IET Generation, Transmission & Distribution

Research Article

Distributed balancing energy activation and exchange optimisation

ISSN 1751-8687
Received on 30th October 2018
Revised 27th February 2019
Accepted on 19th March 2019
E-First on 16th August 2019
doi: 10.1049/iet-gtd.2018.6819
www.ietdl.org

Martin Håberg[1] ✉, *Gerard Doorman*[2]

[1] *Department of Electric Power Engineering, Norwegian University of Science and Technology, O. S. Bragstads plass 2E, 7034 Trondheim, Norway*

[2] *Market Development, Statnett SF, Nydalen allé 33, 0484 Oslo, Norway*

✉ *E-mail: martin.haberg@ntnu.no*

Abstract: European balancing markets are presently fragmented and mostly national. Integrating them on common platforms requires the ability to identify optimal balancing energy volumes to be activated and exchanged to cover imbalances across the interconnected system, but this leads to new challenges. To avoid balancing energy flows compromising operational security, capacity limitations in the transmission network should be taken into account in the optimisation. However, a zonal market structure and limited computational time inhibit using detailed models in the optimisation. This study describes and demonstrates a distributed formulation for optimisation of activation and exchange of balancing energy. Using Benders decomposition, the optimisation is separated into local activation problems that are smaller and less complex, distributing the computational effort. A single exchange problem identifies the optimal exchange volumes between geographical areas using optimality cuts obtained in the local subproblems. The proposed formulation is capable of optimising balancing decisions across a large, interconnected system, while still keeping a zonal market structure and taking detailed network constraints into account.

Nomenclature

Indices

a, n	balancing area
b	balancing energy bid
i, j	node
k	optimality cut
p	exchange scenario

Parameters

λ_{an}^{k}	shadow cost on exchange from area a to n in cut k
$\overline{X}_{an}$	available cross-zonal capacity from area a to n
$\overline{Y}_{b}$	capacity of bid b
Φ_{il}	PTDF from node i to line l
σ_{a}^{k}	value at zero exchange of cut k for area a
$\underline{F}_{l}, \overline{F}_{l}$	remaining transmission capacities on line l
A_{il}	adjacency of node i to line l
C_{b}	activation price for bid b
D_{a}	balancing energy demand in area a
D_{i}	balancing energy demand in node i
F_{a}^{p}	cost of situation p in upper bound calculation
X_{an}^{p}	exchange from area a to neighbour n in situation p

Sets

$\mathscr{A}$	areas
$\mathscr{B}_{a}$	balancing bids located in area a
$\mathscr{B}_{i}$	balancing bids located at node i
$\mathscr{I}_{a}$	nodes, including external, in area a
$\mathscr{I}_{a}^{\text{ext}}$	external nodes representing neighbours of area a
$\mathscr{K}_{a}$	optimality cuts for area a
$\mathscr{L}_{a}$	lines considered for area a
$\mathscr{N}_{a}$	neighbouring areas to a
$\mathscr{P}_{a}$	exchange situations evaluated for area a

Variables

α^{p}	weight of situation p in upper bound calculation
λ_{i}	marginal energy price in node i
θ_{a}	estimated balancing cost in area a
F_{a}	evaluated balancing cost in area a
f_{l}	power flow on line l
x_{an}	exchange from area a to area n
y_{b}	activated volume from bid b
z_{a}	activated volume in area a

Specifiers

$\downarrow$	downward direction
$\uparrow$	upward direction

1 Introduction

European power markets are in the process of integration, and significant progress has already been made in coupling the different day-ahead and intraday markets. For the balancing markets, platforms for integration of balancing services are currently being developed as part of the implementation of new Network Codes and Guidelines. The new pan-European markets for balancing energy provide opportunities to better coordinate balancing actions and exchange of balancing energy across borders. Among the objectives is increasing social welfare by enhancing competition and utilising balancing resources more efficiently [1]. The platforms being established for each of the standardised balancing products [2–4] largely share the same fundamental operational philosophy. Collecting balancing energy bids from all participating transmission system operators (TSOs) on a continuous basis, the platforms aim to identify the best decisions in terms of activation, exchange, and imbalance netting to cover the expected imbalances in each area of the system. Such an optimisation allows the least expensive balancing bids to be used as far as allowed by transmission capacities.

Since power flows due to balancing are determined by the locations of imbalances and the activated reserves, electricity balancing is inseparably intertwined with congestion management.

The European power market has a zonal structure, where existing day-ahead and intraday markets have been coupled either through net transfer capacities or using the flow-based methodology [5]. Such mechanisms address congestion between – but not within – zones, and without knowledge of the reserve location, the impacts of activations on network congestion are unforeseeable. Consequently, there is a conflict between, on the one hand facilitating a zonal approach to balancing, that allows the integration of large areas and many countries, and on the other hand the necessary consideration of intra-zonal grid constraints. The proposed solution is bid filtering [3], a concept for which methods are yet to be formalised. However, the availability of the bids depends on the final flows, which are unknown. Thus, pre-determination of which bids are available and which not will either be too restrictive, which increases costs, or too liberal, which may threaten system security.

The methodology described in this paper contrasts these earlier proposals. With the hypothesis that the inefficiencies of bid filtering can be avoided while keeping a zonal market structure, this paper aims to demonstrate a new approach to handling intra-zonal congestion in the upcoming pan-European balancing markets. The key idea is the development of a distributed market clearing structure. Decomposing the interconnected power system into smaller network areas, complexity is greatly reduced and the computational effort can be parallelised, allowing for detailed network considerations within each area. Interaction with a central master problem is handled through optimality cuts, without information on individual bids or network constraints. The mathematical techniques used to decompose and solve the distributed problem are well-established in literature, but the European balancing activation optimisation comprises a new context and application of these methods.

The distributed formulation presented in this paper is subject to a few key assumptions. Firstly, the formulation is convex with linear constraints. Some of the European balancing markets being implemented [3, 4] will likely allow indivisible and linked bids, whereas at least one [2] will not. With integer decisions, the solution algorithm would likely perform worse in terms of convergence due to weaker optimality cuts being generated from MILP subproblems. Benchmarking computational performance and convergence rate analysis of different models is outside the scope of this article, but in general, linear programs can be solved quickly, also for systems of considerable size. Still, even if computationally tractable within the limited available operational time, a centralised nodal optimisation on the full system would challenge dispatch autonomy, which is a contentious issue in Europe, and require closer harmonisation between areas, and possibly also require a system-wide transition to nodal pricing. Including integer decisions increases computational effort significantly and inhibits a full nodal approach in real time. Moreover, the decomposed formulation represents a hybrid nodal/zonal approach. Each subsystem considers a simplified representation of the transmission grid in neighbouring areas, and the transport model in the zonal master problem is unable to account for loop flows in a meshed topology. This is a notable shortcoming compared to a full nodal model, nevertheless it is the concept being implemented in the upcoming balancing platforms.

Section 2 describes established design principles for the integrated European balancing markets, as well as earlier approaches to balancing optimisation and distributed market clearing. In Section 3, the details of the optimisation and solution procedure are described, including master and subproblem formulations, and calculation of cuts and bounds. The methodology is applied on a test system in Section 4, demonstrating the steps in the method and providing a numerical example. The implications and unsolved challenges indicating areas for further work are discussed in Section 5, leading to the conclusions in Section 6.

2 Background

In the balancing market, three main parties interact to maintain the balance between generation and consumption in the power system. Balance responsible parties (BRPs) failing to counteract imbalances within their perimeter, thereby not meeting their scheduled positions, impose imbalances on the system. The necessary adjustments to mitigate the resulting system imbalance are coordinated by the TSO through activation of balancing energy. This energy is delivered through bids by balancing service providers (BSPs) holding reserve capacity. [In European Network of Transmission System Operators for Electricity (ENTSO-E) terms, balancing activation is part of the frequency restoration and reserve replacement processes. Reserves activated for balancing energy are correspondingly frequency restoration reserve or replacement reserve products, comparable to secondary and tertiary control reserves.]. To cover imbalances, the TSO activates as many bids as needed from the merit order list of balancing energy bids. In an imbalance settlement mechanism, the BRPs causing imbalances by deviating from their schedules are penalised, and thereby the BRPs indirectly pay BSPs for the activated balancing energy [6].

Several studies have indicated the benefits of European balancing market integration [6–9], and key market design aspects have been set through the development and implementation of new network codes. These codes, developed by the ENTSO-E establish common rules and regulations for the European power markets. Specifically, the Guideline on Electricity Balancing [1] aims to enable and facilitate exchange of balancing energy. It requires balancing services to be harmonised into Standard Products for balancing energy, and outlines a market structure based on a common merit order list, consisting of balancing bids from all TSOs. An activation optimisation function optimises activation and imbalance netting actions to satisfy the imbalance needs of each TSO, subject to available cross-zonal transmission capacities.

The future European balancing market will have a zonal structure, largely following national borders. Three implementation projects [2–4] all currently develop common platforms for cross-border balancing in accordance with this high-level design, with cross-zonal capacities preventing congestion between market areas. The appetite for nodal approaches in Europe is limited, and in any case they would meet challenges due to problem size. Rather, in [1] handling intra-zonal constraints through *bid filtering* is proposed, i.e. TSOs may flag a bid unavailable and withhold it from the common merit order list if its activation would cause internal congestion. As the severity of intra-zonal congestion varies between different European countries, and a common standard bid filtering methodology is not expected.

Bid filtering in zonal balancing markets has not received widespread attention in scientific literature. The algorithm developed in [10] enumerates a range of different requests for balancing energy exchange from neighbouring zones. The method uses power flow analyses to detect whether a bid would need to be skipped to avoid congestion for each of the considered flow outcomes. The algorithm only considers one neighbouring zone at a time, whereas balancing market outcomes would often instruct balancing energy exchange on multiple zonal borders. Such flow combinations (including transit flows) load the network differently, and the method in [11] considers simultaneous balancing energy flow to multiple neighbours. An important finding is that bid filtering alone cannot provide any guarantee against intra-zonal congestion, and that additional measures will sometimes be necessary.

Closely related problems and techniques also been studied for US electricity markets, where traditional reserve requirements are based on deterministic reserve zones. When disregarding the grid location of reserves procured within a zone, there is a risk that the procured reserves are ineffective against intra-zonal congestion [12]. And just as in the zonal European balancing market, all reserves within a zone are assumed to have equal shift factors on critical lines, and the true deliverability of the procured reserves will be imprecise. To ensure adequate volumes and locations of operating reserves, Lyon *et al.* [13] demonstrates a locational reserve disqualification method taking into account a range of distinct contingency scenarios.

Regardless of its implementation, there are obvious drawbacks with the bid filtering approach. The impact of a bid activation on congestion depends on the exchange situation, which is unknown

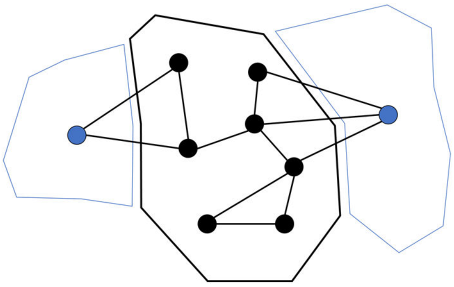

Fig. 1 *Example single-area system consisting of internal nodes and external nodes representing neighbouring areas*

before the outcome of the European activation optimisation. Under unforeseeable flow conditions, TSOs need to evaluate a range of possible exchange flow outcomes. In order to be robust against internal congestion, TSOs would need to withhold a bid even when its activation causes congestion only under specific exchange situations, potentially reducing both social welfare and security of supply. Making such a bid available, however, incurs the risk of congestion, and even when manageable through urgent redispatch, the price signals from the optimisation would still be incorrect. Without an adequate representation of the transmission grid, internal congestion can hardly be prevented efficiently.

Balancing activation optimisation models in literature often comprise linear programs covering a firm imbalance demand at minimum activation cost [8, 9]. Being a form of real-time dispatch, they are sometimes based on DC optimal power flow (DCOPF) formulations [14]. Some models extend the basic formulations into scheduling models, taking into account multiple time periods, which requires more detailed constraints regarding e.g. ramping and duration of delivery [15, 16], or even multiple balancing products or stochastic imbalances [17]. Other models [3, 4, 18] have introduced elasticity in the imbalance demand and resemble more traditional market clearing models. All such models share similar structures with either unit commitment or economic dispatch models, with network representations ranging from copper plate to nodal models.

Distributed algorithms have been developed decades ago to efficiently solve OPF (optimal power flow) problems on very large networks. The most common methodology in literature revolves around decomposing a large economic dispatch problem into smaller subproblems representing different areas in the network, creating duplicate *dummy variables* on the borders. The individual network areas are linked through coupling constraints in a central coordination problem. Often, (augmented) Lagrangian relaxation is applied to these coupling constraints, allowing for solving one smaller economic dispatch problem per area [19–22]. Rather than dualising all coupling constraints in the objective function and solving an auxiliary problem to update the Lagrangian multipliers, the authors in [23, 24] apply optimality condition decomposition on a multiarea OPF problem, a closely related but more sophisticated technique. Here, complicating constraints are kept, but in separate subproblems, resulting in automatic multiplier updates.

Benders decomposition [25] has been widely applied to solve problems with complicating variables. A master problem proposes candidate values for these variables, which are considered fixed parameters in the subproblems. Optimality (and sometimes also feasibility) cuts generated by subproblems are added as constraints to the master problem, iteratively refining the feasibility space until decision variables converge to their optimal values. In the context of power system optimisation, integer decisions such as investment or unit commitment are often considered the complicating variables, and subproblems typically comprise linear programs evaluating the minimum cost given a candidate set of decisions. Thus, the method has been widely used together in combination with OPF to solve two-stage problems, including (stochastic) unit commitment and expansion planning [23].

Even more relevant to this paper are efforts to solve large OPF problems using Benders decomposition. In [26], the total power output in each subnetwork is considered a complicating variable, and subproblems optimising smaller areas are computed in parallel. A decentralised OPF solution is found in [27] by duplicating exchange variables at the subnetwork boundaries and solving individual-area OPFs, while coordinating the values of these variables using a line search in a centralised problem. Another kind of primal decomposition is applied to a large-scale OPF in [28], abstracting neighbouring network areas into so-called marginal equivalents, solving area-dispatch subproblems, and coordinating information on binding constraints and free variables. The decomposition algorithm proposed in this paper is similar, but adopts the methodology to the context of the integrated balancing market, which shares a structure similar to the multi-area OPF problem.

3 Methodology

The objective of the balancing energy activation problem can be seen as finding the cost-minimising set of bid activations to cover the imbalances of all TSOs, while respecting the relevant constraints, including limitations on network flows. Keeping a high level of detail in the network representation, this procedure decomposes the problem into separate subproblems per balancing area. Each subproblem calculates the activation cost in its area taking into account the local imbalance and an assumed set of balancing energy exchange flows, comprising an *exchange scenario*. A master problem aims to minimise total balancing costs by finding the optimal exchange volumes between all pairs of neighbouring areas, subject to cross-zonal capacities. The master problem needs no information on bids or intra-zonal network constraints. Rather, solutions of the subproblems are used to generate optimality cuts in the master problem, providing an approximate supply function representing the true balancing activation in each area under feasible flow conditions. These cuts can be added iteratively, thereby refining the cost function representations close to the exchange volumes proposed by the master problem. They can also be pre-generated by solving the subproblems for a multitude of different exchange situations, thereby reducing the number of iterations or providing a near-optimal solution a single iteration.

Different geographical zones of the interconnected system are represented as single-area systems. For each bidding zone, the single-area system consists of a detailed model of the internal network, while disregarding the network structure inside neighbouring areas. This allows each neighbouring area to be represented by a single external node (as in [20]), and transmission lines between the areas are considered to be connected to this node, cf. Fig. 1.

3.1 Single-area subproblem

The subproblem (**SP**) (1)–(5) comprises a dispatch model, and largely follows a DCOPF structure. The set $\mathscr{I}_a$ contains all internal and external nodes associated with single-area system a. The subproblem assumes knowledge of the need for balancing energy D_i in each node $i \in \mathscr{I}_a$. For external nodes, this value denotes a specific cross-zonal balancing energy exchange program. The corresponding set of lines comprises $\mathscr{L}_a$, and an adjacency matrix contains parameters A_{il}, equalling 1 or -1 if line l is directed out of or into node i, respectively, and zero otherwise. The power transfer distribution factors (PTDFs) Φ_{il} describe the linearised relationship between a power injection in node i on the flow in line

l. Furthermore, each line $l \in \mathscr{L}_a$ has a lower and upper (bidirectional) flow capacity limit, $\underline{F}_l, \overline{F}_l$, given as the residual between the rated capacity and the initial flow on the line. All bids $b \in \mathscr{B}_a$ have an associated capacity limit $\overline{Y}_b$ and also an activation cost C_b, given by the bid prices for upward bids $\mathscr{B}^\uparrow$ and for downward bids $\mathscr{B}^\downarrow$. The set of bids located at node i is denoted by $\mathscr{B}_i^{\uparrow/\downarrow}$.

The objective in (1) is to find optimal activation volumes y_b to minimise the activation cost F_a in area a. The equality constraint in (2) ensures energy balance and flow conservation in each node of the system, while (3) sets the balancing energy flow f_l on each line based on the nodal net injections and the node-to-line PTDFs. These flow restrictions could easily be extended to include combinations of lines or other network elements. Bid capacities and available line capacities are given in (4) and (5). Note that selecting bids through cost minimisation does not inhibit compensating activated bids with a marginal (clearing) price, but it does imply that supply costs – and not only TSO payments – are minimised

$$(\mathbf{SP}) \quad \min_{f,y} F_a(\boldsymbol{D}) = \sum_{b \in \mathscr{B}_a} C_b y_b \tag{1}$$

$$\text{s.t.} \quad \sum_{b \in \mathscr{B}_i^\uparrow} y_b - \sum_{b \in \mathscr{B}_i^\downarrow} y_b - \sum_{l \in \mathscr{L}_a} A_{il} f_l = D_i, \quad i \in \mathscr{I}_a \tag{2}$$

$$f_l - \sum_{i \in \mathscr{I}_a} \Phi_{il} \left(\sum_{b \in \mathscr{B}_i^\uparrow} y_b - \sum_{b \in \mathscr{B}_i^\downarrow} y_b - D_i \right) = 0, \quad l \in \mathscr{L}_a \tag{3}$$

$$0 \le y_b \le \overline{Y}_b, \quad b \in \mathscr{B}_a \tag{4}$$

$$\underline{F}_l \le f_l \le \overline{F}_l, \quad l \in \mathscr{L}_a \tag{5}$$

3.2 Generating cuts

For the problem of identifying optimal exchange volumes, the aim is to build an approximation of balancing costs for each area with respect to exchange volumes to neighbouring areas. Denoting by y^* an optimal balancing dispatch found in (**SP**) for balancing exchange volumes D^0, and by λ_i the shadow prices of the energy balance constraints in (2), a subgradient to the cost function F_a in terms of exchange volumes D_i to neighbouring areas represented by nodes $\mathscr{I}_a^{\text{ext}}$ is given as

$$F_a \ge \sum_{b \in \mathscr{B}_a} C_b y_b^* + \sum_{i \in \mathscr{I}_a^{\text{ext}}} \lambda_i (D_i - D_i^0) \tag{6}$$

Since (**SP**) is convex, an outer approximation of F_a with respect to changes in imbalance needs D_i can be made based on the subgradient. By setting the parameter $\lambda_{an}^k = \lambda_i$ for the relevant corresponding exchange nodes, a supporting hyperplane k to the estimated balancing cost θ_a in the master problem can be expressed as a linear constraint in terms of the exchange variables x_{an}

$$\theta_a \ge \sigma_a^k + \sum_{n \in \mathscr{N}_a} \lambda_{an}^k x_{an}, \tag{7}$$

where

$$\sigma_a^k = \sum_{b \in \mathscr{B}_a} C_b y_b^* - \sum_{i \in \mathscr{I}_a^{\text{ext}}} \lambda_i D_i^0 \tag{8}$$

3.3 Multi-area master problem

The master problem (**MP**) searches for optimal exchange values x_{an} to minimise the sum of all estimated balancing activation costs θ_a in (9). To satisfy its energy balance (10), each area a can cover its demand D_a for balancing energy through import or through local activation z_a. For each area a and its neighbours n, limits on exchange are imposed in (11), and (12) couples exchange decisions in different directions. Finally, (13) applies the supporting hyperplanes (7) calculated from solutions of the subproblems, providing lower bounds on the balancing activation costs in each area

$$(\mathbf{MP}) \quad \min_{\theta, x, z} \sum_{a \in A} \theta_a \tag{9}$$

$$\text{s.t.} \quad z_a - \sum_{n \in \mathscr{N}_a} x_{an} = D_a, \quad a \in \mathscr{A} \tag{10}$$

$$x_{an} \le \overline{X}_{an}, \quad a \in \mathscr{A}, n \in \mathscr{N}_a \tag{11}$$

$$x_{an} + x_{na} = 0, \quad a \in \mathscr{A}, n \in \mathscr{N}_a \tag{12}$$

$$\sigma_a^k + \sum_{n \in \mathscr{N}_a} \lambda_{an}^k x_{an} \le \theta_a, \quad a \in \mathscr{A}, k \in \mathscr{K}_a \tag{13}$$

3.4 Iterative solution procedure

The problem can be solved by iteratively solving the master problem and each of the subproblems. The cost-minimising exchange volumes and estimated balancing costs from (**MP**) will be passed on to the subproblems as parameters, which evaluate the true balancing costs for these volumes, and if necessary add new cuts to refine the cost function representations in the master problem. Thus, the iterative solution procedure adds new balancing cost information in each iteration, until the cost estimate is proved to be valid. This is elegant and efficient, since the complexity of a sufficient balancing cost approximation will be limited.

Step 1: In the first iteration, initialise exchange volumes, $x_{an}^* = 0$, for $a \in \mathscr{A}, n \in \mathscr{N}_a$. Let $\theta_a^* = -\infty$ for $a \in \mathscr{A}$. Create empty sets K_a for $a \in \mathscr{A}$. In later iterations, solve (**MP**). Retrieve proposed exchange x_{an}^* and estimated balancing costs θ_a^*.

Step 2: For each area a: Set demand D_i in exchange nodes $i \in \mathscr{I}_a^{\text{ext}}$ equal to proposed exchange x_{an}^*. Solve (**SP**), let optimal solution be y^* with objective function value F_a^*. If $F_a^* > \theta_a^*$, add a new cut k to $\mathscr{K}_a$ as in (7).

Step 3: If $\sum_a F_a^* \le \sum_a \theta_a^* + \epsilon$, or if no cuts were added, the exchange values x_{an}^* are optimal. Else, go to step 1.

The iterative solution procedure is able to prove optimality with a finite number of cuts. At the same time, communication is required between the coordinating platform and the distributed entities providing new cuts in each iteration. Not only is the iterative method vulnerable to communication delays, also the number of iterations required to converge may be highly uncertain, as it depends on the problem instance (and the choice of tolerance gap ϵ). Finally, the time necessary to solve the single-area subproblems may differ greatly between TSOs and with different input values. An iterative procedure could be vulnerable to one or more slow-solving subproblems, although this can partly be guarded against by making the algorithm asynchronous, as in [29].

3.5 Two-step solution procedure

Another approach would be to pre-generate a set of cuts representing the balancing cost function in each area. By pre-generating cuts and submitting them all at once, the communication is simplified, and the computational workload to solve the subproblems is moved ahead of the platform optimisation. A straightforward cut pre-generation strategy could be to evaluate the F_a for a structured (regular or rectilinear) grid of different sets of values for $x_{an}, n \in \mathscr{N}_a$, and add a new cut whenever F_a exceeds all lower bounds given by previously generated cuts. Using a structured grid also enables calculating

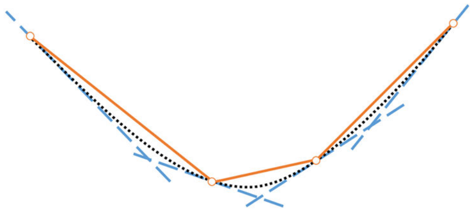

Fig. 2 *Two-dimensional example of lower bounds on F_a (blue), upper bounds on F_a (orange), and the unknown true cost function F_a (black)*

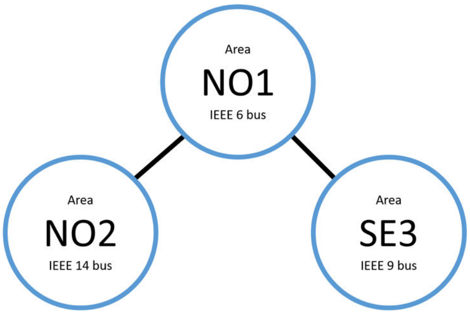

Fig. 3 *Test system topology and network models*

Table 1 Test system data per area

Area	Nodes	Gen. units	Trans. lines	Local imbalance
NO1	6	3	11	38 MW
NO2	14	5	20	120 MW
SE3	9	3	9	189 MW

Table 2 Test system data per tie-line

Tie-line	External node	Exchange cap.
NO1-NO2	Node 4	10 MW
NO2-NO1	Node 14	150 MW
NO1-SE3	Node 6	50 MW
SE3-NO1	Node 9	150 MW

upper bounds on F_a using multivariate interpolation between the evaluated points. Solving (**MP**) with a set of pre-generated cuts will yield proposed exchange values x^*_{an}. However, with incomplete outer approximations of the true cost functions, there is a possibility that the corresponding estimated balancing costs θ_a will be an underestimation. Pre-generating a large amount of cuts for each subsystem will reduce this risk, but without further iterations in the solution procedure, the cost function representation will not be refined, and a global optimum will not be guaranteed.

3.6 Upper bounds

For the system as a whole, the objective function value of (**MP**) in each iteration serves as a lower bound on the true balancing costs. Correspondingly, the sum of all *evaluated* balancing costs $\sum_a F_a$ for any given set of exchange flows $x_{an}, n \in \mathcal{N}_a$ represents the cost of a feasible solution and is thus a valid upper bound for the total balancing costs.

Fig. 2 demonstrates the calculation of upper and lower bounds for the cost function F_a of a single area. For each single-area subproblem, lower bounds on F_a are formed by the so-far generated optimality cuts. Moreover, each cut coincides with the true cost function F_a for the specific exchange flows $x_{an}, n \in \mathcal{N}_a$ for which it is calculated. If denoting by $p \in \mathcal{P}_a$ the different exchange scenarios for which (**SP**) has been evaluated with costs F^p_a and exchange volumes X^p_{an}, upper bounds on F_a can be observed as the facets of the lower convex hull containing points $\mathcal{P}_a$. The bound value $F^{\text{UB}}_a(\boldsymbol{x})$ along the facets can be found as the minimum cost convex combination (14)–(17) of the evaluated scenarios p. For a specific scenario with exchange volumes $x_{an}, n \in \mathcal{N}_a$

$$(\textbf{UB}) \quad \min_{\alpha} F^{\text{UB}}_a = \sum_{p \in \mathcal{P}_a} \alpha^p F^p_a \tag{14}$$

$$\sum_{p \in \mathcal{P}_a} \alpha^p x^p_{an} = x_{an}, \quad n \in \mathcal{N}_a \tag{15}$$

$$\sum_{p \in \mathcal{P}_a} \alpha^p = 1 \tag{16}$$

$$0 \le \alpha^p \le 1, \quad p \in \mathcal{P}_a \tag{17}$$

4 Numerical example

4.1 System description

To improve understanding of the proposed methodology, this section demonstrates the distributed optimisation on a very simplified scale. Three different power system models represent subsystems (corresponding to balancing areas) in a larger, interconnected test system, as shown in Fig. 3. The test system areas carry names NO1, NO2, and SE3, and are indexed as areas 1, 2, and 3.

Area NO1 is represented by a 6-bus system from [30], NO2 by the IEEE 14-bus system, and SE3 by the IEEE 9-bus system. Key information on the subsystem models is summarised in Tables 1 and 2. The models are inhomogeneous, yet this poses no issues under the simplifying assumptions of network decomposition and flow in neighbouring areas. This inhomogeneity also illustrates a flexibility advantage in that TSOs are not required to align the level of detail in the models used for evaluating balancing costs.

Exchange between areas is limited by remaining cross-zonal capacities, which are directional and asymmetric due to utilisation in other markets. Moreover, each area needs to cover a local imbalance, either through exchange or local activation.

4.2 Implementation

The iterative solution procedure has been implemented as a Python library, in particular leveraging modelling functionality from the Pyomo framework [31]. Subproblems are managed and solved with functionality from PyPower, which is based on the Matpower toolbox [32].

4.3 Solution

The iterative solution procedure is used in this example. In each iteration, the exchange volumes proposed by the master problem are added as additional demand at the nodes in each subsystem corresponding to the exchange corridor. The one-area system optimises its balancing actions (using a modified DCOPF calculation in this example) according to the exchange situation, taking into account intra-zonal grid constraints, and passes balancing cost information back to the master problem in the form of a cut.

Iteration 1

Step 1: Initialise all exchange volumes x^*_{an} to 0, and all cost estimates θ^*_a to $-\infty$.

Step 2:

Area 1: Set $D_4 = x_{1,2} = 0$ and $D_6 = x_{1,3} = 0$. Solving (**SP**) gives $F^*_1 = 7246$, and since $F^*_1 > \theta^*_1$, a cut is added

$$\theta_1 \ge 7246 + 31.90x_{1,2} + 31.90x_{1,3}$$

In this case, both marginal exchange costs $\lambda_{1,2}$ and $\lambda_{1,3}$ from area 1 take the value 31.90.

Area 2: Set $D_{14} = x_{2,1} = 0$. Solving (**SP**) gives $F^*_2 = 7643$, and since $F^*_2 > \theta^*_2$, a cut is added

$$\theta_2 \ge 7643 + 39.01x_{2,1}$$

Area 3: Set $D_9 = x_{3,1} = 0$. Solving (**SP**) gives $F_3^* = 5216$, and since $F_3^* > \theta_3^*$, a cut is added

$$\theta_3 \geq 5216 + 24.04x_{3,1}$$

The evaluated balancing costs $\sum_{a \in \mathscr{A}} F_a^* = 20105$ provide an upper bound on the total balancing cost.

Step 3: We have $\sum_a F_a^* > \sum_a \theta_a^* + \epsilon$, and at least one new cut was added, hence the exchange values x_{an}^* were not optimal.

Iteration 2

Step 1: Solving (**MP**) with the updated list of cuts yields $\theta^* = (2780, 7252, 8823)$ for $x_{1,2}^* = 10$ and $x_{3,1}^* = 150$. New lower bound: $\sum_{a \in \mathscr{A}} \theta_a^* = 18856$.

Step 2: Subproblem evaluation yields $F_1^* = 4802$, $F_2^* = 7256$, $F_3^* = 9598$. New cuts

$$\theta_1 \geq 4802 + 0.0x_{1,2} + 0.0x_{1,3}$$

$$\theta_2 \geq 7639 + 38.28x_{2,1}$$

$$\theta_3 \geq 4441 + 34.38x_{3,1}$$

New upper bound: $\sum_{a \in \mathscr{A}} F_a^* = 21656$.

Step 3: We have $\sum_a F_a^* > \sum_a \theta_a^* + \epsilon$, and at least one new cut was added, hence the exchange values x_{an}^* were not optimal.

Iteration 3

Step 1: Solving (**MP**) with the updated list of cuts yields $\theta^* = (5173, 7256, 7019)$ for $x_{1,2}^* = 10$ and $x_{3,1}^* = 75$. New lower bound: $\sum_{a \in \mathscr{A}} \theta_a^* = 19448$.

Step 2: Subproblem evaluation yields $F_1^* = 5191$, $F_2^* = 7256$, $F_3^* = 7213$. New cuts

$$\theta_1 \geq 7220 + 31.22x_{1,2} + 31.22x_{1,3}$$

$$\theta_3 \geq 5022 + 29.21x_{3,1}$$

New upper bound: $\sum_{a \in \mathscr{A}} F_a^* = 19660$.

Step 3: We have $\sum_a F_a^* > \sum_a \theta_a^* + \epsilon$, and at least one new cut was added, hence the exchange values x_{an}^* were not optimal.

Iteration 4

Step 1: Solving (**MP**) with the updated list of cuts yields $\theta^* = (4802, 7256, 7577)$ for $x_{1,2}^* = 10$ and $x_{3,1}^* = 87.5$. New lower bound: $\sum_{a \in \mathscr{A}} \theta_a^* = 19\,635$.

Step 2: Subproblem evaluation yields $F_1^* = 4804$, $F_2^* = 7256$, $F_3^* = 7582$. New cuts

$$\theta_1 \geq 7204 + 31.00x_{1,2} + 31.00x_{1,3}$$

$$\theta_3 \geq 4952 + 30.07x_{3,1}$$

New upper bound: $\sum_{a \in \mathscr{A}} F_a^* = 19642$.

Step 3: We have $\sum_a F_a^* > \sum_a \theta_a^* + \epsilon$, and at least one new cut was added, hence the exchange values x_{an}^* were not optimal.

Iteration 5

Step 1: Solving (**MP**) with the updated list of cuts yields $\theta^* = (4802, 7256, 7584)$ for $x_{1,2}^* = 10$ and $x_{3,1}^* = 87.5$. New lower bound: $\sum_{a \in \mathscr{A}} \theta_a^* = 19\,642$. No new cuts are added in this iteration, and the solution is optimal.

Solution summary: In this example, the first iteration is calculated with zero exchange, thereby providing the balancing costs and marginal prices from using only local balancing. For the global optimum, the procedure needs four full iterations (cf. Figs. 4 and 5) to find optimal exchange values. With a higher tolerance ϵ, three iterations would suffice. Marginal balancing costs on each side of the non-congested link between NO1 and SE3 have almost converged to around 30–31 €/MWh, whereas marginal costs are around 38 €/MWh on the NO2 side of the NO1–NO2 link, thus it is used at full capacity. Also, for illustration, the cuts generated for area SE3 are included in Fig. 6. Cuts generated for area NO1 are shown in Fig. 7, requiring a 3D projection as the area has two exchange dimensions.

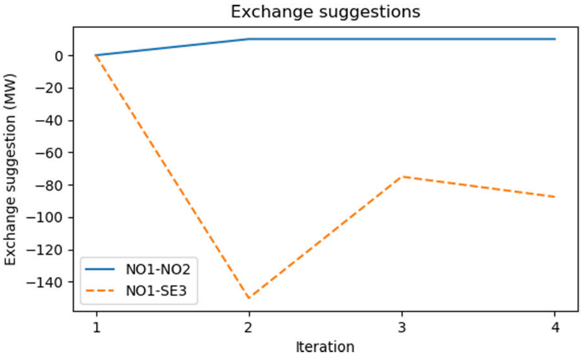

Fig. 4 *Exchange volumes x_{an}^* suggested by (**MP**) in each iteration*

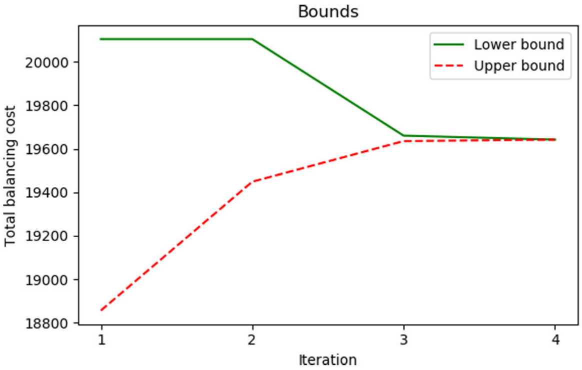

Fig. 5 *Upper and lower bounds on total balancing activation cost*

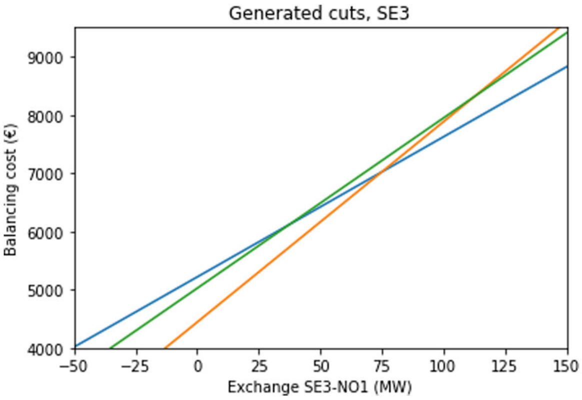

Fig. 6 *Cuts representing balancing costs in SE3 for different exchange volumes*

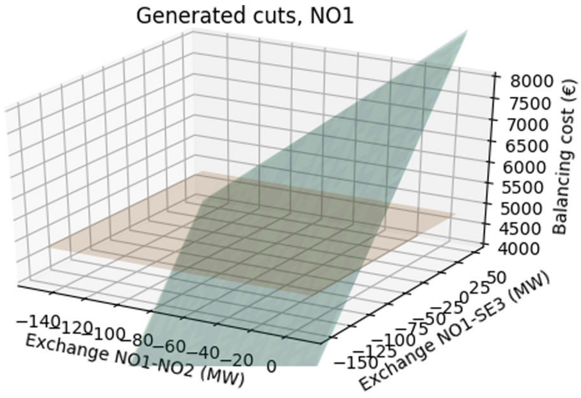

Fig. 7 *Cuts representing balancing costs in NO1 for different exchange volumes*

5 Discussion

The proposed distributed approach serves as an alternative method to a zonal optimisation with bid filtering, with its main advantage being the efficient handling of intra-zonal constraints. Regardless of implementation, pre-filtering balancing bids reduces social welfare. TSOs will in some cases detect balancing energy bids that can be used safely only under specific flow conditions. As exchange flows are unknown, the TSO can decide to make such bids unavailable to avoid internal congestion. When the safe flow conditions occur, these bids are underutilised. If such bids are made available and the safe flow conditions do not occur, the solution to the optimisation problem will be infeasible, and the exchange flows will generally be suboptimal. By distinguishing between an exchange problem and multiple dispatch problems, all resources can be considered available without taking the risk of infeasible dispatch solutions.

The distributed method carries fundamental implications. A distinction between exchange and dispatch gives each TSO an active role in the optimisation. In the distributed balancing optimisation algorithm, the platform's ability to identify optimal exchange volumes depends not on available balancing bids, but rather on TSOs' evaluations of their total and marginal balancing costs under different exchange scenarios. This shifts computational workload and also responsibility to the TSOs.

A distributed, iterative solution procedure is elegant, but also poses potential challenges. The balancing cost cuts provide an approximate representation of the true balancing costs in each area given different exchange volumes. The iterative procedure gradually refines these representations by adding more cuts, until the balancing cost estimation is representative. Moreover, the procedure will produce a feasible solution (as well as a lower bound) in each iteration, since the TSOs' response to a given exchange scenario is evaluated. Although the number of iterations required to reach a global optimum is limited, it is nevertheless unknown. Even if subproblems are evaluated promptly and efficiently, the exchange optimisation algorithm would still be vulnerable to communication delays or numeric instabilities causing slow convergence.

Pre-generating cuts provides a head start, but generally not an optimal solution. By solving the single-area subproblem for a set of different exchange scenarios and submitting the resulting cuts to the exchange optimisation, the balancing cost representation can be pre-refined. Solving the exchange problem using pre-refined cost representations can lead directly to a solution close to the true optimum. The direct solution will generally be an underestimation of the true balancing costs, and unless more cuts can be added, the exchange volumes will not be guaranteed to be optimal. The underestimation is bounded, however. An upper bound on the true balancing costs for a given exchange situation can be found as the minimum cost convex combination of already evaluated exchange scenarios.

Analysing computational performance is not the focus of the case study, however such insight could be obtained by running the algorithm on networks of realistic size. This would also allow direct comparison between the iterative solution and the cut pregeneration procedures, and possibly also against other methods, including bid filtering. Moreover, some markets allow complex bid structures, such as linked or indivisible bids, which introduce integer variables in the subproblems. Investigating methods for efficiently incorporating non-convex subproblems in the distributed balancing optimisation is of therefore also of interest in future work.

6 Conclusions

An integrated European balancing market needs to efficiently identify the optimal balancing actions in terms of activation and cross-border exchange of balancing energy. The European legislation strongly favours a zonal network representation, and there is political opposition against moving to a nodal system. Yet the aggregated network representation fails to effectively manage congestion within market areas, leading to inefficiencies and infeasible network flows. The methodology presented in this article allows for more detailed network representations to be used efficiently in the balancing energy optimisation without abandoning the overarching zonal market structure. Decomposing the system into smaller network areas, the computational effort to identify the optimal bid activation volumes can be parallelised by distributing it across several local subproblems. A master problem searches for optimal cross-zonal balancing exchange flows to minimise total balancing costs, doing so without information on bids or intra-zonal network constraints, but rather using information from the subproblem solutions using a cutting-plane method. The decomposition also allows flexibly integrating areas with network models of different scales and detail levels. The numeric example with a small number of cuts and iterations demonstrates the elegant iterative procedure, although near-optimal solutions can also be identified directly by pre-generating cuts. Open questions remain regarding convergence properties and approximation accuracy of the cutting-plane approach proposed in this paper, but the distributed solution structure gives TSOs a more active role, and handles intra-zonal network constraints more efficiently than alternative methods.

7 References

[1] ENTSO.E: 'Guideline on Electricity Balancing (May 2017 version)', 2017

[2] 50 Hertz, Amprion, APG, Elia, RTE, TenneT et al.: 'Consultation on the design of the platform for automatic Frequency Restoration Reserve (aFRR) of PICASSO region', 2017

[3] MARI: 'First Consultation – Call For Input', 2017

[4] ADMIE, National Grid, REE, REN, RTE, Terna et al.: 'Public consultation document for the design of the TERRE (Trans European Replacement Reserves Exchange) Project solution', 2016

[5] Gebrekiros, Y.: 'Analysis of Integrated Balancing Markets in Northern Europe under Different Market Design Options', Norwegian University of Science and Technology, 2015

[6] van der Veen, R.A.C.: 'Designing Multinational Electricity Balancing Markets', Technische Universiteit Delft, 2012

[7] Vandezande, L.: 'Design and Integration of Balancing Markets in Europe', Katholieke Universiteit Leuven, 2011

[8] Farahmand, H., Aigner, T., Doorman, G., *et al.*: 'Balancing market integration in the Northern European continent: a 2030 case study', *IEEE Trans. Sustain. Energy*, 2012, **3**, (4), pp. 918–930

[9] Jaehnert, S., Doorman, G.: 'Assessing the benefits of regulating power market integration in Northern Europe', *Int. J. Electr. Power Energy Syst.*, 2012, **43**, (1), pp. 70–79

[10] Guntermann, C., Gunderson, N.W., Lindeberg, E., *et al.*: 'Detecting unavailable balancing energy bids due to risk of internal congestions'. In: 18th Int. Conf. Environ. Electr. Eng., Palermo, Italy, 2018

[11] Håberg, M., Bood, H., Doorman, G.: 'Preventing internal congestion in an integrated european balancing activation optimization', *Energies*, 2019, **12**, (3), p. 490

[12] Lyon, J.D., Hedman, K.W., Zhang, M.: 'Reserve requirements to efficiently manage intra-zonal congestion', *IEEE Trans. Power Syst.*, 2014, **29**, (1), pp. 251–258

[13] Lyon, J.D., Zhang, M., Hedman, K.W.: 'Locational reserve disqualification for distinct scenarios', *IEEE Trans. Power Syst.*, 2015, **30**, (1), pp. 357–364

[14] Gebrekiros, Y., Doorman, G., Jaehnert, S., *et al.*: 'Balancing energy market integration considering grid constraints'. 2015 IEEE Eindhoven PowerTech, Eindhoven, Netherlands, 2015, pp. 1–6

[15] Abbaspourtorbati, F., Scherer, M., Ulbig, A., *et al.*: 'Towards an optimal activation pattern of tertiary control reserves in the power system of Switzerland'. 2012 Am. Control Conf., Montréal, Canada, 2012, pp. 3629–3636

[16] Malík, O., Havel, P.: 'Decision support tool for optimal dispatch of tertiary control reserves', *Int. J. Electr. Power Energy Syst.*, 2012, **42**, (1), pp. 341–349

[17] Håberg, M., Doorman, G.: 'A stochastic mixed integer linear programming formulation for the balancing energy activation problem under uncertainty'. 2017 IEEE Manchester PowerTech, Manchester, UK, 2017

[18] Håberg, M., Doorman, G.: 'Proactive planning and activation of manual reserves in sequentially cleared balancing markets'. 2017 IEEE Electr. Power Energy Conf., Saskatoon, Canada, 2017

[19] Batut, J., Renaud, A.: 'Daily generation scheduling optimization with transmission constraints: a new class of algorithms', *IEEE Trans. Power Syst.*, 1992, **7**, (3), pp. 982–989

[20] Kim, B.H., Baldick, R.: 'Coarse-grained distributed optimal power flow', *IEEE Trans. Power Syst.*, 1997, **12**, (2), pp. 932–939

[21] Gatsis, N., Giannakis, G.B.: 'Decomposition algorithms for market clearing with large-scale demand response', *IEEE Trans. Smart Grid*, 2013, **4**, (4), pp. 1976–1987

[22] Zhang, Y., Gatsis, N., Giannakis, G.B.: 'Disaggregated bundle methods for distributed market clearing in power networks'. 2013 IEEE Glob. Conf. Signal Inf. Process., Austin, TX, USA, 2013, pp. 835–838

[23] Conejo, A.J., Castillo, E., Minguez, R., *et al.*: '*Decomposition techniques in mathematical programming*' (Springer-Verlag, Berlin/Heidelberg, 2006)

[24] Ahmadi Khatir, A., Conejo, A.J., Cherkaoui, R.: ‘Multi-area energy and reserve dispatch under wind uncertainty and equipment failures’, *IEEE Trans. Power Syst.*, 2013, **28**, (4), pp. 4373–4383

[25] Benders, J.F.: ‘Partitioning procedures for solving mixed-variables programming problems’, *Numer. Math.*, 1962, **4**, (1), pp. 238–252

[26] Yuan, Z., Hesamzadeh, M.R.: ‘A modified benders decomposition algorithm to solve second-order cone AC optimal power flow’, *IEEE Trans. Smart Grid*, 2017, **3053**, (c), pp. 1–11

[27] Min, L., Abur, A.: ‘A decomposition method for multi-area OPF problem’. 2006 IEEE PES Power Syst Conf Expo PSCE 2006 - Proc, Atlanta, GA, USA, 2006, pp. 1689–1693

[28] Zhao, F., Litvinov, E., Zheng, T.: ‘A marginal equivalent decomposition method and its application to multi-area optimal power flow problems’, *IEEE Trans. Power Syst.*, 2014, **29**, (1), pp. 53–61

[29] Aravena, I., Papavasiliou, A.: ‘A distributed asynchronous algorithm for the two-stage stochastic unit commitment problem’. 2015 IEEE Power Energy Soc. Gen. Meet., Denver, CO, USA, 2015, pp. 1–5

[30] Wood, A.J., Wollenberg, B.F., Sheblé, G.B.: ‘*Power generation, operation, and control*’ (John Wiley & Sons, USA, 2014, 3rd edn.)

[31] Hart, W.E., Laird, C.D., Watson, J.P.*, et al.*: ‘Pyomo — optimization modeling in python’, in ‘*Vol. 67 of springer optimization and its applications*’ (Springer International Publishing, Cham, 2017, 2nd edn.), pp. 1–277

[32] Zimmerman, R.D., Murillo Sanchez, C.E., Thomas, R.J.: ‘MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education’, *IEEE Trans. Power Syst.*, 2011, **26**, (1), pp. 12–19