Stronger cuts for Benders' decomposition for stochastic Unit Commitment Problems based on interval variables

Mathis Azéma[1*†], Vincent Leclère[1†] and Wim van Ackooij[2†]

[1*]CERMICS, ENPC, Institut Polytechnique de Paris, 6 et 8 avenue Blaise Pascal, Champs sur Marne, 77455, France.
[2]EDF Lab Paris-Saclay, 7 Boulevard Gaspard Monge, Palaiseau, 91120, France.

*Corresponding author(s). E-mail(s): mathis.azema@enpc.fr;
Contributing authors: vincent.leclere@enpc.fr; wim.van-ackooij@edf.fr;
[†]These authors contributed equally to this work.

Abstract

The Stochastic Unit Commitment (SUC) problem models the scheduling of power generation units under uncertainty, typically using a two-stage stochastic program with integer first-stage and continuous second-stage variables. We propose a new Benders decomposition approach that leverages an extended formulation based on interval variables, enabling decomposition by both unit and time interval under mild technical assumptions. This formulation leads to provably stronger Benders cuts than those derived from the standard 3-bin representation.

To improve computational efficiency, we introduce a heuristic based on weak duality to rapidly approximate the cut coefficients for units subject to ramping and capacity constraints. This allows us to retain cut strength while significantly reducing per-iteration cost.

Extensive computational experiments on large-scale risk neutral and risk averse instances — including the IEEE 118-bus system and SMS++ benchmarks — demonstrate that our method consistently outperforms existing approaches. On the most challenging test cases, it is up to five times faster than the best alternative and solves instances where no baseline closes the optimality gap within one hour.

Keywords: Unit Commitment, Stochastic Optimization, Benders' decomposition

1 Introduction

One of the main challenges for energy producers and network operators is the *Unit Commitment* (UC) problem [1, 2], which involves determining the optimal production schedule for a set of electricity generation units (thermal power plants, hydroelectric units, ...) while considering both technical and economic constraints. In this paper, we focus on technical constraints that satisfy a specific decomposition assumption, which states that the technical constraints between two non-overlapping intervals are independent. In particular, this assumption (detailed in Assumption 3, Sect. 3.3) is met by most common operational constraints, including initial conditions, capacity limits, and ramping constraints. In the Stochastic Unit Commitment (SUC) framework, net demand is considered random, primarily due to the increasing difficulty of forecasting net demand resulting from the growing integration of renewable energy sources, such as wind and solar power. To manage this uncertainty, we adopt a standard two-stage framework, where the first stage involves integer variables representing the commitment status of units, while the second stage involves only continuous (dispatch) variables. Building on this, we present a novel Benders' Decomposition (BD) approach to solve the two-stage SUC problem with uncertainty in the right-hand side of the constraints and binary variables representing the commitment status of units in the first stage. In particular, our method differs mainly from existing approaches in the literature through the modeling of first stage decisions. Indeed, the most widely used formulation for solving SUC is the *3-bin formulation*, which introduces three sets of binary decision variables to represent whether a unit is on, starts-up, or shuts-down. In recent years, several solution techniques have been developed using the *3-bin formulation*, which can be categorized into three main approaches. We refer to [3] for an exhaustive literature review on SUC.

The first approach, known as the extensive formulation, directly tackles the SUC problem by reformulating it as a large-scale deterministic UC problem and solving it using commercial optimization solvers. However, this method does not scale well, and generally fails to solve large-scale stochastic UCs with more than 10 scenarios within an hour. For example, [4] examined an SUC model incorporating gas emissions constraints and applied the extensive formulation to a modified IEEE 118-bus system with 25 thermal units over a 24-hour planning horizon. Their findings indicate that for 30 scenarios, the problem can be solved in approximately 1.7 hours.

To overcome the scalability limitations of the extensive formulation, researchers have investigated decomposition techniques such as Lagrangian Relaxation (LR) and Benders' Decomposition (BD) or a combination of both (see *e.g.,* [5]). In LR approaches, the specific constraints chosen for relaxation play a critical role. Typically, either non-anticipativity constraints or demand satisfaction constraints are relaxed, allowing the problem to be decomposed across scenarios or generation units, respectively. For instance, [6] proposed relaxing the non-anticipativity constraints, which enabled the decomposition of the SUC into independent UC subproblems, one per scenario, while iteratively updating dual multipliers. Their experiments on an instance with 130 thermal units and a 24-hour horizon suggest that solving instances with more than 10 scenarios can take several hours. A more detailed investigation of LR methods is presented by [7], who analyzed different relaxation strategies. Their computational results, based on instances with around 10 thermal and 10 hydro units over 24 time steps and 9 scenarios, indicate that solving the SUC with LR methods can require over an hour.

Despite their potential for parallelization, LR methods suffer from a critical drawback: they may yield infeasible solutions due to the relaxation of key constraints. It is, in fact, well known that one can at best solve an appropriately convexified version of a problem through Lagrangian relaxation [8]. This is a result of the fundamental fact that the Lagrangian dual is invariant to a convexification of the primal, e.g., [9]. However, "Lagrangian Heuristics" can allow one to retrieve excellent solutions close to optimality by exploiting the information generated while solving the Lagrangian dual, e.g., [10, 11]. In contrast, BD methods, also known as L-shaped methods [12], offer theoretical guarantees of convergence, typically, in a finite number of iterations. Yet, practical implementations using the standard *3-bin formulation* often suffer from slow convergence, impacting overall efficiency. For example, [13] applied BD to the SUC problem on a system with 39 buses, 10 generating units, and a 24-hour horizon, showing that the problem could be solved in approximately 10 minutes for 4 scenarios. Similarly, [14] developed a BD method incorporating the average value-at-risk in the second stage objective. Their method successfully solved a modified IEEE 118-bus system with 54 thermal generators, 24 time steps, and 100 scenarios in about one hour. Since BD is essentially just Kelley's cutting planes approach, it can be enhanced using similar methodologies. These can come in the form of selecting better or improving cuts, e.g., [15], stabilizing the computations by incorporating ideas from bundle methods, e.g., [16], or using inexact computations for the subproblems, e.g., [17, 18]. We refer to [19] for a recent review of Bender's decomposition.

Recently, some studies have focused on new formulations using interval binary variables to indicate whether a unit is on during a given interval $[a : b]$ within the time horizon. Referred to as the *extended formulation*, these formulations have thus generally more binary variables than the *3-bin formulation* but can produce tighter cuts. The *extended formulation* was first introduced by [20], who proved that it enables the derivation of valid cuts linking the 3-bin variables with production variables in a deterministic framework. Moreover, [21] proposed an alternative formulation that strikes a balance between the *3-bin formulation* and the *extended formulation*, offering new perspectives for enhancing computational efficiency. In particular, they showed that restricting attention to intervals $[a : b]$ of limited length can significantly reduce the number of variables, while preserving the tightness of the *extended formulation* and outperforming the *3-bin formulation* in terms of relaxation strength. Furthermore, the *extended formulation* has proven helpful in addressing the SUC problem involving quick-start generators, where startup and shutdown decisions are deferred to the second stage. In particular, [22] showed that the *extended formulation* can be used to determine the convex hull of a single quick-start generator in the second stage. It is worth noting that earlier work, such as [23], also tackled this challenge by adapting BD with the *3-bin formulation*.

In this paper, we study how to incorporate interval variables into the master problem of BD to solve the SUC problem, with the following key contributions:

1. We propose a novel Benders decomposition based on interval-based *extended formulations*, which yields provably stronger cuts than the classical 3-bin approach. Under mild technical assumptions, this enables decomposition both by unit and by time interval.
2. We design an efficient heuristic—based on weak duality—that approximates cut coefficients for units with ramping and capacity constraints, significantly reducing the cost of each Benders iteration without compromising solution quality.
3. Our algorithm outperforms state-of-the-art methods across all tested non-deterministic instances. For example, it is 5× faster than the best alternative on the IEEE 118-bus instance with 54 units and 100 scenarios, and solves SMS++ instances with 100 units and 100 scenarios in 220 seconds on average while all baselines fail to close the optimality gap within one hour (Table 1).

To the best of our knowledge, this approach has not yet been explored in the literature. One possible reason is that while the *3-bin formulation* introduces $O(NT)$ first stage variables, the *extended formulation* requires $O(NT^2)$ first stage variables, where N is the number of units and T is the length of the time horizon.

The rest of the paper is organized as follows. In Sect. 2, we recall how the SUC problem is solved using the *3-bin formulation* (extensive approach and BD). We then introduce our proposed integration of the *extended formulation* into the BD framework and discuss how it enhances the strength of Benders' cuts in Sect. 3. In Sect. 4, we develop a heuristic for efficiently computing a lower bound, yielding valid but not necessarily tight cuts. Finally, in Sect. 5, we present computational experiments evaluating the performance of our formulation on large-scale instances and compare it with existing approaches based on the *3-bin formulation*.

2 The 3-bin formulation of the unit-commitment problem

In this section, we describe the mathematical formulation of SUC using the *3-bin formulation*. Specifically, we present the extensive formulation in Sect. 2.1 and the corresponding BD in Sect. 2.2.

2.1 A description of the *3-bin formulation*

Before introducing the constraints and variables used in the *3-bin formulation*, we first establish the underlying assumption regarding the treatment of uncertainty.

Assumption 1 (Stochastic framework) In this study, we adopt a two-stage linear model with right-hand side uncertainty (the net demand). More precisely, the uncertainty is captured through a finite set of scenarios denoted by $\mathcal{S}$. Each scenario $s \in \mathcal{S}$ has a realization ξ_s of net demand and an associated probability π_s. Furthermore, first stage decisions (usually commitment variables, possibly integer) are made before knowing which scenario is realized, and thus are the same across all scenarios. The second stage decisions (usually dispatch variables) are made after the scenario is revealed and are thus indexed by the scenarios $s \in \mathcal{S}$. We also assume that second stage decisions are continuous variables.

The classical *3-bin formulation* models the first stage decisions using the variables $x_j = (u_j, v_j, w_j)$, where u_j, v_j, and w_j are vectors of binary variables of length $T + 1$, representing the commitment status of unit j over the time horizon $[0 : T] = \{0, \dots, T\}$. Specifically, the binary variable $u_{j,t}$ equals 1 if and only if unit j is on at time t. The variable $v_{j,t}$ (respectively $w_{j,t}$) equals 1 if unit j starts up (respectively shuts down) at time t. These three sets of binary variables are linked by the following constraints:

$$u_{j,t} - u_{j,t-1} = v_{j,t} - w_{j,t} \qquad \forall j \in \mathcal{J}, \forall t \in [1 : T]. \tag{1a}$$

Here $\mathcal{J}$ is the set of units. Moreover, additional constraints are imposed on the commitment decisions. These include the minimum uptime, minimum downtime constraints, and initial conditions specifying the initial commitment status $(u_{j,0}, v_{j,0}, w_{j,0})$ for each unit j. Since we later introduce in Sect. 3 an alternative formulation that specifically enforces the minimum uptime constraints, we separate these constraints on the decision variables x into two distinct polyhedra: X_j^U, which represents the minimum uptime constraints, and X_j, which contains all the remaining constraints. For example, X_j^U is defined by the following constraint on x_j:

$$\sum_{k \in [t - T_j^U + 1, t]} v_{j,k} \leq u_{j,t} \quad \forall t \in [L_j + 1, T], \tag{1b}$$

where T_j^U is the minimum uptime parameter and L_j is the earliest time step at which the minimum uptime constraint must be enforced. As a result, the first stage decision variables x must satisfy:

$$x_j \in X_j^U \cap X_j \qquad \forall j \in \mathcal{J}. \tag{1c}$$

In the second stage, the uncertainty is represented by a set $\mathcal{S}$ of scenarios, with all recourse decisions indexed by scenario. We first consider the power generation variables $p_{j,s}$ over the time horizon $\{0, \dots, T\}$ for each unit j and each scenario s. These decisions must satisfy technical constraints specific to each unit. All the common constraints in the unit commitment problem, such as ramping constraints, capacity limits, and initial conditions, can be compactly expressed as:

$$W_j x_j \leq A_j p_{j,s} \qquad \forall j \in \mathcal{J}, \forall s \in \mathcal{S}. \tag{1d}$$

We explicitly define the matrices A_j and W_j in Sect. 4.1 for the ramping constraints, capacity limits, and initial conditions.

Next, we consider the network constraints and variables. The units $\mathcal{J}$ are distributed across a network with load demands at each node, also known as a bus, and power flows between buses. Let $\mathcal{N}$ denote the set of buses in the network. For each bus $n \in \mathcal{N}$, let Λ_n^G be the set of units located at bus n, and Λ_n^L the set of buses connected to bus n. Regarding the network variables, we define $f_{n,m,s}$ as the power flow from bus n to bus m across the time steps $\{0, \cdots, T\}$ in scenario s, and $\theta_{n,s}$ as the voltage angle at bus n over the same time horizon in scenario s. Under the DC power flow approximation, the network constraints are expressed as:

$$f_{n,m,s} = B_{nm}(\theta_{n,s} - \theta_{m,s}) \qquad \forall (n,m) \in \mathcal{L}, \forall s \in \mathcal{S} \tag{1e}$$

$$-\overline{P}_{nm}e \leq f_{n,m,s} \leq \overline{P}_{nm}e \qquad \forall (n,m) \in \mathcal{L}, \forall s \in \mathcal{S}, \tag{1f}$$

where e is the $T+1$-dimensional vector of ones, $\mathcal{L}$ is the set of lines in the network, $\overline{P}_{nm}$ is the transmission capacity of line (n,m), and B_{nm} is the susceptance of that line.

Finally, the last constraint to express is the satisfaction of load demand under each scenario, which can be written as:

$$\sum_{j \in \Lambda_n^G} p_{j,s} = \xi_{n,s} + \sum_{m \in \Lambda_n^L} f_{n,m,s} \qquad \forall n \in \mathcal{N}, \forall s \in \mathcal{S}, \tag{1g}$$

where $\xi_{n,s}$ denotes the load demand at bus n over the time horizon $\{0, \dots, T\}$ in scenario s. Thus, the SUC problem can be formulated in an extensive form using the 3-bin variables x for first stage decisions:

$$\begin{aligned} \min_{x,p,f,\theta} \quad & \sum_{j \in \mathcal{J},\, t \in [0:T]} C_{j,t}^\top x_{j,t} + \sum_{s \in \mathcal{S},\, j \in \mathcal{J},\, t \in [0:T]} \pi_s\, c_{j,t}\, p_{j,t,s} & \text{(1h)} \\ \text{s.t.} \quad & \text{(1a)} - \text{(1g)} \\ & x_j \in \{0,1\}^{3(T+1)},\ p_{j,s} \in \mathbb{R}^{T+1},\ f_{n,m,s} \in \mathbb{R}^{T+1},\ \theta_{n,s} \in \mathbb{R}^{T+1}. \end{aligned}$$

In the objective function (1h), π_s denotes the probability of scenario s, $C_{j,t}$ is a 3-dimensional vector representing the commitment costs of unit j at time t, and $c_{j,t}$ is the production cost of unit j at time t.

2.2 Benders decomposition based on the *3-bin formulation*

To introduce BD, we begin with the extensive formulation (1) and reformulate it to highlight the recourse function Q, where $Q(x,\xi)$ represents the optimal cost of the second stage knowing the first stage decision x and the realization of uncertainty ξ. Thus, problem (1) is equivalent to:

$$\min_{x,z} \quad z + \sum_{j \in \mathcal{J}, t \in [0:T]} C_{j,t}^\top x_{j,t} \tag{2a}$$

$$\text{s.t.} \quad \sum_{s \in \mathcal{S}} \pi_s Q(x, \xi_s) \leq z \tag{2b}$$

$$\text{(1a)} - \text{(1c)},\ x_j \in \{0,1\}^{3(T+1)},\ z \in \mathbb{R}_+,$$

where $Q(x,\xi)$ is the optimal value of the problem (3):

$$
\begin{aligned}
Q(x,\xi) = \min_{p,f,\theta} \quad & \sum_{j\in\mathcal{J},t\in[0:T]} c_{j,t}p_{j,t} && && \text{(3a)}\\
\text{s.t.} \quad & W_j x_j \le A_j p_j && \forall j \in \mathcal{J} \quad [\mu_j] && \text{(3b)}\\
& \sum_{j\in\Lambda_n^G} p_j = \xi_n + \sum_{m\in\Lambda_n^L} f_{n,m} && \forall n \in \mathcal{N} \quad [\nu_n] && \text{(3c)}\\
& f_{n,m} = B_{nm}(\theta_n - \theta_m) && \forall (n,m) \in \mathcal{L} \quad [\lambda_{n,m}^1] && \text{(3d)}\\
& f_{n,m} \le \overline{P}_{nm} e && \forall (n,m) \in \mathcal{L} \quad [\lambda_{n,m}^2] && \text{(3e)}\\
& f_{n,m} \ge -\overline{P}_{nm} e && \forall (n,m) \in \mathcal{L} \quad [\lambda_{n,m}^3] && \text{(3f)}\\
& p_j \in \mathbb{R}^{T+1},\ f_{n,m} \in \mathbb{R}^{T+1},\ \theta_n \in \mathbb{R}^{T+1}.
\end{aligned}
$$

To derive feasibility and optimality Benders cuts, we study the dual problem:

$$
\begin{aligned}
\max_{\nu,\mu,\lambda} \quad & g(\xi,\nu,\lambda) + \sum_{j\in\mathcal{J}} \mu_j^\top W_j x_j && && \text{(4a)}\\
\text{s.t.} \quad & -\nu_n + \lambda_{n,m}^1 - \lambda_{n,m}^2 + \lambda_{n,m}^3 = 0 && \forall n \in \mathcal{N} && \text{(4b)}\\
& \sum_{m\in\Lambda_n^L} (B_{mn}\lambda_{m,n}^1 - B_{nm}\lambda_{n,m}^1) = 0 && \forall n \in \mathcal{N} && \text{(4c)}\\
& A_j^\top \mu_j + \nu_{j_n} = c_j && \forall j \in \mathcal{J} && \text{(4d)}\\
& \mu_j \in \mathbb{R}_+^q,\ \nu_n \in \mathbb{R}^{T+1},\ \lambda_{n,m}^i \in \mathbb{R}^{T+1}, && && \text{(4e)}
\end{aligned}
$$

where j_n denotes the bus to which unit j belongs and the function g,introduced to improve readability, is defined as:

$$
g(\xi,\nu,\lambda) := \sum_{n\in\mathcal{N}} \xi_n^\top \nu_n - \sum_{n,m} \overline{P}_{nm} e^\top (\lambda_{n,m}^2 + \lambda_{n,m}^3). \tag{5}
$$

Since problem (3) is a linear program, strong duality holds if and only if at least one of the primal or dual problems is feasible. Throughout this paper, we assume that strong duality holds between problems (3) and (4) under any admissible first stage decision:

Assumption 2 (Strong duality) For any first stage decision x satisfying constraints (1a)–(1c), strong duality holds between problems (3) and (4).

As a result, the recourse function $Q(x,\xi)$ equals the optimal value of the dual problem (4).

A well-known sufficient condition ensuring this strong duality is the relatively complete recourse assumption, which requires that for any first stage decision x satisfying constraints (1a)–(1c), the second stage problem remains feasible under all scenarios:

$$Q(x, \xi_s) < +\infty \quad \forall s \in \mathcal{S}. \tag{6}$$

In the context of the UC problem, this relatively complete recourse assumption is typically enforced by introducing load shedding or curtailment. These artificial units provide flexibility, ensuring the feasibility of the second stage problem in all scenarios.

Let us highlight a few key observations regarding the structure of the dual problem (4). First, the dual polyhedron, denoted by $\mathcal{D}$ and defined by constraints (4b)-(4e), is independent of the first stage decision variables x, as we are in a fixed recourse setting [24]. Consequently, any dual solution of (4) provides a feasible cut for the master problem. Second, when the variables ν and λ are fixed, the problem becomes separable at the unit level. These two insights are fundamental for our strategy for deriving new cuts in Sect. 3.

Now, we derive an equivalent formulation of problem (1) by using the extreme points $\mathcal{F}$ and extreme rays $\mathcal{O}$ of the feasible set $\mathcal{D}$ of the dual problem (4):

$$\min_{x,z} \quad z + \sum_{j\in\mathcal{J}, t\in[0:T]} C_{j,t}^\top x_{j,t} \tag{7a}$$

$$\text{s.t.} \quad \sum_{s\in\mathcal{S}} \pi_s z_s \leq z \tag{7b}$$

$$g(\xi_s, \nu^o, \lambda^o) + \sum_{j\in\mathcal{J}} (\mu_j^o)^\top W_j x_j \leq z_s \qquad \forall s \in \mathcal{S}, \forall o \in \mathcal{O} \tag{7c}$$

$$g(\xi_s, \nu^f, \lambda^f) + \sum_{j\in\mathcal{J}} (\mu_j^f)^\top W_j x_j \leq 0 \qquad \forall s \in \mathcal{S}, \forall f \in \mathcal{F} \tag{7d}$$

$$(1a) - (1c),\ x_j \in \{0,1\}^{3(T+1)}, z_s \in \mathbb{R}_+, z \in \mathbb{R}.$$

Constraints (7c) and (7d) represent the optimality and feasibility cuts, respectively. Since the sets $\mathcal{O}$ and $\mathcal{F}$ are finite, problem (7) is an MILP and can, in principle, be solved using standard commercial solvers. However, the sets $\mathcal{O}$ and $\mathcal{F}$ may contain a prohibitively large number of elements, many of which are not known a priori. Thus, the complete enumeration of all cuts (7c) and (7d) may be computationally infeasible. To address this issue, we adopt a classical BD approach, in which cuts are generated iteratively. In each iteration, a subproblem is solved to identify a new cut. Specifically, this subproblem corresponds to solving the dual problem (4). If the dual is bounded, we obtain an extreme point $o \in \mathcal{O}$, which gives rise to an optimality cut. Conversely, if the dual is unbounded, an extreme ray $f \in \mathcal{F}$ is identified, indicating infeasibility of the corresponding primal problem, and a feasibility cut is generated.

3 Extended formulation for efficient Benders' decomposition

In this section, we present an enhanced Benders' decomposition algorithm, utilizing an alternative formulation for the first stage decisions that leads to strengthened Benders' cuts and improved convergence. We begin in Sect. 3.1 by introducing the variables and constraints of the extended formulation. Sect. 3.2 discusses a technical assumption that holds for most common constraints. In Sect. 3.3, we outline the process for deriving the novel Benders' cuts, which represent the main contribution of our work. Finally, in Sect. 3.4, we study the link between these novel Benders' cuts and the ones in the *3-bin formulation*.

3.1 Extended formulation of SUC problem with interval variables

Let $\mathcal{T}_j$ denote the set of all feasible continuous operating intervals for generator j, determined by its minimum uptime requirement. Specifically, it contains two categories of intervals. The first includes intervals $[a : b] \subseteq [1 : T+1]$. The second includes intervals of the form $[-a_0, b]$, where $b \in [1 : T+1]$ and a_0 denotes the duration the unit j had already been operating before the start of the time horizon. Among these, we only retain intervals whose lengths satisfy the minimum uptime requirement, except when $b = T+1$, which we allow by convention to account for units that remain on beyond the end of the planning horizon. We introduce binary variables $\gamma_{j,a,b}$, which equal 1 if and only if unit j is on from a to $b-1$, with $[a : b] \in \mathcal{T}_j$. The relationship between these interval variables and the 3-bin variables $x_j = (u_j, v_j, w_j)$ (defined in Sect. 2.1) can be made by introducing the vectors $x_{j,a,b} = (u_{j,a,b}, v_{j,a,b}, w_{j,a,b}) \in \{0,1\}^{3(T+1)}$ defined as follows:

$$u_{j,a,b,t} = \mathbb{1}_{t\in[a:b-1]} \tag{8a}$$

$$v_{j,a,b,t} = \mathbb{1}_{t=a} \tag{8b}$$

$$w_{j,a,b,t} = \mathbb{1}_{t=b}. \tag{8c}$$

Each vector $x_{j,a,b}$ captures the 3-bin commitment decisions if the unit is on only during the interval $[a : b-1]$. The overall commitment decision $x_j \in \{0,1\}^{3(T+1)}$ is then expressed as a convex combination of these vectors $x_{j,a,b}$ weighted by the γ-variables:

$$x_{j,t} = \sum_{[a:b]\in\mathcal{T}_j} \gamma_{j,a,b} x_{j,a,b,t} \quad \forall j \in \mathcal{J}, \forall t \in [0 : T]. \tag{9}$$

Using these new first stage variables, the SUC problem reads as follows.

$$\min_{\gamma,z} \quad z + \sum_{j\in\mathcal{J},t\in[0:T]} C_{j,t}^\top x_{j,t} \tag{10a}$$

$$\text{s.t.} \quad \sum_{s \in \mathcal{S}} \pi_s Q(x, \xi_s) \leq z \tag{10b}$$

$$\sum_{[a:b] \in \mathcal{T}_j \,|\, t \in [a:b]} \gamma_{j,a,b} \leq 1 \qquad \forall j \in \mathcal{J}, \quad \forall t \in [0:T] \tag{10c}$$

$$x_j = \sum_{[a:b] \in \mathcal{T}_j} \gamma_{j,a,b} x_{j,a,b} \qquad \forall j \in \mathcal{J} \tag{10d}$$

$$x_j \in X_j \qquad \forall j \in \mathcal{J} \tag{10e}$$

$$\gamma_{j,a,b} \in \mathbb{R}_+, x_j \in \{0,1\}^{3(T+1)}, z \in \mathbb{R}_+. \tag{10f}$$

Let us describe this formulation in detail. First, constraint (10c) ensures that a unit cannot operate in two overlapping intervals simultaneously. This restriction guarantees a valid scheduling of the unit's operational periods. Second, constraint (10e) involves the polyhedron X_j, which represents all unit commitment constraints except the minimum uptime constraints for each unit. The latter are already implicitly incorporated through the definition of the index set $\mathcal{T}_j$ used for the variables γ. Note that due to constraint (10d), if the variables x are binary, then the variables γ must also be binary. Therefore, it is sufficient to impose the binary constraint on x only. Finally, as in the previous section, Q represents the recourse cost function, capturing the dispatch's cost in response to uncertainties and is determined as the value of problem (3) or (4).

3.2 Technical decomposition assumption

Before diving into the advanced cut generation methodology, we discuss a technical decomposition assumption on the technical constraints associated with each generation unit.

Assumption 3 (Decomposition Assumption) Let j be the index of any arbitrary generation unit. We will say that the decomposition assumption holds if for any given sub-vector (γ, x) that satisfies constraints (10c)–(10f) and $\gamma_{j,a,b} = 1$ for some interval $[a:b] \in \mathcal{T}_j$, the following implication holds:

$$p_j \in \mathbb{R}^{T+1} \text{ s.t. } W_j x_j \leq A_j p_j \implies W_j x_{j,a,b} \leq A_j \tilde{p}_{j,a,b} \text{ for } \tilde{p}_{j,a,b} = (p_{j,t} x_{j,a,b,t})_{t \in [0:T]}, \tag{11}$$

where $\tilde{p}_{j,a,b}$ is a vector in $\mathbb{R}^{T+1}$ which depends on (j, a, b).

Assumption 3 implies that the technical constraints associated with two non-overlapping intervals are independent. This condition is satisfied when each unit is subject to any standard technical constraints described in Sect. 4.1, including ramping limits, capacity bounds, and initial conditions. This follows from two key observations. First, these constraints only introduce temporal coupling between consecutive time steps t and $t-1$. Second, constraint (10c) ensures that on-intervals do not overlap. We emphasize that all constraints involving commitment variables, such as minimum uptime, minimum downtime, or a maximum number of startups, are handled either through the definition of the feasible interval set $\mathcal{T}_j$ or via the polyhedron X_j in constraint (10e).

We now present two similar examples of constraint sets: one that violates Assumption 3, and one that satisfies it.

Example 1 (Constraints the do not satisfy Assumption 3) Consider a small time horizon $T = 2$ and a unit j subject to the following technical constraints:

$$10u_{j,t} \leq p_{j,t} \leq 20u_{j,t} \quad \forall t \in [0:2], \tag{12a}$$

$$p_{j,2} - p_{j,0} \leq 10u_{j,0}. \tag{12b}$$

Now consider the case where the first stage commitment vector is $u_j = (1, 0, 1)$, *i.e.,* the unit is committed in the intervals $[0, 1]$ and $[2, 3]$. Suppose the power generation vector is $p_j = (20, 0, 20)$. This vector satisfies all constraints in (12) globally. However, the subvector $\tilde{p}_{j,2,3} = (0, 0, 20)$, corresponding to the second active interval, violates constraint (12b).

This illustrates that even though the intervals $[0 : 1]$ and $[2 : 3]$ are disjoint, the constraint (12b) introduces a coupling between them. As a result, the assumption of independence between non-overlapping intervals (Assumption 3) is violated.

Example 2 (Constraints Satisfying Assumption 3) Consider again a unit j with time horizon $T = 2$, but now subject to a different set of constraints:

$$10u_{j,t} \leq p_{j,t} \leq 20u_{j,t} \quad \forall t \in [0:2], \tag{13a}$$

$$p_{j,1} - p_{j,0} \leq 10u_{j,0} \tag{13b}$$

$$p_{j,2} - p_{j,1} \leq 10u_{j,1}. \tag{13c}$$

Consider the same commitment vector as in Example 1 $u_j = (1, 0, 1)$. In this case, the power generation vector $p_j = (20, 0, 20)$ does not satisfy constraint (13c) due to the zero value of $u_{j,1}$. However, the vector $p_j = (10, 0, 10)$, for example, satisfies all constraints in (13), and so does the subvector $\tilde{p}_{j,2,3} = (0, 0, 10)$.

More generally, constraints linking dispatch decisions across time, *e.g.,* ramping constraints between time steps t and $t + 2$ or maximal total energy produced during the horizon (fuel limits), do not satisfy Assumption 3. However, such constraints are uncommon in UC problems, and could be dualized as well, at the price of additional dual variables.

3.3 Advanced cut generation in Benders' decomposition

In this section, we outline the process for deriving the novel formulation (23) for solving SUC under the decomposition Assumption 3.

When computing the recourse function Q by solving problem (4) using the x variables, we cannot directly derive stronger Benders' cuts than in the *3-bin formulation*. To achieve this, we introduce the recourse function $\mathcal{Q}$ formulated using the γ variables and the dispatch variables $\tilde{p}_{j,a,b,t}$, representing the power output of unit j at time t, conditional on $\gamma_{j,a,b} = 1$.

These variables $\tilde{p}_{j,a,b,t}$ are handy for expressing the production constraints differently, rather than in the *3-bin formulation* (3). Indeed, we consider that the variables $\tilde{p}_{j,a,b}$ must respect the following constraints:

$$
\begin{aligned}
& W_j x_{j,a,b} \le A_j \tilde{p}_{j,a,b} && \forall j \in \mathcal{J},\, \forall [a:b] \in \mathcal{T}_j && \text{(14a)}\\
& \sum_{j \in \Lambda_n^G, [a:b] \in \mathcal{T}_j \,|\, t \in [a:b]} \gamma_{j,a,b} \tilde{p}_{j,a,b} = \xi_n + \sum_{m \in \Lambda_n^L} f_{n,m} && \forall n \in \mathcal{N} && \text{(14b)}\\
& f_{n,m} = B_{nm}(\theta_n - \theta_m) && \forall (n,m) \in \mathcal{L} && \text{(14c)}\\
& f_{n,m} \le \overline{P}_{nm} e && \forall (n,m) \in \mathcal{L} && \text{(14d)}\\
& -\overline{P}_{nm} e \le f_{n,m} \le \overline{P}_{nm} e && \forall (n,m) \in \mathcal{L}. && \text{(14e)}
\end{aligned}
$$

Constraints (14a) represent the technical constraint on the interval $[a:b]$, and constraints (14b)-(14e) represent the linking constraints between the units, i.e, the satisfaction of the demand and the network constraints.

Thus, we define the recourse value $\mathcal{Q}(\gamma, \xi)$ as the optimal value of the minimization problem (14).

$$
\begin{aligned}
\mathcal{Q}(\gamma, \xi) = \min_{\tilde{p}, f, \theta} \quad & \sum_{j \in \mathcal{J}, [a:b] \in \mathcal{T}_j, t \in [0:T]} c_{j,t} \gamma_{j,a,b} \tilde{p}_{j,a,b,t} && \text{(14f)}\\
\text{s.t.} \quad & (14a) - (14e).\\
& \tilde{p}_{j,a,b} \in \mathbb{R}^{T+1},\; f_{n,m} \in \mathbb{R}^{T+1},\; \theta_n \in \mathbb{R}^{T+1}.
\end{aligned}
$$

Proposition 1 *Let Assumption 3 hold true. At any (γ, x) satisfying the constraints (10c)-(10f), the recourse functions $Q(x, .)$ and $\mathcal{Q}(\gamma, .)$ coincide.*

We give the proof of Proposition 1, which relies on Assumption 3, in Appendix A. Proposition 1 confirms that the two recourse functions Q and $\mathcal{Q}$ describe the same problem. As a result, constraint (10b) can be equivalently reformulated as:

$$
\sum_{s \in \mathcal{S}} \pi_s \mathcal{Q}(\gamma, \xi_s) \le z. \tag{15}
$$

As in Sect. 2.2, we denote by ν and λ the dual variables associated with the demand and network constraints (14b)-(14e). Instead of dualizing all the constraints as in the *3-bin formulation*, we only dualize the network constraints (14b)-(14e). We obtain the following dual problem, where g is the function defined in (5) and $\mathcal{D}_1$ is the polyhedron linking the variables ν and λ (*i.e.,* constraints (4b) and (4c)):

$$
\begin{aligned}
\max_{(\nu,\lambda) \in \mathcal{D}_1} g(\xi, \nu, \lambda) + \min_{\tilde{p}} & \sum_{j \in \mathcal{J}, [a:b] \in \mathcal{T}_j} \gamma_{j,a,b} \tilde{p}_{j,a,b}^{\top} (c_j - \nu_{j_n}) && \text{(16a)}\\
\text{s.t.}\; & W_j x_{j,a,b} \le A_j \tilde{p}_{j,a,b} \quad \forall j \in \mathcal{J},\, \forall [a:b] \in \mathcal{T}_j && \text{(16b)}\\
& \tilde{p}_{j,a,b} \in \mathbb{R}^{T+1}. && \text{(16c)}
\end{aligned}
$$

The following proposition proves that strong duality holds between problems (14) and (16), showing that $\mathcal{Q}(\gamma, \xi)$ is also the optimal value of problem (16).

Proposition 2 (Strong duality between problems (14) and (16)) *Let Assumptions 2 and 3 hold true. Moreover, let* (γ, x) *satisfy constraints* (10c)-(10f), $(\nu, \lambda) \in \mathcal{D}_1$, *and* $\xi \in \mathbb{R}^{T+1}$. *Then, strong duality between problems* (14) *and* (16) *holds.*

Proof To prove the proposition, consider the following dual problem of (3), where only the constraints (4b) and (4c) are dualized:

$$\max_{(\nu,\lambda)\in\mathcal{D}_1} g(\xi,\nu,\lambda) + \min_{p} \sum_{j\in\mathcal{J}} p_j^\top (c_j - \nu_{j_n}) \tag{17a}$$

$$\text{s.t. } W_j x_j \le A_j p_j \quad \forall j \in \mathcal{J}, \tag{17b}$$

$$p_j \in \mathbb{R}^{T+1}. \tag{17c}$$

By applying the same reasoning as in the proof of Proposition 1 [see Appendix A], the inner minimization problems in (16) and (17) are equivalent and yield the same optimal value.

By Assumption 2, strong duality holds for this partially dualized problem. Hence, the optimal value of the dual problem (17) is equal to $Q(x,\xi)$. Moreover, from Proposition 1, we have $Q(x,\xi) = \mathcal{Q}(\gamma,\xi)$.

Therefore, $\mathcal{Q}(\gamma,\xi)$ is also the optimal value of the dual problem (16), which concludes the proof that strong duality holds between problems (14) and (16). □ □

Finally, by the structure of its objective function and constraints, the inner minimization problem in (14) is separable over the units $j \in \mathcal{J}$ and the intervals $[a:b] \in \mathcal{T}_j$. Moreover, since the variables γ are fixed in the recourse problem (16), we obtain the equivalent formulation (18):

$$\mathcal{Q}(\gamma,\xi) = \max_{(\nu,\lambda)\in\mathcal{D}_1} \quad g(\xi,\nu,\lambda) + \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b} \hat{Q}_{j,a,b}(c_j - \nu_{j_n}), \tag{18}$$

where $\hat{Q}_{j,a,b}(y)$ represents the optimal value of problem (19), *i.e.,* the optimal operating cost of unit j on the interval $[a:b]$ when its production cost is equal to $y \in \mathbb{R}^{T+1}$.

$$\hat{Q}_{j,a,b}(y) = \min_{p\in\mathbb{R}^{T+1}} \quad y^\top p \tag{19a}$$

$$W_j x_{j,a,b} \le A_j p. \tag{19b}$$

We highlight that formulation (18) enables a decomposition of the Unit Commitment dispatch problem across both the set of units $\mathcal{J}$ and the time intervals $\mathcal{T}_j$, once the linking constraints are dualized. This decomposition property, valid under Assumption 3, is not exploited with the *3-bin formulation*.

Then, to perform BD, it is now necessary to derive a finite set of feasibility and optimality cuts for formulation (18).

The proof of the following Proposition 3 is provided in Appendix B.

Proposition 3 *Let Assumptions 2 and 3 hold. Let* (γ, x) *satisfy constraints* (10c)-(10f), $(\nu, \lambda) \in \mathcal{D}_1$, *and* $\xi \in \mathbb{R}^{T+1}$. *With* μ^* *an optimal solution of the following problem:*

$$\max_{\mu} \quad \sum_{j\in\mathcal{J}} \mu_j^\top W_j x_j \tag{20a}$$

$$s.t. \quad A_j^\top \mu_j + \nu_{j_n} = c_j \quad \forall j \in \mathcal{J} \tag{20b}$$

$$\mu_j \in \mathbb{R}_+^q. \tag{20c}$$

it follows that, the objective function value of problem (4) at (μ^*, ν, λ) is equal to the objective function value of problem (18) at (ν, λ).

Proposition 3 enables us to derive a finite set of optimality and feasibility cuts from the extreme points $\mathcal{O}$ and extreme rays $\mathcal{F}$ of the feasible set $\mathcal{D}$ of problem (4). According to Proposition 1, the two subproblems (4) and (18) are either both bounded or both unbounded. Consider the case in which these problems are bounded. In this setting, there exists $o \in \mathcal{O}$ such that the extreme point $(\mu^o, \nu^o, \lambda^o)$ is optimal for problem (4). Since this triplet satisfies the conditions of Proposition 3, it follows that (ν^o, λ^o) is also optimal for problem (18). Similarly, we prove in Appendix C that problems (4) and (18) are unbounded if and only if:

$$\exists f \in \mathcal{F}, \quad g(\xi, \nu^f, \lambda^f) + \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\hat{Q}_{j,a,b}(-\nu_{j_n}^f) > 0. \tag{21}$$

Consequently, the recourse value $\mathcal{Q}(\gamma, \xi)$ is equal to:

$$\begin{cases} +\infty \quad \text{if } \exists f \in \mathcal{F}, g(\xi, \nu^f, \lambda^f) + \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\hat{Q}_{j,a,b}(-\nu_{j_n}^f) > 0 \\ \max_{o\in\mathcal{O}} \left(g(\xi, \nu^o, \lambda^o) + \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\hat{Q}_{j,a,b}(c_j - \nu_{j_n}^o) \right) \text{otherwise.} \end{cases} \tag{22}$$

As a result, we derive formulation (23) for the SUC, which, to the best of our knowledge, is novel.

$$\min_{\gamma,z} \quad z + \sum_{j\in\mathcal{J},\, t\in[0:T]} C_{j,t}^\top x_{j,t} \tag{23a}$$

$$\text{s.t.} \quad \sum_{[a:b]\in\mathcal{T}_j | t\in[a:b]} \gamma_{j,a,b} \leq 1 \qquad \forall j \in \mathcal{J}, \forall t \in [0:T] \tag{23b}$$

$$\sum_{s\in\mathcal{S}} \pi_s z_s \leq z \tag{23c}$$

$$g(\xi_s, \nu^o, \lambda^o) + \sum_{\substack{j\in\mathcal{J} \\ [a:b]\in\mathcal{T}_j}} \gamma_{j,a,b}\hat{Q}_{j,a,b}(c_j - \nu_{j_n}^o) \leq z_s \qquad \forall s \in \mathcal{S},\, \forall o \in \mathcal{O} \tag{23d}$$

$$g(\xi_s, \nu^f, \lambda^f) + \sum_{\substack{j\in\mathcal{J} \\ [a:b]\in\mathcal{T}_j}} \gamma_{j,a,b}\hat{Q}_{j,a,b}(-\nu_{j_n}^f) \leq 0 \qquad \forall s \in \mathcal{S},\, \forall f \in \mathcal{F} \tag{23e}$$

$$x_j = \sum_{[a:b] \in \mathcal{T}_j} \gamma_{j,a,b} x_{j,a,b} \qquad \forall j \in \mathcal{J} \tag{23f}$$

$$x_j \in X_j \qquad \forall j \in \mathcal{J} \tag{23g}$$

$$\gamma_{j,a,b} \in \{0,1\}, \tag{23h}$$

where we recall that g is given in (5). Once again, as before, the idea is to generate the optimality cuts (23d) and the feasibility cuts (23e) iteratively. By Proposition 3 and the way we derive these new optimality and feasibility cuts, this generation can be done through the resolution of the same subproblem (4) as in the *3-bin formulation*. This remark is advantageous in practice, as problem (3) involves significantly fewer variables than problem (14).

3.4 Link between Benders' cuts in the two formulations

In this section, we establish the relationship between the Benders' cuts (7c) and (7d) derived from the *3-bin formulation* and those from the *extended formulation* (23d) and (23e), proving that the latter dominate the former. Throughout this part, we consider a fixed unit $j \in \mathcal{J}$, an operating interval $[a : b] \in \mathcal{T}_j$, and a given vector ν.

We recall that $\hat{Q}_{j,a,b}(y)$ corresponds to the optimal value of problem (19). By strong duality, which holds by Assumption (2), this optimal value can also be obtained from the dual problem (24):

$$\hat{Q}_{j,a,b}(y) = \max_{\mu \in \mathbb{R}^q_+} \quad \mu^\top W_j x_{j,a,b} \tag{24a}$$

$$\text{s.t.} \quad A_j^\top \mu = y. \tag{24b}$$

Since $\hat{Q}_{j,a,b}(y)$ is the optimal value of problem (24) and γ is positive, then for any element $o \in \mathcal{O}$ (resp. $f \in \mathcal{F}$), and for any collection $(\mu^o_{j,a,b})_{j \in \mathcal{J}, [a:b] \in \mathcal{T}_j}$ (resp. $(\mu^f_{j,a,b})_{j \in \mathcal{J}, [a:b] \in \mathcal{T}_j}$) of solutions to problem (24) with $y = c_j - \nu^o_{j_n}$ (resp. $y = -\nu^f_{j_n}$), the following cuts are valid for problem (23) and weaker than cuts (23d) and (23e):

$$g(\xi_s, \nu^o, \lambda^o) + \sum_{j \in \mathcal{J}, [a:b] \in \mathcal{T}_j} \gamma_{j,a,b} (\mu^o_{j,a,b})^\top W_j x_{j,a,b} \le z_s \tag{25a}$$

$$g(\xi_s, \nu^f, \lambda^f) + \sum_{j \in \mathcal{J}, [a:b] \in \mathcal{T}_j} \gamma_{j,a,b} (\mu^f_{j,a,b})^\top W_j x_{j,a,b} \le 0. \tag{25b}$$

This weak duality-based approach reveals that the Benders cut (7c) and (7d) in the *3-bin formulation* is a special case of the extended cut (25). In particular, for any extreme point $(\nu^o, \lambda^o, \mu^o)$ (resp. extreme ray $(\nu^f, \lambda^f, \mu^f)$) of the feasible set of problem(4), the optimality cut (7c) (resp. feasibility cut (7d)) can be recovered from (25) by setting $\mu^o_{j,a,b} = \mu^o_j$ (resp. $\mu^f_{j,a,b} = \mu^f_j$) for all unit j and intervals $[a : b] \in \mathcal{T}_j$. This correspondence is established through the relations (9), which connect the γ variables, the 3-bin variables x, and the coefficients $x_{j,a,b}$.

However, the solution μ_j^o (resp. μ_j^f) can be highly suboptimal in problem (24) with $y = c_j - \nu_{j_n}^o$ (resp. $y = -\nu_{j_n}^f$) for some of the intervals $[a : b] \in \mathcal{T}_j$, which explains that the convergence of the Benders' Decomposition with the *3-bin formulation* can be very slow.

4 Heuristic for fast valid cuts

In Sect. (3.4), we showed that the new cuts (23d) and (23e) are stronger than the classical Benders' cuts (7c) and (7d) of the *3-bin formulation*. Nevertheless, they require to solve problem (19), for each unit $j \in \mathcal{J}$ and each feasible operating interval $[a : b] \in \mathcal{T}_j$, to compute $\hat{Q}_{j,a,b}(y)$, with $y \in (c_j - \nu_{j_n}^o)_{o\in\mathcal{O}}$ or $y \in (-\nu_{j_n}^f)_{f\in\mathcal{F}}$. Although problem (19) is very small (only $T + 1$ variables), the number of triplets (j, a, b) can be very large, making the overall procedure computationally expensive.

To address this issue, we use the weak duality-based approach, described in Sect. 3.4, to produce valid cuts of the form (25) without solving all the problems (19) at optimality. The heuristic uses deterministic formulas to construct a near-optimal dual multiplier $\hat{\mu}_{j,a,b}$ for the dual problem (24), starting from any point $(\nu^i, \lambda^i, \mu^i)$, with $i \in \mathcal{O} \cup \mathcal{F}$, returned by subproblem (4). These formulas depend on the specific technical constraints encoded in W_j and A_j. In this paper, we focus on the most common constraints: ramping limits, capacity bounds, and initial conditions. Section 4.1 details how these constraints are formulated, and Section 4.2 presents the corresponding heuristic formulas for computing $\hat{\mu}_{j,a,b}$.

4.1 Technical Constraints

This section outlines the commonly used technical constraints that a generation unit j must satisfy. Each of these constraints can be represented in the general linear form:

$$W_j x_j \leq A_j p_j. \tag{26}$$

First, initial conditions can be imposed and the corresponding constraint is:

$$p_{j,0} = P^{\text{init}} u_{j,0}, \tag{27a}$$

where P^{init} denotes the initial power output of unit j.

Next, the power generation must generally satisfy minimum and maximum capacity limits when the unit is on:

$$\underline{P}_j u_j \leq p_j \leq \overline{P}_j u_j, \tag{27b}$$

where $\underline{P}_j$ and $\overline{P}_j$ are the minimum and maximum capacity, respectively.

Finally, Unit Commitment generally considers ramping up and down constraints, which can be expressed as:

$$p_{j,t} - p_{j,t-1} \leq -R_j^U v_{j,t} + (S_j^U + R_j^U) u_{j,t} - S_j^U u_{j,t-1}, \quad \forall t \in [1 : T], \tag{27c}$$

$$p_{j,t-1} - p_{j,t} \leq -R_j^D w_{j,t} + (S_j^D + R_j^D) u_{j,t-1} - S_j^D u_{j,t}, \quad \forall t \in [1 : T]. \tag{27d}$$

Here, S_j^U and S_j^D represent startup and shutdown ramping parameters, respectively, while R_j^U and R_j^D denote the ramping limits during normal operation.

4.2 Heuristic Formulas

In this section, we consider a unit j whose generation must satisfy the initial conditions (27a), capacity limits (27b), and ramping constraints (27c). We denote by μ_j^{init}, μ_j^{max}, μ_j^{min}, μ_j^{up}, and μ_j^{down} the dual variables associated, respectively, with the initial conditions, maximum and minimum capacity constraints, and the ramping up and ramping down constraints. Under these assumptions, problem (24) for an interval $[a:b]$ and a vector $y \in \mathbb{R}^{T+1}$ can be reformulated as follows:

$$
\begin{aligned}
\max_{\mu_j} \quad & \mu_j^{init} P_j^{init} u_{j,a,b,0} - \sum_{t=0}^{T} \mu_{j,t}^{max} \overline{P}_j u_{j,a,b,t} + \sum_{t=0}^{T} \mu_{j,t}^{min} \underline{P}_j u_{j,a,b,t} && \text{(28a)} \\
& - \sum_{t=1}^{T} \mu_{j,t}^{up} \mathcal{R}_{j,a,b,t}^{U} - \sum_{t=1}^{T} \mu_{j,t}^{down} \mathcal{R}_{j,a,b,t}^{D} \\
\text{s.t.} \quad & \mu_j^{init} + \mu_{j,0}^{max} - \mu_{j,0}^{min} - \mu_{j,1}^{up} + \mu_{j,1}^{down} = y_0 && \text{(28b)} \\
& \mu_{j,t}^{max} - \mu_{j,t}^{min} - \mu_{j,t+1}^{up} + \mu_{j,t}^{up} + \mu_{j,t+1}^{down} - \mu_{j,t}^{down} = y_t \quad \forall t \in [1:T-1] && \text{(28c)} \\
& \mu_{j,T}^{max} - \mu_{j,T}^{min} - \mu_{j,T}^{up} - \mu_{j,T}^{down} = y_T && \text{(28d)} \\
& \mu_j^{max} \in \mathbb{R}_+^{T+1}, \mu_j^{min} \in \mathbb{R}_+^{T+1}, \mu_j^{up} \in \mathbb{R}_+^{T}, \mu_j^{down} \in \mathbb{R}_+^{T}, \mu_j^{init} \in \mathbb{R}, && \text{(28e)}
\end{aligned}
$$

where the parameters $\mathcal{R}_{j,a,b,t}^{U}$ and $\mathcal{R}_{j,a,b,t}^{D}$ are defined as follows:

$$
\begin{aligned}
\mathcal{R}_{j,a,b,t}^{U} &= -R_j^U v_{j,a,b,t} + (S_j^U + R_j^U) u_{j,a,b,t} - S_j^U u_{j,a,b,t-1} && \text{(29a)} \\
\mathcal{R}_{j,a,b,t}^{D} &= -R_j^D w_{j,a,b,t} + (S_j^D + R_j^D) u_{j,a,b,t-1} - S_j^D u_{j,a,b,t}. && \text{(29b)}
\end{aligned}
$$

We now propose a heuristic to compute a feasible solution $\hat{\mu}_{j,a,b}$ to problem (28). Given the structure of constraints (28b)-(28e), the heuristic proceeds by first fixing the values of the variables $\hat{\mu}_{j,a,b}^{up}$ and $\hat{\mu}_{j,a,b}^{down}$, after which the remaining variables can be determined accordingly. Specifically, once feasible vectors $\hat{\mu}_{j,a,b}^{up}$ and $\hat{\mu}_{j,a,b}^{down}$ are chosen, it is possible to construct a feasible solution for the dual problem by computing $\hat{\mu}_{j,a,b}^{max}$, $\hat{\mu}_{j,a,b}^{min}$, and $\hat{\mu}_{j,a,b}^{init}$ using the following expressions:

$$
\begin{aligned}
\hat{\mu}_{j,a,b}^{init} &= y_0 + \hat{\mu}_{j,a,b,0}^{up} - \hat{\mu}_{j,a,b,0}^{down} && \text{(30a)} \\
\hat{\mu}_{j,a,b,0}^{max} &= 0 && \text{(30b)} \\
\hat{\mu}_{j,a,b,0}^{min} &= 0 && \text{(30c)} \\
\hat{\mu}_{j,a,b,t}^{max} &= (-y_t + \hat{\mu}_{j,a,b,t+1}^{up} - \hat{\mu}_{j,a,b,t}^{up} - \hat{\mu}_{j,a,b,t+1}^{down} + \hat{\mu}_{j,a,b,t}^{down})^+ \qquad \forall t \in [1:T-1] && \text{(30d)}
\end{aligned}
$$

$$\hat{\mu}^{min}_{j,a,b,t} = (y_t - \hat{\mu}^{up}_{j,a,b,t+1} + \hat{\mu}^{up}_{j,a,b,t} + \hat{\mu}^{down}_{j,a,b,t+1} - \hat{\mu}^{down}_{j,a,b,t})^+ \qquad \forall t \in [1 : T-1] \tag{30e}$$

$$\hat{\mu}^{max}_{j,a,b,T} = (-y_T - \hat{\mu}^{up}_{j,a,b,T} + \hat{\mu}^{down}_{j,a,b,T})^+ \tag{30f}$$

$$\hat{\mu}^{min}_{j,a,b,T} = (y_T + \hat{\mu}^{up}_{j,a,b,T} - \hat{\mu}^{down}_{j,a,b,T})^+, \tag{30g}$$

where $z^+ = \max(0, z)$ denotes the positive part of z.

Next, considering a unit j, an interval $[a : b]$, first stage decisions (x, γ), and the point $(\nu^i, \lambda^i, \mu^i)$, with $i \in \mathcal{O} \cup \mathcal{F}$, returned by subproblem (4), we provide specific formulas to $\hat{\mu}^{up}_{j,a,b}$ and $\hat{\mu}^{down}_{j,a,b}$. We first note that by complementary conditions, we know that at the optimum $\mu^{up}_{j,t}$ (resp. $\mu^{down}_{j,t}$) is non-zero if and only if the ramping up (resp. ramping down) constraint is active. The number of consecutive steps where these constraints are active is limited by two things: the capacity limits of the unit j and the length of the interval $[a : b]$ which imposes that the production generation is zero at time steps a and b. Thus, we first introduce the parameters T^{RU}_j and T^{RD}_j defined as follows:

$$T^{RU}_j(a) = \begin{cases} 1 + \lfloor \frac{\overline{P}_j - \underline{P}_j}{R^U_j} \rfloor & \text{if } a \geq 0 \\ \lfloor \frac{\overline{P}_j - P^{init}_j}{R^U_j} \rfloor & \text{otherwise,} \end{cases} \tag{31a}$$

$$T^{RD}_j(a) = \begin{cases} 1 + \lfloor \frac{\overline{P}_j - \underline{P}_j}{R^D_j} \rfloor & \text{if } a \geq 0 \\ 1 + \lfloor \frac{\overline{P}_j - P^{init}_j}{R^D_j} \rfloor & \text{otherwise.} \end{cases} \tag{31b}$$

Then, we consider the time $t_{j,a,b}$ as the intersection point of two lines with slopes $-R^D_j$ and R^U_j, as illustrated in Figure 1:

$$t_{j,a,b} = \begin{cases} \frac{S^D_j - S^U_j + (b-1)R^D_j + aR^U_j}{R^D_j + R^U_j} & \text{if } a \geq 0 \\ \frac{S^D_j - P^{init}_j + (b-1)R^D_j}{R^D_j + R^U_j} & \text{otherwise.} \end{cases} \tag{32}$$

For example, the instant $t_{j,a,b}$ defines the latest time step up to which the ramping-up constraint can be continuously active without violating the required shutdown condition at time b. As illustrated in Figure 1, if the ramping-up constraint is binding throughout the interval $[a : a+2]$, the unit cannot reduce its output to reach the shutdown threshold S^D_j by time $b-1$.

It is important to note that if $t_{j,a,b} \notin [a : b-1]$, then no feasible generation trajectory exists on the interval $[a : b]$ that satisfies the technical constraints given in (27). In such cases, the interval $[a : b]$ can be removed from the set $\mathcal{T}_j$. Hence, we assume without loss of generality that $a \leq t_{j,a,b} \leq b-1$ holds for all units $j \in \mathcal{J}$ and all intervals $[a : b] \in \mathcal{T}_j$.

We finally we define the parameters $\varphi^U_j(a, b)$ (resp.$\varphi^D_j(a, b)$) that represent the number of consecutive time steps where the ramping-up (resp. ramping-down)

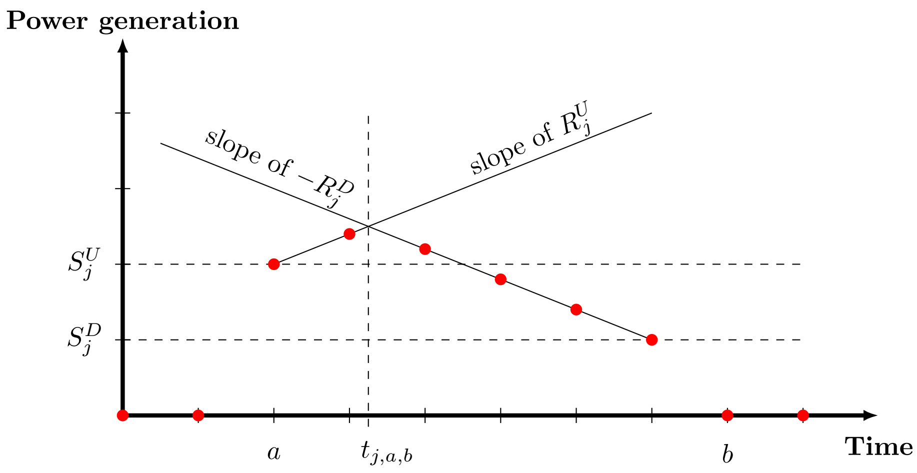

Fig. 1 Definition of $t_{j,a,b}$. Red points indicate the maximum power generation at each time step when $\gamma_{j,a,b} = 1$.

constraint can be active as:

$$\varphi_j^U(a,b) = \min\left(T_j^{RU}(a), \lfloor t_{j,a,b} \rfloor - a + 1\right) \tag{33a}$$

$$\varphi_j^D(a,b) = \min\left(T_j^{RD}(a), b - 1 - \lfloor t_{j,a,b} \rfloor\right) \tag{33b}$$

To leverage the information provided by the dual solution μ^i (optimal solution of problem (4) with first stage decisions (x,γ)), our heuristic distinguishes two specific situations for a given time step t:

(i) The time step t belongs to the interval $[a:b]$ which is selected in the first stage decisions, *i.e.*, $\gamma_{j,a,b} = 1$;
(ii) The interval $[a:b]$ is sufficiently long, specifically $b - a \geq T_j^{RU} + T_j^{RD}$, and there exists an interval $[c:d]$ such that $\gamma_{j,c,d} = 1$ and the time step t belongs to the interval $[\max(a,c) + T_j^{RU}\mathbb{1}_{a=c} : \min(b,d) - T_j^{RD}\mathbb{1}_{b=d}]$.

Based on these cases, the heuristic defines the values of $\hat{\mu}^{up}_{j,a,b,t}$ and $\hat{\mu}^{down}_{j,a,b,t}$ using the following explicit formulas:

$$\hat{\mu}^{up}_{j,a,b,t} = \begin{cases} (\mu^i)^{up}_{j,t} & \text{if condition (i) or (ii) holds,} \\ 0 & \text{if } t \notin [a:b], \\ \left(-\sum_{k=t}^{a+\varphi_j^U(a,b)-1} y_k\right)^+ & \text{otherwise,} \end{cases} \tag{34a}$$

$$\hat{\mu}^{down}_{j,a,b,t} = \begin{cases} (\mu^i)^{down}_{j,t} & \text{if condition (i) or (ii) holds,} \\ 0 & \text{if } t \notin [a:b] \text{ or } b = T+1, \\ \left(-\sum_{k=b-\varphi_j^D(a,b)}^{t-1} y_k\right)^+ & \text{otherwise,} \end{cases} \tag{34b}$$

where $y = (c_j - \nu^i_{j_n})$ if $i \in \mathcal{O}$ and $y = (-\nu^i_{j_n})$ if $i \in \mathcal{F}$.

Condition (i) exploits the fact that if $\gamma_{j,a,b} = 1$, then the dual solution μ^i is already optimal for problem (28). Condition (ii) aims to approximate the behavior of the unit j in the central portion of $[a : b]$, which can be assumed to mirror the optimal behavior in an interval $[c : d]$ where $\gamma_{j,c,d} = 1$, and the time step t is sufficiently distant from the boundaries of both $[a : b]$ and $[c : d]$.

Finally, our heuristic consists of selecting, for each unit j and each interval $[a : b] \in \mathcal{T}_j$, the best solution between the dual solution μ^i obtained from the subproblem (4) and the solution computed using equations (30) and (34) for problem (28). This selection ensures that the resulting Benders' cut of the form (25) is stronger than the one obtained directly from μ^i in the *3-bin formulation* (constraint (7c) or (7d)), ensuring the convergence of the BD.

5 Computational experiments

In this section, we present extensive computational experiments conducted on various instances drawn from two different sources in the literature. First, we use open-source instances from the SMS++ project, previously studied in [25]. These instances exclude network constraints and feature between 10 and 100 thermal generators. Second, we also consider a modified IEEE instance that includes network constraints, with 54 generators, 118 buses, and 186 lines. This instance is commonly used in the literature; however, it is essential to note that uncertainty modeling varies across different studies. Details on our scenario generation methodology for representing uncertainty are provided in the Appendix D.

For each size category in the SMS++ dataset (10, 50, or 100 thermal units), five different instances are available. For each of these instances and each test with S scenarios, we generate three independent scenario sets of size S. As a result, all reported results for each size category are averaged over fifteen tests. Similarly, for the IEEE 118-bus instance, three independent scenario sets are generated for each test, with S scenarios, yielding averaged results across the three tests.

To guarantee that Assumption 2 holds, we enforce a relatively complete recourse structure by introducing, in each instance, a load shedding unit and a renewable curtailment unit. Both are assigned a cost of 700\$/MWh, more than an order of magnitude higher than the maximum production cost of any other unit, ensuring they are used only as a last resort. Finally, all instances consider ramping constraints, capacity limits, and initial conditions as constraints on the power generation variables, and minimum uptime and downtime constraints on the commitment variables. Therefore, Assumption 3 holds.

These instances serve as our benchmark for evaluating the performance of our extended formulation and the effectiveness of the generated cuts. Our computational experiments are divided into three parts. In Sect. 5.1, we compare five approaches described in the previous sections: the *3-bin formulation* combined with either an extensive formulation (`3-bin extensive`) or BD (`3-bin BD`), and the extended formulation combined with either an extensive formulation (`Extended extensive`) or BD (`Extended BD-O` where the subproblems (16) defining the cut coefficients are solved at optimality and `Extended BD-H` where the heuristics described in Sect 4 is used). Sect. 5.2 presents an in-depth analysis of our approach `Extended BD-H` focusing on the computational time required for convergence under different parameter settings, such as varying the optimality gap. Finally, in Sect. 5.3, we explore the performance of our approach under alternative conditions, including the use of a different risk measure, such as average value-at-risk, and the multi-cut strategy. In all result tables, outcomes are reported in the format `a (b/c)`, where `b` indicates the number of instances successfully solved out of the `c` total instances. If `b` = 0, then `a` denotes the average optimality gap over the `c` instances. Otherwise, `a` represents the average solution time in seconds for the `b` successfully solved instances. In cases where the method fails to compute even an optimality gap, we denote $a = \times$.

Unless otherwise specified, all BD approaches are implemented using the single-cut strategy. For more information on single-cut and multi-cut strategies, we refer the reader to [24, 26]. Additionally, all BD approaches are implemented using a callback mechanism that enables solving the problem within a single search tree, with Benders cuts added when new incumbent solutions are found. This approach corresponds to the *Branch-and-Benders-cut* method, for which further details can be found in [19].

All the results have been obtained on an Intel(R) Xeon(R) CPU E5-2667 (3.30GHz) computer with 189 GigaBytes of RAM. The LP and MILP problems are solved using the 12.0.1 Gurobi optimization solver. The algorithms are coded with JuMP (v. 1.26.0, see [27]) in Julia (v. 1.11.5) and executed in a sequential setting, *i.e.,* using a single computational core without any parallelization.

5.1 Comparison between 3-bin and extended formulation

Table 1 presents the results of computational experiments conducted on the five approaches: `3-bin extensive`, `3-bin BD`, `Extended extensive`, `Extended BD-O`, and `Extended BD-H`. The performance of each method is evaluated across various instances, differing in the number of units and scenarios $|\mathcal{S}|$. All experiments were conducted with a fixed time limit of one hour and a target optimality gap of 0.1%. For each instance, the method achieving the best performance in terms of solving time is indicated in bold.

From Table 1, we observe that for small instances, particularly those with only one scenario, the extensive approaches `3-bin extensive` and `Extended extensive` perform best. Moreover, the `3-bin extensive` consistently outperforms `Extended extensive` on nearly all instances. This result is not surprising, as the two formulations differ only in how they model the minimum uptime constraints. Specifically, `3-bin extensive` uses the constraint set (1b), while `Extended extensive` models the same requirements using the interval-based γ variables and the linking constraints (10d). As a result, `Extended extensive` introduces significantly more variables while maintaining a similar number of constraints. However, these two extensive approaches struggle to scale as the number of scenarios increases. In contrast, the `Extended BD-H` approach exhibits computational times that are nearly independent of the number of scenarios, making it more effective for most instances, especially large-scale instances with more than 50 units or 50 scenarios. Using this approach, all instances are solved in under ten minutes, whereas the `3-bin extensive` fails to solve some of the largest instances in

Table 1 Solution approaches' comparison for a 0.1% gap with a time limit of 3600s, on SMS++ instances with various number of units and on the IEEE 118-bus instance with 54 units.

$\|\mathcal{J}\|$	$\|\mathcal{S}\|$	`3-bin BD`	`3-bin extensive`	`Extended extensive`	`Extended BD-O`	`Extended BD-H`
10	1	1201s (13/15)	**0s (15/15)**	0s (15/15)	18s (15/15)	2s (15/15)
10	25	2290s (1/15)	8s (15/15)	13s (15/15)	481s (15/15)	**5s (15/15)**
10	50	3208s (2/15)	22s (15/15)	32s (15/15)	1039s (15/15)	**7s (15/15)**
10	75	0.84% (0/15)	43s (15/15)	58s (15/15)	1467s (15/15)	**9s (15/15)**
10	100	1.06% (0/15)	69s (15/15)	96s (15/15)	2130s (15/15)	**13s (15/15)**
50	1	4.66% (0/15)	**7s (15/15)**	13s (15/15)	182s (15/15)	93s (15/15)
50	25	5.52% (0/15)	435s (15/15)	299s (15/15)	2903s (11/15)	**64s (15/15)**
50	50	4.88% (0/15)	1122s (15/15)	2343s (15/15)	3496s (1/15)	**69s (15/15)**
50	75	5.56% (0/15)	2402s (12/15)	2080s (2/15)	3.07% (0/15)	**115s (15/15)**
50	100	5.23% (0/15)	3075s (5/15)	2483s (6/15)	11.1% (0/15)	**109s (15/15)**
100	1	8.74% (0/15)	**7s (15/15)**	14s (15/15)	249s (15/15)	53s (15/15)
100	25	7.09% (0/15)	1758s (15/15)	1787s (15/15)	0.94% (0/15)	**89s (15/15)**
100	50	8.20% (0/15)	3033s (6/15)	99.01% (0/15)	35.69% (0/15)	**114s (15/15)**
100	75	7.01% (0/15)	92.65% (0/15)	99.09% (0/15)	61.61% (0/15)	**138s (15/15)**
100	100	7.47% (0/15)	99.24% (0/15)	99.24% (0/15)	84.35% (0/15)	**220s (15/15)**
54	1	45s (3/3)	**3s (3/3)**	8s (3/3)	82s (3/3)	8s (3/3)
54	25	6.34% (0/3)	207s (3/3)	190s (3/3)	1181s (3/3)	**104s (3/3)**
54	50	1.47% (0/3)	619s (3/3)	646s (3/3)	3114s (2/3)	**204s (3/3)**
54	75	4.98% (0/3)	1315s (3/3)	3058s (3/3)	30.31% (0/3)	**404s (3/3)**
54	100	1.92% (0/3)	2161s (3/3)	× (0/3)	39.31% (0/3)	**435s (3/3)**

one hour. Table 1 also highlights the crucial role of the heuristic developed in Sect. 4.2 in enhancing the performance of `Extended BD-O`. In `Extended BD-O`, the subproblems (19) are solved to optimality. However, since the number of these subproblems grows linearly with the number of scenarios and units and quadratically with the time horizon, this approach becomes computationally expensive as the number of scenarios increases. Consequently, the scalability of `Extended BD-O` is significantly limited compared to `Extended BD-H`. Finally, the `3-bin BD` method fails to converge within one hour, even for the smallest instances.

To further summarize and visualize the performance of the different solution methods, we employ the widely used performance profile tool introduced by [28]. These profiles show the proportion of instances solved (y-axis) by a method in less than K times the fastest method (x-axis, in log scale). Figure 2 displays two performance profiles: the left plot corresponds to instances with a single scenario (i.e., deterministic cases), while the right plot pertains to larger instances involving multiple scenarios. These profiles confirm that our `Extended BD-H` method significantly outperforms all the other methods for the larger instances. In particular, the second-best method, `3-bin extensive`, requires approximately 2^4 times longer to solve half of the instances and can solve only 75% of them in one hour.

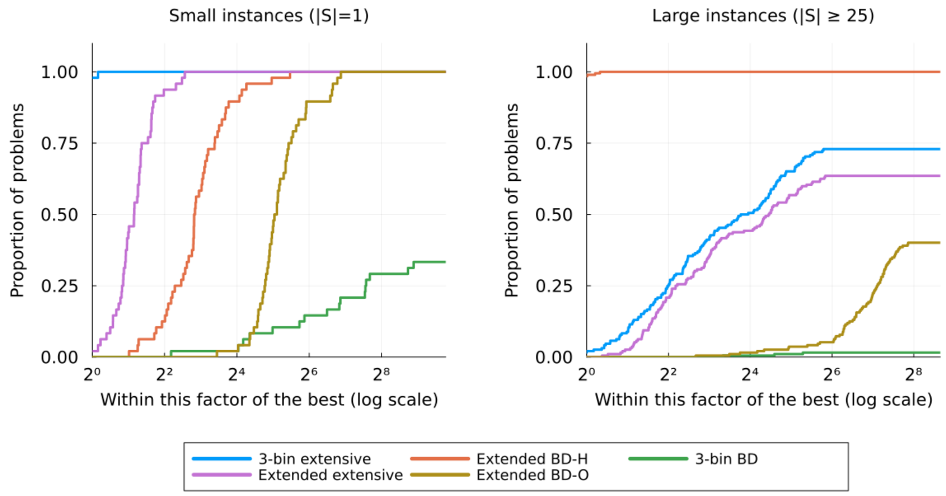

Fig. 2 Performance profiles comparing solution approaches on the IEEE 118-bus instance and SMS++ instances with 10, 50, and 100 thermal units.

In contrast, for single-scenario instances, the `3-bin extensive` method remains the most efficient. Specifically, `Extended extensive` requires roughly 2^2 times longer to solve all instances, while `Extended BD-H` and `Extended BD-O` need about 2^5 and 2^7 times longer, respectively.

5.2 Analysis of the computational efficiency of the extended Benders formulation

This section focuses on the two best approaches, `3-bin extensive` and `Extended BD-H`, and analyzes their performance with respect to the gap tolerance and the standard deviation of uncertainty in the instances. The analysis is presented in two tables. Table 2 first shows the computational times required to reach different gap tolerances (0.5%, 0.1%, and 0.01%), with a time limit of one hour and considering instances with 50 scenarios.

Table 2 shows that, contrary to `3bin extensive`, `Extended BD-H` quickly finds high-quality solutions with a 0.5% optimality gap, which may already be sufficient for decision-makers operating under uncertainty. Moreover, if a tighter gap is required, such as 0.01%, the method still achieves rapid convergence, further confirming the strength and effectiveness of the Benders' cuts of `Extended BD-H`. In particular, none of the instances with 100 units and a 0.01% optimality gap are solved by `3-bin extensive`, whereas `Extended BD-H` successfully solves all such instances, with an average solution time of less than five minutes.

The scenario generation methodology may vary depending on the case study. To evaluate the robustness of our method `Extended BD-H` with respect to the level of

Table 2 Computational performance for different gap tolerances using 50 scenarios.

$\lvert\mathcal{J}\rvert$	Gap (%)	`3-bin extensive`	`Extended BD-H`
10	0.5	15s (15/15)	**6s (15/15)**
10	0.1	22s (15/15)	**7s (15/15)**
10	0.01	31s (15/15)	**7s (15/15)**
50	0.5	614s (15/15)	**39s (15/15)**
50	0.1	1122s (15/15)	**69s (15/15)**
50	0.01	2680s (11/15)	**147s (15/15)**
100	0.5	2054s (12/15)	**78s (15/15)**
100	0.1	3033s (6/15)	**114s (15/15)**
100	0.01	33.32% (0/15)	**363s (15/15)**
54	0.5	615s (3/3)	**142s (3/3)**
54	0.1	619s (3/3)	**204s (3/3)**
54	0.01	2280s (2/3)	**232s (3/3)**

uncertainty, we introduce a scaling parameter σ, which multiplies the standard deviation and thereby controls the dispersion of the scenarios. For example, setting $\sigma = 0$ results in identical scenarios, eliminating uncertainty. Additional details are provided in Appendix D. In particular, all the other results reported in this Sect. 5 use $\sigma = 1$.

Table 3 presents the computational times (in seconds) required to solve instances with 50 scenarios for various values of the σ parameter, using a gap tolerance of 0.1%. The results on the SMS++ instances indicate that the computational times of `Extended BD-H` are not impacted by the increased level of uncertainty introduced by the σ factor. In contrast, the `3-bin extensive` approach appears to benefit from higher uncertainty levels. However, it still requires roughly ten times longer to solve the instances compared to `Extended BD-H`. For the IEEE 118-bus instances, however, both methods fail to solve any instance when $\sigma > 2$. A likely explanation lies in the scenario generation process: our modified IEEE 118-bus model includes ten wind farms distributed across the network, and increasing σ effectively increases the generation capacity of these wind farms. This may cause system destabilization, leading to numerical difficulties and poor performance for both `3-bin extensive` and `Extended BD-H` on these instances.

5.3 Comparison AVaR and single or multi-cuts strategies

In this section, we investigate two additional aspects of interest. First, since all previous results were obtained using a single-cut strategy, we compare its performance (number of iterations and total running time) with a multi-cut strategy in Table 4. Second, Table 5 shows that our method remains effective when applied to alternative risk measures, such as the Average Value-at-Risk (AVaR).

The results reported in Table 4 show that the computational time of the multi-cut strategy is significantly longer on the SMS++ instances, particularly as the number of scenarios increases. This can be attributed to the structure of the new cuts (23e) added at each iteration: they are dense and involve all first stage variables γ, each

Table 3 Impact of increasing uncertainty volatility on running time (in seconds), using 50 scenarios.

$\mathcal{J}$	σ	`3bin extensive`	`Extended BD-H`
10	0.5	30s (15/15)	**7s (15/15)**
10	1	22s (15/15)	**7s (15/15)**
10	2	16s (15/15)	**6s (15/15)**
10	3	20s (15/15)	**8s (15/15)**
50	0.5	1402s (15/15)	**92s (15/15)**
50	1	1122s (15/15)	**69s (15/15)**
50	2	696s (15/15)	**65s (15/15)**
50	3	615s (15/15)	**59s (15/15)**
100	0.5	3177s (5/15)	**160s (15/15)**
100	1	3033s (6/15)	**114s (15/15)**
100	2	2486s (13/15)	**108s (15/15)**
100	3	2137s (9/15)	**112s (15/15)**
54	0.5	1294s (3/3)	**156s (3/3)**
54	1	619s (3/3)	**204s (3/3)**
54	2	13.99% (0/3)	4.44% (0/3)
54	3	× (0/3)	1.56% (0/3)

with different coefficients. While adding a single constraint per iteration (as in the single-cut strategy) remains computationally manageable, the multi-cut approach adds many such constraints at once. This results in a master problem with a constraint matrix containing a large number of non-zero elements, which becomes increasingly challenging to solve as the algorithm progresses. Note that this conclusion does not hold for the IEEE 118-bus instance, as it includes network constraints, which make the subproblems longer to solve. As a result, the multi-cut approach is more effective in this case, since it typically leads to convergence with fewer nodes explored in the Branch-and-Benders-cut algorithm.

Table 5 presents the performance of our new approach, `Extended BD-H`, alongside the `3-bin extensive` method, under the AVaR risk measure with a confidence level of 0.8. Before analyzing the results, we recall the formulation of the stochastic UC problem with AVaR, as defined in problem (35), for a general confidence level α:

$$\min_{\gamma,x,z,\theta} \quad \theta + \frac{z}{1-\alpha} + \sum_{j\in\mathcal{J},\, t\in[0:T]} C_{j,t}^\top x_{j,t} \tag{35a}$$

$$\text{s.t.} \quad \sum_{s\in\mathcal{S}} \pi_s \max(0, Q(x,\xi_s) - \theta) \le z \tag{35b}$$

$$\sum_{[a:b]\in\mathcal{T}_j \,|\, t\in[a:b]} \gamma_{j,a,b} \le 1 \qquad \forall j \in \mathcal{J}, \quad \forall t \in [0:T] \tag{35c}$$

$$x_j = \sum_{[a:b]\in\mathcal{T}_j} \gamma_{j,a,b} x_{j,a,b} \qquad \forall j \in \mathcal{J} \tag{35d}$$

$$x_j \in X_j \qquad \forall j \in \mathcal{J} \tag{35e}$$

Table 4 Comparison Single-Cut and Multi-Cut

$\mathcal{J}$	$\mathcal{S}$	`Extended BD-H` Single-Cut	`Extended BD-H` Multi-Cut
10	25	**5s (15/15)**	13s (15/15)
10	50	**7s (15/15)**	24s (15/15)
10	75	**9s (15/15)**	41s (15/15)
10	100	**13s (15/15)**	55s (15/15)
50	25	**64s (15/15)**	154s (15/15)
50	50	**69s (15/15)**	246s (15/15)
50	75	**115s (15/15)**	413s (15/15)
50	100	**109s (15/15)**	399s (15/15)
100	25	**89s (15/15)**	233s (15/15)
100	50	**114s (15/15)**	393s (15/15)
100	75	**138s (15/15)**	495s (15/15)
100	100	**220s (15/15)**	614s (15/15)
54	25	104s (3/3)	**49s (3/3)**
54	50	204s (3/3)	**246s (3/3)**
54	75	404s (3/3)	**308s (3/3)**
54	100	435s (3/3)	**326s (3/3)**

$$\gamma_{j,a,b} \in \mathbb{R}_+, x_j \in \{0,1\}^{3(T+1)}, z \in \mathbb{R}_+. \tag{35f}$$

To formulate the feasibility and optimality cuts corresponding to the AVaR single-cut model, we define the set of scenarios $\mathcal{S}(x) \subseteq \mathcal{S}$ that contribute to the AVaR tail:

$$\mathcal{S}(x) = \{s \in \mathcal{S} \mid Q(x,\xi_s) - \theta \geq 0\}. \tag{36}$$

Using this set, constraint (35b) can equivalently be reformulated as:

$$z \geq \sum_{s\in\mathcal{S}(x)} \pi_s(Q(x,\xi_s) - \theta), \tag{37}$$

Moreover, for any pair of first stage decisions x and y, it holds that:

$$\sum_{s\in\mathcal{S}(x)} \pi_s(Q(x,\xi_s) - \theta) \geq \sum_{s\in\mathcal{S}(y)} \pi_s(Q(x,\xi_s) - \theta). \tag{38}$$

As a result, at each iteration k of the extended Benders' decomposition, given the first stage decisions (x^k, γ^k), we add the following optimality cut to the master problem if all subproblems are feasible:

$$\sum_{s\in\mathcal{S}(x^k)} \pi_s \left(g(\hat{\nu}_s^k, \hat{\lambda}_s^k, \xi_s) + \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\hat{Q}_{j,a,b}(c_j - \hat{\nu}_{s,j_n}^k) - \theta \right) \leq z, \tag{39}$$

where $(\hat{\nu}_s^k, \hat{\lambda}_s^k)$ are the dual variables obtained by solving the subproblem for each scenario s with the first stage decisions x^k. If there exists a scenario $s \in \mathcal{S}$ for which the subproblem is infeasible, we instead add the same feasibility cut as in the risk-neutral setting:

$$g(\hat{\nu}_s^k, \hat{\lambda}_s^k, \xi_s) + \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\hat{Q}_{j,a,b}(-\hat{\nu}_{s,j_n}^k) \leq 0. \tag{40}$$

Finally, it is important to note that the efficiency enhancements proposed in Sect. 4 remain fully compatible with the AVaR formulation and are also applied in this context.

Accordingly, Table 5 reports the performance of the `3-bin extensive` and `Extended BD-H` methods under the AVaR risk measure, following the same format as Table 1, which presents results for the risk-neutral case.

Table 5 Performance with AVaR risk measure

$\lvert\mathcal{J}\rvert$	$\lvert\mathcal{S}\rvert$	`3-bin extensive`	`Extended BD-H`
10	25	9s (15/15)	**7s (15/15)**
10	50	25s (15/15)	**12s (15/15)**
10	75	50s (15/15)	**16s (15/15)**
10	100	69s (15/15)	**18s (15/15)**
50	25	740s (15/15)	**189s (15/15)**
50	50	1960s <u>(13/15)</u>	**197s (15/15)**
50	75	2992s <u>(6/15)</u>	**252s (15/15)**
50	100	2830s <u>(3/15)</u>	**187s (15/15)**
100	25	2188s <u>(14/15)</u>	**375s (15/15)**
100	50	99.7% <u>(0/15)</u>	**437s (15/15)**
100	75	99.7% <u>(0/15)</u>	**601s (15/15)**
100	100	99.72% <u>(0/15)</u>	**645s (15/15)**
54	25	424s (3/3)	**352s (3/3)**
54	50	1139s (3/3)	**408s (3/3)**
54	75	2005s (3/3)	**726s (3/3)**
54	100	2524s (3/3)	**777s (3/3)**

First, we observe that for both approaches, solving the AVaR problem takes slightly longer than solving the risk-neutral counterpart, whose results are reported in Table 1. Second, our approach, `Extended BD-H`, remains more efficient than `3-bin extensive` for tests involving more than 25 scenarios.

6 Conclusion

We have presented a new Benders decomposition method for solving the stochastic unit commitment (SUC) problem, leveraging an extended formulation based on interval variables. This approach yields provably stronger cuts than the classical 3-bin-based formulation and enables decomposition across both units and time intervals under mild assumptions. To address the computational cost of solving the many small subproblems, we have developed a heuristic based on weak duality that computes valid cut coefficients at a fraction of the cost.

Extensive experiments on both standard and large-scale instances demonstrate that the proposed method consistently outperforms existing approaches. In particular, it solves SUC instances with 100 units and 100 scenarios in under four minutes on average—while baseline methods fail to close the optimality gap within one hour.

These results highlight the practical potential of exploiting extended formulations in Benders decomposition. Future work includes relaxing the decomposition assumption (3), which requires that technical constraints associated with two non-overlapping intervals be independent. This would allow for additional features such as fuel constraints. One direction is to develop new tailored heuristics to accommodate these constraints; alternatively, they can be dualized and handled similarly to the linking constraints discussed in this paper, at the cost of introducing additional dual variables.

Declarations

Funding This research benefited from the support of the FMJH Program Gaspard Monge for optimization and operations research and their interactions with data science.
Conflict of interest The authors declare that they have no conflict of interest.
Data and Code Availability The data and code used in this paper are available at: https://github.com/MathisAzema/BDSUC.

Appendix A Proof of Proposition 1

Let (γ, x) a solution satisfying constraints (10c)-(10f) and $\xi \in \mathbb{R}^{T+1}$. We aim to show that $Q(x, \xi) = \mathcal{Q}(\gamma, \xi)$, even in the absence of relatively complete recourse, where the value may be $+\infty$. To this end, we first prove that for any feasible solution p of problem (3), there exists a corresponding feasible solution $\tilde{p}$ for problem (14) with the same objective value.

We define $\tilde{p}$ as follows:

$$\tilde{p}_{j,a,b,t} = \begin{cases} p_{j,t} x_{j,a,b,t} & \text{if } \gamma_{j,a,b} = 1 \\ q_{j,a,b,t} & \text{otherwise.} \end{cases} \tag{A1}$$

where $q_{j,a,b}$ is a solution satisfying the technical constraints of unit j over the interval $[a : b]$. Such a solution necessarily exists, otherwise, it would be meaningless to include the interval $[a : b]$ in the set of feasible intervals $\mathcal{T}_j$.

Moreover, by Assumption 3, the vector $(p_{j,t} x_{j,a,b,t})_{t \in [0:T]}$ is feasible for problem (14). Then, the vector $\tilde{p}$ is feasible for problem (14), and its objective value coincides with that of p in problem (3).

Conversely, given a feasible solution $\tilde{p}$ of problem (14), we define p as:

$$p_j = \sum_{[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\tilde{p}_{j,a,b}. \tag{A2}$$

By aggregating the constraints (14a) satisfied by $\tilde{p}$, we observe that p is feasible for problem (3), and that the objective value of p in problem (3) is equal to that of $\tilde{p}$ in problem (14).

Therefore, the feasibility of one problem implies the feasibility of the other, and the optimal values are identical. This establishes that $Q(x,\xi) = \mathcal{Q}(\gamma,\xi)$ for any (γ, x) a solution satisfying constraints (10c)-(10f), even in the absence of relatively complete recourse.

Appendix B Proof of Proposition 3

Let (γ, x) be a vector satisfying constraints (10c)-(10f), $(\nu,\lambda) \in \mathcal{D}_1$, and $\xi \in \mathbb{R}^{T+1}$. Let μ^* be an optimal solution to the following problem:

$$\max_{\mu} \quad \sum_{j\in\mathcal{J}} \mu_j^\top W_j x_j \tag{B3a}$$

$$\text{s.t.} \quad A_j^\top \mu_j + \nu_{j_n} = c_j \quad \forall j \in \mathcal{J} \tag{B3b}$$

$$\mu_j \in \mathbb{R}^q_+. \tag{B3c}$$

We denote by $v^{\text{3bin}}(x,\xi,\mu^*,\nu,\lambda)$ the objective value of the solution (μ^*,ν,λ) in problem (4), and by $v^{\text{ext}}(\gamma,\xi,\nu,\lambda)$ the objective value of the solution (ν,λ) in problem (18). By definition of these objective values, the following equality holds:

$$v^{\text{3bin}}(x,\xi,\mu^*,\nu,\lambda) - v^{\text{ext}}(\gamma,\xi,\nu,\lambda) = \sum_{j\in\mathcal{J}} (\mu_j^*)^\top W_j x_j - \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\hat{Q}_{j,a,b}(c_j - \nu_{j_n}). \tag{B4}$$

To prove this difference is zero and establish Proposition 3, we consider the dual formulation of problem (B3). By strong duality (Assumption 2), the value $\sum_{j\in\mathcal{J}}(\mu_j^*)^\top W_j x_j$ is equal to the optimal value of the following primal problem:

$$\min_{p} \quad \sum_{j\in\mathcal{J}} p_j^\top (c_j - \hat{\nu}_{j_n}) \tag{B5a}$$

$$\text{s.t.} \quad W_j x_j \leq A_j p_j \quad \forall j \in \mathcal{J} \tag{B5b}$$

$$p_j \in \mathbb{R}^{T+1}. \tag{B5c}$$

Indeed, under Assumption 2, either the primal problem (3) or the dual problem (4) admits a feasible solution. Since the feasible sets of problems (B3) and (B5) are subsets of those of problems (3) and (4), respectively, this guarantees that at least one of the problems (B3) or (B5) is also feasible.

On the other hand, the term $\sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j}\gamma_{j,a,b}\hat{Q}_{j,a,b}(c_j-\nu_{j_n})$ corresponds to the optimal value of the following extended primal problem:

$$\min_{\tilde{p}} \quad \sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j} \gamma_{j,a,b}\,\tilde{p}_{j,a,b}^{\top}(c_j-\hat{\nu}_{j_n}) \tag{B6a}$$

$$\text{s.t.} \quad W_j x_{j,a,b} \leq A_j \tilde{p}_{j,a,b} \qquad \forall j\in\mathcal{J},\,\forall[a:b]\in\mathcal{T}_j \tag{B6b}$$

$$\tilde{p}_{j,a,b}\in\mathbb{R}^{T+1}. \tag{B6c}$$

The equivalence between problems (B5) and (B6) follows from arguments similar to those presented in the proof of Proposition 1. Therefore, the two objective values are equal, which completes the proof.

Appendix C Feasibility Cuts

Let (γ,x) be a solution satisfying constraints (10c)-(10f), and let $\xi\in\mathbb{R}^{T+1}$ be a realization of uncertainty. We denote by $v^{\text{3bin}}(x,\xi,\mu,\nu,\lambda)$ the objective value of a solution (μ,ν,λ) in problem (4), and by $v^{\text{ext}}(\gamma,\xi,\nu,\lambda)$ the objective value of the solution (ν,λ) in problem (18). In this appendix, we prove that problems (4) and (18) are unbounded if and only if:

$$\exists f\in\mathcal{F},\quad v^{\infty}(\gamma,\xi,\nu^f,\lambda^f)=g(\xi,\nu^f,\lambda^f)+\sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j}\gamma_{j,a,b}\hat{Q}_{j,a,b}(-\nu^f_{j_n})>0. \tag{C7}$$

Since problem (4) is linear, it is unbounded if and only if:

$$\exists f\in\mathcal{F},\quad v^{3bin}(x,\xi,\mu^f,\nu^f,\lambda^f)>0. \tag{C8}$$

Let consider an extreme ray (μ^f,ν^f,λ^f) of $\mathcal{D}$. For any $(\hat{\mu},\hat{\nu},\hat{\lambda})\in\mathcal{D}$ and any $N\in\mathbb{N}$, the point $(\hat{\mu}+N\mu^f,\hat{\nu}+N\nu^f,\hat{\lambda}+N\lambda^f)\in\mathcal{D}$. Consequently, $(\hat{\nu}+N\nu^f,\hat{\lambda}+N\lambda^f)\in\mathcal{D}_1$ for all $N\in\mathbb{N}$.

To prove condition (C7), we study the limit of:

$$\frac{1}{N}v^{\text{ext}}(\gamma,\xi,\hat{\nu}+N\nu^f,\hat{\lambda}+N\lambda^f) \tag{C9}$$

$$=\frac{1}{N}g(\xi,\hat{\nu}+N\nu^f,\hat{\lambda}+N\lambda^f)+\sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j}\gamma_{j,a,b}\cdot\frac{1}{N}\hat{Q}_{j,a,b}(c_j-\hat{\nu}_{j_n}+N\nu^f_{j_n}). \tag{C10}$$

Let us first consider the limit of the first term, for which, by the linearity of g, we have:

$$\frac{1}{N}g(\xi,\hat{\nu}+N\nu^f,\hat{\lambda}+N\lambda^f)=\frac{1}{N}g(\xi,\hat{\nu},\hat{\lambda})+g(\xi,\nu^f,\lambda^f)\xrightarrow[N\to\infty]{}g(\xi,\nu^f,\lambda^f). \quad \text{(C11)}$$

We now analyze the limit of $\frac{1}{N}\hat{Q}_{j,a,b}(c_j-\hat{\nu}_{j_n}+N\nu^f_{j_n})$, which is the optimal value of problem (C12):

$$\min_{p\in\mathbb{R}^{T+1}} \quad \left(\frac{c_j-\hat{\nu}}{N}-\nu^f_{j_n}\right)^\top p \quad \text{(C12a)}$$

$$\text{s.t.} \quad W_j x_{j,a,b}\leq A_j p. \quad \text{(C12b)}$$

The feasible set is independent of N, and the objective coefficients converge to $-\nu^f_{j_n}$. Thus:

$$\frac{1}{N}\hat{Q}_{j,a,b}(c_j-\hat{\nu}_{j_n}+N\nu^f_{j_n})\xrightarrow[N\to\infty]{}\hat{Q}_{j,a,b}(-\nu^f_{j_n}), \quad \text{(C13)}$$

It follows that the sequence $\left(\frac{1}{N}v^{\text{ext}}(\gamma,\xi,\hat{\nu}+N\nu^f,\hat{\lambda}+N\lambda^f)\right)_{N\in\mathbb{N}}$ converges to the limit:

$$v^\infty(\xi,\nu^f,\lambda^f)=g(\xi,\nu^f,\lambda^f)+\sum_{j\in\mathcal{J},[a:b]\in\mathcal{T}_j}\gamma_{j,a,b}\hat{Q}^\infty_{j,a,b}(\nu^f). \quad \text{(C14)}$$

Therefore, we proved that if:

$$\exists f\in\mathcal{F},\quad v^\infty(\gamma,\xi,\nu^f,\lambda^f)>0, \quad \text{(C15)}$$

then the subproblem (18) is unbounded.

To prove the converse, we assume that:

$$\forall f\in\mathcal{F},\quad v^\infty(\gamma,\xi,\nu^f,\lambda^f)\leq 0, \quad \text{(C16)}$$

and aim to show that problem (18) is bounded. By Proposition (1), it suffices to prove that problem (4) is bounded, *i.e.*,

$$\forall f\in\mathcal{F},\quad v^{3bin}(\gamma,\xi,\nu^f,\lambda^f)\leq 0. \quad \text{(C17)}$$

Let (μ^f,ν^f,λ^f) an extreme ray of $\mathcal{D}$, the feasible set of problem (4) and let μ^f_N be an optimal solution to:

$$\max_{\mu} \quad \sum_{j\in\mathcal{J}}\mu_j^\top W_j x_j \quad \text{(C18a)}$$

$$\text{s.t.} \quad A_j^\top\mu_j+\hat{\nu}+N\nu^f=c_{j_n},\quad \forall j\in\mathcal{J}, \quad \text{(C18b)}$$

$$\mu_j\in\mathbb{R}^q_+. \quad \text{(C18c)}$$

By Proposition 3, it follows that:

$$v^{\text{ext}}(\gamma, \xi, \hat{\nu} + N\nu^f, \hat{\lambda} + N\lambda^f) = v^{3\text{bin}}(x, \xi, \mu^f_N, \hat{\nu} + N\nu^f, \hat{\lambda} + N\lambda^f). \tag{C19}$$

Moreover, since $\hat{\mu} + N\mu^f$ is a feasible solution of (C18), and $v^{3\text{bin}}$ is linear in its arguments, we have:

$$v^{3\text{bin}}(x, \xi, \mu^f_N, \hat{\nu} + N\nu^f, \hat{\lambda} + N\lambda^f) \geq v^{3\text{bin}}(x, \xi, \hat{\mu} + N\mu^f, \hat{\nu} + N\nu^f, \hat{\lambda} + N\lambda^f) \tag{C20a}$$

$$= v^{3\text{bin}}(x, \xi, \hat{\mu}, \hat{\nu}, \hat{\lambda}) + N \cdot v^{3\text{bin}}(x, \xi, \mu^f, \nu^f, \lambda^f). \tag{C20b}$$

Combining (C19) and (C20), dividing by N, and taking the limit as $N \to \infty$, we obtain:

$$v^{3\text{bin}}(x, \xi, \mu^f, \nu^f, \lambda^f) \leq v^{\infty}(\gamma, \xi, \nu^f, \lambda^f), \tag{C21}$$

which implies that constraint (C16) dominates constraint (C17). As a result, problem (4) is bounded, and by Proposition 1, so is problem (18).

Appendix D Scenario Generation Methodology

Both the SMS++ instances and the IEEE 118-bus test case originally include only the predicted hourly demand $\hat{\xi}_{n,t}$ at each bus n over a 24-hour horizon. To evaluate the performance of the stochastic methods, we generate net demand scenarios using two different approaches: one for the SMS++ instances and another for the IEEE 118-bus instance. First, for the SMS++ instances, which involve only a single bus (and hence a single net load), we use the publicly available dataset from https://data.open-power-system-data.org/time_series/, which provides both day-ahead forecasts and actual hourly electricity demand in France.

For each day j in the dataset, we compute the relative forecast error $\varepsilon^j = (\varepsilon^j_t)_{t\in[0:T]}$ as defined in Equation (D22), where $d^j_{\text{actual},t}$ and $d^j_{\text{forecast},t}$ represent the actual and forecasted demand at time t, respectively:

$$\varepsilon^j_t = \frac{d^j_{\text{actual},t} - d^j_{\text{forecast},t}}{d^j_{\text{forecast},t}}, \qquad d^j_{\text{actual},t} = d^j_{\text{forecast},t}(1 + \varepsilon^j_t). \tag{D22}$$

Using these relative errors, we construct scenario j by modifying the predicted demand $\hat{\xi}_{1,t}$ according to:

$$\xi^j_{1,t} = \hat{\xi}_{1,t}\left(1 + \sigma \cdot \varepsilon^j_t\right), \tag{D23}$$

where the parameter σ allows us to scale the variability of the scenarios. All the scenarios used in this paper were randomly sampled from the period between 2019 and 2024.

Second, for the IEEE 118-bus instance, we use wind generation data from [29], *i.e.,* we consider ten buses with wind generation. We denote by $w_{n,t}^j$ the wind generation at bus n and time t in scenario j. For all buses except these ten, $w_{n,t}^j = 0$.

Accordingly, the net demand ξ^j for scenario j is defined as:

$$\xi_{n,t}^j = \hat{\xi}_{n,t} - \sigma w_{n,t}^j \quad \forall n \in \mathcal{N}, \forall t \in [0:T], \tag{D24}$$

where σ is still a scaling parameter that controls the level of variability in the scenarios and can be interpreted as the degree of wind power integration into the system. The source of data for the wind power generation is the Global Energy Competition (GEFCom) 2014 [30].

In Table 3, we report the impact of different values of the σ parameter on the performance of the solution methods. Unless stated otherwise, all results in the paper use $\sigma = 1$.

References

[1] Anjos, M.F., Conejo, A.J., *et al.*: Unit commitment in electric energy systems. Foundations and Trends in Electric Energy Systems **1**(4), 220–310 (2017)

[2] van Ackooij, W., Danti Lopez, I., Frangioni, A., Lacalandra, F., Tahanan, M.: Large-scale unit commitment under uncertainty: an updated literature survey. Annals of Operations Research **271**(1), 11–85 (2018)

[3] Montero, L., Bello, A., Reneses, J.: A review on the unit commitment problem: Approaches, techniques, and resolution methods. Energies **15**(4), 1296 (2022)

[4] Geng, Z., Conejo, A.J., Chen, Q., Kang, C.: Power generation scheduling considering stochastic emission limits. International Journal of Electrical Power & Energy Systems **95**, 374–383 (2018)

[5] van Ackooij, W., Malick, J.: Decomposition algorithm for large-scale two-stage unit-commitment. Annals of Operations Research **238**(1), 587–613 (2016)

[6] Papavasiliou, A., Oren, S.S., Rountree, B.: Applying high performance computing to transmission-constrained stochastic unit commitment for renewable energy integration. IEEE Transactions on Power Systems **30**(3), 1109–1120 (2014)

[7] Scuzziato, M.R., Finardi, E.C., Frangioni, A.: Comparing spatial and scenario decomposition for stochastic hydrothermal unit commitment problems. IEEE Transactions on Sustainable Energy **9**(3), 1307–1317 (2017)

[8] Geoffrion, A.M.: Lagrangean Relaxation for Integer Programming, pp. 82–114. Springer, Berlin, Heidelberg (1974)

[9] Lemaréchal, C., Renaud, A.: A geometric study of duality gaps, with applications. Mathematical Programming **90**, 399–427 (2001)

[10] Borghetti, A., Frangioni, A., Lacalandra, F., Nucci, C.A.: Lagrangian heuristics based on disaggregated bundle methods for hydrothermal unit commitment. IEEE Transactions on Power Systems **18**(1), 313–323 (2003)

[11] Dubost, L., Gonzalez, R., Lemaréchal, C.: A primal-proximal heuristic applied to the french unit-commitment problem. Mathematical Programming **104**, 129–151 (2005)

[12] Van Slyke, R.M., Wets, R.: L-shaped linear programs with applications to optimal control and stochastic programming. SIAM Journal on Applied Mathematics **17**(4), 638–663 (1969)

[13] Vatanpour, M., Yazdankhah, A.S.: The impact of energy storage modeling in coordination with wind farm and thermal units on security and reliability in a stochastic unit commitment. Energy **162**, 476–490 (2018)

[14] Huang, Y., Zheng, Q.P., Wang, J.: Two-stage stochastic unit commitment model including non-generation resources with conditional value-at-risk constraints. Electric Power Systems Research **116**, 427–438 (2014)

[15] Magnanti, T.L., Wong, R.T.: Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. Operations Research **29**(3), 464–484 (1981)

[16] Zaourar, S.: Optimisation convexe non-différentiable et méthodes de décomposition en recherche opérationnelle. PhD thesis, University of Grenoble (November 2014)

[17] Zakeri, G., Philpott, A.B., Ryan, D.M.: Inexact cuts in benders decomposition. SIAM Journal on Optimization **10**(3), 643–657 (2000)

[18] Oliveira, W., Sagastizábal, C.: Level bundle methods for oracles with on-demand accuracy. Optimization Methods and Software **29**(6), 1180–1209 (2014)

[19] Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W.: The benders decomposition algorithm: A literature review. European Journal of Operational Research **259**(3), 801–817 (2017)

[20] Knueven, B., Ostrowski, J., Wang, J.: The ramping polytope and cut generation for the unit commitment problem. INFORMS Journal on Computing **30**(4), 739–749 (2018)

[21] Knueven, B., Ostrowski, J., Watson, J.-P.: A novel matching formulation for startup costs in unit commitment. Mathematical Programming Computation **12**(2), 225–248 (2020)

[22] Lu, Y., Zhang, T., Guan, Y.: An extended formulation for two-stage stochastic unit commitment with reserves. Operations Research Letters **50**(3), 235–240 (2022)

[23] Zheng, Q.P., Wang, J., Pardalos, P.M., Guan, Y.: A decomposition approach to the two-stage stochastic unit commitment problem. Annals of Operations Research **210**, 387–410 (2013)

[24] Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer, New York (1997)

[25] Frangioni, A., Gentile, C., Lacalandra, F.: Solving unit commitment problems with general ramp constraints. International Journal of Electrical Power & Energy Systems **30**(5), 316–326 (2008)

[26] Birge, J.R., Louveaux, F.V.: A multicut algorithm for two-stage stochastic linear programs. European Journal of Operational Research **34**(3), 384–392 (1988)

[27] Lubin, M., Dowson, O., Dias Garcia, J., Huchette, J., Legat, B., Vielma, J.P.: JuMP 1.0: Recent improvements to a modeling language for mathematical optimization. Mathematical Programming Computation **15**(3), 581–589 (2023)

[28] Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Mathematical Programming **91**, 201–213 (2002)

[29] Gamboa, C.A., Valladao, D.M., Street, A., Homem-de-Mello, T.: Decomposition methods for wasserstein-based data-driven distributionally robust problems. Operations Research Letters **49**(5), 696–702 (2021)

[30] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J.: Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. International Journal of Forecasting **32**(3), 896–913 (2016)