Modeling and solving a stochastic generation and transmission expansion planning problem with a “Loss Of Load Expectation” reliability criterion

Xavier Blanchot, François Clautiaux, Aurélien Froger, Manuel Ruiz

Modeling and solving a stochastic generation and transmission expansion planning problem with a "Loss Of Load Expectation" reliability criterion

Xavier Blanchot[1,2], François Clautiaux[1], Aurélien Froger[1], Manuel Ruiz[2]

[1] Université de Bordeaux, CNRS, INRIA, Bordeaux INP, IMB, UMR 5251, Talence, France
[2] RTE, Paris La Défense, France

Abstract

In this paper, we study how a regulatory constraint limiting a measure of unserved demand, called *Loss Of Load Expectation* (LOLE), can be incorporated into a strategic version of a stochastic generation and transmission expansion planning problem. This problem is tackled by the French Transmission System Operator RTE for producing prospective reports on the evolution of the electricity network. We show that a direct inclusion of the constraint into the extensive form of the two-stage stochastic problem leads to a formulation that violates the time-consistency principle. To obtain a valid model, we use bilevel programming and introduce a formulation of the problem in which the leader and follower have the same objective function. To solve this formulation, we propose a matheuristic that embeds a Benders decomposition algorithm in a binary search on the total investment cost. We performed computational experiments to study the practical difficulty of the problem and validate the proposed solution method. Our experiments show that solving the single-level reformulation of the problem obtained using the KKT complementary conditions is intractable in practice, even for small size instances, and that a simple heuristic procedure is not sufficient to compute feasible solutions for all test cases. This is not the case for our matheuristic, which finds a feasible solutions for all instances of our test bed.

Keywords— Stochastic optimization, Mixed integer linear programming, Bilevel programming, Benders decomposition, Matheuristic, Expansion planning

1 Introduction

RTE, France's Transmission System Operator, is a public utility in charge of the high and ultra-high voltage electricity transmission network in France and of the electricity exchanges with the neighbouring countries. Its main role is to guarantee in real time the balance between electricity production and consumption.

RTE produces prospective reports on the evolution of the electricity network to horizons from five to thirty years (e.g., RTE (2019, 2021)). These reports aim to measure the risk of imbalance between electricity production and demand, and to analyze investment strategies on the network in order to ensure the security of supply depending on different technical, economical, environmental and political contexts referred to as *systemic scenarios*. For each systemic scenario, an independent prospective study is performed to assess the overall adequacy of the electricity system to supply the projected electricity demand. More specifically, the objective is to make an approximate assessment of the location of cost-effective investments for new generation units, and most of all for the reinforcement of the electricity transmission network to satisfy the projected electricity demand in the distant future of 10 to 30 years. Potential weaknesses identified in the electrical system can then be addressed proactively.

In its prospective study, RTE considers an integrated (centralized) planner approach, meaning that generation and transmission capacity expansion decisions are simultaneously made, and does not rely on a precise physical modeling of fine-grained operational constraints. It formulates the generation and transmission expansion problem as a two-stage stochastic problem, where discrete scenarios are used to model the uncertain parameters (e.g., demand, resource availability). The first stage problem is an investment problem, while the second stage deals with aggregated operational decisions and constraints. An important feature of the second stage problem is that unserved demand incurs a cost penalty, set equal to the so-called *value of lost load* (VOLL), that is "*an estimation in €/MWh, of the maximum electricity price that customers are willing to pay to avoid an outage*"[1]. Using VOLL when taking generation and transmission capacity expansion decisions is aimed at finding a compromise between a guaranteed electricity demand satisfaction at all times and a sustainable investment cost.

Additionally, RTE takes into account a so-called reliability constraint imposed by regulation. Specifically, European regulation[1] imposes each member state to define a necessary level of security of supply referred to as a reliability standard. Using a probabilistic approach that takes into account appropriate scenarios of projected demand and supply, the reliability standard is expressed as a limit on the loss of load expectation (LOLE) which is "*the expected number of hours, in a given geographic area and in a given time period, during which capacity resources are insufficient to meet the demand and hence positive energy not served (ENS) occurs*"[2]. In France, the number of hours with ENS in a year should be in expectation less than or equal to 3 hours[3]. In other European countries, identical or close values have been defined[4]. These values are used in particular to perform capacity adequacy assessment in prospective studies. Note that a positive LOLE does not necessarily translate into blackouts as there exist mechanisms to lower the impacts on consumers. Hereafter, we refer to a constraint that imposes an upper bound on the LOLE as a *LOLE reliability criterion*.

In its prospective study to determine capacity expansion decisions, RTE adopts a two-phase approach. In a first phase, it solves an expansion planning problem by ignoring the LOLE reliability criterion. In a second phase, RTE relies on expert knowledge to obtain a solution that satisfies the criterion.

In this work, we study the integration of the LOLE reliability criterion into the generation and transmission expansion planning problem tackled by RTE. Compared to the existing approach, we propose a single-phase approach, meaning that investment decisions must be made by anticipating the outcome of the aggregated operational decisions to ensure that the criterion is met.

The contribution of this work is twofold. The first and foremost contribution is the mathematical modelling of the stochastic generation and transmission expansion planning problem with a LOLE reliability criterion. We show that directly introducing the new constraint into the extensive form of the two-stage stochastic program is time-inconsistent. Specifically, second-stage decisions taken when evaluating a first-stage decision may be suboptimal in the sense that they would not be implemented if the uncertainty were to revealed, because at that time it is possible to take strictly better second-stage decisions. This is because the LOLE reliability criterion couples the decisions made in the different scenarios. We show that the cost difference induced by the evaluation of first-stage decisions based on suboptimal second-stage decisions is unbounded. To ensure time-consistency, we formulate the problem with the LOLE reliability criterion as a bilevel problem. We are not aware of any work on stochastic generation and transmission expansion planning problem that highlights this possible inconsistency in mathematical models and uses bilevel programming to overcome it. The second contribution is a practical method for producing (bilevel) feasible solutions to the problem. The method is iterative and based on a Benders

[1] Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity: https://eur-lex.europa.eu/eli/reg/2019/943/oj

[2] ACER - Methodology for calculating the value of lost load, the cost of new entry and the reliability standard - https://acer.europa.eu/sites/default/files/documents/Decisions_annex/ACER%20Decision%2023-2020%20on%20VOLL%20CONE%20RS%20-%20Annex%20I.pdf

[3] Article D141-12-6 from the French Energy Code: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044622322) and the current associated decree from 2022: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000046180401 (in French)

[4] ACER - Security of EU electricity supply in 2021: Report on Member States approaches to assess and ensure adequacy - October 2022 - https://acer.europa.eu/Publications/ACER_Security_of_EU_Electricity_Supply_2021.pdf

decomposition algorithm and a binary search on the total investment cost. We report computational experiments conducted on a set of randomly generated instances. While solving the KKT reformulation of the problem is already challenging for the the smallest instances, our matheuristic always computes a feasible solution and provides better results in comparison to three simple procedures.

The paper is organized as follows. In Section 2, we describe the two-stage generation and transmission capacity expansion planning problem into which we aim to introduce the LOLE reliability criterion. We present our main modeling hypotheses, how the problem relates to the literature, and mathematical models. In Section 3 we describe in details how we introduce the LOLE reliability criterion into the problem. We start by positioning our work with regard to the literature, before discussing the time-consistency concept when presenting mathematical models. To find feasible solutions to the problem, we present in Section 4 a matheuristic based on a Benders decomposition and a binary search on the total investment cost. In Section 5, we present numerical results. We finally conclude and present several perspectives to this work in Section 6.

2 A two-stage generation and transmission capacity expansion planning problem

In this section, we describe a classical two-stage stochastic generation and transmission capacity expansion planning problem in which the LOLE reliability criterion is not taken into account. It is similar to the problem considered by RTE in the first phase of its two-phase approach. We first describe our main modeling hypotheses and the corresponding literature. We then present a mathematical formulation of the problem.

2.1 Modeling hypotheses and related literature

In the literature, several works aim at determining long-term optimal generation and transmission capacity investments using an integrated planner approach, see, e.g., Lara et al. (2018, 2020); Micheli et al. (2020); Li et al. (2022) for recent ones (the first two only deal with generation capacity expansion planning). We make a number of assumptions similar to those made in these works.

We focus on the power grid of a given territory and use, like in Lara et al. (2018, 2020), Micheli et al. (2020), Li et al. (2022), an aggregated representation of the grid to reduce the spatial complexity of the problem. Specifically, we assume that the studied territory is divided in a few regions, with areas within each region sharing similarities such as climate and load profiles. In RTE's study, the division is performed by a priori identifying (based on expert knowledge at RTE) locations where transmission line reinforcements are almost inevitable (see Figure 1 and RTE (2019), page 30). In the end, France is divided in a few regions (see Figure 1) and neighbouring countries are represented with one to four regions. In each region, we consider a single *representative* generating unit and a single load demand. A finer representation aggregating only units that share similar characteristics, as presented in Palmintier and Webster (2013), Lara et al. (2018) or Micheli et al. (2020) and done by RTE, would only require very few adaptations of the work we present here, although it would increase the computational burden. We do not go down that road in order to focus on the methodological aspect. As done by RTE, we also consider at most a single *representative* transmission line between regions.

We follow most of the modeling hypotheses used in the Antares-Simulator project[5] developed by RTE. The first-stage problem is to simultaneously make generation and transmission capacity expansion decisions for a given target year. We assume that all investment variables are continuous, which is a common assumption for investments in renewable generating units such as solar panels or wind turbines (Lara et al., 2018, 2020; Micheli et al., 2020; Li et al., 2022). As we consider a single *representative* generating unit for each region, the assumption still gives a reasonable estimate even when considering small thermal generating units. Investment capacities in generating units with large power capacity (such as nuclear power plants) are not considered. They are taken as inputs in the prospective study conducted at RTE because they are not decided on the basis of techno-economic criteria alone, but are the result of political decisions. Between two neighbouring regions, we do not assume that

[5] https://antares-simulator.readthedocs.io/en/latest/reference-guide/11-modeling/

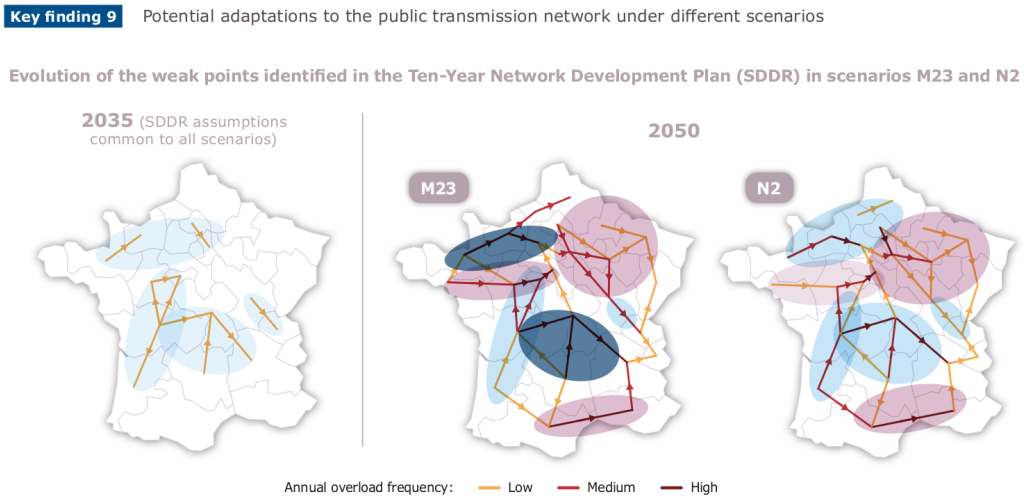


Figure 1: A representation of the French transmission network used by RTE in a prospective study. (RTE (2021), page 39).

we invest in an integer number of transmission lines whose capacity is given as input like in Micheli et al. (2020) and Li et al. (2022), but we rather consider that investments continuously increase the nominal capacity of the representative transmission line. Note that the introduction of integer investment variables would require little change in our models and algorithms, but would make the problem much more numerically challenging.

The second-stage problem decomposes by scenario and involves solving for the targeted year a simplified operational problem taking load balance and ramping limit constraints into account. The uncertain parameters are restricted to the load demand and the production and transmission costs. Similar to modeling operational decisions for a set representative days with an hourly time resolution as is generally the case for long-term generation and transmission expansion planning problems (see Nahmmacher et al. (2016)), we use a set of *representative periods* to capture the temporal correlation between decisions of successive time units. Like in Lara et al. (2018) and Micheli et al. (2020) and like RTE, we model the transmission network using a flow-based model (also called transport model). Contrary to the direct current power flow representation, Kirchhoff's voltage law is ignored, since we only need to evaluate strategic decisions in a long-term horizon. A more accurate representation of the power flow can be used in a post-processing phase to evaluate short-term investment decisions. As done in RTE's prospective study, we penalize ENS with the VOLL. This latter value[6] is the result of an economic trade-off between the cost of power cuts during load shedding and the cost of constructing additional peak capacity to reduce these cuts. We note that Micheli et al. (2020) also penalized ENS with a large value (10000 €/MWh).

Finally, similarly to the literature (Lara et al., 2018, 2020; Micheli et al., 2020; Li et al., 2022), the operational subproblem is such that the two-stage stochastic problem has relatively complete recourse, i.e. there is always a feasible solution for any investment decision.

In the literature, such problem is usually tackled with methods based on Benders decomposition. Lara et al. (2018) considered a deterministic multi-period capacity generation investment problem where the demand must be strictly met. They used a Lagrangian decomposition approach to obtain independent problems for every year and then took inspiration from stochastic dual dynamic integer programming to derive a nested Benders decomposition algorithm. The stochastic version of the problem is considered in Lara et al. (2020). A similar problem, but including transmission expansion decisions and a DC power flow model, is tackled by Li et al. (2022) with a nested Benders decomposition algorithm. Micheli et al. (2020) studied a generation and transmission expansion planning problem with high shares of renewable energy technologies and solved it using a multicut Benders decomposition algorithm.

[6]In France, the VOLL has been set to 33000 €/MWh in a decree[3] published on August 5, 2022.

2.2 Mathematical model

We now give a formal definition of the two-stage stochastic expansion planning problem. Let $G = (\mathcal{N}, \mathcal{A})$ be an undirected graph representing the power grid where nodes stand for regions and arcs stand for representative transmission lines. We denote by $\mathcal{N}^\star \subset \mathcal{N}$ the set of nodes that contain a representative generating unit or where a generating unit can be built. The initial nominal capacity (MW) of each node $i \in \mathcal{N}^\star$ is $P_i^{\text{init}} \in \mathbb{R}_+$. The cost of investing in production capacity (€/MW) at each node $i \in \mathcal{N}^\star$ is denoted by $c_i^{\text{prod}} \in \mathbb{R}_+$. Similarly, the cost of investing in transmission capacity (€/MW) for each arc $a \in \mathcal{A}$ is denoted by $c_a^{\text{line}} \in \mathbb{R}_+$. The initial transmission capacity (MW) of each arc a is denoted by $F_a^{\text{init}} \in \mathbb{R}_+$. Although we said that G is undirected, we use for convenience the term "arc", but $\mathcal{A}$ is such that $(i,j) \in \mathcal{A} \Rightarrow (j,i) \notin \mathcal{A}$ for every $i, j \in \mathcal{N}$. For each node $i \in \mathcal{N}$, we denote respectively by $\Gamma^+(i)$ and $\Gamma^-(i)$ the sets of ingoing and outgoing arcs.

Let $\mathcal{T}$ be the set of time units in the time period. We model the uncertain parameters with a finite set $\mathcal{S}$ of scenarios. The probability associated with scenario $s \in \mathcal{S}$ is denoted by $p_s \in [0,1]$. For each scenario $s \in \mathcal{S}$, each time unit $t \in \mathcal{T}$, the production operating (e.g., fuel, labor, maintenance) cost (€/MW) at node $i \in \mathcal{N}^\star$ is denoted by $g_{s,i,t}^{\text{prod}} \in \mathbb{R}_+$ and the transmission operating (e.g., maintenance, degradation) cost (€/MW) for arc $a \in \mathcal{A}$ by $g_{s,a,t}^{\text{flow}} \in \mathbb{R}_+$. The load demand of node $i \in \mathcal{N}$ during time unit t in scenario s (MWh) is denoted by $d_{s,i,t}$. The maximum and minimum ramp-up rate for the production at node $i \in \mathcal{N}^\star$ is denoted by $\Delta_i \in [0,1]$. The VOLL is set equal to a constant value $V \geq 0$ and applies to the demand unsatisfied at nodes from $\mathcal{N}$.

We formulate the problem using the following decisions variables. For $i \in \mathcal{N}^\star$, $P_i \in \mathbb{R}_+$ is the increase of the nominal capacity (MW) of the representative generating unit at node i. For $a \in \mathcal{A}$, $F_a \in \mathbb{R}_+$ is the increase of the nominal capacity (MW) of the transmission line associated to arc a. For each node $i \in \mathcal{N}^\star$, each time unit $t \in \mathcal{T}$, and each scenario $s \in \mathcal{S}$, $h_{s,i,t} \in \mathbb{R}_+$ is the power output (MW) of the generating unit. For each arc $a \in \mathcal{A}$, each time unit t, and each scenario s, we represent the power (MW) transferred through line a with the algebraic flow variable $f_{s,a,t} \in \mathbb{R}$ and the absolute flow variable $\underline{f}_{s,a,t} \in \mathbb{R}_+$. The continuous variable $u_{s,i,t} \in \mathbb{R}_+$ indicates the energy not served (MWh) at node $i \in \mathcal{N}$ in scenario s during time unit t. In the following, we denote vectors of variables with a bold font (e.g., $\mathbf{F} \in \mathbb{R}_+^{|\mathcal{A}|}$ is the vector of variables related to the increase of the nominal capacity for every transmission line).

For a scenario $s \in \mathcal{S}$, we define the operational cost $\phi_s(\mathbf{P}, \mathbf{F})$ according to the value of the investment variables $\mathbf{P}$ and $\mathbf{F}$ as follows:

$$
\phi_s(\mathbf{P}, \mathbf{F}) = \begin{cases}
\min & \displaystyle\sum_{i \in \mathcal{N}^\star} \sum_{t \in \mathcal{T}} g_{s,i,t}^{\text{prod}} h_{s,i,t} + \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} g_{s,a,t}^{\text{flow}} \underline{f}_{s,a,t} + V \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} u_{s,i,t} & \text{(1a)} \\
\text{s.t.} & h_{s,i,t} \leq P_i^{\text{init}} + P_i \quad \forall i \in \mathcal{N}^\star, \forall t \in \mathcal{T} & \text{(1b)} \\
& h_{s,i,t} \leq h_{s,i,t-1} + \Delta_i (P_i^{\text{init}} + P_i) \quad \forall i \in \mathcal{N}^\star, \forall t \in \mathcal{T} \setminus \{0\} & \text{(1c)} \\
& h_{s,i,t} \geq h_{s,i,t-1} - \Delta_i (P_i^{\text{init}} + P_i) \quad \forall i \in \mathcal{N}^\star, \forall t \in \mathcal{T} \setminus \{0\} & \text{(1d)} \\
& h_{s,i,0} \leq h_{s,i,|\mathcal{T}|} + \Delta_i (P_i^{\text{init}} + P_i) \quad \forall i \in \mathcal{N}^\star & \text{(1e)} \\
& h_{s,i,0} \geq h_{s,i,|\mathcal{T}|} - \Delta_i (P_i^{\text{init}} + P_i) \quad \forall i \in \mathcal{N}^\star & \text{(1f)} \\
& \underline{f}_{s,a,t} \leq F_a^{\text{init}} + F_a \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} & \text{(1g)} \\
& \underline{f}_{s,a,t} \geq f_{s,a,t} \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} & \text{(1h)} \\
& \underline{f}_{s,a,t} \geq -f_{s,a,t} \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} & \text{(1i)} \\
& h_{s,i,t} + \displaystyle\sum_{a \in \Gamma^+(i)} f_{s,a,t} - \sum_{a \in \Gamma^-(i)} f_{s,a,t} + u_{s,i,t} \geq d_{s,i,t} \quad \forall i \in \mathcal{N}^\star, \forall t \in \mathcal{T} & \text{(1j)} \\
& \displaystyle\sum_{a \in \Gamma^+(i)} f_{s,a,t} - \sum_{a \in \Gamma^-(i)} f_{s,a,t} + u_{s,i,t} \geq d_{s,i,t} \quad \forall i \in \mathcal{N} \setminus \mathcal{N}^\star, \forall t \in \mathcal{T} & \text{(1k)} \\
& h_{s,i,t} \in \mathbb{R}_+ \quad \forall i \in \mathcal{N}^\star, \forall t \in \mathcal{T} & \text{(1l)} \\
& u_{s,i,t} \in \mathbb{R}_+ \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T} & \text{(1m)} \\
& f_{s,a,t} \in \mathbb{R} \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} & \text{(1n)} \\
& \underline{f}_{s,a,t} \in \mathbb{R}_+ \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} & \text{(1o)}
\end{cases}
$$

Constraints (1b) and (1g) respectively ensure that the production at each node and the flow on each arc are not larger than the capacities decided in the investment phase. Constraints (1h) and (1i) ensure that each variable $\underline{f}_{s,a,t}$ is the absolute value of $f_{s,a,t}$. Constraints (1c)-(1f) state that in a given scenario the difference in production at node i between two time units is not larger than Δ_i times the installed capacity. Constraints (1j) and (1k) define the flow conservation constraints. The remaining constraints define the domain definition of the variables.

The two-stage stochastic generation and transmission expansion planning problem can then be written as follows:

$$\min \quad \sum_{i\in\mathcal{N}^\star} c_i^{\text{prod}} P_i + \sum_{a\in\mathcal{A}} c_a^{\text{line}} F_a + \sum_{s\in\mathcal{S}} p_s \phi_s(\mathbf{P},\mathbf{F}) \tag{2a}$$

$$\text{s.t.} \quad P_i \in \mathbb{R}_+ \quad \forall i \in \mathcal{N}^\star \tag{2b}$$

$$F_a \in \mathbb{R}_+ \quad \forall a \in \mathcal{A} \tag{2c}$$

For every $(\mathbf{P},\mathbf{F}) \in \mathbb{R}_+^{|\mathcal{N}^\star|} \times \mathbb{R}_+^{|\mathcal{A}|}$, we define $\mathcal{R}_s(\mathbf{P},\mathbf{F}) = \{(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) : (1b) - (1o)\}$ as the set of feasible solutions to the operational problem associated with scenario s. For every solution $(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \mathcal{R}_s(\mathbf{P},\mathbf{F})$, we denote its cost as $\psi_s(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) = \sum_{i\in\mathcal{N}^\star}\sum_{t\in\mathcal{T}} g_{s,i,t}^{\text{prod}} h_{s,i,t} + \sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}} g_{s,a,t}^{\text{flow}} \underline{f}_{s,a,t} + V \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}} u_{s,i,t}$. Using this notation, problem (2) can equivalently be written in its extensive form as follows:

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{N}^\star} c_i^{\text{prod}} P_i + \sum_{a\in\mathcal{A}} c_a^{\text{line}} F_a + \sum_{s\in\mathcal{S}} p_s \psi_s(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \\ \text{s.t.} \quad & (2b) - (2c) \\ & (\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \mathcal{R}_s(\mathbf{P},\mathbf{F}) \quad \forall s \in \mathcal{S} \end{aligned}$$

As the operational subproblem decomposes into as many independent continuous subproblems as there are scenarios when the investment decisions $(\mathbf{P},\mathbf{F})$ are fixed, the problem is typically solved with a Benders decomposition algorithm (see §2.1). This is the method currently used by RTE.

3 Introducing the LOLE reliability criterion into the problem

Optimal solutions to problem (2) may, in some scenarios, have multiple nodes with ENS at multiple time units, as investing in less nominal capacity can be cheaper than avoiding ENS. We illustrate this situation in the example of Figure 2. In the optimal solution to problem (2) (see Figure 3), there is no capacity investment and there is ENS for node 2 in the first scenario (Figure 3a) and for nodes 2 and 3 in the second scenario (Figure 3b). This is due to the presence of VOLL in the objective function, which in this particular example makes ENS less expensive than capacity investments. But this weakens the *reliability* of the network, understood here as the ability to ensure a balance between supply and demand in optimal solutions to the generation and transmission expansion planning problem.

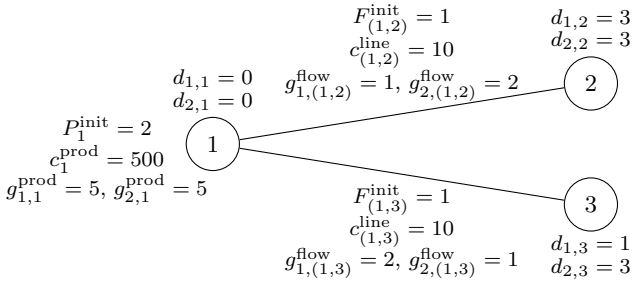
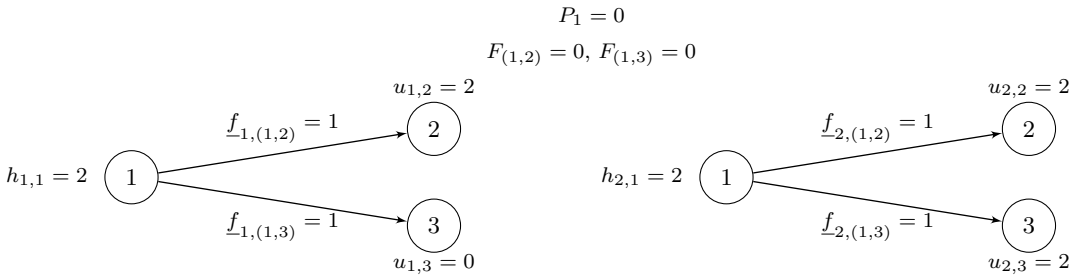


Figure 2: An instance with 3 nodes, 2 arcs, 2 scenarios ($p_1 = 0.95$, $p_2 = 0.05$), 1 time unit, and such that $V = 50$.



(a) Scenario 1

(b) Scenario 2

Figure 3: Optimal solution to problem (2) for the example of Figure 2

Introducing the LOLE reliability criterion into the problem is designed to address the aforementioned issue. Recently, the IEEE Resource Adequacy Working Group clarified how one should define a LOLE risk metric and how it should be interpreted (Stephen et al., 2022). LOLE is a counting measure, specifically "*the expected count of event-periods per horizon*". In our work, the horizon is the considered time period ($\mathcal{T}$ is the ordered set resulting from the homogeneous partition of the time period into time units) and an event-period represents the existence of ENS above a given threshold ϵ at a given node and time unit. We define the LOLE as the number of nodes and time units at which the ENS exceeds ϵ, and we require it to be, in expectation over the scenarios, less than or equal to α.

3.1 Related literature and positioning of this work

Although the issue of reliability in transmission expansion planning has been explored in the literature (Leite da Silva et al., 2010; Hemmati et al., 2013; Lei et al., 2018; Cho et al., 2022), there is no consensus on how reliability should be factored into generation and transmission expansion planning problems. Cho et al. (2022) defined the system adequacy as the installation of sufficient generation and transmission capacity to meet the load demand considering failure scenarios. In a deterministic approach, the power system can be designed either to meet a load demand above the expected peak, or to withstand an unexpected failure of the largest production unit. At the operational level, reserve constraints can be imposed (Lara et al., 2018; Micheli et al., 2020; Li et al., 2022). In a probabilistic approach, failure scenarios can be explicitly identified and taken into account in the calculation of reliability indices. Choi et al. (2005) developed a branch-and-bound method to find transmission capacity expansion decisions that minimize costs and satisfy LOLE constraints. The LOLE is analytically computed based on a given load duration curve at each node, the total installed capacity and probability distribution functions of outage capacity. Alizadeh and Jadid (2011) considered the LOLE and the loss of load probability as reliability criteria. To achieve the level of reliability required when considering forced outage and failure rates of generation units and transmission lines, they proposed an ad hoc iterative heuristic method that strengthens the system by iteratively investing in additional capacities. Slipac et al. (2019) described a methodology to satisfy a reliability criterion assuming that, based on expert knowledge, a function indicating the total installed capacity required for each possible value of LOLE is available. Depending on the desired LOLE value, a capacity expansion problem is then solved by taking as input the total capacity to be installed. The objective is to minimize costs, one element of which is the penalization of ENS by the VOLL. Cho et al. (2023) calculated LOLE and ENS using power plant failure probabilities given as inputs and penalized the value of these indices in the objective function.

In this work, our objective is to derive a mathematical model for the problem resulting from the introduction of the LOLE reliability criterion (i.e., an upper bound on the LOLE) into problem (2). We do not assume binary investment decisions, which makes the method of Choi et al. (2005) inapplicable, since it is based on the enumeration of possible investments. We also minimize expected operational costs in addition to the investment cost, which also adds significant complexity to the problem. In addition, the uncertain parameters are given by

a discrete set of scenarios and are not limited to transmission line or production unit failures, but also include, for example, demand and operational costs. To our knowledge, no similar problems have been studied in the literature. Finally, even if at first sight our work is only a question of introducing a constraint into the two-stage generation and transmission capacity expansion planning problem described in Section 2, taking into account the LOLE reliability criterion has a significant impact on the structure of the problem as we will show in the next subsection, and this differentiates significantly our work from those listed in §2.1.

3.2 A time-inconsistent mathematical modeling of the LOLE reliability criterion

A first intuitive adaptation of problem (2) leads to the following mathematical model, where binary variable $\delta_{s,i,t}$ is equal to one if and only if the ENS at node i during time unit $t \in \mathcal{T}$ in scenario $s \in \mathcal{S}$ is larger than ϵ, and C_s is a large positive real number verifying $C_s \geq \max_{t\in\mathcal{T}}\{\sum_{i\in\mathcal{N}} d_{s,i,t}\}$.

$$\min \quad \sum_{i\in\mathcal{N}^\star} c_i^{\text{prod}} P_i + \sum_{a\in\mathcal{A}} c_a^{\text{line}} F_a + \sum_{s\in\mathcal{S}} p_s \psi_s(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \tag{3a}$$

$$\text{s.t.} \quad (2b) - (2c) \tag{3b}$$

$$(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \mathcal{R}_s(\mathbf{P}, \mathbf{F}) \quad \forall s \in \mathcal{S} \tag{3c}$$

$$C_s \delta_{s,i,t} \geq u_{s,i,t} - \epsilon \qquad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \tag{3d}$$

$$\epsilon \delta_{s,i,t} \leq u_{s,i,t} \qquad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \tag{3e}$$

$$\sum_{s\in\mathcal{S}} p_s \sum_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} \delta_{s,i,t} \leq \alpha \tag{3f}$$

$$\delta_{s,i,t} \in \{0,1\} \qquad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \tag{3g}$$

Constraints (3d) and (3e) couple the $u_{s,i,t}$ and $\delta_{s,i,t}$ variables. Constraint (3f) ensures the satisfaction of the LOLE reliability criterion.

We now rewrite problem (3) to highlight its decomposable structure.

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{N}^\star} c_i^{\text{prod}} P_i + \sum_{a\in\mathcal{A}} c_a^{\text{line}} F_a + \widetilde{\phi}(\mathbf{P}, \mathbf{F}) \\ \text{s.t.} \quad & (2b) - (2c) \end{aligned} \tag{4}$$

where $\widetilde{\phi}(\mathbf{P}, \mathbf{F})$, the operational cost according to the value of the investment variables $\mathbf{P}$ and $\mathbf{F}$, is defined as follows:

$$\widetilde{\phi}(\mathbf{P}, \mathbf{F}) = \begin{cases} \min & \displaystyle\sum_{s\in\mathcal{S}} p_s \psi_s(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \\ \text{s.t.} & (1b) - (1o) \quad \forall s \in \mathcal{S} \\ & (3d) - (3g) \end{cases}$$

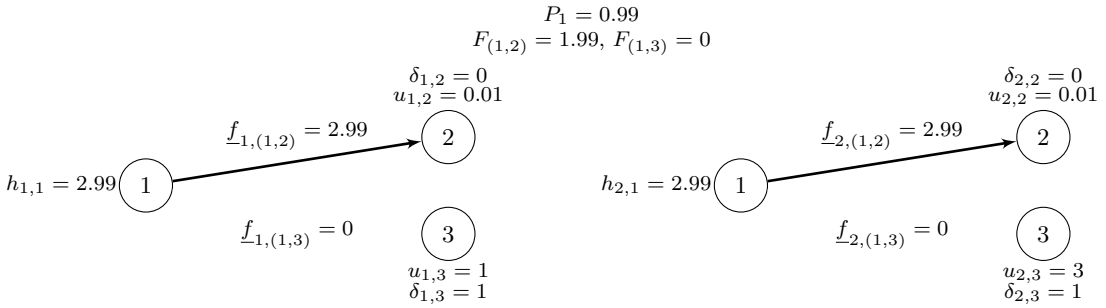
Observe that problem (3) is much more complex to solve than problem (2) for two reasons. First, constraint (3f) prevents the decomposition of the operational subproblem into independent subproblems related to the different scenarios. Note that the methods described in the previously cited works (Lara et al., 2018, 2020; Micheli et al., 2020; Li et al., 2022) all rely on this decomposition. Second, the introduced integer variables $\delta_{s,i,t}$ belong to the subproblem. Although Micheli et al. (2020) and Li et al. (2022) considered integer variables in the operational subproblems, they relaxed the integrality constraints when solving the problem. Mixed integer subproblems are solved only once at the end, taking the best computed investment decisions as input, and a duality gap is usually observed. To handle integer variables in the operational subproblems, Lara et al. (2018) and Lara et al. (2020) developed an algorithm based on the Stochastic Dual Dynamic Integer Programming algorithm introduced by Zou et al. (2019). This algorithm is not exact as there is a duality gap in general. For two-stage stochastic mixed-integer programs, recent exact algorithms are based on Lagrangian dual decomposition (Atakan and Sen, 2018; Li et al., 2022) and are currently successful for solving problems with hundred of variables and constraints.

The size of the problem targeted in this work is much larger, the subproblem is not decomposable according to the scenarios, and most importantly, as we shall see, problem (3) suffers from a major flaw.

The issue with problem (3) is the possibility of making suboptimal operational decisions in one or several scenarios of $\mathcal{S}$ to meet the LOLE reliability criterion. Let us illustrate this. Figure 4 displays the optimal solution to problem (3) for the example of Figure 2.

The optimal investment decision is such that $\mathbf{P} = (0.99)$ and $\mathbf{F} = (1.99, 0)^\top$. The cost of the operational decision made for scenario 2 is $g^{\text{prod}}_{2,1} h_{2,1} + g^{\text{flow}}_{2,(1,2)} \underline{f}_{2,(1,2)} + V(u_{2,2} + u_{2,3}) = 171.43$. However, the minimum operational cost that could be achieved for scenario 2 when $\mathbf{P} = (0.99)$ and $\mathbf{F} = (1.99, 0)^\top$ is $\phi_2(\mathbf{P}, \mathbf{F}) = 170.43 < 171.43 = \psi_2(\mathbf{h}_2, \underline{\mathbf{f}}_2, \mathbf{f}_2, \mathbf{u}_2)$ (see Figure 5). After the uncertainty is revealed (i.e., we know which scenario has been realized), there is no reason to make a suboptimal operational decision. In the example, for the investment decision $\mathbf{P} = (0.99)$ and $\mathbf{F} = (1.99, 0)^\top$, the operational decision taken if scenario 2 realizes will therefore be the one displayed in Figure 5 and not in Figure 4b. Note that there is not a similar issue with scenario 1 for this specific example.

If we consider the operation decision presented in Figure 5 for scenario 2 rather than the one of Figure 4b, we observe that the LOLE reliability criterion is not satisfied. The LOLE is equal to $0.95 \times 1 + 0.05 \times 2 = 1.05$ that is larger than $\alpha = 1$. In order to satisfy the LOLE reliability criterion, problem (3) accepts an operational decision that is not optimal for scenario 2. We argue that this situation should not happen, and that the investment decision $\mathbf{P} = (0.99)$ and $\mathbf{F} = (1.99, 0)^\top$ should not be considered feasible. We conclude that problem (3) does not adequately model the LOLE reliability criterion.



(a) Scenario 1

(b) Scenario 2

Figure 4: Optimal solution to problem (3) for the example of Figure 2 with $\alpha = 1$ and $\epsilon = 0.01$

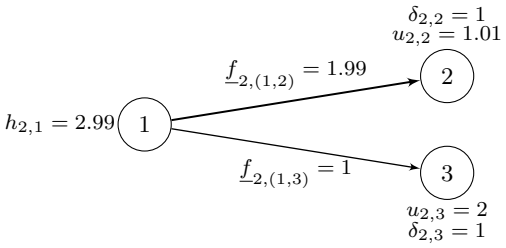


Figure 5: Optimal operational decision for scenario 2 when $\mathbf{P} = (0.99)$ and $\mathbf{F} = (1.99, 0)^\top$

The issue pointed out here arises when one or more constraints linking operational decisions taken in different scenarios are introduced into a "risk-neutral" two-stage stochastic problem. Takriti and Ahmed (2004) and Liu et al. (2016) pointed out similar issues for a special class of two-stage stochastic programs where the objective function addresses second-stage cost variability and for two-stage chance-constrained programs, respectively. All of these fall under the concept of *time-consistency*. Shapiro (2009) defined time-consistent optimal policies for multistage stochastic problems in the following way: *"At every state of the system our optimal decisions should not depend on scenarios which we already know cannot happen in the future"*. Rudloff et al. (2014) introduced

a more general definition for dynamic multistage models: *"a policy is time-consistent if and only if the future planned decisions are actually going to be implemented"*. For a two-stage stochastic problem, time-consistency means that second-stage decisions taken in the evaluation of the first stage decisions should be implemented when the uncertainty is revealed. Specifically, if $(\widehat{\mathbf{h}}_s, \widehat{\underline{\mathbf{f}}}_s, \widehat{\mathbf{f}}_s, \widehat{\mathbf{u}}_s)_{s\in\mathcal{S}}$ are the second-stage decisions taken when evaluating investment decision $(\widehat{\mathbf{P}}, \widehat{\mathbf{F}})$, then, when the uncertainty reveals that we are in scenario $s' \in \mathcal{S}$, $(\widehat{\mathbf{h}}_{s'}, \widehat{\underline{\mathbf{f}}}_{s'}, \widehat{\mathbf{f}}_{s'}, \widehat{\mathbf{u}}_{s'})$ should be an optimal operational decision, i.e., $\phi_{s'}(\widehat{\mathbf{P}}, \widehat{\mathbf{F}}) = \psi_{s'}(\widehat{\mathbf{h}}_{s'}, \widehat{\underline{\mathbf{f}}}_{s'}, \widehat{\mathbf{f}}_{s'}, \widehat{\mathbf{u}}_{s'})$. The above example demonstrates that problem (3) is not *time-consistent*.

3.3 A time-consistent mathematical modeling of the LOLE reliability criterion using bilevel programming

To enforce time-consistency in the mathematical modeling of the LOLE reliability criterion , we replace (3c) in problem (3) with the following constraint:

$$(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \underset{(\mathbf{h}'_s, \underline{\mathbf{f}}'_s, \mathbf{f}'_s, \mathbf{u}'_s)\in\mathcal{R}_s(\mathbf{P},\mathbf{F})}{\arg\min} \left\{\psi_s(\mathbf{h}'_s, \underline{\mathbf{f}}'_s, \mathbf{f}'_s, \mathbf{u}'_s)\right\} \qquad \forall s \in \mathcal{S} \tag{5}$$

Constraints (5) ensure that the operational decisions in every scenario are optimal and that they only depend on the investment decisions. Note that they can be equivalently written as $\psi_s(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) = \phi_s(\mathbf{P}, \mathbf{F})$ for every $s \in \mathcal{S}$.

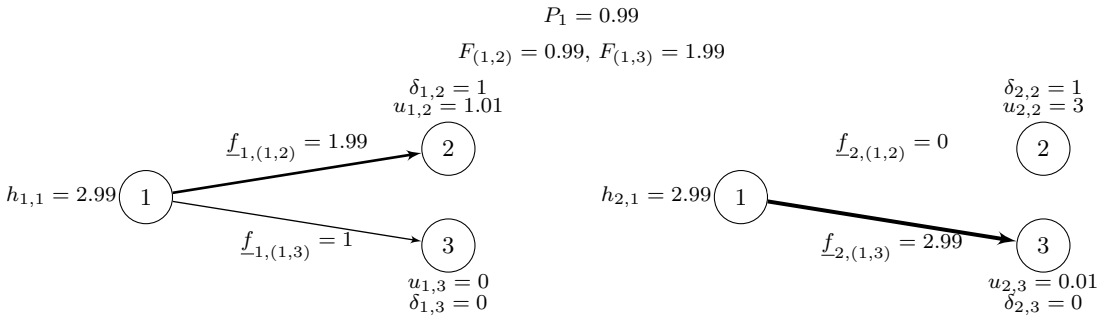
This leads to the following time-consistent formulation of the problem:

$$\begin{aligned}
\min \quad & \sum_{i\in\mathcal{N}^\star} c_i^{\text{prod}} P_i + \sum_{a\in\mathcal{A}} c_a^{\text{line}} F_a + \sum_{s\in\mathcal{S}} p_s \psi_s(\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \\
s.t. \quad & P_i \in \mathbb{R}_+ \quad \forall i \in \mathcal{N}^\star \quad (2b) \\
& F_a \in \mathbb{R}_+ \quad \forall a \in \mathcal{A} \quad (2c) \\
& (\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \underset{(\mathbf{h}'_s, \underline{\mathbf{f}}'_s, \mathbf{f}'_s, \mathbf{u}'_s)\in\mathcal{R}_s(\mathbf{P},\mathbf{F})}{\arg\min} \left\{\psi_s(\mathbf{h}'_s, \underline{\mathbf{f}}'_s, \mathbf{f}'_s, \mathbf{u}'_s)\right\} \qquad \forall s \in \mathcal{S} \quad (5) \\
& C_s \delta_{s,i,t} \geq u_{s,i,t} - \epsilon \qquad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (3d) \\
& \epsilon \delta_{s,i,t} \leq u_{s,i,t} \qquad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (3e) \\
& \sum_{s\in\mathcal{S}} p_s \sum_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} \delta_{s,i,t} \leq \alpha \quad (3f) \\
& \delta_{s,i,t} \in \{0,1\} \qquad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (3g)
\end{aligned} \tag{6}$$

The formulation (6) defines an optimistic bilevel problem with one leader and $|\mathcal{S}|$ followers. Using the terminology of bilevel programming, problem (3) is the *high-point relaxation* (HPR) of problem (6). Hereafter, we refer to (6), as the bilevel stochastic generation and transmission expansion planning problem that we abbreviate as Bilevel-α-Stoch-GTEPP.

Remark 1. *As the investment cost can be inserted into the lower-level objective functions without modifying the optimal solutions to the lower-level problems, the upper and lower levels have the same objective function.*

Figure 6 shows the optimal solution to problem (6) for the example of Figure 2. Its value is 599.19, that is strictly larger than the optimal value of the HPR (3) which is equal to 588.4895 (see Figure 4). This is not surprising as we described in §3.2 that the optimal solution to the HPR is not bilevel feasible. Specifically, the operational decision for scenario 2 taken when evaluating the optimal investment decision found solving the HPR is suboptimal, with an absolute difference with respect to optimal operational decisions equal to $171.43 - 170.43 = 1$. We show in Proposition 1 that, more generally, this absolute difference cannot be bounded by a constant.



(a) Scenario 1

(b) Scenario 2

Figure 6: Optimal solution to problem (6) for the example of Figure 2 with $\alpha = 1$ and $\epsilon = 0.01$

Proposition 1. *For any* $\mathsf{A} > 0$*, there exists an instance of Bilevel-*α*-Stoch-GTEPP such that for any optimal solution* $\left(\mathbf{P}^*, \mathbf{F}^*, (\mathbf{h}_s^*, \underline{\mathbf{f}}_s^*, \mathbf{f}_s^*, \mathbf{u}_s^*)_{s\in\mathcal{S}}\right)$ *to its high-point relaxation, we have* $\psi_s(\mathbf{h}_s^*, \underline{\mathbf{f}}_s^*, \mathbf{f}_s^*, \mathbf{u}_s^*) > \mathsf{A} \cdot \phi_s(\mathbf{P}^*, \mathbf{F}^*)$ *for a scenario* $s \in \mathcal{S}$*.*

Proof. We show this result by exhibiting an instance of Bilevel-α-Stoch-GTEPP parameterized by $\beta \geq 2$ (see Figure 7). For readability, we omit the time unit index.

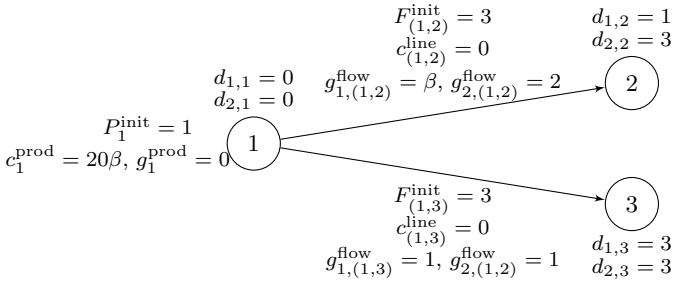
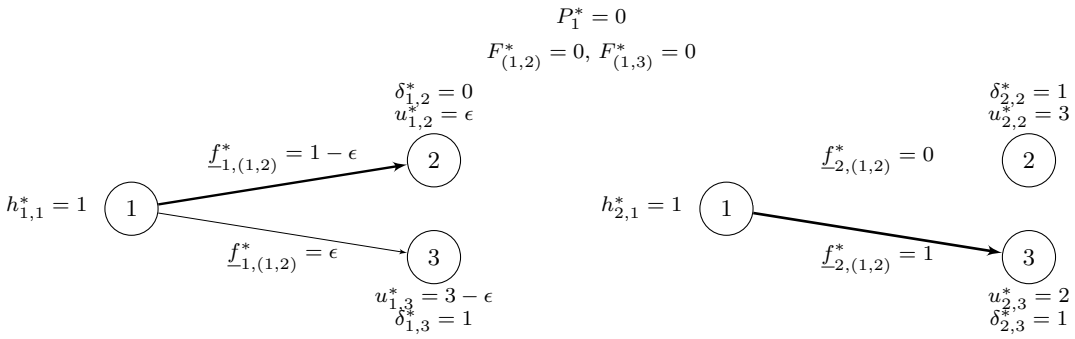


Figure 7: An instance with 3 nodes, 2 arcs, 2 scenarios ($p_1 = 0.9$ and $p_2 = 0.1$), 1 time unit, and such that $V = 10\beta$, $\epsilon = 0.1$, $\alpha = 1.5$.

The optimal solution to the HPR is such that there are no investments in generation and transmission capacity, i.e., $(\mathbf{P}^*, \mathbf{F}^*) = (\mathbf{0}, \mathbf{0})$. We also have $\underline{f}^*_{1,(1,2)} = 1 - \epsilon$ ($1 - \epsilon$ unit of flow is sent to node 2 during time unit 1 in scenario 1), $f^*_{1,(1,3)} = \epsilon$, $\underline{f}^*_{2,(1,3)} = 0$, and $\underline{f}^*_{2,(1,3)} = 1$. We show this solution in Figure 8. Note that the expected value of the sum of the number of time units with ENS larger than ϵ at each node (i.e., the LOLE) is equal to $0.9 + 0.1 \times 2 = 1.1 < 1.5 = \alpha$.

In the HPR, the cost of the operational decision $(\mathbf{h}_1^*, \underline{\mathbf{f}}_1^*, \mathbf{f}_1^*, \mathbf{u}_1^*)$ for scenario 1 is $\psi_1(\mathbf{h}_1^*, \underline{\mathbf{f}}_1^*, \mathbf{f}_1^*, \mathbf{u}_1^*) = (1-\varepsilon)\beta + \varepsilon + 3V$. Figure 9 shows the optimal operational decision for scenario 1 when the investment decision is $(\mathbf{P}^*, \mathbf{F}^*)$. We have $\phi_1(\mathbf{P}^*, \mathbf{F}^*) = 1 + 3V$, as it is cheaper to send all the energy to node 3. We finally obtain:

$$\lim_{\beta\to\infty} \frac{\psi_1(\mathbf{h}_1^*, \underline{\mathbf{f}}_1^*, \mathbf{f}_1^*, \mathbf{u}_1^*)}{\phi_1(\mathbf{P}^*, \mathbf{F}^*)} = +\infty$$



(a) Scenario 1 (b) Scenario 2

Figure 8: Optimal solution to the HPR (3) for the example of Figure 7

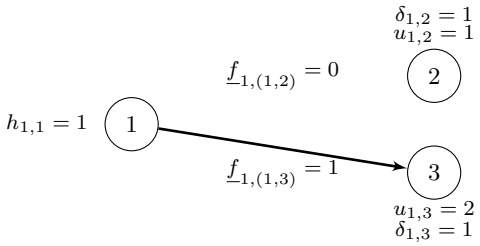


Figure 9: Optimal operational decision for scenario 1 when the investment decision is $(\mathbf{P}^*, \mathbf{F}^*)$

□

Proposition 1 exhibits an instance of Bilevel-α-Stoch-GTEPP, such that, for an optimal solution to the HPR, there exists a scenario for which the cost of the operational decision taken during the evaluation of the optimal investment decision is A times higher than the cost of the optimal operational decision, and this for any value $\mathsf{A} > 0$. The time-inconsistency of the HPR (3) is the fundamental difference with respect to the bilevel problem (6).

We conclude with a more general remark that contradicts a result stated in Grimm et al. (2016)[7].

Remark 2. *An optimal solution of the HPR of a bilevel problem with identical objective functions in the upper and lower levels is not a feasible bilevel solution in general.*

4 A Benders-based heuristic solution method

We have established that problem (6) is a special case of a MIP-LP bilevel problem and is not equivalent to its HPR. This type of problem is known to be theoretically and practically hard to solve. A recent survey by Kleinert et al. (2021) describes existing mixed-integer programming techniques to tackle such problems. Most of the techniques use the KKT conditions of the lower-level problem to derive a non-linear single-level formulation of the bilevel problem. This reformulation is obtained by replacing the optimality conditions of the lower-level variables in the upper-level problem by lower level's KKT conditions. Fortuny-Amat and McCarl (1981) proposed to linearize the KKT complementarity conditions to obtain a mixed-integer linear single-level program, which can then be solved by branch-and-bound approaches, provided the suitable value is found for a "big-M" parameter (Kleinert et al., 2020). Branching directly on the KKT complementarity constraints overcomes the previously

[7] We contacted the authors, who agreed that Proposition 2 in Grimm et al. (2016) is valid only if there are no upper-level constraints depending on lower-level variables.

mentioned drawback. A solver can perform this branching automatically when the complementarity conditions are modeled using special ordered sets of type 1 (SOS1) (see e.g., Siddiqui and Gabriel (2013)). A non linear single-level reformulation of a MIP-LP bilevel problem can alternatively be obtained by using strong duality, although to our knowledge it has never led to any practical method for solving large bilevel problems. Tailored solution techniques have also been proposed in the literature for problems satisfying particular assumptions, such as the absence of linking constraints (Kleinert and Schmidt, 2019), the integrality of all linking upper-level variables (Fischetti et al., 2018), the compactness of the high-point relaxation feasible region (Saharidis and Ierapetritou, 2009), or the presence of only integer variables in the upper-level (Lozano and Smith, 2017).

Many applications of bilevel programming arise in energy and electricity markets (Wogrin et al., 2020). Generation and transmission expansion planning in market environments has been studied in (Garces et al., 2009; Jin and Ryan, 2011) and these authors reformulate the bilevel problem using either KKT conditions (Jin and Ryan, 2011) or strong duality conditions (Garces et al., 2009). Kleinert and Schmidt (2019) study the optimal price zones in electricity markets using a trilevel model, where investments in transmission lines are made by the regulator at the first level and investments in production capacity are made by competitive companies at the second level, before computing the optimal energy dispatch at the third level. They solve the problem using a generalized Benders decomposition approach using the weak coupling between the three levels of their problem (no first and second level constraints contain lower-level variables and the objective function of the first level can be expressed as a linear combination of the objective functions of the lower-level problems). The special structure of their problem also allows them to reformulate it as an equivalent bilevel problem. Using KKT conditions, a large equivalent single-level problem is then obtained. To our knowledge and according to Cho et al. (2022), bilevel optimization for generation and transmission capacity expansion problem has never been used when considering an integrated planner approach.

The Bilevel-α-Stoch-GTEPP combines several difficulties that make the methods described above unsuitable. First, the upper-level variables that appear in the lower-level problem are continuous, which prevent the use of algorithms from the literature based on the implicit enumeration of the coupling variables (Xu and Wang, 2014; Lozano and Smith, 2017; Fischetti et al., 2018). Then, there are constraints in the upper-level problem that involve lower-level variables which invalidate an approach similar to Kleinert and Schmidt (2019). Finally, the upper-level problem involves $|\mathcal{S}| \times |\mathcal{N}| \times |\mathcal{T}|$ binary variables. With such a large number of binary variables, we do not expect that solving the KKT reformulation of the bilevel problem provides an efficient approach for real-size instances. As pointed out in §3.2, solving the HPR of the problem is already challenging.

In the light of what has been described, we present a Benders-based heuristic to compute feasible solutions to Bilevel-α-Stoch-GTEPP. Our approach is based on the observation that if we impose a large investment cost in problem (2), optimal solutions are more likely to have a lower ENS. We therefore define a sequence of single-level problems obtained by relaxing the LOLE reliability criterion and imposing a lower bound on the total investment cost. We implemented a heuristic binary search strategy for determining the value of this lower bound. A quite similar idea was used by Grimm et al. (2016) for solving a transmission and generation investment problem in a competitive environment. These authors designed a binary search algorithm based on the value of the revenues coming from network fees to be able to remove the single constraint that coupled the decisions of two subproblems. Under a realistic assumption stated below, our procedure always computes a bilevel feasible solution.

For readability purposes, we introduce a more concise representation of the problem: vector of variables $\mathbf{x}$ is the concatenation of the vectors of variables $\mathbf{P}$ and $\mathbf{F}$ (i.e., $\mathbf{x} = (\mathbf{P}, \mathbf{F})$) and vector $\mathbf{c}$ is the concatenation of vectors $\mathbf{c}^{\text{prod}}$ and $\mathbf{c}^{\text{line}}$. For each scenario s, the vector of variables $\mathbf{y}_s$ is the concatenation of the vector of operational variables $\mathbf{h}_s$, $\mathbf{f}_s$, and $\underline{\mathbf{f}}_s$ (i.e., $\mathbf{y}_s = (\mathbf{h}_s, \underline{\mathbf{f}}_s, \mathbf{f}_s)$) and $\mathbf{g}_s$ is the associated vector of operational costs. Hereafter we denote by $\mathbf{1}$ (resp. $\mathbf{0}$) a vector of ones (resp. zeros) of dimension dictated by the context. The computation of the operational cost $\phi_s(\mathbf{x})$ for a scenario $s \in \mathcal{S}$ according to the value of the investment variables $\mathbf{x}$ (defined in (1)) can be reformulated as follows (where W, W' and T are matrices, $\ell_s = (|\mathcal{N}^\star| + |\mathcal{A}|)|\mathcal{T}|$, and

$\ell'_s = |\mathcal{A}||\mathcal{T}|$):

$$\phi_s(\mathbf{x}) = \begin{cases} \min & \mathbf{g}_s^\top \mathbf{y}_s + V(\mathbf{1}^\top \mathbf{u}_s) & (7a) \\ \text{s.t.} & W\mathbf{y}_s + W'\mathbf{u}_s \geq \mathbf{b}_s - T\mathbf{x} & (7b) \\ & \mathbf{y}_s \in \mathbb{R}_+^{\ell_s} \times \mathbb{R}^{\ell'_s}, \mathbf{u}_s \in \mathbb{R}_+^{|\mathcal{N}|\times|\mathcal{T}|} & (7c) \end{cases}$$

We refer to problem (7) as $(SP_s(\mathbf{x}))$. This leads to a concise form of the Bilevel-α-Stoch-GTEPP (problem (6)), where matrices D and E and vector $\mathbf{q}_s$ are appropriately defined to represent constraints (3d)-(3e) (these elements depend on ϵ and C_s).

$$\min \quad \mathbf{c}^\top \mathbf{x} + \sum_{s\in\mathcal{S}} p_s\big(\mathbf{g}_s^\top \mathbf{y}_s + V(\mathbf{1}^\top \mathbf{u}_s)\big) \tag{8a}$$

$$s.t. \quad \mathbf{x} \in \mathbb{R}_+^{|\mathcal{N}^\star|+|\mathcal{A}|} \tag{8b}$$

$$(\mathbf{y}_s, \mathbf{u}_s) \in \underset{\substack{(\mathbf{y}'_s,\mathbf{u}'_s)\in\{\mathbf{y}''_s\in\mathbb{R}_+^{\ell_s}\times\mathbb{R}^{\ell'_s}, \mathbf{u}''_s\in\mathbb{R}_+^{|\mathcal{N}|\times|\mathcal{T}|} \\ : W\mathbf{y}''_s + W'\mathbf{u}''_s \geq \mathbf{b}_s - T\mathbf{x}\}}}{\arg\min} \{\mathbf{g}_s^\top \mathbf{y}'_s + V(\mathbf{1}^\top \mathbf{u}'_s)\} \qquad \forall s \in \mathcal{S} \tag{8c}$$

$$D\boldsymbol{\delta}_s + E\mathbf{u}_s \geq \mathbf{q}_s \qquad \forall s \in \mathcal{S} \tag{8d}$$

$$\sum_{s\in\mathcal{S}} p_s(\mathbf{1}^\top \boldsymbol{\delta}_s) \leq \alpha \tag{8e}$$

$$\boldsymbol{\delta}_s \in \{0,1\}^{|\mathcal{N}|\times|\mathcal{T}|} \qquad \forall s \in \mathcal{S} \tag{8f}$$

Constraints (8c) simply states that $(\mathbf{y}_s, \mathbf{u}_s)$ should be an optimal solution to $(SP_s(\mathbf{x}))$. As the values of variables $\boldsymbol{\delta}$ are directly deductible from the values of variables $\mathbf{u}$, we refer to a (bilevel) solution to (8) as $(\mathbf{x}, \mathbf{y}, \mathbf{u})$.

In the remainder of this paper, we assume that the VOLL V (the penalty cost for each unit of ENS) is large enough to prefer satisfying a demand whenever it is possible if only operational costs are considered in Bilevel-α-Stoch-GTEPP. This assumption, realistic at a strategical level, is formalized in Assumption 1, which says that if investing in generation and transmission capacity is free, then there is no optimal solution to problem (8) with a positive ENS value.

Definition 1. *The* investment free expansion planning problem *is defined as follows:*

$$\begin{cases} \min & \sum_{s\in\mathcal{S}} p_s \phi_s(\mathbf{x}) \\ s.t. & (8b) \end{cases} \tag{9}$$

Assumption 1. *We have* $\mathbf{u} = \mathbf{0}$ *in any optimal solution to the investment free expansion planning problem* (9).

Assumption 1 implies that an optimal solution to the investment free expansion planning problem (9) satisfies the LOLE reliability criterion. This leads to the result formulated in Proposition 2.

Proposition 2. *Let* $\mathbf{x}^*$ *be an optimal solution to the investment free expansion planning problem* (9). *Every bilevel feasible solution* $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{u}})$ *to problem* (8) *with a higher total investment cost (i.e.,* $\mathbf{c}^\top \hat{\mathbf{x}} \geq \mathbf{c}^\top \mathbf{x}^*$*) has a larger total cost (i.e.,* $\mathbf{c}^\top \hat{\mathbf{x}} + \sum_{s\in\mathcal{S}} p_s \phi_s(\hat{\mathbf{x}}) \geq \mathbf{c}^\top \mathbf{x}^* + \sum_{s\in\mathcal{S}} p_s \phi_s(\mathbf{x}^*)$*).*

Proof. Let $(\mathbf{y}^*_s, \mathbf{u}^*_s)$ be an optimal solution to $(SP_s(\mathbf{x}^*))$ for every $s \in \mathcal{S}$. Assumption 1 implies that $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{u}^*)$ induces a bilevel feasible solution to problem (8). As $\mathbf{x}^*$ is an optimal solution to the investment free expansion planning problem, we know that $\sum_{s\in\mathcal{S}} p_s \phi_s(\mathbf{x}^*) \leq \sum_{s\in\mathcal{S}} p_s \phi_s(\hat{\mathbf{x}})$. The result is then derived from the fact that $\mathbf{c}^\top \hat{\mathbf{x}} \geq \mathbf{c}^\top \mathbf{x}^*$. □

Proposition 2 implies that an optimal solution to the Bilevel-α-Stoch-GTEPP has necessarily a total investment cost lower than the total investment cost, denoted by Λ_{free}, of an optimal solution to the investment free expansion planning problem. We use this result to restrict the search of a good-quality solution to Bilevel-α-Stoch-GTEPP.

In the following, we use Iverson brackets[8] to express the number of time units with ENS from variables $\mathbf{u}$.

[8] $[\mathsf{P}] = 1$ if a statement P is true, and $[\mathsf{P}] = 0$ otherwise

4.1 General description of the method

In order to formally describe our Benders-based heuristic, we introduce the *minimum investment expansion planning problem*, in which the LOLE reliability criterion is not taken into account.

Definition 2. *Let* $\Lambda \geq 0$, *the* minimum investment expansion planning problem, *denoted as* (Inv(Λ)), *is defined as follows:*

$$\begin{cases} \min\ \mathbf{c}^\top \mathbf{x} + \sum_{s \in \mathcal{S}} p_s \phi_s(\mathbf{x}) \\ s.t.\ (8b) \\ \quad \mathbf{c}^\top \mathbf{x} \geq \Lambda \end{cases} \tag{10}$$

Observe that Λ is a lower bound on the total investment cost and that (Inv(Λ)) is a two-stage stochastic single-level linear program.

The core idea of our heuristic is to solve (Inv(Λ)) for every value of Λ selected during a binary search and to check if the computed solutions are feasible solutions to the Bilevel-α-Stoch-GTEPP or if it is possible to derive such feasible solutions from the computed solutions. Algorithm 1 outlines the general structure of our heuristic. We start by solving the investment free expansion planning problem (9) and by initializing our binary search strategy (lines 2 to 4 in Algorithm 1). Then, given a lower bound $\Lambda_{\min}$ (initially 0) and an upper bound $\Lambda_{\max}$ (initially Λ_{free}, the total investment cost of an optimal solution to problem (9)) on the total investment cost, we solve (Inv(Λ)) with a Benders decomposition algorithm (lines 7 to 21 - see §4.2) for $\Lambda = \tau\Lambda_{\min} + (1-\tau)\Lambda_{\max}$ (Λ is calculated with another formula in some particular cases specified in line 27), with $\tau \in (0,1)$ the binary search parameter. If, when solving (Inv(Λ)) for a given Λ, we do not find a bilevel feasible solution with a cost lower than the cost of the best bilevel feasible solution computed so far, then $\Lambda_{\min}$ is set to Λ (this operation is motivated by the fact that by imposing a higher total investment cost, we are more likely to compute solutions that satisfy the LOLE reliability criterion). Otherwise $\Lambda_{\max}$ is set to the current value of Λ which ensures exploring in the next iteration a new region of the search space which is likely to contain a bilevel feasible solution with a lower cost.

During the first iteration of the binary search, the minimum total investment cost is set to 0, which means that we solve the two-stage stochastic generation and transmission expansion planning problem (2). As this problem is a relaxation of Bilevel-α-Stoch-GTEPP, if an optimal solution to this problem is bilevel feasible, then it is an optimal solution to Bilevel-α-Stoch-GTEPP (line 25). In that case, we stop the algorithm.

To increase the chances of finding a feasible bilevel solution during the course of our method, we implemented two main strategies. Let $\bar{\mathbf{x}}$ be a given first-stage solution (i.e., given investment decisions) to problem (10), we can check if $(\bar{\mathbf{x}}, \mathbf{y}, \mathbf{u})$ is a bilevel feasible solution, where $(\mathbf{y}_s, \mathbf{u}_s)$ is the optimal solution to $(SP_s(\bar{\mathbf{x}}))$ returned by the solver for every $s \in \mathcal{S}$, by computing the number of variables $u_{s,i,t}$ with a value larger than ϵ. However, if $(\bar{\mathbf{x}}, \mathbf{y}, \mathbf{u})$ is not bilevel feasible, there might exist another optimal recourse solution $(\mathbf{y}''_s, \mathbf{u}''_s)$ to $(SP_s(\bar{\mathbf{x}}))$ for every s such that $(\bar{\mathbf{x}}, \mathbf{y}'', \mathbf{u}'')$ is bilevel feasible. Based on this observation, we introduce $\varphi_s(\bar{\mathbf{x}})$ as the optimal value of an auxiliary subproblem $(\text{Aux}_s(\bar{\mathbf{x}}))$ that is defined as follows:

$$\varphi_s(\bar{\mathbf{x}}) = \begin{cases} \min\ \mathbf{1}^\top \boldsymbol{\delta}_s \\ s.t.\ W\mathbf{y}_s + W'\mathbf{u}_s \geq \mathbf{b}_s - T\bar{\mathbf{x}} \\ \quad D\boldsymbol{\delta}_s + E\mathbf{u}_s \geq \mathbf{q}_s \\ \quad \mathbf{g}_s^\top \mathbf{y}_s + V\mathbf{1}^\top \mathbf{u}_s \leq \phi_s(\bar{\mathbf{x}}) \\ \quad \mathbf{y}_s \in \mathbb{R}_+^{\ell_s} \times \mathbb{R}^{\ell'_s}, \mathbf{u}_s \in \mathbb{R}_+^{|\mathcal{N}|\times|\mathcal{T}|}, \boldsymbol{\delta}_s \in \{0,1\}^{|\mathcal{N}|\times|\mathcal{T}|} \end{cases}$$

This auxiliary subproblem searches, among all optimal solutions to $(SP_s(\bar{\mathbf{x}}))$, one solution that minimizes the sum of the number of time units with ENS larger than ϵ at each node. Solving all the auxiliary subproblems for a given $\bar{\mathbf{x}}$ allows to determine if there exists a bilevel feasible solution with these investment decisions. We formalize this statement in the following proposition.

Proposition 3. *Let $\bar{\mathbf{x}} \in \mathbb{R}_+^{|\mathcal{N}^\star|+|\mathcal{A}|}$. There exists $\bar{\mathbf{y}}, \bar{\mathbf{u}}$ such that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{u}})$ is a bilevel feasible solution if and only if $\sum_{s\in\mathcal{S}} p_s\varphi_s(\bar{\mathbf{x}}) \leq \alpha$.*

Proof. ($\Rightarrow$) Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{u}})$ be a bilevel feasible solution. For every $s \in \mathcal{S}$, let $(\mathbf{y}_s^*, \mathbf{u}_s^*, \boldsymbol{\delta}_s^*)$ be an optimal solution to $(\text{Aux}_s(\bar{\mathbf{x}}))$. We have $\varphi_s(\bar{\mathbf{x}}) = \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}} \delta_{s,i,t}^*$, or equivalently $\varphi_s(\bar{\mathbf{x}}) = \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}}[u_{s,i,t}^* > \epsilon]$. By definition of every auxiliary subproblem associated with a scenario $s \in \mathcal{S}$, we also have $\varphi_s(\bar{\mathbf{x}}) \leq \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}}[\bar{u}_{s,i,t} > \epsilon]$. Because $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{u}})$ is bilevel feasible, we have $\sum_{s\in\mathcal{S}} p_s \sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}}[\bar{u}_{s,i,t} > \epsilon] \leq \alpha$ and therefore $\sum_{s\in\mathcal{S}} p_s\varphi_s(\bar{\mathbf{x}}) \leq \alpha$.

($\Leftarrow$) Let $(\mathbf{y}_s^*, \mathbf{u}_s^*, \boldsymbol{\delta}_s^*)$ be an optimal solution to $(\text{Aux}_s(\bar{\mathbf{x}}))$. By assumption, we have $\sum_{s\in\mathcal{S}} p_s\varphi_s(\bar{\mathbf{x}}) \leq \alpha$, which is equivalent to $\sum_{s\in\mathcal{S}} p_s\mathbf{1}^\top\boldsymbol{\delta}_s^* \leq \alpha$. As $\mathbf{g}_s^\top\mathbf{y}_s^* + V\mathbf{1}^\top\mathbf{u}_s^* \leq \phi_s(\bar{\mathbf{x}})$ ensures that $(\mathbf{y}_s^*, \mathbf{u}_s^*)$ is an optimal solution to $(SP_s(\bar{\mathbf{x}}))$, the solution $(\bar{\mathbf{x}}, \mathbf{y}_s^*, \mathbf{u}_s^*)$ is bilevel feasible. □

Solving $|\mathcal{S}|$ auxiliary subproblems formulated as mixed-integer linear programs, each involving $|\mathcal{N}|\times|\mathcal{T}|$ binary variables, is time-consuming. To accelerate this procedure, we stop if, after solving the auxiliary subproblems for a subset $\mathcal{S}_0 \subset \mathcal{S}$ of scenarios, we have $\sum_{s\in\mathcal{S}_0} p_s\varphi_s(\bar{\mathbf{x}}) > \alpha$.

In Algorithm 1, we define a parameter named `acc` as the *accuracy* of the bilevel feasibility checking. This parameter is either set to **Rand** if the algorithm checks the bilevel feasibility of a first-stage solution combined with the arbitrary optimal solutions to the subproblems returned by the solver, or set to **MinENS** if the algorithm checks the bilevel feasibility of a first-stage solution combined with the optimal solutions to the auxiliary subproblems. An intermediate version is to set parameter `acc` to **MinENS-Opt**, the difference with **Rand** being that the algorithm also checks whether optimal first-stage solutions to the current minimum investment expansion planning problem are bilevel feasible when combined with the optimal solutions to the auxiliary subproblems.

Checking whether a given first-stage solution can produce a bilevel solution can be performed at different places in our method. Preliminary experiments showed that the version of the algorithm that checks only the bilevel feasibility of optimal solutions to the current minimum investment expansion planning problem is outperformed by the version that checks bilevel feasibility for all first-stage solutions computed in the Benders decomposition algorithm described in §4.2. We therefore adopt this latter strategy.

4.2 A Benders decomposition algorithm for solving (Inv(Λ))

As problem (Inv(Λ)) decomposes into as many independent subproblems as there are scenarios when the investment decisions $\mathbf{x}$ are fixed, solving it with a Benders decomposition algorithm is suitable. For a scenario $s \in \mathcal{S}$, the set Π_s of solutions to the dual of $(SP_s(\mathbf{x}))$ does not depend on the first-stage solution $\mathbf{x} \in \mathbb{R}_+^{|\mathcal{N}^\star|+|\mathcal{A}|}$ and we denote by $\text{Vert}(\Pi_s)$ the set of its extreme points. The multicut Benders reformulation of problem (Inv(Λ)) is defined as follows:

$$\min\ \mathbf{c}^\top\mathbf{x} + \sum_{s\in\mathcal{S}} p_s\theta_s \tag{11a}$$

$$s.t.\ (8b) \tag{11b}$$

$$\mathbf{c}^\top\mathbf{x} \geq \Lambda \tag{11c}$$

$$\theta_s \geq \boldsymbol{\pi}_s^\top(\mathbf{b}_s - T\mathbf{x}),\ \forall s \in \mathcal{S},\ \forall\boldsymbol{\pi}_s \in \text{Vert}(\Pi_s) \tag{11d}$$

Constraints (11d) are Benders (optimality) cuts. Because the subproblem (7) is feasible for any investment decisions $\mathbf{x}$, Benders feasibility cuts are not necessary. The relaxation of (11) defined with a subset of the Benders cuts is called the relaxed master program and we denote this program at iteration k by $(\text{RMP})^{(k)}$. In the Benders decomposition algorithm, we iteratively solve the relaxed master program and generate Benders cuts until convergence is proven.

It is important to observe that the Benders cuts (11d) computed when solving (Inv(Λ)) for a given $\Lambda \geq 0$ are still valid when solving (Inv(Λ')) for any $\Lambda' \geq 0$. The initial relaxed master program for a given Λ can therefore include all the previously generated Benders cuts. We adopt this approach in our method.

Algorithm 1: The Benders-based heuristic solution method

Parameters: $\epsilon_{\text{opt}} > 0$, $\epsilon_{\text{inv}} > 0$, $\tau \in (0,1)$, $\texttt{acc} \in \{\textbf{Rand}, \textbf{MinENS-Opt}, \textbf{MinENS}\}$

$k \leftarrow 0$, $\texttt{x\_best\_updated} \leftarrow \texttt{False}$, $\Lambda \leftarrow 0$, $\Lambda_{\min} \leftarrow 0$

Solve the investment free expansion planning problem (9) and retrieve $\mathbf{x}_{\text{best}}$ an optimal first-stage solution, and the values $\big(\phi_s(\mathbf{x}_{\text{best}})\big)_{s\in\mathcal{S}}$

$z_{\text{best}} \leftarrow \mathbf{c}^\top \mathbf{x}_{\text{best}} + \sum_{s\in\mathcal{S}} p_s \phi_s(\mathbf{x}_{\text{best}})$

$\Lambda_{\max} \leftarrow \mathbf{c}^\top \mathbf{x}_{\text{best}}$

while $\Lambda_{\max} - \Lambda_{\min} > \epsilon_{\text{Inv}}$ **do**

  Set the total minimum investment cost to Λ in $(\text{RMP})^{(k)}$ ▷ Selecting a new bound on the total investment cost

  $\text{UB}^{(k)} \leftarrow +\infty$, $\text{LB}^{(k)} \leftarrow -\infty$

  $\texttt{x\_best\_updated} \leftarrow \texttt{False}$

  while $UB^{(k)} > LB^{(k)} + \epsilon_{\text{opt}}$ **do** ▷ Solving $\big(\text{Inv}(\lambda)\big)$ with a Benders decomposition algorithm

    $k \leftarrow k + 1$

    Solve $(\text{RMP})^{(k)}$ and retrieve an optimal solution $\big(\check{\mathbf{x}}^{(k)}, \check{\boldsymbol{\theta}}^{(k)}\big)$

    $\text{LB}^{(k)} \leftarrow \mathbf{c}^\top \check{\mathbf{x}}^{(k)} + \sum_{s\in\mathcal{S}} p_s \check{\theta}_s^{(k)}$

    for $s \in \mathcal{S}$ **do**

      Solve $(SP_s(\check{\mathbf{x}}^{(k)}))$ and retrieve an optimal solution $\big(\mathbf{y}_s, \mathbf{u}_s\big)$, its value $\phi_s(\check{\mathbf{x}}^{(k)})$, and $\boldsymbol{\pi}_s \in \text{Vert}(\Pi_s)$

      Add $\theta_s \geq \boldsymbol{\pi}_s^\top (\mathbf{b}_s - T\mathbf{x})$ to $(\text{RMP})^{(k)}$

    $\text{UB}^{(k)} \leftarrow \min\big\{\text{UB}^{(k-1)}, \mathbf{c}^\top \check{\mathbf{x}}^{(k)} + \sum_{s\in\mathcal{S}} p_s \phi_s(\check{\mathbf{x}}^{(k)})\big\}$

    $(\text{RMP})^{(k+1)} \leftarrow (\text{RMP})^{(k)}$

    if $\texttt{updateBestBilevelSolution}\big(\texttt{acc}, \mathbf{x}^{(k)}, \mathbf{u}, z_{\text{best}}, \text{LB}^{(k)}, \text{UB}^{(k)}\big)$ **then**

      $\mathbf{x}_{\text{best}} \leftarrow \check{\mathbf{x}}^{(k)}$, $z_{\text{best}} \leftarrow \mathbf{c}^\top \check{\mathbf{x}}^{(k)} + \sum_{s\in\mathcal{S}} p_s \phi_s(\check{\mathbf{x}}^{(k)})$

      $\texttt{x\_best\_updated} \leftarrow \texttt{True}$

      $\Lambda_{\max} \leftarrow \Lambda$

  if $\neg\texttt{x\_best\_updated}$ **then**

    $\Lambda_{\min} \leftarrow \Lambda$

  else

    if $\Lambda = 0$ **then return** $\mathbf{x}_{\text{best}}$ ▷ $\mathbf{x}_{\text{best}}$ is an optimal solution

  if $\Lambda_{max} > 100 * \Lambda_{min}$ **then**

    if $\Lambda_{min} \neq 0$ **then** $\Lambda \leftarrow 10 * \Lambda_{\min}$ **else** $\Lambda \leftarrow 0.1 * \Lambda_{\max}$

  else

    $\Lambda \leftarrow \tau\Lambda_{\min} + (1-\tau)\Lambda_{\max}$

return $\mathbf{x}_{best}$ ▷ $\mathbf{x}_{\text{best}}$ is a heuristic solution

Function $\texttt{updateBestBilevelSolution}\big(\texttt{acc}, \mathbf{x}, \mathbf{u}, z, LB, UB\big)$**:**

  if $\mathbf{c}^\top \mathbf{x} + \sum_{s\in\mathcal{S}} p_s \phi_s(\mathbf{x}) \geq z$ **then** ▷ Check if the given solution is improving

    return $\texttt{False}$

  if $\sum_{s\in\mathcal{S}} p_s \sum_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} [u_{s,i,t} > \epsilon] \leq \alpha$ **then** ▷ Check if the given solution is bilevel feasible

    return $\texttt{True}$

  else

    if $\texttt{acc} = \textbf{MinENS}$ or $\big(\texttt{acc} = \textbf{MinENS-Opt}$ and $UB - LB \leq \epsilon_{\text{opt}}\big)$ **then**

      for $s \in \mathcal{S}$ **do**

        Solve $(\text{Aux}_s(\mathbf{x}))$ and retrieve $\varphi_s(\mathbf{x})$ its optimal value

      if $\sum_{s\in\mathcal{S}} p_s \varphi_s(\mathbf{x}) \leq \alpha$ **then** ▷ Check if there exists a bilevel feasible solution with the same lower-level cost

        return $\texttt{True}$

  return $\texttt{False}$

The Benders decomposition algorithm may suffer, as every cutting-plane algorithm, from strong oscillations of the first-stage variables, and thus may compute, in the early iterations, cuts that exclude solutions that are far away from the optimal solution. To speed-up the solution of (Inv(Λ)), we use an in-out stabilization scheme (Ben-Ameur and Neto, 2007). At each iteration, we solve the subproblems for a first-stage solution $\mathbf{x}^{(k)}$ (hereafter referred to as the *separation* point) that is different from the current optimal first-stage solution $\check{\mathbf{x}}^{(k)}$ to (RMP$^{(k)}$). Specifically, we have $\mathbf{x}^{(k)} = \kappa\check{\mathbf{x}}^{(k)} + (1-\kappa)\hat{\mathbf{x}}^{(k)}$ where $\kappa \in (0,1]$ and $\hat{\mathbf{x}}^{(k)} = \arg\min_{j\in[\![0,k-1]\!]} \{\mathbf{c}^\top\mathbf{x}^{(j)} + \sum_{s\in\mathcal{S}} p_s\phi_s(\mathbf{x}^{(j)})\}$ is the separation point (among those calculated so far) with the smallest objective function value. The point $\hat{\mathbf{x}}^{(k)}$ is referred to as a *stability center*. The *in-out approach* creates a sequence of stability centers with decreasing objective values converging to an optimal solution to the problem. We use a dynamic strategy to update the stabilization parameter κ. If $\mathbf{c}^\top\mathbf{x}^{(k)} + \sum_{s\in\mathcal{S}} p_s\phi_s(\mathbf{x}^{(k)}) < \mathbf{c}^\top\hat{\mathbf{x}}^{(k)} + \sum_{s\in\mathcal{S}} p_s\phi_s(\hat{\mathbf{x}}^{(k)})$, then the separation point has a lower cost than the current stability center. If we had separated farther, we could have found an even better point, so we increase κ with the rule $\kappa \leftarrow \min\{1.0, 1.2\kappa\}$. If $\mathbf{c}^\top\mathbf{x}^{(k)} + \sum_{s\in\mathcal{S}} p_s\phi_s(\mathbf{x}^{(k)}) \geq \mathbf{c}^\top\hat{\mathbf{x}}^{(k)} + \sum_{s\in\mathcal{S}} p_s\phi_s(\hat{\mathbf{x}}^{(k)})$, we did not stabilize enough, and we therefore decrease the stabilization parameter κ with the rule $\kappa \leftarrow \max\{0.1, 0.8\kappa\}$. In our experiments, we initialize κ to 0.5.

To avoid overloading Algorithm 1, it does not contain the in-out stabilization scheme previously described. For the sake of reproducibility, let us indicate that the only modification in the algorithm would be to compute $\mathbf{x}^{(k)}$ before line 12, to replace $\check{\mathbf{x}}^{(k)}$ by $\mathbf{x}^{(k)}$ from line 12 to line 16, and to update κ when computing UB$^{(k)}$ in line 16.

5 Computational experiments

We performed computational experiments on a set of small-sized and a set of large-sized instances. These instances were randomly generated according to the procedure described in §5.1.

In this section, we refer to our matheuristic with the parameter `acc` equal to **Rand**, **MinENS-Opt**, and **MinENS** to **H_Rand**, **H_MinENS-Opt**, and **H_MinENS**, respectively.

The first aim of our experiments is to justify the relevance of the general scheme we use in our matheuristic on the basis of empirical results (see §5.2).

The second aim of our computational experiments is to assess the quality of the solutions to Bilevel-α-Stoch-GTEPP computed by our matheuristic presented in Section 4. For this purpose, we compare the results of the matheuristic with those given by the following algorithms:

KKT_SOS1: We solve the KKT-reformulation of the bilevel program (6) (or equivalently (8)) using SOS1 to enforce complementarity constraints.

HPR-Aux: We solve the HPR of the bilevel problem. Specifically, we solve (3) (or equivalently program (8) where for each scenario $s \in \mathcal{S}$ constraints (8c) are replaced by constraints $\mathbf{y}_s \in \mathbb{R}_+^{\ell_s} \times \mathbb{R}^{\ell'_s}$, $\mathbf{u}_s \in \mathbb{R}_+^{|\mathcal{N}|\times|\mathcal{T}|}$, and $W\mathbf{y}_s + W'\mathbf{u}_s \geq \mathbf{b}_s - T\mathbf{x}$). Denoting $\mathbf{x}^{\text{HPR}}$ the investment decisions in the best solution computed within the time limit, we compute $\sum_{s\in\mathcal{S}} p_s\varphi_s(\mathbf{x}^{\text{HPR}}) \leq \alpha$ to check whether $\mathbf{x}^{\text{HPR}}$ induces a feasible solution to Bilevel-α-Stoch-GTEPP or not.

woLOLE-Aux: We solve the two-stage problem without the LOLE reliability criterion using a multicut Benders decomposition algorithm similar to the one presented in §4.2 except for the absence of constraints (11c). Specifically, we solve (2) (or equivalently problem (11) that is the multicut Benders reformulation of program (8) in which constraints (8d), (8e) and (8f) are dropped and where for each scenario $s \in \mathcal{S}$ constraints (8c) are replaced by constraints $\mathbf{y}_s \in \mathbb{R}_+^{\ell_s} \times \mathbb{R}^{\ell'_s}$, $\mathbf{u}_s \in \mathbb{R}_+^{|\mathcal{N}|\times|\mathcal{T}|}$, and $W\mathbf{y}_s + W'\mathbf{u}_s \geq \mathbf{b}_s - T\mathbf{x}$). Denoting $\mathbf{x}^{\text{woLOLE}}$ the investment decisions in the best solution computed within the time limit, we compute $\sum_{s\in\mathcal{S}} p_s\varphi_s(\mathbf{x}^{\text{woLOLE}}) \leq \alpha$ to check whether $\mathbf{x}^{\text{woLOLE}}$ induces a feasible solution to Bilevel-α-Stoch-GTEPP or not.

0_LOLE: We solve the two-stage problem with the additional restriction that the demand should always be satisfied. Solutions to this problem are feasible solutions to Bilevel-α-Stoch-GTEPP .

Algorithms **HPR-Aux** and **woLOLE-Aux** rely on solving a relaxation of the Bilevel-α-Stoch-GTEPP. If the investment decisions in an optimal solution to these problems induce a feasible solution to Bilevel-α-Stoch-GTEPP , then this latter is an optimal solution to Bilevel-α-Stoch-GTEPP. When these two algorithms do not converge within the time limit, we retrieve the best lower bound they computed. The results of this comparison are presented in §5.3. We only considered the small-sized instances due to the computational complexity of optimally solving the bilevel problem.

The third aim of our experiments is to determine which configuration of our matheuristic, among **H_Rand**, **H_MinENS-Opt**, or **H_MinENS** is best suited for computing solutions for large instances. We discuss this in §5.4.

We implemented all algorithms with the Julia programming language version 1.6.1. We used the package JuMP version 1.11.1 as a modeler, and CPLEX 22.1.1 as the mixed integer linear programming solver. We set $\tau = 0.5$ for the binary search parameter, $\epsilon_{\text{opt}} = 10^{-6}$ and $\epsilon_{\text{Inv}} = 10^{-6}$. We used the package BilevelJuMP version 0.6.0 (Garcia et al., 2022) with CPLEX 22.1.1 for the implementation of **KKT_SOS1**. We ran all our experiments on a 2 Deca-core Ivy-Bridge Haswell Intel Xeon E5-2680 v3 server running at 2.50 GHz. The 128 GB of available RAM was shared between six copies of the algorithm running in parallel on the server. Each instance is solved by one copy of the algorithm using a single thread. The CPU time limit was fixed to **two hours** for every algorithm.

The detailed results for every experiment are available in C. In all the tables, the CPU time is given in seconds.

5.1 Instances

Our instance generation procedure is divided into two parts. First, we generate the undirected graph G representing the approximated power grid. Then, on this graph, we generate investment and operational data.

In preliminary experiments, we noticed that classical graph generation algorithms such as the ones proposed in (Erdos and Rényi, 1960; Hakimi, 1962; Viger and Latapy, 2005) did not produce satisfying results. Either the produced graphs did not model well the classical structure of power grids, or they were not similar to networks generated based on geographical positions. To address this issue, we propose a dedicated graph generation method. This method is composed of three parts. First, we sample node positions in a plane. Then we add edges between the nodes until the graph has a single connected component, and we finally remove edges until the given graph density is reached. We report in B our graph generation procedure.

We generated two sets of instances using different values for the number of nodes, number of scenarios, and number of time units (see Table 1). For each combination of parameters ($|\mathcal{N}|$, $|\mathcal{S}|$, $|\mathcal{T}|$), we generate one random instance, denoted $\mathsf{N}|\mathcal{N}|\text{-}\mathsf{S}|\mathcal{S}|\text{-}\mathsf{T}|\mathcal{T}|$, using the parameters of Table 2. In our experiments, we set V either to 1000 or to 5000.

Parameter	Value
$\lvert\mathcal{N}\rvert$	$\{5, 10\}$
$\lvert\mathcal{S}\rvert$	$\{5, 10, 20\}$
$\lvert\mathcal{T}\rvert$	$\{3, 6, 12\}$

(a) 18 small-sized instances

Parameter	Value
$\lvert\mathcal{N}\rvert$	$\{20, 30\}$
$\lvert\mathcal{S}\rvert$	$\{100, 200\}$
$\lvert\mathcal{T}\rvert$	$\{24, 48\}$

(b) 8 large-sized instances

Table 1: Our two sets of instances

Parameter	Value
c_i^{prod}	$U(1000, 8000)$
c_a^{line}	$U(1000, 8000)$
ϵ	10^{-3}
α	$\begin{cases} 1 & \text{if } \lvert\mathcal{T}\rvert \leq 6 \\ 2 & \text{if } \lvert\mathcal{T}\rvert = 12 \\ 3 & \text{if } \lvert\mathcal{T}\rvert = 24 \\ 4 & \text{if } \lvert\mathcal{T}\rvert = 48 \end{cases}$
$g_{s,i,t}^{\text{prod}}$	$U(20, 80)$
$g_{s,a,t}^{\text{flow}}$	1
P^{init}	$U(0, 100)$
F_a^{init}	$U(0, 50)$
Δ_i	0.3
$d_{s,i,t}$	$U(50, 200)$

Table 2: Parameters for the data generation

5.2 Empirical study of the LOLE evolution as a function of the minimum investment cost constraint

Our matheuristic uses a binary search on the total investment cost Λ and and is based on the intuitive idea that increasing capital expenditure reduces the LOLE. Let $\Phi(\Lambda)$ be the minimum value of the LOLE for an optimal solution to (Inv(Λ)). It is important to note that function Φ may not be monotonous in general (see A).

To illustrate that function Φ is close to being monotonically non-increasing, we report the evolution of the LOLE (the expected number of time units with ENS) according to the minimum investment value Λ on two instances (with V=1000). In Figure 10, for a large number of Λ values, we show the value $\Phi(\Lambda) = \sum_{s\in\mathcal{S}} \varphi_s(\mathbf{x}^*)$ ("Minimum" scatter plot) for an optimal solution $\mathbf{x}^*$ to the minimum investment problem ($Inv(\Lambda)$). Denoting $(\mathbf{y}_s^*, \mathbf{u}_s^*)$ the optimal solution to $(SP_s(\mathbf{x}^*))$ returned by the solver for every $s \in \mathcal{S}$, we also report the value $\sum_{s\in\mathcal{S}} p_s(\sum_{i\in\mathcal{N}}\sum_{t\in\mathcal{T}}[u_{s,i,t}^* > \epsilon])$ ("Random"[9] scatter plot) and the optimal value of $(\text{Aux}_s(\mathbf{x}^*))$ if we change the optimization direction ("Maximum" scatter plot). Although we observe that the function Φ is not perfectly a monotonically non-increasing function, there is a clear correlation between a larger value of Λ and a smaller LOLE value. We believe this provides a rationale for the use of the proposed matheuristic in order to find a good feasible solution to the bilevel problem. We also observe that for many investment values, there exists several equivalent optimal solutions, as the difference between the minimum and the maximum expected number of time units with ENS is strictly positive and usually large. Interestingly, the solution returned by the solver has a LOLE close to the smallest possible LOLE but the difference is not zero most of the times. This shows that solving the auxiliary subproblems may help to find better solutions and motivates the introduction of the configuration $\texttt{acc} = \textbf{minENS}$ and $\texttt{acc} = \textbf{minENS} - \textbf{Opt}$ in our heuristic.

[9]The lower-level solution used to compute the LOLE is a random optimal solution to the lower-level program returned by the solver.

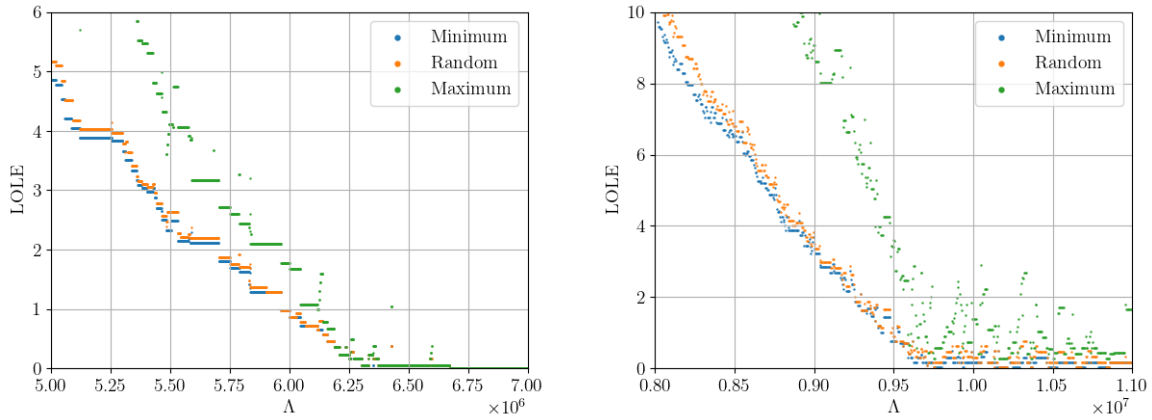


Figure 10: Evolution of the LOLE according to the minimum investment cost Λ, for instance N5-S10-T6 (left) and instance N10-S10-T12 (right).

5.3 Numerical results on small-sized instances

To assess the quality of the solutions to Bilevel-α-Stoch-GTEPP computed by our matheuristic, we considered the small-sized instances and compare the results of the matheuristic with those given by the algorithms **KKT_SOS1**, **HPR-Aux**, and **woLOLE-Aux**.

Let M be the set of methods we compare. We denote by $\mathrm{LB}_{\mathsf{p},\mathsf{m}}$ and $z_{\mathsf{p},\mathsf{m}}$ the best lower bound value and the best feasible solution to Bilevel-α-Stoch-GTEPP computed within the time limit with method m on instance p ($z_{\mathsf{p},\mathsf{m}} = \infty$ when no feasible solution to Bilevel-α-Stoch-GTEPP is computed). We set $\mathrm{LB}_{\mathsf{p},\mathsf{m}} = 0$ when m is our matheuristic as this latter does not provide a lower bound value. For each instance p, we compute $\overline{\mathrm{LB}}_{\mathsf{p}} = \max_{\mathsf{m}\in\mathsf{M}}\{\mathrm{LB}_{\mathsf{p},\mathsf{m}}\}$ and $\underline{z}_{\mathsf{p}} = \min_{\mathsf{m}\in\mathsf{M}}\{z_{\mathsf{p},\mathsf{m}}\}$ that correspond to the best lower bound value and to the value of the best feasible solution to Bilevel-α-Stoch-GTEPP obtained among all methods. When $\overline{\mathrm{LB}}_{\mathsf{p}} = \underline{z}_{\mathsf{p}}$, then $\underline{z}_{\mathsf{p}}$ is the optimal value of instance p and we denote it z^*_{p}. When $z_{\mathsf{p},\mathsf{m}} \neq \infty$, we compute the gap with respect to the best lower bound value as $\frac{z_{\mathsf{p},\mathsf{m}} - \overline{\mathrm{LB}}_{\mathsf{p}}}{\overline{\mathrm{LB}}_{\mathsf{p}}}$, the gap with respect to the best computed solution as $\frac{z_{\mathsf{p},\mathsf{m}} - \underline{z}_{\mathsf{p}}}{\underline{z}_{\mathsf{p}}}$, and the gap to the optimal solution as $\frac{z_{\mathsf{p},\mathsf{m}} - z^*_{\mathsf{p}}}{z^*_{\mathsf{p}}}$ when z^*_{p} exists.

Table 3 displays the results of the comparison. For each instance p and two values of V, it reports the CPU time and the gap between the value of the computed solution and $\overline{\mathrm{LB}}_{\mathsf{p}}$ (the gap is underlined when $\overline{\mathrm{LB}}_{\mathsf{p}} = z^*_{\mathsf{p}}$ and the symbol '×' means that the solution produced by the algorithm is not a feasible solution to Bilevel-α-Stoch-GTEPP). First, we observe that **KKT_SOS1** is able to compute an optimal solution within the two hour time-limit for 1 or 3 instances depending on the value of V. For all other instances, it does not even return a feasible solution. This shows the complexity of exactly solving the bilevel problem. We also observe that solving the HPR of this problem is already challenging as around half of the instances are not optimally solved.

Second, calculating feasible solutions is not straightforward for our set of instances. Method **woLOLE-Aux** compute feasible solutions for very few instances which shows that taking the LOLE reliability criterion a posteriori is not a suitable approach. Interestingly, we see that **HPR-Aux** compute feasible solutions for very few instances, which empirically confirms Remark 2. In contrast, our matheuristic always computes feasible solution. This makes it attractive in an operational setting. Finally, we note that forcing investment decisions to fully satisfy demands significantly increases the total cost and leads to low-quality solutions to Bilevel-α-Stoch-GTEPP.

Third, we observe that our matheuristic computes the optimal solution for one instance and solutions with a gap that is reasonably small but still significant for the other 7 instances for which the optimum is known. The gap is smaller when V is larger. This may be explained by the fact that investments decisions made by ignoring the LOLE reliability criterion but considering a minimum investment constraint tend to have a smaller number of nodes with ENS, which facilitates the identification of a feasible solution. For the other instances, we do not expect the lower bound to be strong, but the last column of the table gives a first evaluation of the quality of the solutions returned by our heuristic. The average gap is very large when V is equal to 1000, but answering

the question of whether this is only due to the quality of our heuristic or related to the weakness of this bound is out of our reach at this moment.

		CPU Time					Gap to $\overline{\text{LB}}_{\mathsf{p}}$				
V	Instance	**KKT_SOS1**	**HPR-Aux**	**woLOLE-Aux**	**0_LOLE**	**H_MinENS**	**KKT_SOS1**	**HPR-Aux**	**woLOLE-Aux**	**0_LOLE**	**H_MinENS**
1000	N5-S5-T3	T.L.	22	16	16	23	-	×	×	22.8	**17.9**
	N5-S5-T6	95	43	15	16	21	**0.0**	**0.0**	×	8.2	3.8
	N5-S5-T12	T.L.	24	16	15	24	-	**0.0**	×	5.8	1.6
	N5-S10-T3	T.L.	3469	16	15	23	-	×	×	8.8	**2.1**
	N5-S10-T6	T.L.	115	16	17	23	-	**0.0**	×	10.7	5.2
	N5-S10-T12	T.L.	T.L.	17	16	27	-	×	×	23.1	**10.3**
	N5-S20-T3	T.L.	T.L.	16	16	22	-	**3.3**	×	26.2	7.1
	N5-S20-T6	T.L.	T.L.	17	16	26	-	×	×	27.8	**13.5**
	N5-S20-T12	T.L.	T.L.	15	16	28	-	×	×	31.3	**10.3**
	N10-S5-T3	T.L.	61	15	16	22	-	×	×	8.7	**2.2**
	N10-S5-T6	T.L.	821	16	15	25	-	×	×	9.4	**2.4**
	N10-S5-T12	T.L.	T.L.	16	15	33	-	×	×	4.9	**1.6**
	N10-S10-T3	T.L.	T.L.	16	16	25	-	×	×	33.3	**24.9**
	N10-S10-T6	T.L.	T.L.	16	17	28	-	×	×	21.5	**18.6**
	N10-S10-T12	T.L.	T.L.	18	16	37	-	×	×	19.6	**11.7**
	N10-S20-T3	T.L.	T.L.	16	16	29	-	×	×	27.6	**18.0**
	N10-S20-T6	T.L.	T.L.	17	16	33	-	×	×	18.3	**10.8**
	N10-S20-T12	T.L.	T.L.	17	19	67	-	×	×	18.2	**11.5**
5000	N5-S5-T3-13	T.L.	19	17	16	23	-	×	×	8.5	**6.0**
	N5-S5-T6	34	25	18	16	22	**0.0**	**0.0**	×	6.6	2.8
	N5-S5-T12	64	20	16	15	24	**0.0**	**0.0**	×	3.0	0.2
	N5-S10-T3	T.L.	23	16	15	22	-	×	×	5.0	**0.2**
	N5-S10-T6	T.L.	26	17	17	24	-	**0.0**	×	6.6	1.9
	N5-S10-T12	T.L.	36	16	16	31	-	×	×	8.0	**0.2**
	N5-S20-T3	T.L.	1389	16	16	24	-	**0.0**	×	19.4	2.9
	N5-S20-T6	T.L.	T.L.	16	16	27	-	×	×	12.1	**1.4**
	N5-S20-T12	T.L.	T.L.	16	16	36	-	×	×	20.5	**2.9**
	N10-S5-T3	T.L.	23	16	16	24	-	×	×	7.0	**1.2**
	N10-S5-T6	T.L.	29	17	15	26	-	×	×	6.0	**1.0**
	N10-S5-T12	167	19	16	15	24	**0.0**	**0.0**	**0.0**	2.0	**0.0**
	N10-S10-T3	T.L.	235	17	16	26	-	×	×	7.9	**1.7**
	N10-S10-T6	T.L.	3696	17	17	35	-	×	×	6.8	**2.2**
	N10-S10-T12	T.L.	T.L.	17	16	37	-	×	×	8.4	**1.5**
	N10-S20-T3	T.L.	T.L.	17	16	31	-	×	×	10.0	**3.2**
	N10-S20-T6	T.L.	T.L.	17	16	33	-	×	×	8.4	**1.8**
	N10-S20-T12	T.L.	T.L.	17	19	67	-	×	×	7.0	**1.2**

T.L.: time limit reached

Table 3: Results comparison for methods **KKT_SOS1**, **HPR-Aux**, **0_LOLE**, **woLOLE-Aux**, and **H_MinENS** for all small-sized instances and for a VOLL equal to 1000 and 5000.

5.4 Comparison of the three versions of the matheursitic

Table 4 and 5 display the results of the comparison between **H_Rand**, **H_MinENS-Opt**, and **H_MinENS** for the small-sized and large-sized instances. For each instance p and two values of V, it reports the gap between the value of the computed solution and $\underline{z}_{\mathsf{p}}$. For the largest instances, it also reports the CPU Time.

For the small-sized instances, similar results are obtained with **H_MinENS-Opt**, and **H_MinENS** (except for 2 instances), while **H_Rand** provides solutions that are 0.2% and 0.3% worse on average when V is equal to 1000 and 5000. For the large-sized instances, **H_MinENS** provides significantly better results than **H_MinENS-Opt**, especially when the VOLL is 1000, but at the cost of a much larger CPU time. **H_MinENS-Opt** is a compromise (compared to the two other versions) between solution quality and calculation time. Finally, comparing the results of **H_Rand** with those of the other two versions shows that it is advantageous in terms of solution quality to compute for each subproblem the solution with the minimum number of time units with ENS, and the gain increases with the size of the instance.

		Gap to $\underline{z}_{\mathsf{p}}$		
VOLL	Instance	**H_Rand**	**H_MinENS-Opt**	**H_MinENS**
1000	N5-S5-T3	2.2	**0.0**	**0.0**
	N5-S5-T6	**3.8**	**3.8**	**3.8**
	N5-S5-T12	2.0	**1.6**	**1.6**
	N5-S10-T3	0.1	**0.0**	**0.0**
	N5-S10-T6	**5.2**	**5.2**	**5.2**
	N5-S10-T12	0.4	0.3	**0.0**
	N5-S20-T3	3.8	**3.7**	**3.7**
	N5-S20-T6	**0.0**	**0.0**	**0.0**
	N5-S20-T12	0.8	**0.0**	**0.0**
	N10-S5-T3	0.8	**0.0**	**0.0**
	N10-S5-T6	**0.0**	**0.0**	**0.0**
	N10-S5-T12	**0.0**	**0.0**	**0.0**
	N10-S10-T3	**0.0**	**0.0**	**0.0**
	N10-S10-T6	0.2	0.1	**0.0**
	N10-S10-T12	**0.0**	**0.0**	**0.0**
	N10-S20-T3	**0.0**	**0.0**	**0.0**
	N10-S20-T6	0.3	**0.0**	**0.0**
	N10-S20-T12	0.6	**0.0**	**0.0**
	Average	1.1	0.8	0.8
5000	N5-S5-T3	0.9	**0.0**	**0.0**
	N5-S5-T6	**2.8**	**2.8**	**2.8**
	N5-S5-T12	**0.2**	**0.2**	**0.2**
	N5-S10-T3	0.1	**0.0**	**0.0**
	N5-S10-T6	**1.9**	**1.9**	**1.9**
	N5-S10-T12	0.2	**0.0**	**0.0**
	N5-S20-T3	**2.8**	**2.8**	**2.8**
	N5-S20-T6	**0.0**	**0.0**	**0.0**
	N5-S20-T12	0.6	**0.0**	**0.0**
	N10-S5-T3	**0.0**	**0.0**	**0.0**
	N10-S5-T6	**0.0**	**0.0**	**0.0**
	N10-S5-T12	**0.0**	**0.0**	**0.0**
	N10-S10-T3	0.1	**0.0**	**0.0**
	N10-S10-T6	0.6	**0.0**	**0.0**
	N10-S10-T12	**0.0**	**0.0**	**0.0**
	N10-S20-T3	**0.0**	**0.0**	**0.0**
	N10-S20-T6	**0.0**	**0.0**	**0.0**
	N10-S20-T12	0.2	**0.0**	**0.0**
	Average	0.6	0.4	0.4

Table 4: Results comparison for **H_Rand**, **H_MinENS-Opt**, and **H_MinENS** for all small-sized instances and for a VOLL equal to 1000 and 5000.

		CPU Time			Gap to $\underline{z}_{\mathsf{p}}$		
VOLL	Instance	**H_Rand**	**H_MinENS-Opt**	**H_MinENS**	**H_Rand**	**H_MinENS-Opt**	**H_MinENS**
1000	N20-S100-T24	324	599	1408	0.6	0.3	**0.0**
	N20-S100-T48	437	1280	3721	0.1	**0.0**	**0.0**
	N20-S200-T24	597	1203	2976	0.5	**0.0**	**0.0**
	N20-S200-T48	1131	2950	7859	0.6	0.2	**0.0**
	N30-S100-T24	693	1293	3282	1.1	1.0	**0.0**
	N30-S100-T48	1157	2343	11124	0.8	0.8	**0.0**
	N30-S200-T24	1647	3029	7069	0.4	**0.0**	**0.0**
	N30-S200-T48	3529	7327	21225	1.0	0.2	**0.0**
	Average	1189	2503	7333	0.6	0.3	0.0
5000	N20-S100-T24	321	552	1114	0.3	0.1	**0.0**
	N20-S100-T48	491	1103	2631	0.1	**0.0**	**0.0**
	N20-S200-T24	571	1135	2291	0.4	**0.0**	**0.0**
	N20-S200-T48	1510	2863	6015	0.1	**0.0**	**0.0**
	N30-S100-T24	783	1245	2658	**0.0**	**0.0**	**0.0**
	N30-S100-T48	1300	2700	7607	0.1	**0.0**	**0.0**
	N30-S200-T24	2142	2990	5639	0.3	0.1	**0.0**
	N30-S200-T48	4101	6576	16011	0.1	**0.0**	**0.0**
	Average	1403	2395	5496	0.2	0.0	0.0

Table 5: Results comparison for methods **H_Rand**, **H_MinENS-Opt**, and **H_MinENS** for all large-sized instances and for a VOLL equal to 1000 and 5000.

6 Conclusion

We presented in this paper how to model a stochastic transmission and generation investment planning problem with a LOLE reliability criterion. We showed that an intuitive addition of the constraint into the extensive form of the two-stage stochastic program may lead to a time-inconsistent model. Time-inconsistency arises when second-stage decisions taken in the evaluation of the first stage decisions are not the one that would be implemented when the uncertainty is revealed. To overcome this issue and produce a valid model, we presented a bilevel model, where the upper level problem include investment decisions and the LOLE reliability criterion, and the lower level problem decomposes into an operational subproblem for every scenario. We showed that the ratio between the value of an optimal solution to the high-point relaxation and the optimal value of the lower-level problem obtained by fixing the upper-level variables to their optimal value cannot be bounded by a constant value.

We tested solving the KKT reformulation of the bilevel model: even small-sized instances proved challenging for this method. To compute feasible solutions, we introduced a matheuristic based on iteratively solving stochastic expansion planning problems with a minimum investment constraint. We compared it with two alternative methods which solve a relaxation of the problem and then check whether it is possible to deduce a feasible solution from the computed solution. Our experiments showed that these approaches are often unable to produce feasible solutions. We also empirically observed that the solutions to a restriction of the problem obtained by imposing the satisfaction of the demand are significantly worse than the solutions computed by our matheuristic.

Future research on this problem should focus on solving exactly the high-point relaxation in a time efficient way and on designing advanced decomposition-based methods that solve exactly the problem.

Another important path to explore is the study of alternative criteria for ensuring power system reliability. The LOLE reliability criterion only guarantees that on average (in expectation) the number of time units with ENS is below a given threshold. However, this latter number can be large for scenarios with low probability (i.e., intense cold wave combined with low wind and low nuclear availability). Chance constrained optimization or robust optimization with recourse are challenging avenues to explore for defining a more risk-averse criterion. A comparative study could lead to a change in regulations if a new criterion would prove more suitable.

Acknowledgments

This project has been funded by RTE (Réseau de Transport d'Electricité), the French transmission system operator, through the projects Antares and Antares Xpansion: https://github.com/AntaresSimulatorTeam/antares-xpansion, which are used for long-term adequacy studies. The experiments presented in this paper were carried out using the Federative Platform for Research in Computer Science and Mathematics (PlaFRIM), created under the Inria PlaFRIM development action with support from Institut Polytechnique de Bordeaux, Laboratoire Bordelais de Recherche en Informatique, Institut de Mathématiques de Bordeaux, Conseil Régional d'Aquitaine, Université de Bordeaux, Centre National de la Recherche Scientifique, and Agence Nationale de la Recherche in accordance with the "Programme d'Investissements d'Avenir" (see www.plafrim.fr/en/home).

References

Alizadeh, B. and Jadid, S. (2011). Reliability constrained coordination of generation and transmission expansion planning in power systems using mixed integer programming. *IET Generation, Transmission & Distribution*, 5(9):948–960.

Atakan, S. and Sen, S. (2018). A progressive hedging based branch-and-bound algorithm for mixed-integer stochastic programs. *Computational Management Science*, 15(3-4):501–540.

Ben-Ameur, W. and Neto, J. (2007). Acceleration of cutting-plane and column generation algorithms: Applications to network design. *Networks*, 49(1):3–17.

Cho, S., Li, C., and Grossmann, I. E. (2022). Recent advances and challenges in optimization models for expansion planning of power systems and reliability optimization. *Computers & Chemical Engineering*, page 107924.

Cho, S., Tovar-Facio, J., and Grossmann, I. E. (2023). Disjunctive optimization model and algorithm for long-term capacity expansion planning of reliable power generation systems. *Computers & Chemical Engineering*, page 108243.

Choi, J., Tran, T., El-Keib, A., Thomas, R., Oh, H., and Billinton, R. (2005). A method for transmission system expansion planning considering probabilistic reliability criteria. *IEEE Transactions on Power Systems*, 20(3):1606–1615.

Erdos, P. and Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–61.

Fischetti, M., Ljubić, I., Monaci, M., and Sinnl, M. (2018). On the use of intersection cuts for bilevel optimization. *Mathematical Programming*, 172(1):77–103.

Fortuny-Amat, J. and McCarl, B. (1981). A Representation and Economic Interpretation of a Two-Level Programming Problem. *Journal of the Operational Research Society*, 32(9):783–792.

Garces, L. P., Conejo, A. J., Garcia-Bertrand, R., and Romero, R. (2009). A Bilevel Approach to Transmission Expansion Planning Within a Market Environment. *IEEE Transactions on Power Systems*, 24(3):1513–1522.

Garcia, J. D., Bodin, G., and Street, A. (2022). BilevelJuMP.jl: Modeling and Solving Bilevel Optimization in Julia. *arXiv:2205.02307*.

Grimm, V., Martin, A., Schmidt, M., Weibelzahl, M., and Zöttl, G. (2016). Transmission and generation investment in electricity markets: The effects of market splitting and network fee regimes. *European Journal of Operational Research*, 254(2):493–509.

Hakimi, S. L. (1962). On Realizability of a Set of Integers as Degrees of the Vertices of a Linear Graph. I. *Journal of the Society for Industrial and Applied Mathematics*, 10(3):496–506.

Hemmati, R., Hooshmand, R.-A., and Khodabakhshian, A. (2013). State-of-the-art of transmission expansion planning: Comprehensive review. *Renewable and Sustainable Energy Reviews*, 23:312–319.

Jin, S. and Ryan, S. M. (2011). Capacity Expansion in the Integrated Supply Network for an Electricity Market. *IEEE Transactions on Power Systems*, 26(4):2275–2284.

Kleinert, T., Labbé, M., Ljubić, I., and Schmidt, M. (2021). A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization. *EURO Journal on Computational Optimization*, 9:100007.

Kleinert, T., Labbé, M., Plein, F., and Schmidt, M. (2020). Technical Note—There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization. *Operations Research*, 68(6):1716–1721.

Kleinert, T. and Schmidt, M. (2019). Global optimization of multilevel electricity market models including network design and graph partitioning. *Discrete Optimization*, 33:43–69.

Lara, C. L., Mallapragada, D. S., Papageorgiou, D. J., Venkatesh, A., and Grossmann, I. E. (2018). Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. *European Journal of Operational Research*, 271(3):1037–1054.

Lara, C. L., Siirola, J. D., and Grossmann, I. E. (2020). Electric power infrastructure planning under uncertainty: stochastic dual dynamic integer programming (sddip) and parallelization scheme. *Optimization and Engineering*, 21(4):1243–1281.

Lei, Y., Zhang, P., Hou, K., Jia, H., Mu, Y., and Sui, B. (2018). An Incremental Reliability Assessment Approach for Transmission Expansion Planning. *IEEE Transactions on Power Systems*, 33(3):2597–2609.

Leite da Silva, A. M., Rezende, L. S., Manso, L. A. F., and Anders, G. J. (2010). Transmission expansion planning: A discussion on reliability and "N-1" security criteria. In *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, pages 244–251.

Li, C., Conejo, A. J., Liu, P., Omell, B. P., Siirola, J. D., and Grossmann, I. E. (2022). Mixed-integer linear programming models and algorithms for generation and transmission expansion planning of power systems. *European Journal of Operational Research*, 297(3):1071–1082.

Liu, X., Küçükyavuz, S., and Luedtke, J. (2016). Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming*, 157(1):219–243.

Lozano, L. and Smith, J. C. (2017). A Value-Function-Based Exact Approach for the Bilevel Mixed-Integer Programming Problem. *Operations Research*, 65(3):768–786.

Micheli, G., Vespucci, M. T., Stabile, M., Puglisi, C., and Ramos, A. (2020). A two-stage stochastic MILP model for generation and transmission expansion planning with high shares of renewables. *Energy Systems*.

Nahmmacher, P., Schmid, E., Hirth, L., and Knopf, B. (2016). Carpe diem: A novel approach to select representative days for long-term power system modeling. *Energy*, 112:430–442.

Palmintier, B. S. and Webster, M. D. (2013). Heterogeneous unit clustering for efficient operational flexibility modeling. *IEEE Transactions on Power Systems*, 29(3):1089–1098.

RTE (2019). French transmission network development plan. Technical report, RTE, Report. https://assets.rte-france.com/prod/public/2020-07/Sch%C3%A9ma%20d%C3%A9cennal%20de%20d%C3%A9veloppement%20de%20r%C3%A9seau%202019%20-%20Synth%C3%A8se%20%E2%80%93%20English%20version.pdf.

RTE (2021). Energy pathways to 2050 - key results (executive summary). Technical report, RTE, Report. https://assets.rte-france.com/prod/public/2022-01/Energy%20pathways%202050_Key%20results.pdf.

Rudloff, B., Street, A., and Valladão, D. M. (2014). Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences. *European Journal of Operational Research*, 234(3):743–750.

Saharidis, G. K. and Ierapetritou, M. G. (2009). Resolution method for mixed integer bi-level linear problems based on decomposition technique. *Journal of Global Optimization*, 44(1):29–51.

Shapiro, A. (2009). On a time consistency concept in risk averse multistage stochastic programming. *Operations Research Letters*, 37(3):143–147.

Siddiqui, S. and Gabriel, S. A. (2013). An SOS1-Based Approach for Solving MPECs with a Natural Gas Market Application. *Networks and Spatial Economics*, 13(2):205–227.

Slipac, G., Zeljko, M., and Šljivac, D. (2019). Importance of reliability criterion in power system expansion planning. *Energies*, 12(9):1714.

Stephen, G., Tindemans, S. H., Fazio, J., Dent, C., Acevedo, A. F., Bagen, B., Crawford, A., Klaube, A., Logan, D., and Burke, D. (2022). Clarifying the interpretation and use of the LOLE resource adequacy metric. In *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–4. IEEE.

Takriti, S. and Ahmed, S. (2004). On robust optimization of two-stage systems. *Mathematical Programming*, 99(1):109–126.

Viger, F. and Latapy, M. (2005). Efficient and Simple Generation of Random Simple Connected Graphs with Prescribed Degree Sequence. In Wang, L., editor, *Computing and Combinatorics*, pages 440–449, Berlin, Heidelberg. Springer Berlin Heidelberg.

Wogrin, S., Pineda, S., and Tejada-Arango, D. A. (2020). Applications of Bilevel Optimization in Energy and Electricity Markets. In Dempe, S. and Zemkoho, A., editors, *Bilevel Optimization: Advances and Next Challenges*, Springer Optimization and Its Applications, pages 139–168. Springer International Publishing, Cham.

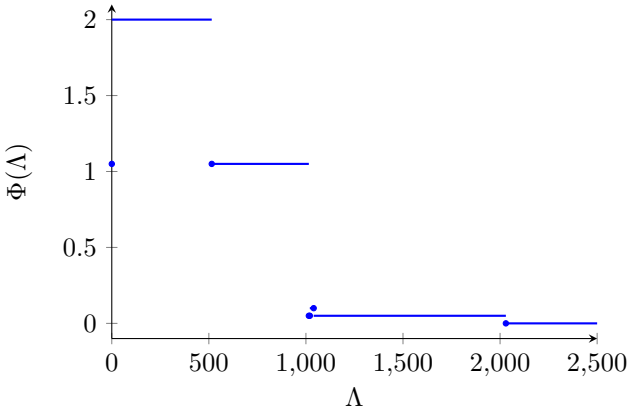
Xu, P. and Wang, L. (2014). An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions. *Computers & Operations Research*, 41:309–318.

Zou, J., Ahmed, S., and Sun, X. A. (2019). Stochastic dual dynamic integer programming. *Mathematical Programming*, 175:461–502.

A The non-monotonicity of function Φ

For the instance of Figure 2, the function Φ is as follows:

$$\Phi(\Lambda) = \begin{cases} 1.05 & 0 \leq \Lambda \leq 0.1 \\ 2 & 0.1 < \Lambda < 514.9 \\ 1.05 & 514.9 \leq \Lambda < 1015 \\ 0.05 & 1015 \leq \Lambda \leq 1020.1 \\ 0.1 & 1020.1 < \Lambda \leq 1039.9 \\ 0.05 & 1039.9 < \Lambda < 2029.9 \\ 0 & 2029.9 \leq \Lambda \end{cases}$$



B Graph generation procedure

Algorithm 2: Undirected graph generation procedure

Parameters: Number of nodes to generate N, Grid ratio $\mathsf{r} > 0$, Proportion of nodes $\mathsf{p} \in (0,1)$, desired graph density $\mathsf{d} \in (0,1)$

Initialization: $\mathcal{N} \leftarrow \varnothing$, $\mathcal{A} \leftarrow \varnothing$

Compute $\mathsf{X}_{\max} = \sqrt{\frac{\mathsf{N}\cdot\mathsf{r}}{\mathsf{p}}}$ and $\mathsf{Y}_{\max} = \sqrt{\frac{\mathsf{N}}{\mathsf{p}\cdot\mathsf{r}}}$

Sample N integer points in $[1, \mathsf{X}_{\max}] \times [1, \mathsf{Y}_{\max}]$ and add them to $\mathcal{N}$

$\mathsf{D}_{\max} \leftarrow 0$

while G is not connected **do** ▷ Adding edges until connectivity

  $\mathsf{D}_{\max} \leftarrow \mathsf{D}_{\max} + 1$

  $\mathcal{A} \leftarrow \{(i,j), 1 \leq i \leq \mathsf{N}, i+1 \leq j \leq \mathsf{N} \mid \text{EuclDist}(i,j) \leq \mathsf{D}_{\max}\}$

$\mathcal{A}^{\text{cand}} \leftarrow \mathcal{A}$

while $\frac{2|\mathcal{A}|}{N(N-1)} > \mathsf{d}$ **do** ▷ Removing edges to reach the required density

  Choose randomly (i,j) in $\mathcal{A}^{\text{cand}}$

  $\mathcal{A}^{\text{cand}} \leftarrow \mathcal{A}^{\text{cand}} \setminus \{(i,j)\}$

  while graph $\big(\mathcal{N}, \mathcal{A} \setminus \{(i,j)\}\big)$ is not connected **do**

    if $\mathcal{A}^{cand} = \emptyset$ **then return** $G = (\mathcal{N}, \mathcal{A})$ ▷ Unable to reach density

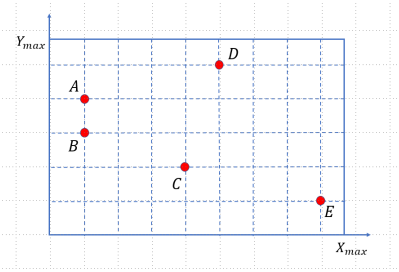
    Choose randomly (i,j) in $\mathcal{A}^{\text{cand}}$

    $\mathcal{A}^{\text{cand}} \leftarrow \mathcal{A}^{\text{cand}} \setminus \{(i,j)\}$

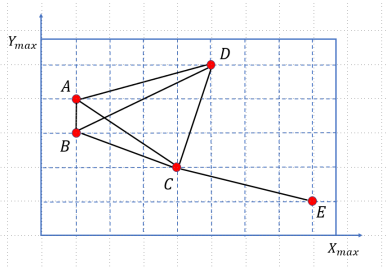
  $\mathcal{A} \leftarrow \mathcal{A} \setminus \{(i,j)\}$

return $G = (\mathcal{N}, \mathcal{A})$

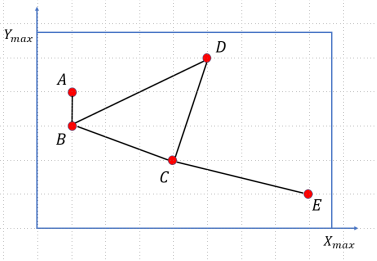
Generating graph $G = (\mathcal{N}, \mathcal{A})$ is performed in three steps in Algorithm 2. In the first step, we sample integer positions on a rectangular grid as follows. Let r be the ratio between the horizontal size and the vertical size of the rectangle, and p be the proportion of integer points on which there will be a node of the graph. We draw N uniformly distributed integer positions on the grid defined by $[1, \mathsf{X}_{\max}] \times [1, \mathsf{Y}_{\max}]$, where $\mathsf{X}_{\max} = \sqrt{\frac{\mathsf{N} \cdot \mathsf{r}}{\mathsf{p}}}$ and $\mathsf{Y}_{\max} = \sqrt{\frac{\mathsf{N}}{\mathsf{p} \cdot \mathsf{r}}}$. Figure 11a shows a possible result of such a sampling procedure.



(a) Graph obtained after generating N nodes on the grid



(b) First connected graph, obtained after adding all the edges between points at an Euclidean distance of at most 5



(c) Final graph obtained after removing edges to reach a density of 0.5

Figure 11: A possible result of Algorithm 2 for $\mathsf{N} = 5$, $\mathsf{r} = 1.5$, $\mathsf{p} = 0.1$ and $\mathsf{d} = 0.5$.

The second step of the graph generation procedure consists of adding edges to the graph so that it becomes connected. Let $\mathsf{D}_{\max} = 1$, we add all the edges between nodes which are at an Euclidean distance of at most $\mathsf{D}_{\max}$. If the resulting graph is connected, we stop, otherwise $\mathsf{D}_{\max}$ increases by one and we restart a similar process. In the example presented in Figure 11a, this leads to the connected graph of Figure 11b.

The third step of the procedure consists in removing random edges letting the graph connected until it reaches the required density d, or it fails at reaching it. For example, in the graph of Figure 11b, edge (C, E) cannot be removed as it would lead to a disconnected graph.

In our experiments, we use $\mathsf{r} = 1.5$, $\mathsf{p} = 0.005$. We set d equal to 0.35, 0.25, 0.15, and 0.10 when N is equal to 5, 10, 20, and 30.

C Detailed results

		KKT_SOS1			HPR-Aux				woLOLE-Aux				0_LOLE	
VOLL	Instance	CPU Time	$LB_{p,m}$	$z_{p,m}$	CPU Time	$LB_{p,m}$	$z_{p,m}$	LOLE/α	CPU Time	$LB_{p,m}$	$z_{p,m}$	LOLE/α	CPU Time	$z_{p,m}$
1000	N5-S5-T3	T.L.	2962618	-	22	3283061	3283061	1.64	16	1272988	1272988	13.92	16	4033125
	N5-S5-T6	95	10667056	10668007	43	10666930	10667964	0.97	15	3498432	3498432	28.92	16	11536931
	N5-S5-T12	T.L.	4073742	-	24	4091316	4091683	0.98	16	3480524	3480524	12.46	15	4326439
	N5-S10-T3	T.L.	2566003	-	3469	2819146	2819427	1.66	16	1428352	1428352	12.21	15	3066149
	N5-S10-T6	T.L.	5007835	-	115	5204401	5204651	0.99	16	2495985	2495985	17.71	17	5761275
	N5-S10-T12	T.L.	5268357	-	T.L.	5250173	5654376	1.30	17	4620821	4620821	13.85	16	6487261
	N5-S20-T3	T.L.	6002632	-	T.L.	6946607	7172858	0.96	16	1583596	1583596	14.20	16	8769617
	N5-S20-T6	T.L.	3796325	-	T.L.	4109166	4539288	1.39	17	2769770	2769770	21.89	16	5251740
	N5-S20-T12	T.L.	10715829	-	T.L.	11379291	11999101	1.01	15	6341938	6341938	25.34	16	14942645
	N10-S5-T3	T.L.	8766936	-	61	9886016	9886911	1.15	15	3190004	3190004	26.94	16	10747287
	N10-S5-T6	T.L.	9238208	-	821	9704890	9705857	1.75	16	5870244	5870244	46.99	15	10618377
	N10-S5-T12	T.L.	9860123	-	T.L.	10541966	10562925	1.03	16	9272311	9272311	20.15	15	11062946
	N10-S10-T3	T.L.	6164377	-	T.L.	6805493	8155134	1.59	16	2773569	2773569	26.32	16	9068878
	N10-S10-T6	T.L.	8133538	-	T.L.	8450348	9416378	1.37	16	5539715	5539715	43.33	17	10264836
	N10-S10-T12	T.L.	11227711	-	T.L.	11297800	12038840	1.25	18	10234679	10234679	22.39	16	13507268
	N10-S20-T3	T.L.	5475653	-	T.L.	5714524	6523120	1.61	16	3024938	3024938	27.14	16	7292817
	N10-S20-T6	T.L.	5507062	-	T.L.	5686166	6094955	1.10	17	3724795	3724795	25.60	16	6725466
	N10-S20-T12	T.L.	7884798	-	T.L.	8050358	8659707	1.20	17	7052416	7052416	22.40	19	9511747
5000	N5-S5-T3	T.L.	3717588	-	19	3685806	3685806	1.78	17	3598564	3598564	2.92	16	4033125
	N5-S5-T6	34	10826947	10827997	25	10826978	10827976	0.97	18	10064868	10064868	7.20	16	11536931
	N5-S5-T12	64	4201967	4201967	20	4201848	4201967	0.92	16	4175245	4175245	2.07	15	4326439
	N5-S10-T3	T.L.	2917082	-	23	2918913	2919190	1.26	16	2905346	2905346	2.04	15	3066149
	N5-S10-T6	T.L.	5386700	-	26	5402903	5403365	0.99	17	4990517	4990517	3.81	17	5761275
	N5-S10-T12	T.L.	6007605	-	36	6008203	6008804	1.16	16	6007563	6007563	1.46	16	6487261
	N5-S20-T3	T.L.	7019239	-	1389	7343697	7344431	0.96	16	6070102	6070102	7.31	16	8769617
	N5-S20-T6	T.L.	4615119	-	T.L.	4683514	4709460	1.39	16	4599075	4599075	2.89	16	5251740
	N5-S20-T12	T.L.	12316701	-	T.L.	12402455	12484043	1.01	16	12158117	12158117	2.96	16	14942645
	N10-S5-T3	T.L.	10048284	-	23	10030497	10031374	1.33	16	9799836	9799836	6.44	16	10747287
	N10-S5-T6	T.L.	10017988	-	29	10014413	10015387	1.83	17	9812700	9812700	4.60	15	10618377
	N10-S5-T12	167	10844693	10844693	19	10844693	10844693	0.94	16	10844693	10844693	0.94	15	11062946
	N10-S10-T3	T.L.	8188373	-	235	8405222	8406062	2.05	17	8044849	8044849	5.06	16	9068878
	N10-S10-T6	T.L.	9424211	-	3696	9610917	9611878	1.37	17	9341446	9341446	6.07	17	10264836
	N10-S10-T12	T.L.	12437407	-	T.L.	12454596	12472215	1.36	17	12432263	12432263	2.73	16	13507268
	N10-S20-T3	T.L.	6566976	-	T.L.	6629666	6716681	1.34	17	6505940	6505940	5.56	16	7292817
	N10-S20-T6	T.L.	6121878	-	T.L.	6204603	6243139	1.09	17	6041945	6041945	4.57	16	6725466
	N10-S20-T12	T.L.	8860263	-	T.L.	8885649	8920777	1.20	17	8853219	8853219	2.93	19	9511747

LOLE/α: a value less than or equal to one means that the the computed solution is a solution to Bilevel-α-Stoch-GTEPP.

Table 6: Detailed results comparison for methods **H_Rand**, **H_MinENS-Opt**, and **H_MinENS** for all small-sized instances and for a VOLL equal to 1000 and 5000.

		CPU Time			$z_{p,m}$		
VOLL	Instance	**H_Rand**	**H_MinENS-Opt**	**H_MinENS**	**H_Rand**	**H_MinENS-Opt**	**H_MinENS**
1000	N5-S5-T3	21	22	23	3956574	3869431	3869431
	N5-S5-T6	20	21	21	11075424	11075388	11075388
	N5-S5-T12	22	22	24	4173623	4157259	4157259
	N5-S10-T3	21	22	23	2880284	2878165	2878165
	N5-S10-T6	22	20	23	5472613	5472613	5472613
	N5-S10-T12	22	24	27	5835823	5825857	5810488
	N5-S20-T3	21	21	22	7441483	7441475	7441475
	N5-S20-T6	21	23	26	4664390	4664376	4664376
	N5-S20-T12	21	24	28	12655975	12552919	12552919
	N10-S5-T3	21	22	22	10183783	10098612	10098612
	N10-S5-T6	20	22	25	9940402	9940402	9940402
	N10-S5-T12	23	23	33	10716018	10715806	10715806
	N10-S10-T3	22	24	25	8502400	8502400	8502400
	N10-S10-T6	22	25	28	10039274	10032582	10023093
	N10-S10-T12	23	26	37	12619790	12619737	12619737
	N10-S20-T3	23	24	29	6743735	6743735	6743735
	N10-S20-T6	25	29	33	6321331	6302881	6302881
	N10-S20-T12	28	38	67	9030803	8974545	8974545
5000	N5-S5-T3	21	22	23	3978817	3941756	3941756
	N5-S5-T6	21	21	22	11131405	11131398	11131398
	N5-S5-T12	21	23	24	4209411	4209409	4209409
	N5-S10-T3	21	22	22	2928332	2926343	2926343
	N5-S10-T6	21	21	24	5506318	5506318	5506318
	N5-S10-T12	22	26	31	6032835	6019347	6019347
	N5-S20-T3	21	24	24	7553449	7553445	7553445
	N5-S20-T6	22	24	27	4749790	4749782	4749782
	N5-S20-T12	21	26	36	12841985	12765004	12765004
	N10-S5-T3	22	23	24	10174383	10174373	10174373
	N10-S5-T6	21	21	26	10119266	10119266	10119266
	N10-S5-T12	21	22	24	10844693	10844693	10844693
	N10-S10-T3	22	22	26	8550814	8545095	8545095
	N10-S10-T6	21	27	35	9879981	9818616	9818616
	N10-S10-T12	22	26	37	12639012	12639012	12639012
	N10-S20-T3	23	25	31	6839908	6839903	6839903
	N10-S20-T6	24	26	33	6316101	6316101	6316101
	N10-S20-T12	28	38	67	9003280	8990049	8990049

Table 7: Detailed results for methods **H_Rand**, **H_MinENS-Opt**, and **H_MinENS** for all small-sized instances and for a VOLL equal to 1000 and 5000.

VOLL	Instance	$\overline{\text{LB}}_{\text{p}}$	$\underline{z}_{\text{p}}$	Optimal?*
1000	N5-S5-T3	3283061	3869431	no
	N5-S5-T6	10667056	10667964	yes
	N5-S5-T12	4091316	4091683	yes
	N5-S10-T3	2819146	2878165	no
	N5-S10-T6	5204401	5204651	yes
	N5-S10-T12	5268357	5810488	no
	N5-S20-T3	6946607	7172858	no
	N5-S20-T6	4109166	4664376	no
	N5-S20-T12	11379291	12552919	no
	N10-S5-T3	9886016	10098612	no
	N10-S5-T6	9704890	9940402	no
	N10-S5-T12	10541966	10715806	no
	N10-S10-T3	6805493	8502400	no
	N10-S10-T6	8450348	10023093	no
	N10-S10-T12	11297800	12619737	no
	N10-S20-T3	5714524	6743735	no
	N10-S20-T6	5686166	6302881	no
	N10-S20-T12	8050358	8974545	no
5000	N5-S5-T3	3717588	3941756	no
	N5-S5-T6	10826978	10827976	yes
	N5-S5-T12	4201967	4201967	yes
	N5-S10-T3	2918913	2926343	no
	N5-S10-T6	5402903	5403365	yes
	N5-S10-T12	6008203	6019347	no
	N5-S20-T3	7343697	7344431	yes
	N5-S20-T6	4683514	4749782	no
	N5-S20-T12	12402455	12765004	no
	N10-S5-T3	10048284	10174373	no
	N10-S5-T6	10017988	10119266	no
	N10-S5-T12	10844693	10844693	yes
	N10-S10-T3	8405222	8545095	no
	N10-S10-T6	9610917	9818616	no
	N10-S10-T12	12454596	12639012	no
	N10-S20-T3	6629666	6839903	no
	N10-S20-T6	6204603	6316101	no
	N10-S20-T12	8885649	8990049	no

*: an instance is labeled as optimally solved whenever **KKT_SOS1**, **HPR-Aux** or **woLOLE-Aux** has computed an optimal solution considering an optimality gap equal to 10^{-4} (the solution needs to be a feasible to Bilevel-α-Stoch-GTEPP for **HPR-Aux** or **woLOLE-Aux**)

Table 8: Lower bound and best know solution values for all small-sized instances.

		CPU Time			$z_{p,m}$		
VOLL	Instance	**H_Rand**	**H_MinENS-Opt**	**H_MinENS**	**H_Rand**	**H_MinENS-Opt**	**H_MinENS**
1000	N20-S100-T24	324	599	1408	21892582	21832373	21766572
	N20-S100-T48	437	1280	3721	27446124	27409725	27409725
	N20-S200-T24	597	1203	2976	24377297	24247989	24245180
	N20-S200-T48	1131	2950	7859	21599634	21513344	21479569
	N30-S100-T24	693	1293	3282	30419971	30362809	30077407
	N30-S100-T48	1157	2343	11124	39825737	39825699	39524573
	N30-S200-T24	1647	3029	7069	29890955	29775781	29775781
	N30-S200-T48	3529	7327	21225	43326380	42986436	42914655
5000	N20-S100-T24	321	552	1114	21664965	21621352	21593821
	N20-S100-T48	491	1103	2631	27100645	27071971	27071971
	N20-S200-T24	571	1135	2291	24430226	24337267	24337267
	N20-S200-T48	1510	2863	6015	21577954	21561443	21561443
	N30-S100-T24	783	1245	2658	29251690	29247289	29247289
	N30-S100-T48	1300	2700	7607	37983714	37957921	37956418
	N30-S200-T24	2142	2990	5639	28979570	28918204	28897672
	N30-S200-T48	4101	6576	16011	39197760	39187954	39168047

Table 9: Detailed results for methods **H_Rand**, **H_MinENS-Opt**, and **H_MinENS** for all large-sized instances and for a VOLL equal to 1000 and 5000.