Global Optimization of Multilevel Electricity Market Models Including Network Design and Graph Partitioning

Thomas Kleinert, Martin Schmidt

Abstract. We consider the combination of a network design and graph partitioning model in a multilevel framework for determining the optimal network expansion and the optimal zonal configuration of zonal pricing electricity markets, which is an extension of the model discussed in [25] that does not include a network design problem. The two classical discrete optimization problems of network design and graph partitioning together with nonlinearities due to economic modeling yield extremely challenging mixed-integer nonlinear multilevel models for which we develop two problem-tailored solution techniques. The first approach relies on an equivalent bilevel formulation and a standard KKT transformation thereof including novel primal-dual bound tightening techniques, whereas the second is a tailored generalized Benders decomposition. For the latter, we strengthen the Benders cuts of [25] by using the structure of the newly introduced network design subproblem. We prove for both methods that they yield global optimal solutions. Afterward, we compare the approaches in a numerical study and show that the tailored Benders approach clearly outperforms the standard KKT transformation. Finally, we present a case study that illustrates the economic effects that are captured in our model.

1. Introduction

In this paper we combine network design and graph partitioning with connectivity constraints in a multilevel framework for optimizing the design of electricity markets. Network design and graph partitioning are both important problems of discrete optimization with many different applications. However, according to our knowledge, a combination of both problems within a multilevel framework has not yet been discussed in the literature.

In particular, we analyze the problem of an optimal expansion of electricity transmission networks in a zonal pricing market environment with redispatch under a long-run perspective. This market design is implemented, e.g., in parts of Europe, Australia, or Latin America. The system works as follows. Given an existing electricity network, a regulatory authority specifies an optimal network extension and partitions the market area into price zones. Electricity generators react to these decisions by investment in generation capacities. Afterward, several periods of day-ahead spot-market trading followed by redispatch take place. In a zonal pricing framework, spot-market trading only has to account for inter-zonal transmission lines that are determined by the specified price zones. Due to this incomplete consideration of network constraints, possibly infeasible outcomes have to be redispatched ex post after each trading period. This redispatch mechanism may be expensive and harms welfare outcomes. The overall objective of the regulator is to maximize total social

Date: October 29, 2018.

2010 *Mathematics Subject Classification.* 90-08, 90B10, 90C11, 90C35, 90C90.

Key words and phrases. Network design, Graph partitioning, Multilevel optimization, Mixed-integer optimization, Electricity market design.

welfare, i.e., net welfare from spot-market trading minus capacity investment costs of firms, less redispatch costs and transmission line investment.

This paper builds upon the model and the solution approaches introduced by us in [25], where only the problem of computing the optimal zonal configuration is considered. Here, we also include a network design problem. This is of high importance in practice if one wants to weigh the costs of building new transmission lines against the possibility to modify the market design. Moreover, a direct application of the techniques developed in [25] is not possible for solving the problems discussed in this paper.

Many contributions discuss the problem of optimal transmission and generation expansion, i.e., optimal network design under a long-run perspective, in different frameworks; see, e.g., [26, 34, 51, 54]. All of them, however, do not consider decisions on price zones. Zonal pricing for fixed network topologies is widely discussed in the literature as well; see, e.g., [6, 19, 25, 32]. However, only our ancestor paper [25] provides a model that incorporates both a long run-perspective and endogenous decisions on price zones. To the best of our knowledge, there is no literature analyzing a simultaneous decision on network design and prize zone specification.

Rigorous mathematical modeling of the above-mentioned economic setting yields a very challenging mixed-integer nonlinear trilevel model that extends the model provided in [25]. In this setting, the first level incorporates a network design and a graph partitioning problem with connectivity constraints. Traditional network design problems (NDPs) typically involve decisions on adding capacity to existing networks at minimum cost in order to meet point-to-point requirements. Such problems are discussed extensively in the literature; see [43] and the references given therein for an overview or [22, 45] for exemplary variants of the problem. NDPs arise in many applications, e.g., in power systems [4, 5, 41], in telecommunication networks [28, 39], or in logistics networks [40]. Graph partitioning problems ask for k subsets that partition the node set of a given graph. Out of the many references, we refer to [16] as well as to [2, 8] for the case $k = 2$ and to [12, 13, 20, 50] for general values of k. In contrast to the general case, graph partitioning problems with connectivity constraints are considered much less. Some applications of a related problem, the maximum-weight connected subgraph problem, can be found in forest planning [10], in wildlife corridor design [17], and in designing fiber-optic networks [37].

As discussed earlier, our problem is a multilevel optimization problem. Such problems are inherently hard to solve due to the interdependencies of several nested optimization problems. Even bilevel problems with a linear upper and lower level are strongly NP-hard; see [15, 31]. The more general case of mixed-integer linear bilevel problems is Σ_p^2-hard; cf. [38]. In our application, we are faced with a mixed-integer nonlinear trilevel problem. For this kind of problem, no general-purpose solvers exist and problem-tailored solution approaches thus need to be developed.

Our contribution is the following. We extend the trilevel model provided in [25] for determining optimal price zones in order to also incorporate decisions on optimal network design. To this end, we newly introduce another discrete problem to the overall model that couples all three levels of our multilevel framework. Since the hardness of multilevel models mainly depends on the coupling of the different levels, the integration of the network design problem leads to a significant structural change w.r.t. what has been considered in [25]. We discuss two different solution approaches. Both approaches are inspired by the techniques developed in [25] but are significantly extended and improved in order to cope with the extended modeling. The first approach is a problem-specific reduction of the trilevel problem to a bilevel problem followed by a KKT reformulation to obtain a large single-level mixed-integer quadratic problem. The KKT approach typically suffers from weak bounds, which is why we enhance the model with new primal-dual bound tightenings that improve the solution approach and also give additional economic insights. The second approach is a variant of a generalized Benders decomposition, for which we introduce two types of problem-specific optimality cuts. These cuts are shown to yield a correct algorithm for computing global optimal solutions. The cuts of the first type extend the cuts stated in [25] to be able to handle the network design subproblem. The second type of cuts is novel and much stronger than the first type due to the utilization of the structural model modifications that are introduced by the network design subproblem on the first level. We evaluate the developed techniques and show the effectiveness of our proposed solution techniques. Finally, we present a case study that illustrates the economic effects of combined optimal decisions for network expansion and for modifying the market design, i.e., the zonal optimal configuration.

The paper is structured as follows. First, we provide a detailed description of the model of all three levels in Section 2. In Section 3, we present the single-level reformulation approach as well as primal-dual bound tightening techniques to enhance the solution approach. Afterward, in Section 4, we develop a generalized Benders method. Section 5 contains a numerical study of both solution approaches and Section 6 briefly presents two case studies that highlight the economic effects captured by our model. Finally, the paper closes with a comparison of the two solution techniques and some concluding remarks in Section 7.

2. The Mixed-Integer Multilevel Market Model

The market model presented in this work is mainly based on the trilevel formulation of [25], which models the decision on welfare-maximizing price zones as well as spot-market and redispatch behavior. In this paper, we extend this problem by incorporating decisions on welfare-maximizing network expansions. Hence, the first-level problem is a combined network design and graph partitioning problem with a multi-commodity flow formulation, that models connectivity within zones, as a substructure. The second- and third-level problems are continuous convex-quadratic problems without genuine integer variables and model spot-market trading as well as a cost-based redispatch, respectively. Both the second- and the third-level model are coupled to the two discrete subproblems (network design and graph partitioning) of the first level. While the optimal zonal configuration only influences the second-level model, the network design solution is linked both to the second and the third level and also introduces explicit costs in the objective function of the first level. In the following we present the technical setup and every level in detail.

2.1. Basic Economic and Technical Setup.

We consider a connected electricity transmission network $\mathcal{G} = (N, L)$ with a finite set of nodes N and a finite set of transmission lines $L \subseteq N \times N$ consisting of already existing lines L^{ex} and candidate lines L^{new}. Thus, $L = L^{\text{ex}} \cup L^{\text{new}}$ holds. Transmission lines l are characterized by their capacity $\bar{f}_l$ and their susceptance B_l. Line investment costs for new lines $l \in L^{\text{new}}$ are denoted by c_l^{inv}. 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:

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

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

At every node $n \in N$ we introduce a finite set of consumers C_n with $0 \leq |C_n| < \infty$ that are located at that node. For a given set of scenarios $T = \{1, \dots, |T|\}$, elastic

max total social welfare (regulator)
s.t. network design,
graph partitioning with connectivity constraints,

> max profits (competitive firms)
> s.t. generation capacity investment,
> production & demand constraints,
> Kirchhoff's 1st law (zonal),
> flow restrictions (inter-zonal),
>
> > min redispatch costs (TSO)
> > s.t. production & demand constraints,
> > lossless DC power flow constraints

FIGURE 1. Structure of the trilevel market model

demand of a consumer $c \in C_n$ in scenario $t \in T$ is modeled by a continuous and strictly decreasing inverse demand function $p_{c,t} = p_{c,t}(d_{c,t})$, where $d_{c,t} \geq 0$ denotes demand. In this case, the gross consumer surplus

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

is a strictly concave function. In particular, we assume that all inverse demand functions are linear to keep the problem tractable, i.e., we have $p_{c,t} = a_{c,t} + b_{c,t} d_{c,t}$ with $a_{c,t} > 0$ and $b_{c,t} < 0$. This is a common assumption in the field of electricity market modeling; see, e.g., [6, 46, 48].

For a given network node $n \in N$, G_n^{all} with $0 \leq |G_n^{\text{all}}| < \infty$ denotes a finite set of generation technologies. We distinguish between existing technologies G_n^{ex} and candidate technologies G_n^{new} that firms can invest in. Hence, $G_n^{\text{all}} = G_n^{\text{new}} \cup G_n^{\text{ex}}$ holds. For all existing generators $g \in G_n^{\text{ex}}$, their capacity $\bar{q}_g^{\text{ex}} > 0$ is given. In contrast, firms can invest in generation capacity $\bar{q}_g^{\text{new}}$ of candidate generators with given maximum capacity $\hat{q}_g^{\text{new}}$. The corresponding investment costs are denoted by $c_g^{\text{inv}} > 0$ for $g \in G_n^{\text{new}}$ and variable costs for production $q_{g,t} \geq 0$ are denoted by $c_g^{\text{var}} > 0$ for all $g \in G_n^{\text{all}}$.

The entire trilevel market model is sketched in technical and economic terms in Fig. 1. We now discuss each level in detail. In the following, 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: Network Design and Price Zoning.** At the first level, the TSO simultaneously decides on a network design, i.e., in building new transmission lines, and the specification of $k \leq |N|$ price zones, in order to maximize social welfare:

$$\begin{aligned}
\max \quad & \psi_1 = \sum_{t\in T}\sum_{n\in N}\left(\sum_{c\in C_n}\int_0^{d^{\text{red}}_{c,t}} p_{c,t}(\omega)\,\mathrm{d}\omega - \sum_{g\in G^{\text{all}}_n} c^{\text{var}}_g q^{\text{red}}_{g,t}\right) \\
& \qquad - \sum_{n\in N}\sum_{g\in G^{\text{new}}_n} c^{\text{inv}}_g \bar{q}^{\text{new}}_g - \sum_{l\in L^{\text{new}}} c^{\text{inv}}_l z_l
\end{aligned} \tag{1a}$$

$$\text{s.t.} \quad z_l \in \{0,1\}, \quad l \in L^{\text{new}}, \tag{1b}$$

$$x_{n,i} \in \{0,1\}, \quad n \in N,\ i \in [k], \tag{1c}$$

$$\sum_{i\in[k]} x_{n,i} = 1, \quad n \in N, \tag{1d}$$

$$s_{n,i} \in \{0,1\}, \quad s_{n,i} \le x_{n,i}, \quad n \in N,\ i \in [k], \tag{1e}$$

$$\sum_{n\in N} s_{n,i} = 1, \quad i \in [k], \tag{1f}$$

$$0 \le u^i_a, \quad a \in \{(n,m),(m,n)\},\ l = (n,m) \in L,\ i \in [k], \tag{1g}$$

$$u^i_a \le z_l, \quad a \in \{(n,m),(m,n)\},\ l = (n,m) \in L^{\text{new}},\ i \in [k], \tag{1h}$$

$$\sum_{a\in\delta^{\text{out}}_n} u^i_a \le M x_{n,i}, \quad n \in N,\ i \in [k], \tag{1i}$$

$$\sum_{a\in\delta^{\text{out}}_n} u^i_a - \sum_{a\in\delta^{\text{in}}_n} u^i_a \ge x_{n,i} - M s_{n,i}, \quad n \in N,\ i \in [k], \tag{1j}$$

$$y_l \in \{0,1\}, \quad x_{n,i} + x_{m,i} + y_l \le 2, \quad l = (n,m) \in L, \tag{1k}$$

$$x_{n,i} - x_{m,i} \le y_l, \quad x_{m,i} - x_{n,i} \le y_l, \quad l = (n,m) \in L. \tag{1l}$$

The objective function (1a) models total social welfare, which is given as the difference of gross consumer surplus from all markets, generation and investment costs of producers after redispatch, and investment costs for network expansion. Network expansion is modeled by the set of binary variables in (1b). The specification of price zones is modeled as a graph partitioning problem in (1c) and (1d). Connectivity within zones is ensured by the multi-commodity flow formulation (1e)–(1j), where M is a sufficiently large number. Note that this partitioning model with connectivity constraints also takes into account the network design decisions. For a discussion on modeling connectivity within a graph partitioning framework, see [10] or [25, 33]. For the spot-market model discussed in Section 2.3, we need additional indicator variables for inter-zonal lines, i.e., lines that connect nodes of different zones. This is taken care of by Constraints (1k) and (1l).

In summary, Model (1) is a mixed-integer quadratic program (MIQP) consisting of a concave-quadratic objective function and a set of mixed-integer linear constraints that contain a network design problem, a graph partitioning problem, as well as a multi-commodity flow problem as substructures. Note finally that, in contrast to the model in [25], the integration of the network design also leads to genuine first-level costs. We will later show in our numerical experiments in Section 5 that this aspect can be exploited to obtain stronger Benders optimality cuts.

2.3. **Second-Level Problem: Generation Investment and Spot-Market Trading.** At the second level we model the behavior of firms and consumers with respect to generation capacity investment and spot-market trading. In this paper, we assume perfect competition on the wholesale electricity market. It is known that in this setting, investment and production decisions are taken in a welfare-maximizing way; see, e.g., [27]. Although this assumption may not be adequate for power systems in general, we need it to keep the multilevel problem tractable. For more details about this assumption see [9, 14, 26], where it is also used for modeling energy markets. The second-level problem reads

$$
\begin{aligned}
\max \quad & \psi_2 = \sum_{t\in T}\sum_{n\in N}\left(\sum_{c\in C_n}\int_0^{d_{c,t}^{\text{spot}}} p_{c,t}(\omega)\,\mathrm{d}\omega - \sum_{g\in G_n^{\text{all}}} c_g^{\text{var}} q_{g,t}^{\text{spot}}\right) \\
& \qquad - \sum_{n\in N}\sum_{g\in G_n^{\text{new}}} c_g^{\text{inv}} \bar{q}_g^{\text{new}} && \text{(2a)}\\
\text{s.t.} \quad & -Mz_l \le f_{l,t}^{\text{spot}} \le Mz_l, \quad l\in L^{\text{new}},\ t\in T, && \text{(2b)}\\
& d_{n,t}^{\text{spot}} = \sum_{c\in C_n} d_{c,t}^{\text{spot}}, \quad q_{n,t}^{\text{spot}} = \sum_{g\in G_n^{\text{all}}} q_{g,t}^{\text{spot}}, \quad n\in N,\ t\in T, && \text{(2c)}\\
& D_{i,t} = \sum_{n\in N} x_{n,i} d_{n,t}^{\text{spot}}, \quad Q_{i,t} = \sum_{n\in N} x_{n,i} q_{n,t}^{\text{spot}}, \quad i\in[k],\ t\in T, && \text{(2d)}\\
& F_{i,t}^{\text{in}} = \sum_{l=(n,m)\in L} (1-x_{n,i}) x_{m,i} f_{l,t}^{\text{spot}}, \quad i\in[k],\ t\in T, && \text{(2e)}\\
& F_{i,t}^{\text{out}} = \sum_{l=(n,m)\in L} x_{n,i}(1-x_{m,i}) f_{l,t}^{\text{spot}}, \quad i\in[k],\ t\in T, && \text{(2f)}\\
& D_{i,t} + F_{i,t}^{\text{out}} = Q_{i,t} + F_{i,t}^{\text{in}}, \quad i\in[k],\ t\in T, && \text{(2g)}\\
& -\bar{f}_l - (1-y_l)M \le f_{l,t}^{\text{spot}} \le \bar{f}_l + (1-y_l)M, \quad l\in L,\ t\in T, && \text{(2h)}\\
& \bar{q}_g^{\text{new}} \le \hat{q}_g^{\text{new}}, \quad g\in G_n^{\text{new}},\ n\in N, && \text{(2i)}\\
& 0 \le q_{g,t}^{\text{spot}} \le \bar{q}_g^{\text{new}}, \quad g\in G_n^{\text{new}},\ n\in N,\ t\in T, && \text{(2j)}\\
& 0 \le q_{g,t}^{\text{spot}} \le \bar{q}_g^{\text{ex}}, \quad g\in G_n^{\text{ex}}\ n\in N,\ t\in T, && \text{(2k)}\\
& 0 \le d_{c,t}^{\text{spot}}, \quad c\in C_n,\ n\in N,\ t\in T. && \text{(2l)}
\end{aligned}
$$

The second-level objective (2a) maximizes total social welfare generated at the spot market. Constraint (2b) accounts for the first-level network design decisions. Constraints (2c)–(2g) correspond to zonal energy balances and can be seen as a zonal version of Kirchhoff's first law; see [25]. We note that power flow within a price zone is unconstrained. In contrast, Constraint (2h) limits power flow on inter-zonal transmission lines. Scarcities on these lines imply price differences between zones. Constraint (2i) models the investment of private firms in generation capacities. Finally, we have bounds on demand and generation (2j)–(2l).

In summary, on the second level we face mixed-integer nonlinear constraints, where all nonlinearities stem from products of binary first-level variables and continuous second-level variables. Hence, the spot-market model, for given first-level decisions, is a concave-quadratic maximization problem over linear constraints.

2.4. **Third-Level Problem: Optimal Cost-Based Redispatch.** At the third level, the TSO decides on cost-based redispatch. Since third-level decisions do not have any impact on second-level decisions, redispatch can be carried out simultaneously for all $|T|$ scenarios of spot-market trading. Reallocation of spot-market outcomes is realized in a way that ensures feasibility with respect to transmission constraints at lowest costs:

$$
\begin{aligned}
\min \quad & \psi_3 = \sum_{t\in T}\sum_{n\in N}\sum_{c\in C_n}\int_{d_{c,t}^{\text{red}}}^{d_{c,t}^{\text{spot}}} p_{c,t}(\omega)\,\mathrm{d}\omega + \sum_{t\in T}\sum_{n\in N}\sum_{g\in G_n^{\text{all}}} c_g^{\text{var}}(q_{g,t}^{\text{red}} - q_{g,t}^{\text{spot}}) && \text{(3a)}\\
\text{s.t.} \quad & \sum_{c\in C_n} d_{c,t}^{\text{spot}} + \sum_{l\in\delta_n^{\text{out}}} f_{l,t}^{\text{spot}} = \sum_{g\in G_n^{\text{all}}} q_{g,t}^{\text{spot}} + \sum_{l\in\delta_n^{\text{in}}} f_{l,t}^{\text{spot}}, \quad n\in N,\ t\in T, && \text{(3b)}
\end{aligned}
$$

$$f^{\text{red}}_{l,t} = B_l(\theta_{n,t} - \theta_{t,m}), \quad l = (n,m) \in L^{\text{ex}},\ t \in T, \tag{3c}$$

$$M(z_l - 1) \le f^{\text{red}}_{l,t} - B_l(\theta_{n,t} - \theta_{t,m}), \quad l = (n,m) \in L^{\text{new}},\ t \in T, \tag{3d}$$

$$M(1 - z_l) \ge f^{\text{red}}_{l,t} - B_l(\theta_{n,t} - \theta_{t,m}), \quad l = (n,m) \in L^{\text{new}},\ t \in T, \tag{3e}$$

$$\theta_{t,\hat{n}} = 0, \quad t \in T \tag{3f}$$

$$-\bar{f}_l \le f^{\text{red}}_{l,t} \le \bar{f}_l, \quad l \in L^{\text{ex}},\ t \in T, \tag{3g}$$

$$-\bar{f}_l z_l \le f^{\text{red}}_{l,t} \le \bar{f}_l z_l, \quad l \in L^{\text{new}},\ t \in T, \tag{3h}$$

$$0 \le q^{\text{red}}_{g,t} \le \bar{q}^{\text{new}}_g, \quad g \in G^{\text{new}}_n,\ n \in N,\ t \in T, \tag{3i}$$

$$0 \le q^{\text{red}}_{g,t} \le \bar{q}^{\text{ex}}_g, \quad g \in G^{\text{ex}}_n,\ n \in N,\ t \in T, \tag{3j}$$

$$0 \le d^{\text{red}}_{c,t}, \quad c \in C_n,\ n \in N,\ t \in T. \tag{3k}$$

As mentioned above, the objective function (3a) denotes redispatch costs. The constraints correspond to a lossless DC power flow formulation; see, e.g., [53]. In particular, Constraint (3b) corresponds to Kirchhoff's first law. Moreover, Constraints (3c)–(3e) model Kirchhoff's second law in dependence of the network design specified at the first level. Again, M is a sufficiently large number. At this point, let us remark that the latter dependence strengthens the link between the levels of our model in comparison to the model discussed in [25] since now the integer decisions of the first level not only connect the first two but all levels. Kirchhoff's second law determines the voltage angles of the network. In order to obtain physically unique solutions, the voltage angle is fixed to zero at an arbitrary node $\hat{n}$; see (3f). Finally, Constraints (3g)–(3k) bound transmission flows, demand, and generation.

The third-level model is a convex-quadratic mixed-integer minimization problem if it is embedded in the trilevel framework. However, all integer variables belong to the first-level. Thus, the redispatch problem (for given first- and second-level decisions) is a convex QP.

2.5. **Discussion of the Structure of the Trilevel Model.** Let us now discuss the structure of the presented trilevel model in detail. It turns out that the coupling of the three levels has the following special structure: No first- or second-level constraint contains any respective lower-level variables. Furthermore, the second-level objective depends on second-level variables and the third-level objective depends on second- and third-level variables. The first-level objective interconnects all three levels and it is easy to see that

$$\psi_1 = \psi_2 - \psi_3 - \sum_{l \in L^{\text{new}}} c^{\text{inv}}_l z_l \tag{4}$$

holds. Thus, the overall structure is the following:

$$\begin{aligned} \max \quad & \psi_1(W_2, W_3, X_1) \\ \text{s.t.} \quad & (W_1, X_1) \in \Omega_1, \\ & \max \quad \psi_2(W_2) \\ & \ \text{s.t.} \quad (W_2, X_1) \in \Omega_2, \\ & \qquad \min \quad \psi_3(W_2, W_3) \\ & \qquad \ \text{s.t.} \quad (W_2, W_3, X_1) \in \Omega_3. \end{aligned} \tag{5}$$

Here, $W_i \in \mathbb{R}^{n_i}$ denotes the continuous variables of level $i \in \{1, 2, 3\}$, $X_1 \in \{0, 1\}^{m_1}$ denotes the binary first-level variables, and Ω_i is the corresponding discrete-continuous feasible set of the ith level. Finally, we note that the solution approaches described in the following section need that the second-level problem has a unique solution. Fortunately, this can be established under certain assumptions that are satisfied in our case; see [27] for a proof. For a detailed equilibrium analysis of related second-level models see [35, 36].

3. A Single-Level MIQP Reformulation

In this section, we briefly recall a solution approach proposed by [25] that reformulates the trilevel model as a large single-level MIQP. This approach can easily be extended to our setting. We then present sophisticated primal and dual bound tightenings together with some economic interpretations to enhance this approach.

The reformulation approach of [25] requires two steps. First, one can equivalently reformulate the trilevel problem (5) to the following bilevel problem:

$$\begin{aligned} \max \quad & \psi_1(W_2, W_3, X_1) \\ \text{s.t.} \quad & (W_1, X_1) \in \Omega_1, \; (W_2, W_3, X_1) \in \Omega_3, \\ & \max \quad \psi_2(W_2) \\ & \;\text{s.t.} \quad (W_2, X_1) \in \Omega_2. \end{aligned} \tag{6}$$

The rationale is based on the weak coupling of the three levels described at the end of the last section; see [26] and [25] for details. The upper level of the bilevel problem (6) maximizes the original first-level objective subject to the original first- and third-level constraints. The lower bilevel problem is the original second-level problem. Since this problem is a concave-quadratic problem, one can replace it by its KKT conditions to obtain a large single-level problem; see [23].

The KKT conditions of the lower level are given by the stationarity conditions

$$p_{c,t}(d^{\text{spot}}_{c,t}) + \sum_{i \in [k]} \varepsilon_{i,t} x_{n,i} + \kappa^-_{c,t} = 0, \quad c \in C_n, \; n \in N, \; t \in T, \tag{7a}$$

$$-c^{\text{var}}_g - \sum_{i \in [k]} \varepsilon_{i,t} x_{n,i} + \pi^-_{g,t} - \pi^+_{g,t} = 0, \quad g \in G^{\text{new}}_n, \; n \in N, \; t \in T, \tag{7b}$$

$$-c^{\text{var}}_g - \sum_{i \in [k]} \varepsilon_{i,t} x_{n,i} + \nu^-_{g,t} - \nu^+_{g,t} = 0, \quad g \in G^{\text{ex}}_n, \; n \in N, \; t \in T, \tag{7c}$$

$$-c^{\text{inv}}_g + \sum_{t \in T} \pi^+_{g,t} - \zeta^+_g = 0, \quad g \in G^{\text{new}}_n, \; n \in N, \tag{7d}$$

$$\eta^-_{l,t} - \eta^+_{l,t} - \sum_{i \in [k]} \varepsilon_{i,t}(x_{m,i} - x_{n,i}) = 0, \quad l = (n,m) \in L^{\text{ex}}, \; t \in T, \tag{7e}$$

$$\eta^-_{l,t} - \eta^+_{l,t} + \rho^-_{l,t} - \rho^+_{l,t} - \sum_{i \in [k]} \varepsilon_{i,t}(x_{m,i} - x_{n,i}) = 0, \quad l = (n,m) \in L^{\text{new}}, \; t \in T, \tag{7f}$$

primal feasibility (2b)–(2l), non-negativity of dual variables of inequality constraints

$$\kappa^-_{c,t} \geq 0, \quad c \in C_n, \; n \in N, \; t \in T, \tag{8a}$$

$$\pi^-_{g,t}, \pi^+_{g,t} \geq 0, \quad g \in G^{\text{new}}_n, \; n \in N, \; t \in T, \tag{8b}$$

$$\nu^-_{g,t}, \nu^+_{g,t} \geq 0, \quad g \in G^{\text{ex}}_n, \; n \in N, \; t \in T, \tag{8c}$$

$$\zeta^+_g \geq 0, \quad g \in G^{\text{new}}_n, \; n \in N, \tag{8d}$$

$$\eta^-_{l,t}, \eta^+_{l,t} \geq 0, \quad l \in L, \; t \in T, \tag{8e}$$

$$\rho^-_{l,t}, \rho^+_{l,t} \geq 0, \quad l \in L^{\text{new}}, \; t \in T, \tag{8f}$$

and KKT complementarity conditions

$$\kappa^-_{c,t} d^{\text{spot}}_{c,t} = 0, \quad c \in C_n,\ n \in N,\ t \in T, \tag{9a}$$
$$\pi^-_{g,t} q^{\text{spot}}_{g,t} = \pi^+_{g,t}(q^{\text{spot}}_{g,t} - \bar{q}^{\text{new}}_g) = 0, \quad g \in G^{\text{new}}_n,\ n \in N,\ t \in T, \tag{9b}$$
$$\nu^-_{g,t} q^{\text{spot}}_{g,t} = \nu^+_{g,t}(q^{\text{spot}}_{g,t} - \bar{q}^{\text{ex}}_g) = 0, \quad g \in G^{\text{ex}}_n,\ n \in N,\ t \in T, \tag{9c}$$
$$\zeta^+_g(\bar{q}^{\text{new}}_g - \hat{q}^{\text{new}}_g) = 0, \quad g \in G^{\text{new}}_n,\ n \in N, \tag{9d}$$
$$\eta^-_{l,t}(-\bar{f}_l - (1 - y_l)M - f^{\text{spot}}_{l,t}) = 0, \quad l \in L,\ t \in T, \tag{9e}$$
$$\eta^+_{l,t}(f^{\text{spot}}_{l,t} - \bar{f}_l - (1 - y_l)M) = 0, \quad l \in L,\ t \in T, \tag{9f}$$
$$\rho^-_{l,t}(-M z_l - f^{\text{spot}}_{l,t}) = 0, \quad l \in L^{\text{new}},\ t \in T, \tag{9g}$$
$$\rho^+_{l,t}(f^{\text{spot}}_{l,t} - M z_l) = 0, \quad l \in L^{\text{new}},\ t \in T. \tag{9h}$$

We note that the dual variables of the primal auxiliary constraints (2c)–(2f) are already eliminated in the stationarity conditions (7). When applying the KKT reformulation we obtain a single-level mixed-integer nonlinear problem (MINLP). All nonlinearities stem from KKT complementarity (9) and products of binary first-level and continuous primal and dual second-level variables; see (2c)–(2g) and (7), respectively. Hence, these nonlinearities can be linearized using standard big-M reformulations; see, e.g., [21]. After all we end up with a single-level mixed-integer quadratic problem that is equivalent to the trilevel market model of Section 2. However, the extensive use of big-M reformulations is challenging from a numerical point of view and often yields weak dual bounds in branch-and-bound algorithms. As a possible remedy we next present strengthened bounds for the dual variables that yield a tighter dual formulation and thus stronger dual bounds. For what follows, we refer to the obtained single-level model as SLMIQP and start by considering the dual variables $\varepsilon_{i,t} \in \mathbb{R}$.

Lemma 1. *Consider an optimal solution of SLMIQP in which $x_{n,i} = 1$ holds for a certain $n \in N$ and $i \in [k]$. Then,*

(a) *for every $g \in G^{\text{all}}_n$ and $t \in T$,*
$$-\varepsilon_{i,t} < c^{\text{var}}_g \implies q^{\text{spot}}_{g,t} = 0$$
holds;

(b) *for every $c \in C_n$ and $t \in T$,*
$$-\varepsilon_{i,t} > a_{c,t} \implies d^{\text{spot}}_{c,t} = 0$$
holds.

Proof. Let $x_{n,i} = 1$ for $n \in N, i \in [k]$. We first consider (a) and generators $g \in G^{\text{ex}}_n$. Then, Equation (7c) simplifies to
$$-\varepsilon_{i,t} + \nu^-_{g,t} - \nu^+_{g,t} = c^{\text{var}}_g.$$
Hence, $-\varepsilon_{i,t} < c^{\text{var}}_g$ implies $\nu^-_{g,t} - \nu^+_{g,t} > 0$. Due to KKT complementarity we obtain $\nu^-_{g,t} > 0$ and $\nu^+_{g,t} = 0$ because lower and upper production bounds cannot be active at the same time. Hence, the lower production bound is active, i.e., $q^{\text{spot}}_{g,t} = 0$. Following the same argumentation, the claim can also be shown for candidate generators $g \in G^{\text{new}}_n$ with $\bar{q}^{\text{new}}_g > 0$. The other case $\bar{q}^{\text{new}}_g = 0$ directly implies $q^{\text{spot}}_{g,t} = 0$.

For (b), let $c \in C_n$. Equation (7a) simplifies to
$$p_{c,t}(d^{\text{spot}}_{c,t}) + \kappa^-_{c,t} = -\varepsilon_{i,t}.$$
With $-\varepsilon_{i,t} > a_{c,t} \geq p_{c,t}(d^{\text{spot}}_{c,t})$ one obtains $\kappa^-_{c,t} > 0$. KKT complementarity then yields $d^{\text{spot}}_{c,t} = 0$. □

From an economic perspective, $-\varepsilon_{i,t}$ denote zonal prices, i.e., $-\varepsilon_{i,t}$ is the price in zone i and scenario t. By intuition, a firm will halt production if the zonal price falls below its variable costs. Similarly, the demand of a consumer will be zero if the zonal price exceeds its maximum willingness to pay. This is formalized in the last lemma. For these zonal prices, we next prove a-priori estimates. To this end, we need the following assumption.

Assumption 1. *For an optimal solution of the trilevel model, the spot-market outcome ψ_2^* less redispatch costs ψ_3^* is strictly positive, i.e., $\psi_2^* - \psi_3^* > 0$ holds.*

We now suppose for the rest of this section that Assumption 1 holds.

Lemma 2. *For an optimal solution of SLMIQP,*

$$\min\left\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\right\} \leq -\varepsilon_{i,t} \leq \max\left\{a_{c,t} \colon c \in C_n, n \in N\right\}$$

holds for all $i \in [k]$ and $t \in T$.

Proof. For proving the first inequality we assume the contrary, i.e., $-\varepsilon_{i,t} < \min\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\}$ and use Lemma 1(a). For every $i \in [k]$, this implies $q_{g,t}^{\text{spot}} = 0$ for all $g \in G_n^{\text{all}}$ and $n \in N$ with $x_{n,i} = 1$. Hence, the zonal production equals zero. As a consequence, every nonzero zonal demand needs to be provided by other zones $j \neq i$. Thus and w.l.o.g., there exist nodes $n \in Z_i$ and $m \in Z_j$ with $0 < f_{l,t} \leq \bar{f}_l$ for $l = (m, n)$. KKT complementarity then implies $\eta_{l,t}^- = 0$, $\eta_{l,t}^+ \geq 0$, and $\rho_{l,t}^- = 0$, $\rho_{l,t}^+ \geq 0$ in case of $l \in L^{\text{new}}$. Hence, Equation (7e) or (7f), respectively, leads to

$$-\varepsilon_{i,t} + \varepsilon_{j,t} \geq 0.$$

Thus, $-\varepsilon_{j,t} \leq -\varepsilon_{i,t} < \min\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\}$ holds. Now, Lemma 1(a) implies $q_{g,t}^{\text{spot}} = 0$ for all generators g in zone j. Following this construction reveals that production is zero in every zone. Thus, the zonal total demand is zero in every zone as well. Both together contradicts Assumption 1.

The correctness of the upper bound can be shown in a similar manner. We assume $-\varepsilon_{i,t} > \max\{a_{c,t} \colon c \in C_n, n \in N\}$ and use Lemma 1(b). For every $i \in [k]$ this implies $d_{c,t}^{\text{spot}} = 0$ for all $c \in C_n$ and $n \in N$ with $x_{n,i} = 1$. Hence, the zonal demand equals zero. As a consequence, nonzero zonal production is consumed in other zones $j \neq i$. Thus and w.l.o.g., there exist nodes $n \in Z_i$ and $m \in Z_j$ with $0 < f_{l,t} \leq \bar{f}_l$ for $l = (n, m)$. KKT complementarity gives $\eta_{l,t}^- = 0$, $\eta_{l,t}^+ \geq 0$, and $\rho_{l,t}^- = 0$, $\rho_{l,t}^+ \geq 0$ in case of $l \in L^{\text{new}}$. Hence, Equations (7e) and (7f) both simplify to

$$-\varepsilon_{i,t} \leq -\varepsilon_{j,t}.$$

Since we assume $\max\{a_{c,t} \colon c \in C_n, n \in N\} < -\varepsilon_{i,t}$, the same holds for zone j. This implies $d_{c,t}^{\text{spot}} = 0$ by Lemma 1(b) for every consumer c in zone j. Again, following this construction shows that total demand and total production equals zero in every zone, which again contradicts Assumption 1. □

This lemma can now be used to tighten other dual bounds. We first consider $\kappa_{c,t}^-$. From KKT complementarity we know that $\kappa_{c,t}^- > 0$ can only be true if $d_{c,t}^{\text{spot}} = 0$ holds. Hence, this variable can be interpreted as the amount by which the zonal price $-\varepsilon_{i,t}$ needs to be reduced such that the consumer c starts to demand.

Lemma 3. *For an optimal solution of SLMIQP,*

$$\kappa_{c,t}^- \leq \max\left\{a_{\tilde{c},\tau} \colon \tilde{c} \in C_n, n \in N\right\} - a_{c,t}$$

holds for every $c \in C_n$, $n \in N$, and $t \in T$.

Proof. Since $\kappa^-_{c,t} = 0$ follows from $d^{\text{spot}}_{c,t} > 0$, it is sufficient to consider the case $d^{\text{spot}}_{c,t} = 0$. Let consumer c be located in zone i. In this case, we obtain

$$\kappa^-_{c,t} = -\varepsilon_{i,t} - p_{c,t}(0) = -\varepsilon_{i,t} - a_{c,t}$$

from Equation (7a). Using the upper bound for $-\varepsilon_{i,t}$ of Lemma 2 yields the claim. □

We turn to the variables $\nu^\pm_{g,t}$. From an economic point of view, they correspond to the amount the zonal price needs to rise to be profitable for an existing generator $g \in G^{\text{ex}}$. This means that they take the opposite role as $\kappa^-_{c,t}$.

Lemma 4. *For an optimal solution of SLMIQP,*

$$\nu^-_{g,t} \leq c^{\text{var}}_g - \min\left\{c^{\text{var}}_{\tilde{g}} \colon \tilde{g} \in G^{\text{ex}}_n, n \in N\right\}$$

and

$$\nu^+_{g,t} \leq \max\left\{a_{c,t} \colon c \in C_n, n \in N\right\} - c^{\text{var}}_g$$

hold for every $g \in G^{\text{ex}}_n$, $n \in N$, and $t \in T$.

Proof. In case of $0 < q^{\text{spot}}_{g,t} < \bar{q}^{\text{ex}}_{g,t}$ we have $\nu^-_{g,t} = \nu^+_{g,t} = 0$. Hence, let us assume $q^{\text{spot}}_{g,t} = 0$. KKT complementarity then delivers $\nu^+_{g,t} = 0$ and $\nu^-_{g,t} \geq 0$. Equation (7c) together with Proposition 2 yields the claim.

Now assume $q^{\text{spot}}_{g,t} = \bar{q}^{\text{ex}}_{g,t}$. In this case, KKT complementarity yields $\nu^-_{g,t} = 0$ and $\nu^+_{c,t} \geq 0$. Again, with Equation (7c) and Proposition 2 we obtain the claimed bound. □

Variables $\pi^\pm_{g,t}$ have the exact same economical meaning as $\nu^\pm_{g,t}$ but for candidate generators $g \in G^{\text{new}}$.

Lemma 5. *For an optimal solution of SLMIQP,*

$$\pi^-_{g,t} \leq c^{\text{var}}_g - \min\left\{c^{\text{var}}_{\tilde{g}} \colon \tilde{g} \in G^{\text{new}}_n, n \in N\right\} + c^{\text{inv}}_g$$

and

$$\pi^+_{g,t} \leq \max\left\{\max\left\{a_{c,t} \colon c \in C_n, n \in N\right\} - c^{\text{var}}_g, c^{\text{inv}}_g\right\}$$

hold for every $g \in G^{\text{new}}_n$, $n \in N$, and $t \in T$.

Proof. For $\bar{q}^{\text{new}}_g > 0$, we obtain

$$\pi^-_{g,t} \leq c^{\text{var}}_g - \min\left\{c^{\text{var}}_g \colon g \in G^{\text{new}}_n, n \in N\right\}$$

and

$$\pi^+_{g,t} \leq \max\left\{a_{c,t} \colon c \in C_n, n \in N\right\} - c^{\text{var}}_g$$

according to Lemma 4. On the other hand, for $\bar{q}^{\text{new}}_g = 0$, we have $\pi^-_{g,t} \geq 0$ and $\pi^+_{g,t} \geq 0$. In this case, $\zeta^+_g = 0$ holds. Hence, we obtain $\pi^+_{g,t} \leq c^{\text{inv}}_g$ with Equation (7d) and get

$$\pi^-_{g,t} \leq c^{\text{var}}_g - \min\left\{c^{\text{var}}_{\tilde{g}} \colon \tilde{g} \in G^{\text{new}}_n, n \in N\right\} + c^{\text{inv}}_g$$

from Equation (7b) and Proposition 2. A valid bound for both cases is then found by taking the maximum. □

Next, we turn to ζ^+_g. These dual variables denote the value of a further increase in the investment's upper bound $\hat{q}^{\text{new}}_g$.

Lemma 6. *For an optimal solution of SLMIQP,*

$$\zeta^+_g \leq \sum_{t \in T} \left(\max\left\{a_{c,t} \colon c \in C_n, n \in N\right\} - c^{\text{var}}_g\right) - c^{\text{inv}}_g$$

holds for all $g \in G^{\text{new}}_n$, $n \in N$.

Proof. This follows from Equation (7d) and Lemma 5. In case of $\bar{q}_g^{\text{new}} < \hat{q}_g^{\text{new}}$, KKT complementarity yields $\zeta_g^+ = 0$. Otherwise, $\bar{q}_g^{\text{new}} = \hat{q}_g^{\text{new}}$ and $\zeta_g^+ \geq 0$ holds. Hence, we can use $\pi_{g,t}^+ \leq \max\{a_{c,t} \colon c \in C_n, n \in N\} - c_g^{\text{var}}$; see the proof of Lemma 5. □

Next, we take a look at the variables $\eta_{l,t}^{\pm}$ that are dual to the flow bounds. If $\bar{f}_l > 0$ holds either both or exactly one of $\eta_{l,t}^-$ and $\eta_{l,t}^+$ are zero. Using Equation (7e) and the discussion after Lemma 1, this shows that $\eta_{l,t}^- - \eta_{l,t}^+$ denotes the zonal price difference in the case of a line $l = (n, m)$ with $n \in Z_i$, $m \in Z_j$, and $i \neq j$.

Lemma 7. *For an optimal solution of SLMIQP,*

$$\eta_{l,t}^{\pm} \leq \max\left\{a_{c,t} \colon c \in C_n, n \in N\right\} - \min\left\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\right\}$$

holds for every $l = (n, m)$ and $t \in T$.

Proof. If $|f_l| < \bar{f}_l$, it follows $\eta_{l,t}^{\pm} = 0$. Thus, we assume $f_l = \bar{f}_l$ and obtain $\eta_l^- = 0$ and $\eta_l^+ \geq 0$ from KKT complementarity. Similarly, we obtain $\eta_l^- \geq 0$ and $\eta_l^+ = 0$ in the case of $f_l = -\bar{f}_l$. In both cases, the claimed bound follows from Equation (7e) and Lemma 2. □

The variables $\rho_{l,t}^{\pm}$ that are dual to the flow bounds of candidate lines have a similar interpretation than the variables $\eta_{l,t}^{\pm}$. In case that a candidate l is built, $\rho_{l,t}^- = \rho_{l,t}^+ = 0$ holds and $\eta_{l,t}^- - \eta_{l,t}^+$ denotes the zonal price difference; see the discussion above. Otherwise, if a candidate line l is not built, $f_{l,t}^{\text{spot}} = 0$ and hence $\eta_{l,t}^- = \eta_{l,t}^+ = 0$ holds. In this case, we have $\rho_{l,t}^{\pm} \geq 0$ and $\rho_{l,t}^- - \rho_{l,t}^+$ denotes the zonal price difference.

Lemma 8. *There exists an optimal solution of SLMIQP that satisfies*

$$\rho_{l,t}^{\pm} \leq \max\left\{a_{c,t} \colon c \in C_n, n \in N\right\} - \min\left\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\right\}$$

for every $l = (n, m) \in L^{\text{new}}$ and $t \in T$.

Proof. Let $l = (n, m) \in L^{\text{new}}$. If $z_l = 1$, it follows $\rho_{l,t}^{\pm} = 0$ from KKT complementarity. Otherwise, $f_{l,t}^{\text{spot}} = 0$ holds and we obtain $\eta_{l,t}^{\pm} = 0$ and $\rho_{l,t}^{\pm} \geq 0$. If $n \in Z_i$ and $m \in Z_i$, Equation (7f) simplifies to $\rho_{l,t}^- - \rho_{l,t}^+ = 0$. Hence, there exists an optimal solution that satisfies $\rho_{l,t}^{\pm} = 0$. Otherwise, we have $n \in Z_i$, $m \in Z_j$, $i \neq j$, and obtain $\rho_{l,t}^- - \rho_{l,t}^+ = \epsilon_{j,t} - \epsilon_{i,t}$. Then, there exists a solution that satisfies the stated bounds; see Lemma 2. □

Before we close this section, we also state some tightened primal bounds.

Lemma 9. *Consider an optimal solution of SLMIQP, a consumer $c \in C_n$, $n \in N$, and a scenario $t \in T$. If $a_{c,t} \geq \min\left\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\right\}$,*

$$d_{c,t}^{\text{spot}} \leq \frac{\min\left\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\right\} - a_{c,t}}{b_{c,t}}$$

holds. Otherwise $d_{c,t}^{\text{spot}} \leq 0$ holds.

In addition to a bound tightening, the latter case can be used for preprocessing purposes. In particular, if $a_{c,t} < \min\left\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\right\}$, then $d_{c,t}^{\text{spot}} = 0$ holds, and c can be deleted from the set of consumers in scenario t.

Proof. Let consumer c be located in zone i and $d_{c,t}^{\text{spot}} > 0$. Equation (7a) and KKT complementarity yield $p_{c,t} + \epsilon_{i,t} = 0$. From Lemma 2, we know

$$p_{c,t} = a_{c,t} + b_{c,t} d_{c,t}^{\text{spot}} \geq \min\left\{c_g^{\text{var}} \colon g \in G_n^{\text{all}}, n \in N\right\}.$$

Since $b_{c,t} < 0$ holds, the claimed bound is correct. □

We will present a computational evaluation of the presented tightened primal-dual bounds later in Section 5.

4. A Generalized Benders Decomposition Approach

The second solution approach also exploits the weak coupling of the three levels. In fact, the three levels can be fully decomposed and solved independently. One way to do so is to apply a generalized Benders decomposition. This method can be seen as an extension of the variable partitioning approach of [3] for solving large linear programs as proposed in [24]; cf. [7] for an introduction. The algorithm proposed in this work is an extension of the generalized Benders decomposition framework for computing optimal price zones [25] and works as follows.

We iteratively compute feasible network designs and graph partitions. After fixing such a network design and graph partition, we subsequently solve the two lower level convex-quadratic problems. In particular, we can solve the second level and then fix its solution in the third level. Finally, we use the solutions of all three levels to compute the objective value of the first level. In principle, these steps need to be performed for every feasible first-level solution. However, valid cuts can be derived to prevent a full enumeration of all feasible first-level solutions. Applying the described decomposition ideas we obtain the following Benders master problem:

$$\max \quad \tau - \sum_{l \in L^{\text{new}}} c_l^{\text{inv}} z_l \tag{10a}$$

$$\text{s.t.} \quad \tau \leq a^\top \begin{pmatrix} x \\ z \end{pmatrix} + b, \quad (a, b) \in O, \tag{10b}$$

$$\text{network design: (1b)}, \tag{10c}$$

$$\text{graph partition with connectivity: (1c)–(1j)}, \tag{10d}$$

$$\text{inter-zonal line indicators: (1k), (1l)}. \tag{10e}$$

Here, O denotes the set of optimality cuts. Typically, in Benders-like frameworks one also requires feasibility cuts F. However, we later show that these cuts are not needed in our case. The master problem is a relaxation of the full trilevel problem that ignores all second- and third-level constraints of the original problem and in which τ overestimates the spot-market outcome less redispatch, i.e., $\tau \geq \psi_2 - \psi_3$. If a solution of the master problem satisfies all ignored constraints, it is also a solution of the original problem. Otherwise, optimality cuts (a, b) are added to the master problem, which is then solved again. Suitable cuts are generated by solving a corresponding subproblem. In our context, this is a bilevel problem consisting 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 subsequently: The zonal spot-market model (2) with fixed line expansions and price zones (given as the result of the master problem) and the redispatch model (3) for fixed line expansions, capacity investment, production, and demand decisions (given as the result of the master problem and the zonal spot-market model).

In [24] it is shown that the generalized Benders decomposition converges under suitable assumptions. However, in [52] it is proven that its naive application to nonconvex problems, like our trilevel problem, may not even lead to local optima. The main reason for this failure is that the standard Benders cuts are not entirely valid in the general nonconvex case. Luckily, we are able to construct optimality cuts that yield a globally optimal solution. For the construction of these optimality cuts we need to compute valid upper bounds for $\psi_2 - \psi_3$. This can be done by solving a single-level MIQP. To see this, we need to collect some more insights.

Lemma 10. *For a bilevel problem*

$$\begin{aligned} \max \quad & f_1(x, y) \\ \text{s.t.} \quad & (x, y) \in Z_1, \\ & y \in \arg\max\{f_2(x, y) \colon (x, y) \in Z_2\}, \end{aligned}$$

the single-level problem

$$\max \quad f_1(x, y) \quad \text{s.t.} \quad (x, y) \in Z_1 \cap Z_2 \tag{11}$$

is a relaxation.

Problem (11) is commonly referred to as the high point relaxation; see, e.g., [44]. The generalization to trilevel or multilevel models is straightforward and can be applied to our trilevel problem. This yields the following lemma.

Lemma 11. *The single-level problem*

$$\max \quad \psi_2 - \psi_3 \quad \text{s.t.} \quad (1b), (2i), (3b)–(3k) \tag{12}$$

is a relaxation of the trilevel market model.

Proof. A generalization of Lemma 10 directly reveals that

$$\max \quad \psi_1 \quad \text{s.t.} \quad (1b)–(1l), (2b)–(2l), (3b)–(3k) \tag{13}$$

is a relaxation of the trilevel market model. Next, we use the special structure of the trilevel model. The consequence of the observed weak coupling is that the optimal solution of the high point relaxation (13) does not depend on any original second-level variables but $\bar{q}_g^{\text{new}}$. Hence, except for Constraints (2i) the constraint block (2b)–(2l) is redundant. Since the graph partitioning of the first level only impacts this redundant constraint block, Constraints (1c)–(1l) are redundant as well. Thus, Model (13) is equivalent to

$$\max \quad \psi_1 \quad \text{s.t.} \quad (1b), (2i), (3b)–(3k). \tag{14}$$

From Equation (4), we know $\psi_2 - \psi_3 \geq \psi_1$. We can use this information, to further relax the objective function of (14) and obtain Model (12). □

Model (12) is a single-level MIQP that contains a network design problem. It is considerably less complex than the original trilevel problem. The optimal objective value of this problem is denoted by ψ_{ub}^*. We further point out that Model (12) is closely related to the economic concept of an integrated planner—a model that is used to determine global system optima; see for example [34]. In our setting, the integrated planner problem models a fictitious integrated generation and transmission company that decides on a welfare maximizing network design, production capacity expansion, as well as production and demand quantities. This problem corresponds exactly to Model (14), i.e., to the high point relaxation of our trilevel problem, which means that the economic concept of an integrated planner and the mathematical concept of high point relaxations coincide. In order to state valid bounds for $\psi_2 - \psi_3$, we need some more notation. We denote a feasible graph partitioning by $x = (x_{n,i})_{n \in N, i \in [k]}$, corresponding inter-zonal indicator variables by $y = (y_l)_{l \in L}$, and a feasible network design by $z = (z_l)_{l \in L^{\text{new}}}$. Similarly, we denote a welfare-maximizing graph partition for a fixed network design $\hat{z}$ by $x(\hat{z})$. The corresponding inter-zonal indicator variables are denoted by $y(\hat{z})$. Furthermore, we denote the optimal spot-market outcome for fixed first-level variables (x, y, z) by $\psi_2^*(x, y, z)$ and minimal redispatch costs for fixed spot-market quantities in dependence of fixed first-level variables by $\psi_3^*(x,y,z)$. Thus, the overall objective value of the trilevel market model for fixed first level variables is given by

$$\psi_1^*(x,y,z) = \psi_2^*(x,y,z) - \psi_3^*(x,y,z) - \sum_{l \in L^{\text{new}}} c_l^{\text{inv}} z_l. \tag{15}$$

We further denote the optimal objective value of Model (12) for a fixed network design by $\psi_{\text{ub}}^*(\hat{z})$. For the remainder of this section, we suppose that the following assumption holds.

Assumption 2. *Let (x,y,z) be part of a feasible first-level solution. Then, the corresponding optimal spot-market outcome less minimal redispatch costs is greater or equal to 0, i.e., $\psi_2^*(x,y,z) - \psi_3^*(x,y,z) \geq 0$.*

This assumption is similar to Assumption 1 and reasonable from an economic perspective. Although Assumption 2 only requires a "$\geq$", Assumption 1 seems to be weaker. It is quite easy to construct instances that can have feasible (but not optimal) zonings with a negative welfare outcome. However, no implications between the two assumptions can be drawn. We can now state the following bounds.

Lemma 12. *Suppose that Assumption 2 holds.*

(1) *Let (x,y,z) be part of a feasible solution of the first-level problem (1). Then,*

$$\psi_{\text{ub}}^* \geq \psi_2^*(x,y,z) - \psi_3^*(x,y,z) \tag{16}$$

holds.

(2) *Let $\hat{z}$ be a feasible network design. Then,*

$$\psi_{\text{ub}}^* \geq \psi_{\text{ub}}^*(\hat{z}) \geq \psi_2^*\left(x(\hat{z}), y(\hat{z}), \hat{z}\right) - \psi_3^*\left(x(\hat{z}), y(\hat{z}), \hat{z}\right) \tag{17}$$

holds.

Proof. The claims follow directly from Assumption 2 and Lemma 11. □

Before we introduce suitable optimality cuts, we need to specify some sets. The feasible set of the first-level problem is denoted by

$$\mathcal{F}_1 := \{(x,y,z,s,u) \colon (x,y,z,s,u) \text{ is feasible for Model (1)}\}$$

and the set of projected feasible first-level solutions by

$$\mathcal{F}_1^{\text{proj}} := \{(x,y,z) \colon \exists (s,u) \text{ with } (x,y,z,s,u) \in \mathcal{F}_1\}.$$

We further denote the feasible set of the master problem for a given set of optimality cuts O by

$$\mathcal{F}_{\text{M}}^{O} := \{x,y,z,s,u,\tau) \colon (x,y,z,s,u,\tau) \text{ satisfies (10b)–(10e)}\}.$$

Finally, we can prove the following lemma.

Lemma 13. *Suppose that Assumption 2 holds. Let $\mathcal{F} \subseteq \mathcal{F}_1^{\text{proj}}$ and let $O = O(\mathcal{F})$ consist of the cuts*

$$\begin{aligned} \tau \leq\ & \psi_2^*(\hat{x},\hat{y},\hat{z}) - \psi_3^*(\hat{x},\hat{y},\hat{z}) \\ & + \psi_{\text{ub}}^* \left(\sum_{i \in [k]} \sum_{n \in N : \hat{x}_{n,i} = 0} x_{n,i} + \sum_{i \in [k]} \sum_{n \in N : \hat{x}_{n,i} = 1} (1 - x_{n,i}) \right) \\ & + \psi_{\text{ub}}^* \left(\sum_{l \in L^{\text{new}} : \hat{z}_l = 0} z_l + \sum_{l \in L^{\text{new}} : \hat{z}_l = 1} (1 - z_l) \right), \end{aligned} \tag{18a}$$

$$\tau \leq \psi_{\text{ub}}^*(\hat{z}) + \psi_{\text{ub}}^* \left(\sum_{l \in L^{\text{new}} : \hat{z}_l = 0} z_l + \sum_{l \in L^{\text{new}} : \hat{z}_l = 1} (1 - z_l) \right), \tag{18b}$$

for all $(\hat{x}, \hat{y}, \hat{z}) \in \mathcal{F}$*. Furthermore, let* O *contain the cut*

$$\tau \le \psi^*_{\text{ub}}. \tag{19}$$

Then for any point $(x, y, z, s, u) \in \mathcal{F}_1$*, there exists a point* $(x, y, z, s, u, \tau) \in \mathcal{F}^O_{\text{M}}$ *that satisfies*

$$\tau - \sum_{l \in L^{\text{new}}} c^{\text{inv}}_l z_l \ge \psi^*_1(x, y, z). \tag{20}$$

As pointed out earlier, this means that τ overestimates the spot-market outcome less redispatch, i.e., $\tau \ge \psi_2 - \psi_3$, and that every master problem is a relaxation of the original trilevel problem.

Proof. Consider $(x, y, z, s, u) \in \mathcal{F}_1$. We distinguish the following three cases.

- C1: If there are no x, y such that $(x, y, z) \in \mathcal{F}$, all cuts in O have a right-hand side greater or equal to ψ^*_{ub}. Hence, Inequality (19) is binding.
- C2: If $(x, y, z) \notin \mathcal{F}$ but if there are x', y' such that $(x', y', z) \in \mathcal{F}$, there exists a cut of the form (18b) that simplifies to $\tau \le \psi^*_{\text{ub}}(z)$. Since all other cuts in O have again a right-hand side greater or equal to ψ^*_{ub} by Lemma 12, this cut is binding.
- C3: If $(x, y, z) \in \mathcal{F}$, then O contains a cut of the form (18a), that simplifies to $\tau \le \psi^*_2(x, y, z) - \psi^*_3(x, y, z)$, and a cut of the form (18b) that simplifies to $\tau \le \psi^*_{\text{ub}}(z)$. Again, all other cuts in O have a right-hand side of at least ψ^*_{ub}. According to Lemma 12, $\tau = \psi^*_2(x, y, z) - \psi^*_3(x, y, z)$ holds.

In all three cases, $\tau \ge \psi^*_2(x, y, z) - \psi^*_3(x, y, z)$ holds. Together with Equation (4), this proves the lemma. □

We remark that both cuts (18a) and (18b) are based on the network design decisions of the first level. Cut (18a) is a direct extension of the cut stated in [25]. Cut (18b) uses the costs introduced by the network design problem and is usually stronger than the cut discussed in [25].

We can now present our problem-specific generalized Benders algorithm; see Algorithm 1. This algorithm yields a globally optimal solution of the trilevel model.

Algorithm 1: Generalized Benders decomposition for the trilevel problem.

Input: The trilevel problem of Section 2.
Output: A globally optimal solution for the trilevel problem.

```
Initialize O ← {(0, ψ*_ub)}, Θ ← 0, φ ← ∞, Z ← ∅, S ← ∅.
while Θ < φ do
    Solve master problem (10).
      Let (x̂, ŷ, ẑ) be part of its optimal solution, set φ to its optimal value.
    Solve the second-level problem (2) with fixed (x̂, ŷ, ẑ).
      Let (q^spot, d^spot, q̄^new) be part of its optimal solution and let ψ*_2(x̂, ŷ, ẑ) be
      its optimal value.
    Solve the third-level problem (3) with fixed (ẑ, q^spot, d^spot, q̄^new).
      Let (q^red, d^red, f^red, θ) be its optimal solution and let ψ*_3(x̂, ŷ, ẑ) be its
      optimal value.
    if ψ = ψ*_2(x̂, ŷ, ẑ) − ψ*_3(x̂, ŷ, ẑ) − Σ_{l∈L^new} c^inv ẑ_l > Θ then
        Set Θ ← ψ and S ← (x̂, ŷ, ẑ, q^spot, d^spot, q̄^new, q^red, d^red, f^red, θ).
    Add cut (18a) to O.
    if ẑ ∉ Z then add cut (18b) to O.
return S.
```

Theorem 1. *Suppose that Assumption 2 holds. Assume further that for every (x, y, z) that is part of a feasible solution for the first-level problem, the second-level solution is unique, and that the network $\mathcal{G} = (N, L)$ is finite and connected. Then, Algorithm 1 terminates within a finite number of iterations and returns a globally optimal solution for the trilevel problem of Section 2.*

Proof. According to Assumption 2, $\Theta \leftarrow 0$ is a valid initialization. In order to obtain a bounded master problem in the first iteration, we initialize O with the cut $\tau \leq \psi^*_{\text{ub}}$. Following Lemma 13 this initialization of O is correct. Since $\mathcal{G}$ is connected, there exists a connected graph partition. Hence, the master problem is always feasible. Furthermore, every second- and third-level problem is feasible. Hence, no feasibility cuts are required. In the following, we assume that the algorithm already performed $s \geq 1$ iterations. Then, Θ denotes the best objective value of the first s iterations; see Lines 6 and 7. Let $(\hat{x}, \hat{y}, \hat{z}, \hat{\tau})$ be part of the solution of the master problem in iteration $s + 1$ with corresponding objective value ϕ.

We first show that Algorithm 1 is finite and therefore assume that the combination of $\hat{x}$, $\hat{y}$, and $\hat{z}$ has already been examined in a previous iteration. This corresponds to the case C3 in the proof of Lemma 13 and $\hat{\tau} = \psi^*_2(\hat{x}, \hat{y}, \hat{z}) - \psi^*_3(\hat{x}, \hat{y}, \hat{z})$ holds. Thus, we obtain

$$\phi = \psi^*_2(\hat{x}, \hat{y}, \hat{z}) - \psi^*_3(\hat{x}, \hat{y}, \hat{z}) - \sum_{l \in L^{\text{new}}} c^{\text{inv}}_l \hat{z}$$

and since Θ denotes the incumbent, $\Theta \geq \phi$ holds. Hence, the algorithm terminates whenever a combination of a network design and a graph partition is examined for the second time. Since $\mathcal{G}$ is finite, the number of feasible first-level solutions and thus the algorithm is finite.

We now show that Algorithm 1 returns a globally optimal solution. Due to Lemma 13 and optimality,

$$\phi = \hat{\tau} - \sum_{l \in L^{\text{new}}} c^{\text{inv}} \hat{z} \geq \psi^*_1$$

holds, where ψ^*_1 denotes the optimal objective value of the original trilevel problem. This means, that the master problem yields a valid bound ϕ for the objective value of the original trilevel problem in every iteration. The algorithm only terminates in case of $\Theta \geq \phi$, i.e., when the incumbent Θ is greater or equal to its bound ϕ. This is only the case, when an optimal solution is found. □

We want to point out that Algorithm 1 still works correctly if we only add the cuts (18a). However, adding cuts (18b) drastically decreases the number of iterations required and hence improves running times. This is shown in the next section, where we present computational results.

5. Computational Study

In this section we present numerical results for both global solution approaches and their enhancements. We first describe the computational setup and the test set in Section 5.1. Afterward, in Section 5.2 we present the results of the single-level MIQP approach and evaluate the effectiveness of the proposed bound tightening. We then discuss the results of the generalized Benders approach in Section 5.3. Two case studies that highlight the economic effects that are captured in our model are presented Section 6.

TABLE 1. Test networks with number of nodes ($|N|$), existing lines ($|L^{\text{ex}}|$), candidate lines ($|L^{\text{new}}|$), and scenarios ($|T|$).

Network	$\|N\|$	$\|L^{\text{ex}}\|$	$\|L^{\text{new}}\|$	$\|T\|$	Reference
Grimm-et-al-2016-3	3	3	1	4	[26]
Chao-Peck-1998	6	6	2	4	[11]
Grimm-et-al-2016-6	6	6	2	52	[26]
Haefner-2017	9	8	1	52	[30]
Kleinert-Schmidt-2018-9	9	12	4	52	App. A
Kleinert-Schmidt-2018-12	12	16	6	52	App. A

5.1. **Instances and Computational Setup.** For our computations we used the instances given in Table 1. The networks vary in the number of nodes, number of existing and candidate transmission lines, and the number of scenarios. The first four networks are taken from the literature; see Table 1 for the references. The last two networks extend the network Grimm-et-al-2016-6. Figure 6 in Appendix A shows the network Kleinert-Schmidt-2018-12, which combines the networks Grimm-et-al-2016-6 and Kleinert-Schmidt-2018-9. Our test set includes all networks with up to 9 nodes for every possible number of zones $k \in \{1, \dots, |N|\}$. For the remaining network Kleinert-Schmidt-2018-12, we exclude the cases $k \in \{4, \dots, 8\}$, since no approach is able to solve these instances within the time limit. In total, our test set contains 40 instances.

All models as well as the generalized Benders framework have been implemented in Python 2.7.12 using the graph library NetworkX [47] as it is contained in Anaconda 2.7 [1]. All models that need to be solved are either MIQPs, MIPs, or convex QPs. In our computational study we used Gurobi 7.5.1 [29] to solve these problem classes. All computations have been executed with a time limit of 6 h on a compute cluster; cf. [49] for the details about the installed hardware. Throughout this section we use log-scaled performance profiles as proposed in [18] to compare running times or branch-and-bound node counts.

We further note that, since the graph partitioning formulation in the first-level problem is inherently symmetric, we equipped both solution approaches with the symmetry breaking constraints proposed in [42]. These additional constraints are shown to be very beneficial in a quite similar setup; see [25].

5.2. **Results for the Single-Level MIQP Reformulation.** We now present the numerical results obtained with the single-level MIQP reformulation. First, we discuss some general observations that hold for both discussed variants, i.e., the standard approach (SLMIQP) and the approach equipped with the new primal-dual bound tightening (SLMIQP-BT) of Section 3.

The reformulation to a single-level problem results in quite large problems, even for considerably small networks. For instance, for $k = 4$ the 6-node network Grimm-et-al-2016-6 leads to an MIQP with 39 649 constraints, 15 119 variables (thereof 1776 binaries), 101 421 nonzeros, and 156 quadratic objective terms. In general, already mid-sized networks with up to 9 nodes can be too challenging for the single-level approach. For example, the network Kleinert-Schmidt-2018-9 can only be solved for the extreme numbers of zones, i.e., for $k = 1$ or $k = 9$. This behavior can be expected, since the number of connected graph partitions is largest for mid-sized k and smallest for extreme values of k.

We now compare the SLMIQP with the SLMIQP-BT approach. To this end, we tested both approaches on the instance set described in Section 5.1. We start by pointing out some numerical issues. The single-level reformulations suffer from many

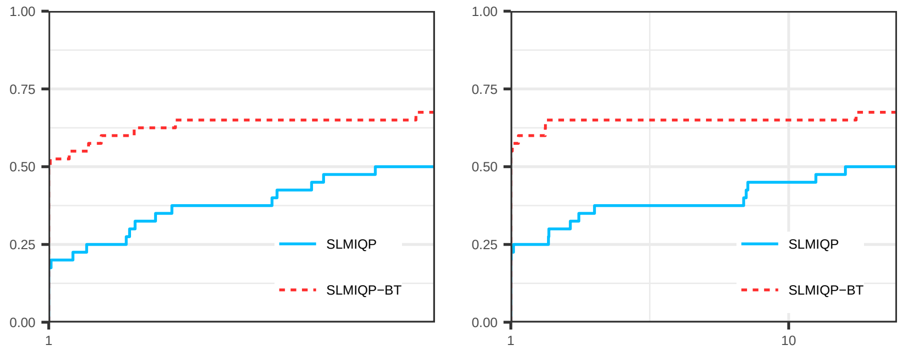

FIGURE 2. Log-scaled performance profiles of the single level MIQP approach with (SLMIQP-BT) and without (SLMIQP) bound tightening. Left: Running times. Right: Node counts.

numerically challenging big-M formulations for linearizing KKT complementarity conditions. As a result, solutions obtained by the solver that fulfill all constraints of the reformulated and linearized single-level model (within the tolerances of the solver) may still violate KKT complementarity conditions. For example, due to insufficient precision of the solver, it may occur that a primal constraint is not binding but the corresponding dual variable is greater than zero. For this reason, we implemented a routine, that checks every solution for KKT complementarity. It turns out that from a numerical point of view, solutions of SLMIQP-BT are way more reliable than solutions of SLMIQP. In fact, for 50 % of the tested instances, solutions obtained by SLMIQP are violating KKT complementarity. This problem is almost completely resolved by using the proposed bound tightening. In this case, only one solution violates KKT complementarity.

Let us now analyze running times and branch-and-bound nodes. In order to have a fair comparison, we consider solutions violating KKT complementarity as infeasible. For every such solution, we recompute the corresponding instance with modified solver settings, i.e., an increased numeric focus and a tighter integer feasibility tolerance. In general, this helps to satisfy KKT complementarity conditions but also results in longer running times. If a solution obtained with modified solver settings still violates KKT complementarity, we label the respective instance as "not solved". The corresponding performance profiles for running times and branch-and-bound nodes are given in Figure 2. The left profile shows, that SLMIQP-BT clearly dominates SLMIQP in terms of running times. Not only is the former method faster for more than 50 % of the instances, whereas the latter is faster for less than 25 % of the instances. It is also more reliable in terms of the number of solved instances. SLMIQP-BT solves around 67 % of the test set within the time limit, compared to 50 % for SLMIQP. We note that 6 out of 20 unsolved instances of SLMIQP result from violations of the KKT complementarity conditions, which could not be resolved even with modified solver settings. A similar result can be drawn from the right profile, which reveals that SLMIQP-BT needs to solve significantly less branch-and-bound nodes.

Altogether, SLMIQP-BT clearly outperforms SLMIQP in terms of running times, reliability, and numerical stability.

5.3. **Results for the Generalized Benders Approach.** In this section, we present numerical results for the generalized Benders approach and evaluate the benefit of using both proposed optimality cuts. We therefore analyze the performance profile in Figure 3, which compares running times of both approaches. It reveals

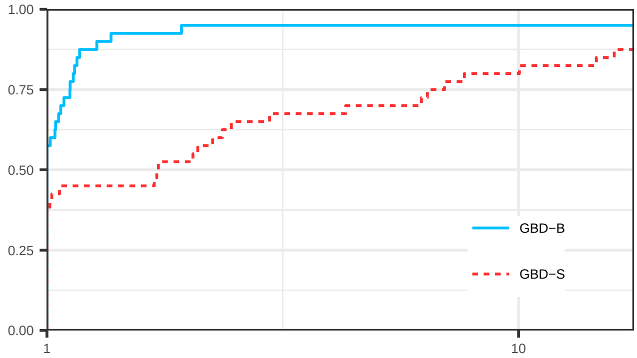

FIGURE 3. Log-scaled performance profiles for the running times of the generalized Benders approaches GBD-S and GBD-B.

that the approach named GBD-B, which uses both optimality cuts (18a) and (18b), clearly dominates the approach GBD-S, which only uses the standard cut (18a). First of all, with 95 % instances solved, GBD-B is more reliable than GBD-S. The latter solves 87.5 % of the tested instances. Moreover, GBD-B is the faster approach for around 60 % of the test set. The reason is that the additional cut significantly reduces the number of iterations that need to be solved. In general, this reduction in iterations linearly correlates to a reduction of the running times. However, one can observe the following trade-off. As seen in Section 4, for each optimality cut (18b) one needs to solve an additional QP in order to obtain $\psi^*_{\text{ub}}(\hat{z})$. In particular for easy instances, this additional effort does not pay off compared to the number of iterations that the additional cut saves. When leaving out considerably easy instances, i.e., all instances of Grimm-et-al-2016-3 and Chao-Peck-1998 as well as instances with $k = 1$ and $k = |N|$ for all other networks, the benefit of the additional cut becomes even more obvious. In this case, GBD-B is the faster approach for 22 of the remaining 23 instances and GBD-B shows speedup factors of up to 16 compared to GBD-S. Altogether, we conclude that GBD-B clearly outperforms GBD-S.

6. Case Studies

In this section we discuss the interplay of welfare maximizing price zones and network expansions for two simplified but realistic instances. For both instances, we compare the case of a single price zone with the one using two optimally configured price zones. All results are obtained with the method GBD-B; see Table 6 in Section 7 for details on the running times.

First, we discuss a variant of the well-known network from [11] with four scenarios. The original network is extended by two candidate lines; see Figure 4, where all relevant network data as well as the optimal price zoning and network expansion for $k = 2$ are depicted. The inverse demand functions, i.e., the willingness to pay of the consumers, vary between the different scenarios to model fluctuating demand. We point out that all three generators (at the nodes 1, 2, and 4) are assumed to be candidate generators. Table 2 sums up welfare outcomes and related costs and Table 3 states relevant spot-market quantities. The interpretation of the results is the following. For a uniform price zone, the generator at node 1 covers the total spot-market demand. The reason is that for a uniform price zone all transmission lines are ignored during spot-market trading, i.e., no scarcities are considered. Hence, only the generator with the cheapest cost structure (as a combination of variable and investment costs) is successful at the spot market. As a result, the candidate generators at node 2 and 4 do not invest in generation capacity. All consumers

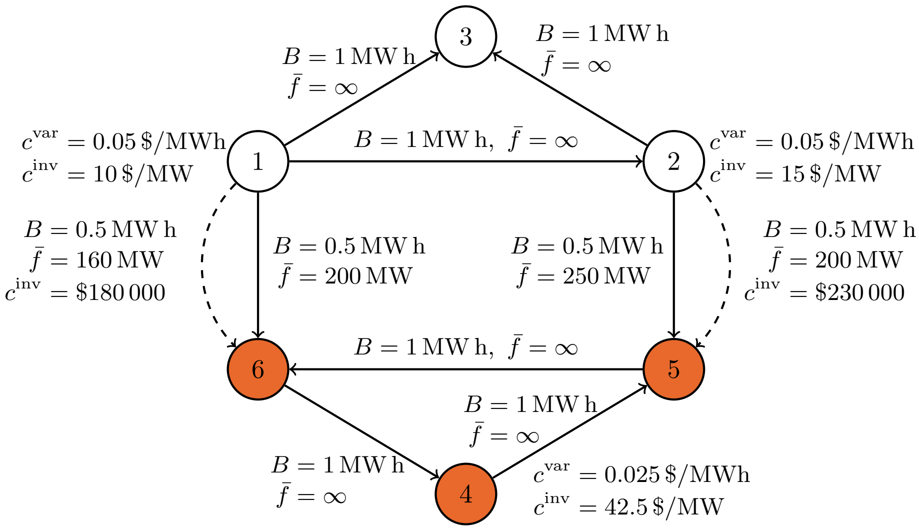

FIGURE 4. The extended Chao-Peck network with two optimally configured price zones and candidate lines (dashed).

TABLE 2. Welfare outcomes and costs for the Chao-Peck network in €.

	Welfare		Costs	
Zones	Total	Spot-Market	Redispatch	Line Investment
1	121 447	273 810	134 363	18 000
2	231 948	242 447	10 499	0

TABLE 3. Aggregated spot-market quantities for all 4 scenarios of the Chao-Peck network in MWh.

	Generation at node			Demand at node		
Zones	1	2	4	3	5	6
1	8792	0	0	2796	2898	3098
2	4596	0	3548	2796	2574	2774

(located at nodes 3, 5, and 6) have non-zero demand. Due to capacity constraints of the transmission lines $(1, 6)$ and $(2, 5)$ that connect the northern with the southern region, the spot-market quantities are infeasible for the actual physical network in every scenario. Since only the generator at node 1 invests in generation capacity, generation cannot be redispatched by the TSO by using other generators. It is therefore beneficial to invest in candidate line $(1, 6)$ to supply the south with energy from the north. However, investment costs for the second candidate line $(2, 5)$ would overcompensate the potential utility of consumers in the south. Thus, the second candidate line is not built. The situation changes when two price zones are considered. The optimal zoning splits north and south and thus makes the scarce transmission capacities of lines $(1, 6)$ and $(2, 5)$ visible at the spot market. Now, the generator at node 1 cannot supply the total spot-market demand of the consumers in the south (located at node 5 and 6). As a consequence, the generator at node 4 (with much higher investment costs compared to the generator at node 1) is also successful at the zonal spot market. The result is a higher price in the southern

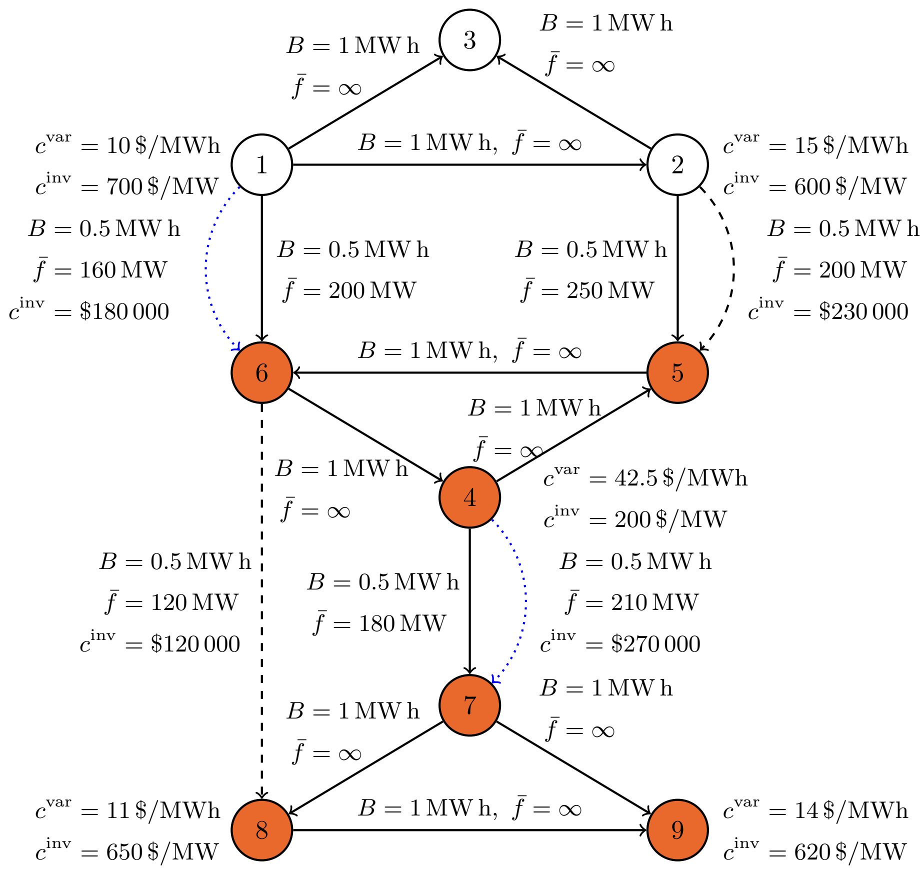

FIGURE 5. The Kleinert-Schmidt-9 network with two optimally configured price zones and optimal network expansion (dotted lines).

zone and it is not beneficial anymore to invest in an additional transmission line linking the northern with the southern zone. This observation is very intuitive and expected: More price zones better reflect scarcities in the network, lead to generation investments closer to demand, and thus reduce the expansion of the network. In other words, this means that expensive network expansion can be substituted by an optimal price zone configuration. As a result, total welfare rises.

We now turn to the instance Kleinert-Schmidt-9, a 9-node network with 52 scenarios that extends the 6-node network published in [26]; see Figure 5. Again, the figure depicts all relevant network data as well as the optimal price zoning and network expansion for $k = 2$ zones. Corresponding welfare outcomes and spot-market quantities are denoted in Table 4 and 5, respectively. For a uniform price zone, the situation is quite similar to the one discussed for the first network. The total spot-market demand is supplied by the cheapest candidate generator (w.r.t. investment and variable costs) at node 1. This is infeasible due to capacity bottlenecks in the network. Hence, it is beneficial to invest in candidate line $(1, 6)$. Again, the investment costs for this line are overcompensated by a decrease in the redispatch costs but investment in an additional line is too expensive. In contrast, for two price zones, it is welfare maximizing to invest in two candidate lines, namely $(1, 6)$ and $(4, 7)$. This is rather counter-intuitive at a first glance. The reason, why it is

Table 4. Welfare outcomes and costs for the Kleinert-Schmidt-9 network in €.

	Welfare		Costs	
Zones	Total	Spot-Market	Redispatch	Line Investment
1	2 556 522	5 609 678	2 873 155	180 000
2	4 219 156	5 536 585	867 429	450 000

Table 5. Aggregated spot-market quantities for all 52 scenarios of the Kleinert-Schmidt-9 network in MWh.

	Generation at node					Demand at node			
Zones	1	2	4	8	9	3	5	6	7
1	138 074	0	0	0	0	32 158	35 579	38 179	32 158
2	63 878	0	0	72 845	0	32 158	35 241	37 841	31 482

beneficial to invest in the additional line $(4, 7)$, is the following. Candidate line $(1, 6)$ is constructed to account for scarcities between north and south. However, not all scarcities across the two zones can be resolved by this additional line. Hence, the cheapest generator in the southern zone (located at node 8) provides the demand of the southern zone that cannot be satisfied by generators of the north via inter-zonal lines. This results in a higher price in the south compared to the north (because the generator at node 8 has higher variable costs than the generator at node 1). Due to scarcities across the upper and lower triangle of the southern zone (that are ignored at the spot market), this supply cannot be transported to all consumers in the south. An investment in candidate line $(4, 7)$ is now reasonable and pays off due to a decrease in redispatch costs.

These two examples demonstrate that it is crucial to evaluate the question of optimal price zone configurations and network expansions carefully. Optimal configurations may be counter-intuitive and may be missed when not using multilevel global optimization techniques like the ones provided in this paper.

7. Conclusion

In this paper we presented a mixed-integer nonlinear trilevel model for computing a combined optimal network design and optimal price zones in electricity markets. Since no general-purpose algorithms exist for this type of problem, we developed two different global solution approaches. In Section 5, we evaluated both approaches and discussed the effectiveness of the proposed enhancements. In this section we compare both approaches in terms of reliability, running times, and numerical stability.

For a proper discussion, we present running times for all tested variants of solution approaches and instances in Table 6. For each instance, the fastest running time is printed in bold. We first compare the reliability of the approaches. The generalized Benders approach solves significantly more instances than the single-level MIQP approach. Moreover, every instance that is solved by one of the single-level approaches is also solved by both generalized Benders approaches. On the other hand, there are many instances that are solved by both generalized Benders approaches but by none of the single-level approaches.

We obtain a similar result, when we analyze running times. In fact, among the 38 instances that are solved by at least one approach, GBD-B is the fastest approach for 27 instances. Furthermore, GBD-S is the fastest approach for 9 of the remaining

11 instances. This leaves 2 instances for which a single-level reformulation approach is faster than both generalized Benders approaches. However, both instances are easy instances with $k = 1$. The reason why the generalized Benders approaches are performing worse than the single-level MIQP approaches on these instances is the

Table 6. Comparison of running times (in s) of the single-level MIQP and generalized Benders approaches.

Network	Zones	SLMIQP	SLMIQP-BT	GBD-S	GBD-B
Grimm-et-al-2016-3	1	0.08	0.12	**0.04**	**0.04**
Grimm-et-al-2016-3	2	0.06	0.07	**0.04**	0.05
Grimm-et-al-2016-3	3	0.05	0.06	**0.01**	**0.01**
Chao-Peck-1998	1	0.34	0.22	**0.06**	**0.06**
Chao-Peck-1998	2	2.43	0.63	**0.12**	0.13
Chao-Peck-1998	3	2.50	4.78	**0.29**	**0.29**
Chao-Peck-1998	4	7.58	6.25	**0.25**	0.28
Chao-Peck-1998	5	3.89	0.96	**0.09**	0.10
Chao-Peck-1998	6	1.39	0.81	**0.02**	**0.02**
Grimm-et-al-2016-6	1	1.92	1.03	**0.16**	0.21
Grimm-et-al-2016-6	2	57.06	371.79	1.40	**0.81**
Grimm-et-al-2016-6	3	—	949.55	4.54	**1.84**
Grimm-et-al-2016-6	4	1843.84	1241.38	3.61	**1.53**
Grimm-et-al-2016-6	5	2018.33	381.15	1.29	**0.62**
Grimm-et-al-2016-6	6	3.10	2.74	0.24	**0.23**
Haefner-2017	1	1.18	0.78	**0.12**	0.24
Haefner-2017	2	—	—	**0.25**	0.26
Haefner-2017	3	—	338.32	0.80	**0.47**
Haefner-2017	4	—	320.19	1.91	**0.12**
Haefner-2017	5	—	372.47	2.44	**0.24**
Haefner-2017	6	—	213.06	1.87	**0.13**
Haefner-2017	7	—	201.83	1.12	**0.15**
Haefner-2017	8	—	10.31	0.31	**0.14**
Haefner-2017	9	3.39	3.35	**0.10**	0.11
Kleinert-Schmidt-2018-9	1	4.65	**1.49**	3.82	4.38
Kleinert-Schmidt-2018-9	2	—	—	37.55	**12.67**
Kleinert-Schmidt-2018-9	3	—	—	1944.97	**303.54**
Kleinert-Schmidt-2018-9	4	—	—	—	**1563.77**
Kleinert-Schmidt-2018-9	5	—	—	—	**2281.35**
Kleinert-Schmidt-2018-9	6	—	—	4320.88	**694.21**
Kleinert-Schmidt-2018-9	7	—	—	350.40	**81.49**
Kleinert-Schmidt-2018-9	8	—	—	17.73	**10.37**
Kleinert-Schmidt-2018-9	9	14.03	18.34	3.94	**3.88**
Kleinert-Schmidt-2018-12	1	7.81	**2.44**	8.35	11.41
Kleinert-Schmidt-2018-12	2	—	—	4770.68	**2338.93**
Kleinert-Schmidt-2018-12	3	—	—	—	—
Kleinert-Schmidt-2018-12	9	—	—	—	—
Kleinert-Schmidt-2018-12	10	—	—	—	**2590.91**
Kleinert-Schmidt-2018-12	11	—	—	305.93	**43.91**
Kleinert-Schmidt-2018-12	12	16.11	16.24	**8.92**	9.98

computation of the optimality cut (18a), which involves the solution of an MIQP; see Section 4. For easier instances, solving this MIQP does not pay off compared to the overall required effort for solving the instance.

In Section 5.2, we discussed that KKT complementarity conditions resulting from the reformulation to a single-level problem are numerically challenging, already for small networks. This results in violated KKT complementarity conditions and/or long running times and thus, in many unsolved instances. In contrast, the generalized Benders approach decomposes the trilevel problem and solves a considerably small MIP and two QPs in every iteration. Hence, this approach is numerically much more stable.

Finally, we compare the columns SLMIQP and SLMIQP-BT as well as the columns GBD-S and GBD-B. One can see that the novel techniques presented in this paper clearly outperform the approaches that are already discussed in [25]. More instances are solved with the novel techniques and, almost always, significant savings in running times can be observed.

As a result of the above discussions, we conclude that (i) the generalized Benders approach outperforms the single-level MIQP approach in every aspect and that (ii) our newly developed techniques outperform the approaches stated in [25]. Following Section 5.3 and Table 6, GBD-B can be considered the best solution approach proposed in this paper.

Acknowledgments

This research has been performed as part of the Energie Campus Nürnberg and is supported by funding of the Bavarian State Government. Moreover, the authors thank the DFG for their support within projects A05 and B08 in CRC TRR 154. The research of the second author has been supported by the Emerging Field Initiative (EFI) of the Friedrich-Alexander-Universität Erlangen-Nürnberg through the project "Sustainable Business Models in Energy Markets". We also thank Frauke Liers, Veronika Grimm, and Gregor Zöttl for many fruitful discussions on the topic of this paper. Finally, we are very grateful to an anonymous reviewer, whose comments on the manuscript greatly helped to improve the quality of the paper.

References

[1] *Anaconda*. https://www.continuum.io/anaconda-overview. Accessed: 2018-02-06.

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

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

[4] D. Bienstock and S. Mattia. "Using mixed-integer programming to solve power grid blackout problems." In: *Discrete Optimization* 4.1 (2007), pp. 115–141. DOI: 10.1016/j.disopt.2006.10.007.

[5] S. Binato, M. V. F. Pereira, and S. Granville. "A new Benders decomposition approach to solve power transmission network design problems." In: *IEEE Transactions on Power Systems* 16.2 (2001), pp. 235–240. DOI: 10.1109/mper.2001.4311380.

[6] 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.

[7] 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.

[8] 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.

[9] J. Boucher and Y. Smeers. "Alternative Models of Restructured Electricity Systems, Part 1: No Market Power." In: *Operations Research* 49.6 (2001), pp. 821–838. DOI: 10.1287/opre.49.6.821.10017.

[10] 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.

[11] 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.

[12] 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.

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

[14] O. Daxhelet and Y. Smeers. "The EU regulation on cross-border trade of electricity: A two-stage equilibrium model." In: *European Journal of Operational Research* 181.3 (2007), pp. 1396–1412. DOI: 10.1016/j.ejor.2005.12.040.

[15] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova. *Bilevel Programming Problems*. Springer, 2015. DOI: 10.1007/978-3-662-45827-3.

[16] 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.

[17] 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.

[18] 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.

[19] A. Ehrenmann and Y. Smeers. "Inefficiencies in European congestion management proposals." In: *Utilities Policy* 13.2 (2005), pp. 135–152. DOI: 10.1016/j.jup.2004.12.007.

[20] J. Fairbrother and A. N. Letchford. "Projection results for the k-partition problem." In: *Discrete Optimization* 26 (2017), pp. 97–111. DOI: 10.1016/j.disopt.2017.08.001.

[21] 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. DOI: 10.2307/2581394.

[22] T. Fukunaga and H. Nagamochi. "Network design with weighted degree constraints." In: *Discrete Optimization* 7.4 (2010), pp. 246–255. DOI: 10.1016/j.disopt.2010.05.004.

[23] S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, and C. Ruiz. *Complementarity modeling in energy markets*. Vol. 180. Springer Science & Business Media, 2012. DOI: 10.1007/978-1-4419-6123-5.

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

[25] V. Grimm, T. Kleinert, F. Liers, M. Schmidt, and G. Zöttl. "Optimal Price Zones in Electricity Markets: A Mixed-Integer Multilevel Model and Global Solution Approaches." In: *Optimization Methods & Software* (2017). DOI: 10.1080/10556788.2017.1401069.

[26] 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.

[27] V. Grimm, L. Schewe, M. Schmidt, and G. Zöttl. "Uniqueness of market equilibrium on a network: A peak-load pricing approach." In: *European Journal of Operational Research* 261.3 (2017), pp. 971–983. DOI: 10.1016/j.ejor.2017.03.036.

[28] M. Grötschel, C. L. Monma, and M. Stoer. "Design of survivable networks." In: *Handbooks in Operations Research and Management Science* 7 (1995), pp. 617–672. DOI: 10.1016/S0927-0507(05)80127-6.

[29] Gurobi Optimization, Inc. *Gurobi Optimizer Reference Manual, Version 7.5*. 2017.

[30] A. Häfner. "Mathematische Modellierung von Stromnetzen: Ein Vergleich von AC- und DC-Modell hinsichtlich Investitionsentscheidungen." MA thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, 2017.

[31] P. Hansen, B. Jaumard, and G. Savard. "New branch-and-bound rules for linear bilevel programming." In: *SIAM Journal on Scientific and Statistical Computing* 13.5 (1992), pp. 1194–1217. DOI: 10.1137/0913069.

[32] W. W. Hogan. "Transmission congestion: the nodal-zonal debate revisited." In: *Harvard University, John F. Kennedy School of Government, Center for Business and Government. Retrieved August* 29 (1999).

[33] C. Hojny, I. Joormann, H. Lüthen, and M. Schmidt. *Mixed-Integer Programming Techniques for the Connected Max-k-Cut Problem*. Tech. rep. 2018. URL: http://www.optimization-online.org/DB_HTML/2018/07/6738.html.

[34] 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 (2013), pp. 2639–2650. DOI: 10.1109/tpwrs.2012.2236110.

[35] V. Krebs, L. Schewe, and M. Schmidt. "Uniqueness and Multiplicity of Market Equilibria on DC Power Flow Networks." In: *European Journal of Operational Research* 271.1 (2018), pp. 165–178. DOI: 10.1016/j.ejor.2018.05.016.

[36] V. Krebs and M. Schmidt. "Uniqueness of market equilibria on networks with transport costs." In: *Operations Research Perspectives* 5 (2018), pp. 169–173. DOI: 10.1016/j.orp.2018.05.002.

[37] 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.

[38] A. Lodi, T. K. Ralphs, and G. J. Woeginger. "Bilevel programming and the separation problem." In: *Mathematical Programming* 146.1-2 (2014), pp. 437–458. DOI: 10.1007/s10107-013-0700-x.

[39] T. L. Magnanti, P. Mirchandani, and R. Vachani. "Modeling and solving the two-facility capacitated network loading problem." In: *Operations Research* 43.1 (1995), pp. 142–157. DOI: 10.1016/0966-8349(95)97847-D.

[40] T. L. Magnanti and R. T. Wong. “Network design and transportation planning: Models and algorithms.” In: *Transportation Science* 18.1 (1984), pp. 1–55. DOI: 10.1287/trsc.18.1.1.

[41] S. A. Malcolm and S. A. Zenios. “Robust optimization for power systems capacity expansion under uncertainty.” In: *Journal of the Operational Research Society* 45.9 (1994), pp. 1040–1049. DOI: 10.2307/2584145.

[42] 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.

[43] M. Minoux. “Networks synthesis and optimum network design problems: Models, solution methods and applications.” In: *Networks* 19.3 (1989), pp. 313–360. DOI: 10.1002/net.3230190305.

[44] J. T. Moore and J. F. Bard. “The mixed integer linear bilevel programming problem.” In: *Operations Research* 38.5 (1990), pp. 911–921. DOI: 10.1287/opre.38.5.911.

[45] E. Morsy. “Approximating the k-splittable capacitated network design problem.” In: *Discrete Optimization* 22.Part B (2016), pp. 328–340. DOI: 10.1016/j.disopt.2016.09.001.

[46] F. H. Murphy and Y. Smeers. “Generation Capacity Expansion in Imperfectly Competitive Restructured Electricity Markets.” In: *Operations Research* 53.4 (Aug. 2005), pp. 646–661. DOI: 10.1287/opre.1050.0211.

[47] *NetworkX. High-productivity software for complex networks.* https://networkx.github.io. Accessed: 2018-02-06.

[48] J.-S. Pang, B. F. Hobbs, and C. J. Day. “Properties of Oligopolistic Market Equilibria in Linearized DC Power Networks with Arbitrage and Supply Function Conjectures.” In: *IFIP Advances in Information and Communication Technology*. Springer, 2003, pp. 113–142. DOI: 10.1007/978-0-387-35699-0_6.

[49] Regionales Rechenzentrum Erlangen. *Woodcrest Cluster.* https://www.rrze.fau.de/dienste/arbeiten-rechnen/hpc/systeme/woodcrest-cluster.shtml. Accessed: 2018-02-06.

[50] G. Reinelt and K. M. Wenger. “Generating partitions of a graph into a fixed number of minimum weight cuts.” In: *Discrete Optimization* 7.1 (2010), pp. 1–12. DOI: 10.1016/j.disopt.2009.07.001.

[51] C. Ruiz and A. J. Conejo. “Robust transmission expansion planning.” In: *European Journal of Operational Research* 242.2 (2015), pp. 390–401. DOI: 10.1016/j.ejor.2014.10.030.

[52] 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.

[53] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn. *Spot pricing of electricity.* Springer Science & Business Media, 1988. DOI: 10.1007/978-1-4613-1683-1.

[54] J. C. Villumsen and A. B. Philpott. “Investment in electricity networks with transmission switching.” In: *European Journal of Operational Research* 222.2 (2012), pp. 377–385. DOI: 10.1016/j.ejor.2012.05.002.

Appendix A. Networks

Figure 6 shows Kleinert-Schmidt-2018-12 that incorporates Kleinert-Schmidt-2018-9 (nodes 1 to 9) and is based on Grimm-et-al-2016-6 (nodes 1 to 6); see [26].

[1]Friedrich-Alexander-Universität Erlangen-Nürnberg, Discrete Optimization, Cauerstr. 11, 91058 Erlangen, Germany; [2]Energie Campus Nürnberg, Fürther Str. 250, 90429 Nürnberg, Germany; [3]Trier University, Department of Mathematics, Universitaetsring 15, 54296 Trier, Germany

E-mail address: `thomas.kleinert@fau.de`

E-mail address: `martin.schmidt@uni-trier.de`

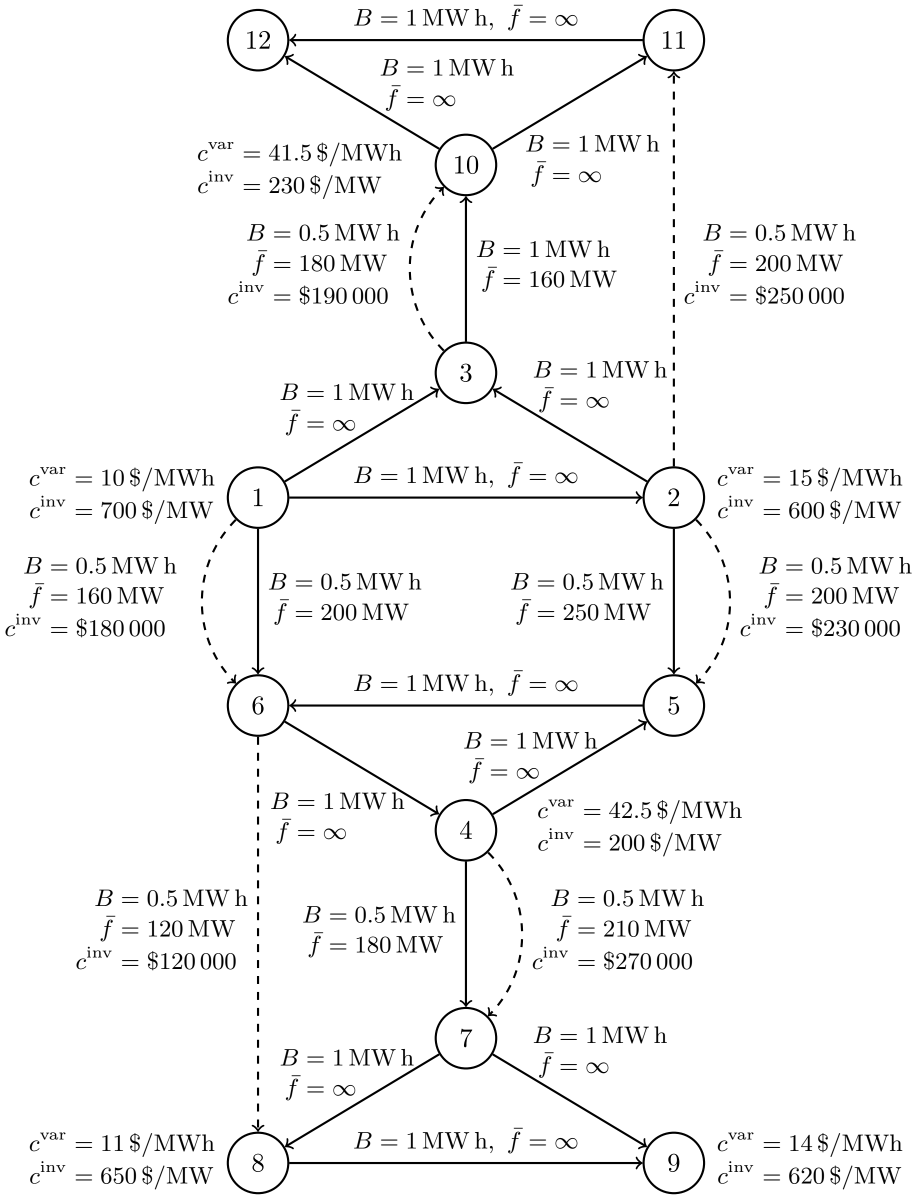

Figure 6. The Kleinert-Schmidt-2018-12 network