moré [22] to compare running times or branch-and-bound node counts.

5.2. **The MIQP Approach.** We now present and discuss the numerical results obtained with the MIQP approach described in Sect. 3.1. Since it turned out that it is very challenging to solve the mixed-integer trilevel problem with this approach, we deleted both the DE-23 and the DE-28 network from the test set used in this section. The academic instances are solved for all possible number of zones $k \in \{1, \ldots, |N|\}$, whereas we reduced the number of zones to $\{1, 2, 3, 7, 8, 9\}$ for the DE-09 network, to $\{1, 2, 10, 11, 12\}$ for the DE-12 network, and to $\{1, 2, 14, 15, 16\}$ for the DE-16 network. This leads to an overall number of 31 test instances.

The number of (connected) graph partitions for a given network obviously is highest for mid-level sizes k whereas the number is small for k close to 1 or close to $|N|$. Since it is not possible to compute global optimal solutions of the trilevel model using the MIQP approach for mid-level k within the given time limit of 2 h we deleted some of these mid-level k test instances in order to get a reasonable test set for comparing the impact of different enhanced techniques used within the MIQP approach.

Before we discuss the numerical results in detail we further note that we did some tuning of Gurobi's parameters but will not discuss the details here.[1]

We start by discussing the impact of different symmetry breaking constraints. We again note that we implemented the symmetry breaking constraints by statically adding all of them to the MIQP that is to be solved. We also tested an implementation of symmetry breaking constraints as lazy constraints in Gurobi. However, this implementation neither yields an improvement in running times nor in the number of solved instances.

Fig. 4 displays the corresponding log-scaled performance profiles. The first and obvious observation is that using "any" type of symmetry breaking constraints is crucial for the performance of the method: The performance profile of Gurobi applied to the MIQP without using symmetry breaking constraints is clearly dominated by all other profiles. Comparing the symmetry breaking constraints with each other, no distinct winner can be designated. Independent of the specific symmetry breaking constraints, one obtains a global optimal solution for slightly less than 50 % of all tested instances

---

[1]The used parameters are `Heuristics = 0`, `Cuts = 0`, `VarBranch = 1`, `MIPFocus = 2`.
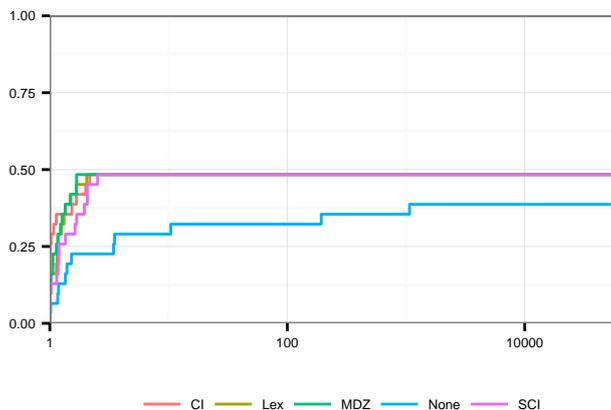
FIGURE 4. Log-scaled performance profiles of all tested types of symmetry breaking constraints within the the MIQP approach; (S)CI: (shifted) column inequalities, Lex: lexicographic ordering, MDZ: Méndez-Díaz and Zabala
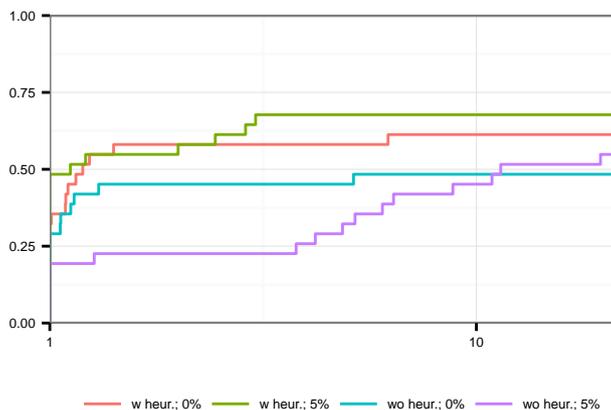

FIGURE 5. Log-scaled performance profiles for the MIQP approach with symmetry breaking constraints of Méndez-Díaz and Zabala, (de-)activated root node heuristic (cf. Sect. 4.2.1), as well as 0 and 5 % MIQP gap

whereas the MIQP without symmetry breaking constraints can only be solved for 38 %. Since the profile corresponding to the constraints (27) of Méndez-Díaz and Zabala seems to dominate the other profiles for the largest range of the profile in which not all constraint types behave equal, we choose to use these constraints in the sequel.

The small amount of solved instances, especially for the MIQP without symmetry breaking constraints, also indicates the hardness of the MIQP instances. First of all, also for small networks and moderate number of scenarios, the resulting MIQP is quite large. For instance, the 6-node network Grimm-et-al-2016-6 with 52 scenarios leads to an MIQP with 20 167 constraints, 8699 variables (out of which 1332 are binary), 49 932 nonzeros, and 156 quadratic objective terms. Moreover, we have many symmetric solutions and a lot of numerically challenging Big-M formulations for linearizing KKT complementarity conditions.

Next, we analyze the impact of the minimum k-cut root node heuristic described in Sect. 4.2.1. The performance profiles are given in Fig. 5, where we compare the MIQP equipped with the symmetry breaking constraints of Méndez-Díaz and Zabala and with and without the root node heuristic. Moreover, we also plot the resulting
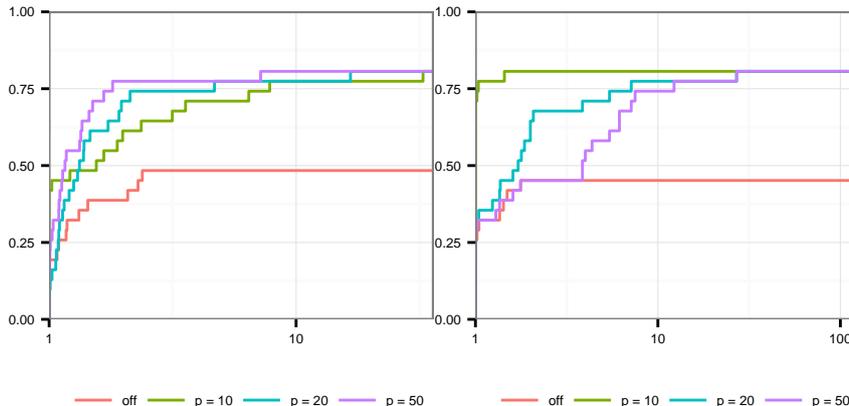
FIGURE 6. Log-scaled performance profiles of the rounding heuristic applied every $p \in \{10, 20, 50, 100\}$ branch-and-bound nodes within the MIQP approach with symmetry breaking constraints of Méndez-Díaz and Zabala and deactivated root node heuristic; cf. Sect. 4.2.2. Left: running times. Right: Node count

profiles if we stop Gurobi whenever the MIQP gap falls below 5 %. Taking first a closer look at the results using a 0 % gap tolerance we see that the usage of the root node heuristic both leads to shorter running times as well as to significantly more solved instances (less than 50 % vs. approximately 62 %). Comparing the profiles for the MIQP runs using a 5 % gap tolerance the spread becomes even more visible. Despite the fact that the root node heuristic is not as lightweight than many other root node heuristics applied in mixed-integer optimization, the application of the heuristic leads to a significant improvement in running times as well as in the overall robustness of the approach. This also indicates that the quality of solutions obtained by the minimum k-cut heuristic is quite good on the considered test set.

We now turn to the evaluation of the rounding node heuristic presented in Sect. 4.2.2. We tested the rounding heuristic for different parameterizations; namely whether the heuristic is applied all 10, 20, or 50 nodes in the branch-and-bound tree. In Fig. 6 we see the corresponding performance profiles both for running times (left) as well as for the number of branch-and-bound nodes (right). We again used the symmetry breaking constraints of Méndez-Díaz and Zabala and de-activated the root node heuristic in order to exclusively measure the impact of the rounding heuristic. The figure shows that the positive impact of the rounding heuristic is significant both w.r.t. running times as well as to the overall number of solved instances. A comparison of both performance profiles also reveals that there is an important trade-off between the number of branch-and-bound nodes solved and the running times. For instance, the parameterization with $p = 10$ solves significantly less branch-and-bound nodes than the one applied every 50 nodes. However, it turns out that the heuristic (and, in particular, its completion steps in which at least one QP has to be solved in order to assign feasible values to all binary variables of the MIQP) is too expensive for being applied as often as in every 10th branch-and-bound node. In summary, the left plot in Fig. 6 shows that $p = 50$ leads to the best method.

We also compared the combined application of the root node heuristic together with the rounding heuristic for different p. However, it turned out that by using the rounding heuristic, the root node heuristic does not have a significant impact on the performance and overall robustness of the MIQP approach. Thus, the parameterization of the MIQP using the symmetry breaking constraints of Méndez-Díaz and Zabala together with the rounding heuristic applied with $p = 50$ leads to the best variant of the MIQP approach. Again looking at the left plot in Fig. 6 we see that this
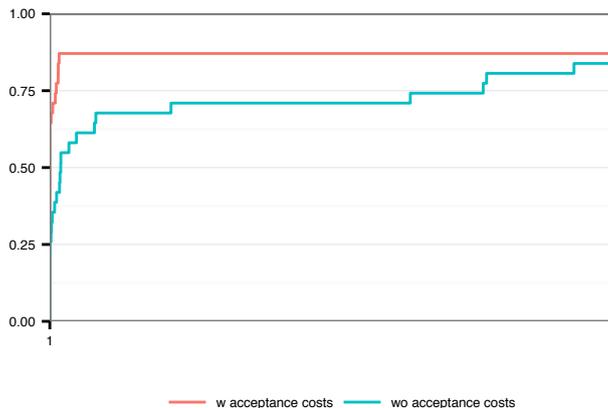
FIGURE 7. Log-scaled performance profiles for the MIQP models with and without acceptance costs solved with deactivated root node heuristic, with the rounding heuristic applied every $p = 50$ nodes, and symmetry breaking constraints of Méndez-Díaz and Zabala

variant solves approximately $81\,\%$ of all test instances. In particular, this means that the enhanced solution techniques improved the plain MIQP approach (which solved approx. $38\,\%$) by a factor larger than 2.

Although we introduced zonal acceptance costs in Sect. 4.3 in order to enhance the performance of the generalized Benders approach, we now finally analyze the impact of these first-level cost terms within the MIQP approach; see Fig. 7 for the corresponding performance profiles. It is apparent that the integration of acceptance costs makes the MIQP comparably easier to solve for Gurobi and that even a larger number of instances can be solved to global optimality at all.

5.3. **Results for the Generalized Benders Decomposition and a Comparison of both Approaches.** During our numerical experiments it turned out that the generalized Benders decomposition approach is by far faster than the MIQP approach. This was to be expected. First of all, the decomposition leads to master and subproblems that are much easier to solve than the single-level MIQP. The master problem is a—for the instances of our test set—quite small mixed-integer linear graph partitioning problem with connectivity constraints and the two successively solvable convex QPs can be solved efficiently.

This is why we extend our test set significantly compared to the analysis of the preceding section. We now consider all test networks listed in Table 1 and also include the instances for all $k \in \{1, \ldots, |N|\}$ for all networks except for DE-23 and DE-28, where we only consider $k \in \{1, 2, 3, 4, 20, 21, 22, 23\}$ for DE-23 and $k \in \{1, 2, 3, 4, 25, 26, 27, 28\}$ for DE-28. That is, we now consider a test set of 68 instances that is more than twice as large as the test set for the MIQP approach in Sect. 5.2 and that is a proper superset of the MIQP test set.

All master problems solved within the generalized Benders approach are equipped with the symmetry breaking constraints (27) of Méndez-Díaz and Zabala. We again tuned Gurobi's parameters for the master problem as well as for both subproblems but again will not discuss the details here.[2]

We now analyze the overall success rate of the generalized Benders approach and compare the impact of acceptance costs as first-level objective function terms. The respective performance profiles can be seen in Fig. 8. First of all, we see that both the

_____

[2]The used parameters for the master problems are `Threads = 4`, `Symmetry = 0`, and `Cuts = 0`. For the zonal spot-market QP we used `Threads = 4`, `PreSolve = 1`, and `NumericFocus 2`. The parameters used for the redispatch QP are `Threads = 4` and `PreSolve = 1`.
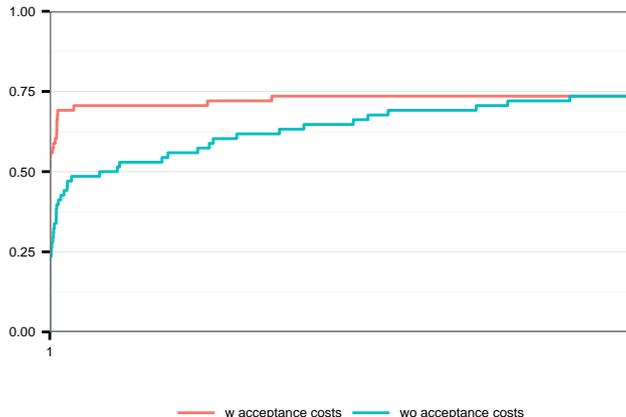
FIGURE 8. Log-scaled performance profiles comparison of the generalized Benders approach with and without acceptance costs

variant with and without acceptance costs leads to an overall success rate of slightly less than 75 %. This means that the incorporation of acceptance costs does not lead to a larger number of solved instances. However, the variant using acceptance costs and thus stronger optimality cuts within the Benders framework clearly outperforms the variant without acceptance costs. Moreover, it is possible to choose acceptance costs that dominate the overall objective more strongly than the costs that we use in our computations. These larger acceptance costs may than, of course, also lead to a larger number of solved instances.

We also remark that the Benders approach is much more stable in terms of numerics than the MIQP approach. When testing the MIQP approach with the global MIQP solvers Gurobi, CPLEX [17], and SCIP [1] we observed almost every imaginable behavior ranging from primal constraints only satisfied with comparably large feasibility tolerances to wrong objective values and false infeasibilities. Moreover, these outcomes do not only vary between different solvers but also between different versions of the same solver. There are two main reasons for these problems: First, the Big-M reformulation of KKT complementarity (20) and, second, the bounding of genuinely free dual variables in the dual feasibility conditions (19) that is required for linearizing the products of binary and continuous variables.

In contrast to this, due to the applied decomposition, the problems that have to be solved within the generalized Benders decomposition method do not include any of these issues and can, thus, be solved in much more stable way.

Finally, we compare the MIQP approach with the generalized Benders decomposition in terms of absolute running times. For both approaches, the results are given in Table 2 and 3 for the academic instances and in Table 4 and 5 for the realistic DE instances. We see that there are already two instances out of the small instances of Table 2 that cannot be solved using the MIQP approach while the generalized Benders approach never needs significantly more than 1 s for solving the instances to global optimality. When using acceptance costs, cf. Table 3, the MIQP approach solves all but one of the smaller instances and the Benders approach solves every instance in less than 0.52 s. Thus, the Benders approach drastically outperforms the MIQP on the small-scale instances. Considering the larger instances in Table 4 and 5 we can make two clear observations. First, the instances are much harder to solve. The MIQP approach only solves the instances with smallest and largest k for the DE-09, DE-12, and the DE-16 network. The mid-level k instances cannot be solved. This is, in principle, also true for the generalized Benders approach. Second, the Benders approach solves more instances and the instances that can be solved by both

TABLE 2. Comparison of absolute running times (in s) of the MIQP and generalized Benders approach (without acceptance costs) on academic instances

Network	Zones	MIQP	Benders
Chao-Peck-1998	1	0.04	0.04
Chao-Peck-1998	2	10.39	0.16
Chao-Peck-1998	3	26.22	0.30
Chao-Peck-1998	4	17.93	0.30
Chao-Peck-1998	5	9.54	0.13
Chao-Peck-1998	6	0.11	0.06
Grimm-et-al-2015-3	1	0.02	0.02
Grimm-et-al-2015-3	2	0.05	0.04
Grimm-et-al-2015-3	3	0.02	0.03
Grimm-et-al-2015-6	1	0.06	0.23
Grimm-et-al-2015-6	2	222.58	0.57
Grimm-et-al-2015-6	3	—	1.06
Grimm-et-al-2015-6	4	—	0.97
Grimm-et-al-2015-6	5	6834.90	0.56
Grimm-et-al-2015-6	6	54.65	0.37

TABLE 3. Comparison of absolute running times (in s) of the MIQP and generalized Benders approach (with acceptance costs) on academic instances

Network	Zones	MIQP	Benders
Chao-Peck-1998	1	0.04	0.03
Chao-Peck-1998	2	9.81	0.10
Chao-Peck-1998	3	25.84	0.24
Chao-Peck-1998	4	18.18	0.21
Chao-Peck-1998	5	9.32	0.11
Chao-Peck-1998	6	0.11	0.05
Grimm-et-al-2015-3	1	0.01	0.03
Grimm-et-al-2015-3	2	0.02	0.03
Grimm-et-al-2015-3	3	0.01	0.04
Grimm-et-al-2015-6	1	0.05	0.15
Grimm-et-al-2015-6	2	201.52	0.29
Grimm-et-al-2015-6	3	3334.64	0.50
Grimm-et-al-2015-6	4	—	0.52
Grimm-et-al-2015-6	5	5261.35	0.45
Grimm-et-al-2015-6	6	17.62	0.24

approaches are solved much faster by the Benders approach. The same holds for the case of acceptance costs, where it can be also seen that both approaches benefit from the additional first-level objective terms.

In summary, the Benders decomposition significantly outperforms the MIQP approach. However, there is still room for improvement not only for the MIQP approach but also for the Benders approach because there are still instances that cannot be solved within the given time limit of 2 h.

TABLE 4. Comparison of absolute running times (in s) of the MIQP and generalized Benders approach with (right) and without (left) acceptance costs on small realistic instances

Network	Zones	MIQP	Benders	Network	Zones	MIQP	Benders
DE-09	1	13.20	0.78	DE-09	1	13.07	0.81
DE-09	2	1656.49	5.20	DE-09	2	1652.11	5.16
DE-09	3	4081.41	33.97	DE-09	3	4164.11	33.82
DE-09	4	—	61.13	DE-09	4	—	60.38
DE-09	5	—	65.13	DE-09	5	—	65.79
DE-09	6	—	36.36	DE-09	6	—	36.27
DE-09	7	3645.49	12.54	DE-09	7	3689.02	12.42
DE-09	8	75.11	3.12	DE-09	8	76.47	3.09
DE-09	9	8.52	1.33	DE-09	9	8.49	1.30
DE-12	1	4.85	0.82	DE-12	1	4.73	0.79
DE-12	2	3470.03	19.60	DE-12	2	3468.23	19.41
DE-12	3	—	434.34	DE-12	3	—	439.33
DE-12	4	—	6097.09	DE-12	4	—	6159.36
DE-12	5	—	—	DE-12	5	—	—
DE-12	6	—	—	DE-12	6	—	—
DE-12	7	—	—	DE-12	7	—	—
DE-12	8	—	3684.91	DE-12	8	—	3428.91
DE-12	9	—	334.37	DE-12	9	—	332.64
DE-12	10	—	38.60	DE-12	10	—	38.35
DE-12	11	303.54	4.99	DE-12	11	297.22	4.99
DE-12	12	10.75	1.66	DE-12	12	10.69	1.67

6. CONCLUSION

In this paper we presented a mixed-integer nonlinear trilevel model for computing welfare-optimal price zones in electricity markets. For problems of this kind no general-purpose solution algorithms exist. Thus, we developed two different global solution approaches. One is based on the reduction of levels using problem-specific insights as well as a standard KKT transformation. The other one is a problem-specific instantiation of generalized Benders decomposition. We then additionally presented enhanced solution techniques like, e.g., symmetry breaking constraints and problem-tailored primal heuristics.

Our computational results show that the generalized Benders framework significantly outperforms the approach yielding a single-level but large and numerically challenging MIQP. Using the techniques developed in this paper, it is now possible to address the application problem on a network size that in principle allows to draw conclusions for, e.g., the German electricity market. However, there is still a lot of room for improvement. The largest network that we could solve within a time limit of 2 h is a 28-node network. Obviously, one wants to further improve the performance of the methods such that it is possible to address problems on much larger graphs; for instance, to address questions of energy market design on the European scale.

To this end, two different branches of research may be followed in the future. First, one could study the actual model and try to develop equivalent formulations that have more desirable properties. For instance, models without IP symmetry are highly preferable. Moreover, numerically more stable mixed-integer linear reformulations of first-order optimality conditions of quadratic problems would lead to much more effective solution approaches. Second, one could study many algorithmic questions. For instance: How is it possible to improve the performance of branch-and-bound

TABLE 5. Comparison of absolute running times (in s) of the MIQP and generalized Benders approach with (right) and without (left) acceptance costs on large realistic instances

Network	Zones	MIQP	Benders	Network	Zones	MIQP	Benders
DE-16	1	64.31	1.22	DE-16	1	64.71	0.87
DE-16	2	—	48.64	DE-16	2	—	39.28
DE-16	3	—	6672.61	DE-16	3	—	5674.01
DE-16	4	—	—	DE-16	4	—	—
DE-16	5	—	—	DE-16	5	—	—
DE-16	6	—	—	DE-16	6	—	—
DE-16	7	—	—	DE-16	7	—	—
DE-16	8	—	—	DE-16	8	—	—
DE-16	9	—	—	DE-16	9	—	—
DE-16	10	—	—	DE-16	10	—	—
DE-16	11	—	—	DE-16	11	—	—
DE-16	12	—	—	DE-16	12	—	—
DE-16	13	—	5286.82	DE-16	13	—	5203.69
DE-16	14	—	129.30	DE-16	14	—	129.38
DE-16	15	350.91	9.73	DE-16	15	343.30	9.68
DE-16	16	12.48	3.26	DE-16	16	12.52	3.30
DE-23	1	—	1.01	DE-23	1	—	1.02
DE-23	2	—	241.18	DE-23	2	—	240.45
DE-23	3	—	—	DE-23	3	—	—
DE-23	4	—	—	DE-23	4	—	—
DE-23	20	—	—	DE-23	20	—	—
DE-23	21	—	705.10	DE-23	21	—	704.54
DE-23	22	—	13.83	DE-23	22	—	13.81
DE-23	23	—	3.35	DE-23	23	—	3.27
DE-28	1	—	1.02	DE-28	1	—	1.02
DE-28	2	—	83.62	DE-28	2	—	84.07
DE-28	3	—	—	DE-28	3	—	—
DE-28	4	—	—	DE-28	4	—	—
DE-28	25	—	—	DE-28	25	—	—
DE-28	26	—	1042.95	DE-28	26	—	1016.68
DE-28	27	—	15.28	DE-28	27	—	15.35
DE-28	28	—	3.74	DE-28	28	—	3.79

methods for solving mixed-integer reformulations of optimality conditions that, due to KKT complementarity, contain a lot combinatorial structure? Going further, another research goal is to enhance decomposition methods for multilevel optimization problems.

References

[1] T. Achterberg. "SCIP: Solving Constraint Integer Programs." In: *Mathematical Programming Computation* 1.1 (2009), pp. 1–41. DOI: `10.1007/s12532-008-0001-1`.

[2] C. Analytics. *Anaconda*. URL: `https://www.continuum.io/anaconda-overview` (visited on 12/16/2016).

[3] M. F. Anjos, B. Ghaddar, L. Hupp, F. Liers, and A. Wiegele. "Solving k-Way Graph Partitioning Problems to Optimality: The Impact of Semidefinite Relaxations and the Bundle Method." In: *Facets of Combinatorial Optimization*. Ed. by M. Jünger and G. Reinelt. Berlin, Heidelberg: Springer, 2013, pp. 355–386. DOI: `10.1007/978-3-642-38189-8_14`.

[4] D. A. Bader, H. Meyerhenke, P. Sanders, and D. Wagner. *Graph partitioning and graph clustering*. Vol. 588. American Mathematical Soc., 2013.

[5] F. Barahona, M. Grötschel, M. Jünger, and G. Reinelt. "An Application of Combinatorial Optimization to Statistical Physics and Circuit Layout Design." In: *Operations Research* 36.3 (1988), pp. 493–513. JSTOR: `170992`.

[6] F. Barahona and A. R. Mahjoub. "On the cut polytope." In: *Mathematical Programming* 36.2 (1986), pp. 157–173. DOI: `10.1007/BF02592023`.

[7] J. F. Benders. "Partitioning procedures for solving mixed-variables programming problems." In: *Numerische Mathematik* 4.1 (1962), pp. 238–252. DOI: `10.1007/bf01386316`.

[8] M. Bjørndal and K. Jørnsten. "Benefits from coordinating congestion management——The Nordic power market." In: *Energy Policy* 35.3 (2007), pp. 1978–1991. DOI: `10.1016/j.enpol.2006.06.014`.

[9] M. Bjørndal and K. Jørnsten. "Zonal Pricing in a Deregulated Electricity Market." In: *The Energy Journal* 22.1 (2001), pp. 51–73. DOI: `10.5547/issn0195-6574-ej-vol22-no1-3`.

[10] J.-F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical Optimization. Theoretical and Practical Aspects*. Springer Science & Business Media, 2006. DOI: `10.1007/978-3-540-35447-5`.

[11] E. Boros and P. L. Hammer. "The max-cut problem and quadratic 0–1 optimization; polyhedral aspects, relaxations and bounds." In: *Annals of Operations Research* 33.3 (1991), pp. 151–180. DOI: `10.1007/BF02115753`.

[12] C. Breuer and A. Moser. "Optimized bidding area delimitations and their impact on electricity markets and congestion management." In: *11th International Conference on the European Energy Market (EEM14)*. May 2014, pp. 1–5. DOI: `10.1109/EEM.2014.6861218`.

[13] R. Carvajal, M. Constantino, M. Goycoolea, J. P. Vielma, and A. Weintraub. "Imposing connectivity constraints in forest planning models." In: *Operations Research* 61.4 (2013), pp. 824–836. DOI: `10.1287/opre.2013.1183`.

[14] H.-P. Chao and S. C. Peck. "Reliability Management in Competitive Electricity Markets." In: *Journal of Regulatory Economics* 14.2 (1998), pp. 189–200. DOI: `10.1023/A:1008061319181`.

[15] S. Chopra and M. R. Rao. "Facets of the k-partition problem." In: *Discrete Applied Mathematics* 61.1 (1995), pp. 27–48. DOI: `10.1016/0166-218X(93)E0175-X`.

[16] S. Chopra and M. R. Rao. "The partition problem." In: *Mathematical Programming* 59.1 (1993), pp. 87–115. DOI: `10.1007/BF01581239`.

[17] CPLEX. *User's Manual for CPLEX*. 12.6. IBM Corporation. Armonk, USA, 2013.

[18] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova. "Bilevel Programming Problems." In: *Energy Systems. Springer, Berlin* (2015).

[19] X. Deng. "Complexity Issues in Bilevel Linear Programming." In: *Multilevel Optimization: Algorithms and Applications*. Ed. by A. Migdalas, P. M. Pardalos, and P. Värbrand. Boston, MA: Springer US, 1998, pp. 149–164. DOI: 10.1007/978-1-4613-0307-7_6.

[20] M. M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Algorithms and Combinatorics 15. Springer, 1997. DOI: 10.1007/978-3-642-04295-9.

[21] B. Dilkina and C. P. Gomes. "Solving Connected Subgraph Problems in Wildlife Conservation." In: *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems: 7th International Conference, CPAIOR 2010, Bologna, Italy, June 14-18, 2010. Proceedings*. Ed. by A. Lodi, M. Milano, and P. Toth. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 102–116. DOI: 10.1007/978-3-642-13520-0_14.

[22] E. D. Dolan and J. J. Moré. "Benchmarking Optimization Software with Performance Profiles." In: *Mathematical Programming* 91 (2 2002), pp. 201–213. DOI: 10.1007/s101070100263.

[23] J. Egerer, J. Weibezahn, and H. Hermann. "Two price zones for the German electricity market — Market implications and distributional effects." In: *Energy Economics* 59 (2016), pp. 365–381. DOI: 10.1016/j.eneco.2016.08.002.

[24] ENTSO-E. *All TSOs' draft proposal for Capacity Calculation Regions (CCRs)*. 2015.

[25] P. Festa, P. M. Pardalos, M. G. Resende, and C. C. Ribeiro. "Randomized heuristics for the Max-Cut problem." In: *Optimization Methods and Software* 17.6 (2002), pp. 1033–1058. DOI: 10.1080/1055678021000090033.

[26] J. Fortuny-Amat and B. McCarl. "A Representation and Economic Interpretation of a Two-Level Programming Problem." In: *The Journal of the Operational Research Society* 32.9 (1981), pp. 783–792. JSTOR: 2581394.

[27] A. Frieze and M. Jerrum. "Improved approximation algorithms for max k-cut and max bisection." In: *Algorithmica* 18.1 (1997), pp. 67–81. DOI: 10.1007/BF02523688.

[28] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co., 1979.

[29] A. M. Geoffrion. "Generalized Benders decomposition." In: *Journal of Optimization Theory and Applications* 10.4 (1972), pp. 237–260. DOI: 10.1007/BF00934810.

[30] A. Ghoniem and H. D. Sherali. "Defeating symmetry in combinatorial optimization via objective perturbations and hierarchical constraints." In: *IIE Transactions* 43.8 (2011), pp. 575–588. DOI: 10.1080/0740817x.2010.541899.

[31] M. X. Goemans and D. P. Williamson. "New $\frac{3}{4}$-approximation algorithms for the maximum satisfiability problem." In: *SIAM Journal of Discrete Mathematics* 7.4 (1994), pp. 656–666. DOI: 10.1137/S0895480192243516.

[32] V. Grimm, G. Zöttl, B. Rückel, and C. Sölch. *Regionale Preiskomponenten im Strommarkt. Gutachten im Auftrag der Monopolkommission in Vorbereitung des 71. Sondergutachtens Energie2015 der Monopolkommission*. Friedrich-Alexander-Universität Erlangen-Nürnberg, 2015.

[33] V. Grimm, A. Martin, M. Schmidt, M. Weibelzahl, and G. Zöttl. "Transmission and Generation Investment in Electricity Markets: The Effects of Market Splitting and Network Fee Regimes." In: *European Journal of Operational Research* 254.2 (2016), pp. 493–509. DOI: 10.1016/j.ejor.2016.03.044.

[34] V. Grimm, L. Schewe, M. Schmidt, and G. Zöttl. *Uniqueness of Market Equilibrium on a Network: A Peak-Load Pricing Approach*. Tech. rep. May 2015.

[35]   Gurobi Optimization, Inc. *Gurobi Optimizer Reference Manual, Version 6.5.*
        2015.

[36]   P. Hansen and N. Mladenović. "An Introduction to Variable Neighborhood
        Search." In: *Meta-Heuristics: Advances and Trends in Local Search Paradigms
        for Optimization.* Ed. by S. Voß, S. Martello, I. H. Osman, and C. Roucairol.
        Boston, MA: Springer US, 1999, pp. 433–458. DOI: `10.1007/978-1-4615-5775-`
        `3_30`.

[37]   P. Holmberg and E. Lazarczyk. "Congestion management in electricity networks:
        Nodal, zonal and discriminatory pricing." In: (2012). Research Institute of
        Industrial Economics (IFN). DOI: `10.2139/ssrn.2055655`.

[38]   M. Jenabi, S. M. Fatemi Ghomi, and Y. Smeers. "Bi-Level Game Approaches
        for Coordination of Generation and Transmission Expansion Planning Within
        a Market Environment." In: *IEEE Transactions on Power Systems* 28.3 (Aug.
        2013), pp. 2639–2650. DOI: `10.1109/tpwrs.2012.2236110`.

[39]   V. Kaibel, M. Peinhardt, and M. E. Pfetsch. "Orbitopal fixing." In: *Integer
        programming and combinatorial optimization.* Springer, 2007, pp. 74–88. DOI:
        `10.1016/j.disopt.2011.07.001`.

[40]   V. Kaibel and M. E. Pfetsch. "Packing and partitioning orbitopes." In: *Mathemat-
        ical Programming* 114.1 (2008), pp. 1–36. DOI: `10.1007/s10107-006-0081-5`.

[41]   T. Kleinert. "A Decomposition Approach for a Multilevel Graph Partitioning
        Model of the German Electricity Market." Master's thesis. Friedrich-Alexander-
        Universität Erlangen-Nürnberg, 2016.

[42]   H. F. Lee and D. R. Dooly. "Decomposition algorithms for the maximum-weight
        connected graph problem." In: *Naval Research Logistics (NRL)* 45.8 (1998),
        pp. 817–837. DOI: `10.1002/(SICI)1520-6750(199812)45:8<817::AID-`
        `NAV4>3.0.CO;2-1`.

[43]   F. Liers, M. Jünger, G. Reinelt, and G. Rinaldi. "Computing Exact Ground States
        of Hard Ising Spin Glass Problems by Branch-and-Cut." In: *New Optimization
        Algorithms in Physics.* Wiley, 2005, pp. 47–69. DOI: `10.1002/3527603794.ch4`.

[44]   A. Lisser and F. Rendl. "Graph partitioning using linear and semidefinite
        programming." In: *Mathematical Programming* 95.1 (2003), pp. 91–101. DOI:
        `10.1007/s10107-002-0342-x`.

[45]   F. Margot. "Exploiting orbits in symmetric ILP." In: *Mathematical Programming*
        98.1-3 (2003), pp. 3–21. DOI: `10.1007/s10107-003-0394-6`.

[46]   F. Margot. "Pruning by isomorphism in branch-and-cut." In: *Mathematical
        Programming* 94.1 (2002), pp. 71–90. DOI: `10.1007/s10107-002-0358-2`.

[47]   F. Margot. "Symmetry in Integer Linear Programming." In: *50 Years of Integer
        Programming 1958-2008: From the Early Years to the State-of-the-Art.* Ed. by
        M. Jünger, T. M. Liebling, D. Naddef, G. L. Nemhauser, W. R. Pulleyblank,
        G. Reinelt, G. Rinaldi, and L. A. Wolsey. Berlin, Heidelberg: Springer Berlin
        Heidelberg, 2010, pp. 647–686. DOI: `10.1007/978-3-540-68279-0_17`.

[48]   I. Méndez-Díaz and P. Zabala. "A Branch-and-Cut algorithm for graph coloring."
        In: *Discrete Applied Mathematics* 154.5 (2006), pp. 826–847. DOI: `10.1016/j.`
        `dam.2005.05.022`.

[49]   I. Méndez-Díaz and P. Zabala. "A Polyhedral Approach for Graph Coloring." In:
        Electronic Notes in Discrete Mathematics 7 (2001), pp. 178–181. DOI: `10.1016/`
        `S1571-0653(04)00254-9`.

[50]   J. Mitchell. *Branch-and-cut for the k-way equipartition problem.* Tech. rep.
        Department of Mathematical Sciences, Rensselaer Polytechnic Institute, 2001.

[51]   *NetworkX. High-productivity software for complex networks.* URL: `https://`
        `networkx.github.io` (visited on 12/16/2016).

[52] J. Ostrowski, J. Linderoth, F. Rossi, and S. Smriglio. "Orbital branching." In: *Integer programming and combinatorial optimization*. Springer, 2007, pp. 104–118. DOI: `10.1007/978-3-540-72792-7_9`.

[53] M. E. Pfetsch and T. Rehn. "A Computational Comparison of Symmetry Handling Methods for Mixed Integer Programs." In: *Optimization Online* (2015). Preprint.

[54] Regionales Rechenzentrum Erlangen. *Woodcrest Cluster*. URL: `https://www.rrze.fau.de/dienste/arbeiten-rechnen/hpc/systeme/woodcrest-cluster.shtml` (visited on 12/16/2016).

[55] F. Rendl, G. Rinaldi, and A. Wiegele. "Solving Max-Cut to Optimality by Intersecting Semidefinite and Polyhedral Relaxations." In: *Mathematical Programming* 121.2 (2008), p. 307. DOI: `10.1007/s10107-008-0235-8`.

[56] N. V. Sahinidis and I. E. Grossmann. "Convergence properties of generalized Benders decomposition." In: *Computers & Chemical Engineering* 15.7 (1991), pp. 481–491. DOI: `10.1016/0098-1354(91)85027-R`.

[57] H. D. Sherali and J. C. Smith. "Improving discrete model representations via symmetry considerations." In: *Management Science* 47.10 (2001), pp. 1396–1407. DOI: `10.1287/mnsc.47.10.1396.10265`.

[58] S. Stoft. "Transmission pricing zones: simple or complex?" In: *The Electricity Journal* 10.1 (1997), pp. 24–31. DOI: `10.1016/s1040-6190(97)80294-1`.

[59] The European Commission. *Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management*. 2015.

[60] K. Trepper, M. Bucksteeg, and C. Weber. "Market splitting in Germany – New evidence from a three-stage numerical model of Europe." In: *Energy Policy* 87 (2015), pp. 199–215. DOI: `10.1016/j.enpol.2015.08.016`.

[61] V. V. Vazirani. *Approximation Algorithms*. New York, NY, USA: Springer, 2001.

[62] L. Vicente, G. Savard, and J. Júdice. "Descent approaches for quadratic bilevel programming." In: *Journal of Optimization Theory and Applications* 81.2 (1994), pp. 379–399. DOI: `10.1007/BF02191670`.

[63] J. J. Ye and D. L. Zhu. "Optimality conditions for bilevel programming problems." In: *Optimization* 33.1 (1995), pp. 9–27. DOI: `10.1080/02331939508844060`.

[1]Veronika Grimm, (a) Friedrich-Alexander-Universität Erlangen-Nürnberg, Chair of Economic Theory, Lange Gasse 20, 90403 Nürnberg, Germany; (b) Energie Campus Nürnberg, Fürther Str. 250, 90429 Nürnberg, Germany, [2]Frauke Liers, Friedrich-Alexander-Universität Erlangen-Nürnberg, Discrete Optimization, Cauerstr. 11, 91058 Erlangen, Germany, [3]Martin Schmidt, (a) Friedrich-Alexander-Universität Erlangen-Nürnberg, Discrete Optimization, Cauerstr. 11, 91058 Erlangen, Germany; (b) Energie Campus Nürnberg, Fürther Str. 250, 90429 Nürnberg, Germany, [4]Gregor Zöttl, Friedrich-Alexander-Universität Erlangen-Nürnberg, Industrial Organization and Energy Markets, Lange Gasse 20, 90403 Nürnberg, Germany

E-mail address: [1]`veronika.grimm@fau.de`
E-mail address: [2]`frauke.liers@math.uni-erlangen.de`
E-mail address: [3]`mar.schmidt@fau.de`
E-mail address: [4]`gregor.zoettl@fau.de`