



Solving a bilevel stochastic generation and transmission expansion planning problem

Xavier Blanchot, François Clautiaux, Aurélien Froger, Manuel Ruiz

► To cite this version:

Xavier Blanchot, François Clautiaux, Aurélien Froger, Manuel Ruiz. Solving a bilevel stochastic generation and transmission expansion planning problem. 2023. hal-03957750v1

HAL Id: hal-03957750

<https://inria.hal.science/hal-03957750v1>

Preprint submitted on 26 Jan 2023 (v1), last revised 21 Jul 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Solving a bilevel stochastic generation and transmission expansion planning problem

Xavier Blanchot^{1,2}, François Clautiaux¹, Aurélien Froger¹, Manuel Ruiz²

¹ Université de Bordeaux, CNRS, INRIA, Bordeaux INP, IMB, UMR 5251, Talence, France

² RTE, Paris La Défense, France

Abstract

We consider a strategic version of a stochastic generation and transmission expansion planning problem. This problem is tackled by the French Transmission System Operator for producing prospective reports on the evolution of the electricity network. We study how a regulatory constraint limiting the loss of load expectation (i.e., a measure of unserved demand) can be incorporated into the problem. We show that the resulting problem is a bilevel problem in which the leader and follower have the same objective function. We propose a heuristic method for this problem that embeds a Benders decomposition algorithm in a binary search on the total investment cost. For each investment value tested, the algorithm disregards the regulatory constraint as well as the bilevel nature of the problem, and tests a posteriori whether the computed solutions can be converted to feasible solutions to the problem. We performed computational experiments on randomly generated instances to study the practical difficulty of the problem and validate the proposed solution method. We show that solving the KKT reformulation of the problem is intractable even for our small instances, whereas our heuristic method produces feasible solutions even for large instances.

Keywords— Bilevel stochastic optimization, Mixed integer linear 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 neighbor countries. Its main role is to guarantee in real time the balance between electricity production and consumption.

RTE also 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 at a given time horizon. Given the nature of the study, RTE does not rely on a precise physical modeling of fine-grained operational constraints.

Email addresses: xavier.blanchot@u-bordeaux.fr (Xavier Blanchot), francois.clautiaux@math.u-bordeaux.fr (François Clautiaux), aurelien.froger@math.u-bordeaux.fr (Aurélien Froger), manuel.ruiz@rte-france.com (Manuel Ruiz)

In the initial version of the problem solved by RTE to produce its prospective reports, the problem is formulated as a two-stage stochastic problem, where discrete scenarios are used to model the uncertain parameters (demand, resource availability). The first stage problem is an investment problem, while the second stage deals with aggregated operational decisions and constraints. Among the costs taken into consideration in the second stage, unserved demand incurs a large penalty, corresponding to a value called the *value of lost load* (VOLL). However this cost is not large enough to ensure that in an optimal solution, all demands of every scenario are served, as this would be overly conservative and could result in a prohibitive investment cost.

To counterbalance this fact, French regulation imposes that the number of hours with energy not served should be *in expectation* less than or equal to three per year¹. We call this constraint the “ α -hours criterion” (with $\alpha = 3$ in practice). Currently, RTE does not include this criterion in the two-stage stochastic problem formulated, but takes it into consideration a posteriori, using various heuristic strategies.

In this work, we study the integration of the “ α -hours criterion” in the expansion planning problem studied by RTE. In this new version, investment decisions must be made by anticipating the outcome of the aggregated operational decisions to ensure that the criterion is met. Since our work focuses on the methodological aspect, we consider a simplified version of the problem by mathematically modeling the electrical network in a coarse way using a simple flow network with additional linear constraints, similar to what RTE does in some of its strategic studies.

The contribution of this work is twofold. First, we show that taking into account the “ α -hours criterion” leads to a bilevel problem where the leader and the follower have the same objective function. Despite this particular structure, the bilevel problem is not equivalent to its high-point relaxation (i.e., the single level problem obtained by relaxing the optimality condition of the lower-level problem). Specifically, we show that optimal solutions to the high-point relaxation (HPR) are not necessarily feasible solutions to the bilevel problem for the reason that the operational recourse decisions in optimal solutions to the HPR are not necessarily optimal solutions of the operational recourse problem. We also show that there are instances for which the gap with respect to the value of optimal solutions to the operational recourse problem is unbounded. Second, we propose an iterative heuristic method based on a Benders decomposition algorithm and a binary search on the investment costs to compute bilevel feasible solutions to the problem. We report computational experiments conducted on randomly generated instances with up to 6 million variables and 8 million constraints. Our heuristic allows to find feasible solutions in a few hours, whereas state-of-the-art exact solvers (Fortuny-Amat and McCarl, 1981; Siddiqui and Gabriel, 2013) are not able to find a feasible solution in six hours, even on the smallest instances.

The paper is organized as follows. In Section 2, we describe the industrial and regulatory context of our problem to show its practical relevance. We also present how the problem relates to the literature and our main modeling hypotheses. In Section 3, we formally define the problem to solve and we formulate it as a bilevel problem. We also highlight two properties of optimal solutions to its high-point relaxation. To find feasible solutions for the problem, we present in Section 4 a matheuristic based on a Benders decomposition and a binary search on the total investment costs. In Section 5, we present numerical results for our heuristic and we compare them to the results obtained by solving the KKT reformulation of the bilevel program. We finally conclude and present several perspectives to this work in Section 6.

2 Industrial context and modeling hypotheses

2.1 Expansion planning problems and the α -hours reliability criterion

Several RTE’s prospective studies currently involve solving a stochastic generation and expansion planning problem. An integrated (centralized) planner approach is considered, meaning that generation and transmission capacity expansion decisions are simultaneously made. The problem aims to determine the investments in generation and transmission capacity in order to satisfy the projected electricity demand in the distant future of 10

¹Article D141-12-6 from the French Energy Code: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044622322 (in French)

to 30 years.

In its initial version, the studied generation and transmission expansion planning problem is formulated as a two-stage stochastic problem, where the uncertain parameters, modeled with a finite set of scenarios, are the load demand, the operational costs, the production and transmission lines availability and actual nominal capacity. The first-stage problem is a static (single-period) investment problem (i.e., investment decisions are not incremental over a multi-year time horizon) and the second-stage problem decomposes into a set of independent problems (one for each scenario) taking operational constraints such as load balance with an hourly resolution. The objective is to minimize the sum of the investment costs and the expected operational costs.

An important feature of the second stage problem is that demand satisfaction is not strictly enforced. Unserved load demand (Energy Not Served (ENS) in the remainder), is penalized in the objective function using the 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”². According to the Regulation (EU) 2019/943 of the European Parliament and of the Council², each member state was required to designate by July 5, 2020, a competent authority to determine a single estimate of the value of lost load for its territory. In France, a recent study has estimated the value of lost load to 33000 €/MWh³.

The objective of using VOLL is to find a compromise between a guaranteed electricity demand satisfaction at all times and sustainable investment costs. Satisfying all hourly demands in each region in all scenarios would lead to costly capacity investments that would not be useful in most scenarios. However, one issue with using VOLL is that it is not sufficient in itself to ensure a very low number of hours with ENS, which 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.

Because in every country, the reliability of the electrical network is essential to the economic dynamism of the territory and the well-being of the population living there, state regulations exist. In France, since 2016, legislation¹ imposes for reliability purposes that the number of hours with energy not served (ENS) should be in expectation less than or equal to three per year. It is referred to as the “three-hours criterion”. Its value is the result of an economic trade-off between the cost of power cuts during load shedding and the cost of measures (e.g., construction of additional peak capacity) to reduce these cuts⁴. In very recent years, each member state of the European Union has been required to define a reliability standard (RS) i.e., a “measure of the necessary level of security of supply”².

In 2020, a methodology for setting the RS, referred to as the “VOLL/CONE/RS methodology”, has been proposed by the ENTSO-E and approved by the European Union Agency for the Cooperation of Energy Regulators⁵. The acronym CONE refers to the cost of new entry (i.e., the annualized cost of the most competitive generating unit that can be used when ENS occurs). The RS is then computed based on the VOLL, the CONE and the loss of load expectation (LOLE)⁶. A very recent application of the VOLL/CONE/RS methodology in France³ has confirmed the “three-hours criterion” as a consistent RS.

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 works). Lara et al. (2018) considered a deterministic multi-period generation investment problem. 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

²<https://eur-lex.europa.eu/eli/reg/2019/943/oj>

³<https://www.cre.fr/Documents/Deliberations/Proposition/valeur-du-critere-de-securite-d-approvisionnement-electrique-pour-la-france-metropolitaine-continentale> (in French)

⁴<https://www.ecologie.gouv.fr/securite-dapprovisionnement-en-electricite> (in French)

⁵https://acer.europa.eu/sites/default/files/documents/Decisions_annex/ACER%20Decision%2023-2020%20on%20VOLL%20CONE%20RS%20-%20Annex%20I.pdf

⁶The LOLE 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 ENS occurs”⁵

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 multi-cut Benders decomposition algorithm.

To our knowledge, the “three-hours criterion” has never been considered in the literature. 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) discussed how the system adequacy, defined as the installation of sufficient generation and transmission capacity to meet the load demand, can be evaluated using deterministic and probabilistic approaches. 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 also be imposed (Lara et al., 2018; Li et al., 2022). In a probabilistic approach, failure scenarios can be explicitly identified and taken into account in the calculation of reliability indices such as LOLE and expected ENS. In (Choi et al., 2005), the reliability of the transmission system is checked by a Monte-Carlo simulation procedure. Alizadeh and Jadid (2011) considered the LOLE and the loss of load probability as reliability criteria and proposed a sequential heuristic method which, while imposing the satisfaction of the system’s energy needs at the end of the planning horizon, guarantees the achievement of the required reliability level. Braga and Saraiva (2005) considered the expected ENS as another objective, and developed a method to compute Pareto-optimal solutions.

2.2 Main modeling hypotheses

We model the problem as a two-stage stochastic problem, with an “ α -time units criterion”. The term time unit refers to any time interval (it does not necessarily last one hour). In the remainder, we assume that the problem has relatively complete recourse if this additional constraint is relaxed, i.e. there is always a feasible solution for any investment decision, but this solution may not meet the “ α -time units criterion”. This assumption comes from the fact that we penalize the ENS with the VOLL.

We focus on the power grid of a given territory and use, like in Lara et al. (2018, 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 and neighbor countries are represented with one to two regions (see Figure 1). 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) and done by RTE, would only require very few adaptations of the work we present here, although it would increase the computational burden. 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 project⁷ developed by RTE. The first-stage problem is to simultaneously make generation and transmission capacity expansion decisions up to 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; 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. This is especially the case when investment capacities in generating units with large power capacity (such as nuclear power plants) are taken as initial assumptions in the studies. Between two neighboring regions, we do not assume that we invest in a precise number of transmission lines, 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.

⁷<https://antares-simulator.readthedocs.io/en/latest/reference-guide/11-modeling/>

Evolution of the weak points identified in the Ten-Year Network Development Plan (SDDR) in scenarios M23 and N2

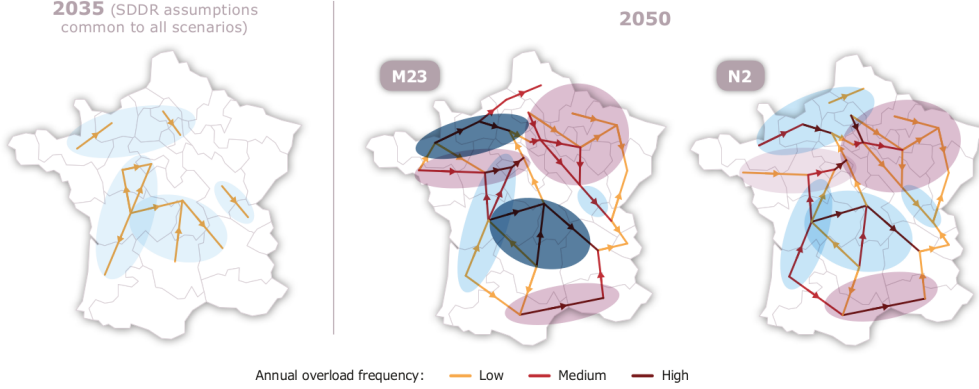


Figure 1: A representation of the French transmission network used by RTE in a strategic study. (RTE (2021), page 39).

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. In long-term generation and transmission expansion planning problems, using representative days with an hourly time resolution for modeling the operational decisions is classic (Nahmmacher et al., 2016). In our work, we use a *representative period*, which allows to capture the temporal correlation between decisions of successive time units. Like in Lara et al. (2018) and Micheli et al. (2020), 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. This is RTE’s current approach.

Finally, the problem must ensure that the sum over all regions of the number of time units with ENS should be in expectation less than or equal to α per time period. This is the “ α -time units criterion” we aim to model.

3 Mathematical models and properties

We now give a formal definition of the 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 $\mathcal{N}^* \subset \mathcal{N}$ 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 $i \in \mathcal{N}^*$ is $P_i^{\text{init}} \in \mathbb{R}_+$. The cost of investing in production capacity (€/MW) at each node $i \in \mathcal{N}^*$ 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 $i \in \mathcal{N}^*$ 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 $i \in \mathcal{N}^*$ is denoted by $\Delta_i \in [0, 1]$.

The VOLL is set equal to a constant value $M \geq 0$ and applies to the demand unsatisfied at nodes from \mathcal{N} .

We model the “ α -time units criterion” using a threshold $\epsilon > 0$, meaning that we only consider that there is ENS at a given node and time step if the ENS is larger than ϵ .

We formulate the problem using the following decisions variables. For $i \in \mathcal{N}^*$, $P_i \in \mathbb{R}^+$ 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 $i \in \mathcal{N}^*$, 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 $\bar{f}_{s,a,t} \in \mathbb{R}_+$. 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).

The two-stage stochastic generation and transmission expansion planning problem without the “ α -time units criterion” is as follows:

$$\min \quad \sum_{i \in \mathcal{N}^*} c_i^{\text{prod}} P_i + \sum_{a \in \mathcal{A}} c_a^{\text{line}} F_a + \sum_{s \in \mathcal{S}} p_s \left(\sum_{i \in \mathcal{N}^*} \sum_{t \in \mathcal{T}} g_{s,i,t}^{\text{prod}} h_{s,i,t} + \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} g_{s,a,t}^{\text{flow}} \bar{f}_{s,a,t} + M \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} u_{s,i,t} \right) \quad (1a)$$

$$\text{s.t.} \quad P_i \in \mathbb{R}_+ \quad \forall i \in \mathcal{N}^* \quad (1b)$$

$$F_a \in \mathbb{R}_+ \quad \forall a \in \mathcal{A} \quad (1c)$$

$$h_{s,i,t} \leq P_i^{\text{init}} + P_i \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \quad (1d)$$

$$\bar{f}_{s,a,t} \leq F_a^{\text{init}} + F_a \quad \forall s \in \mathcal{S}, \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (1e)$$

$$\bar{f}_{s,a,t} \geq f_{s,a,t} \quad \forall s \in \mathcal{S}, \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (1f)$$

$$\bar{f}_{s,a,t} \geq -f_{s,a,t} \quad \forall s \in \mathcal{S}, \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (1g)$$

$$h_{s,i,t} \leq h_{s,i,t-1} + \Delta_i P_i \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \setminus \{0\} \quad (1h)$$

$$h_{s,i,t} \geq h_{s,i,t-1} - \Delta_i P_i \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \setminus \{0\} \quad (1i)$$

$$h_{s,i,0} \leq h_{s,i,|\mathcal{T}|} + \Delta_i P_i \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}^* \quad (1j)$$

$$h_{s,i,0} \geq h_{s,i,|\mathcal{T}|} - \Delta_i P_i \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}^* \quad (1k)$$

$$h_{s,i,t} + \sum_{a \in \Gamma^+(i)} f_{s,a,t} - \sum_{a \in \Gamma^-(i)} f_{s,a,t} + u_{s,i,t} \geq d_{s,i,t} \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \quad (1l)$$

$$\sum_{a \in \Gamma^+(i)} f_{s,a,t} - \sum_{a \in \Gamma^-(i)} f_{s,a,t} + u_{s,i,t} \geq d_{s,i,t} \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N} \setminus \mathcal{N}^*, \forall t \in \mathcal{T} \quad (1m)$$

$$h_{s,i,t} \in \mathbb{R}_+ \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \quad (1n)$$

$$u_{s,i,t} \in \mathbb{R}_+ \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (1o)$$

$$f_{s,a,t} \in \mathbb{R} \quad \forall s \in \mathcal{S}, \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (1p)$$

$$\bar{f}_{s,a,t} \in \mathbb{R}_+ \quad \forall s \in \mathcal{S}, \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (1q)$$

Constraints (1d) and (1e) 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 (1f),(1g) ensure that each variable $\bar{f}_{s,a,t}$ is the absolute value of $f_{s,a,t}$. Constraints (1h)-(1k) 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 (1l) and (1m) define the flow conservation constraints. The remaining constraints define the domain definition of the variables.

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}) = \left\{ \begin{array}{ll} \min & \sum_{i \in \mathcal{N}^*} \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}} \bar{f}_{s,a,t} + M \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} u_{s,i,t} \quad (2a) \\ \text{s.t.} & h_{s,i,t} \leq P_i^{\text{init}} + P_i \quad \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \quad (2b) \\ & h_{s,i,t} \leq h_{s,i,t-1} + \Delta_i P_i \quad \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \setminus \{0\} \quad (2c) \\ & h_{s,i,t} \geq h_{s,i,t-1} - \Delta_i P_i \quad \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \setminus \{0\} \quad (2d) \\ & h_{s,i,0} \leq h_{s,i,|\mathcal{T}|} + \Delta_i P_i \quad \forall i \in \mathcal{N}^* \quad (2e) \\ & h_{s,i,0} \geq h_{s,i,|\mathcal{T}|} - \Delta_i P_i \quad \forall i \in \mathcal{N}^* \quad (2f) \\ & \bar{f}_{s,a,t} \leq F_a^{\text{init}} + F_a \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (2g) \\ & \bar{f}_{s,a,t} \geq f_{s,a,t} \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (2h) \\ & \bar{f}_{s,a,t} \geq -f_{s,a,t} \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (2i) \\ & h_{s,i,t} + \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}^*, \forall t \in \mathcal{T} \quad (2j) \\ & \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}^*, \forall t \in \mathcal{T} \quad (2k) \\ & h_{s,i,t} \in \mathbb{R}_+ \quad \forall i \in \mathcal{N}^*, \forall t \in \mathcal{T} \quad (2l) \\ & u_{s,i,t} \in \mathbb{R}_+ \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (2m) \\ & f_{s,a,t} \in \mathbb{R} \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (2n) \\ & \bar{f}_{s,a,t} \in \mathbb{R}_+ \quad \forall a \in \mathcal{A}, \forall t \in \mathcal{T} \quad (2o) \end{array} \right.$$

For every $(\mathbf{P}, \mathbf{F}) \in \mathbb{R}_+^{|\mathcal{N}^*|} \times \mathbb{R}_+^{|\mathcal{A}|}$, we define $\mathcal{R}_s(\mathbf{P}, \mathbf{F}) = \{(\mathbf{h}_s, \bar{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) : (2b) - (2o)\}$ as the set of feasible solutions to the operational problem associated with scenario s . For every solution $(\mathbf{h}_s, \bar{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \mathcal{R}_s(\mathbf{P}, \mathbf{F})$, we denote $\psi_s(\mathbf{h}_s, \bar{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) = \sum_{i \in \mathcal{N}^*} \sum_{t \in \mathcal{T}} g_{s,i,t}^{\text{prod}} h_{s,i,t} + \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} g_{s,a,t}^{\text{flow}} \bar{f}_{s,a,t} + M \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} u_{s,i,t}$.

Problem (1) can be rewritten as follows:

$$\min \left\{ \sum_{i \in \mathcal{N}^*} c_i^{\text{prod}} P_i + \sum_{a \in \mathcal{A}} c_a^{\text{line}} F_a + \sum_{s \in \mathcal{S}} p_s \phi_s(\mathbf{P}, \mathbf{F}) : (1b) - (1c) \right\}$$

Optimal solutions to problem (1) may lead to several time units with ENS in some scenarios, as investing in less nominal capacity can be cheaper than avoiding ENS in every time unit of each scenario. In A, we illustrate this situation. Introducing the “ α -time units criterion” to limit the expected number of time units with ENS leads to the bilevel formulation (3), 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 scenario s is larger than ϵ , and C_s is a large enough positive real number (e.g., $C_s \geq \max_{t \in \mathcal{T}} \sum_{i \in \mathcal{N}} d_{s,i,t}$).

$$\min \sum_{i \in \mathcal{N}^*} c_i^{\text{prod}} P_i + \sum_{a \in \mathcal{A}} c_a^{\text{line}} F_a + \sum_{s \in \mathcal{S}} p_s \phi_s(\mathbf{P}, \mathbf{F}) \quad (3a)$$

$$\text{s.t.} \quad (1b) - (1c) \quad (3b)$$

$$(1n) - (1q) \quad (3c)$$

$$(\mathbf{h}_s, \bar{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \arg \min_{(\mathbf{h}'_s, \bar{\mathbf{f}}'_s, \mathbf{f}'_s, \mathbf{u}'_s) \in \mathcal{R}_s(\mathbf{P}, \mathbf{F})} \{ \psi_s(\mathbf{h}'_s, \bar{\mathbf{f}}'_s, \mathbf{f}'_s, \mathbf{u}'_s) \} \quad \forall s \in \mathcal{S} \quad (3d)$$

$$C_s \delta_{s,i,t} \geq u_{s,i,t} - \epsilon \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (3e)$$

$$\epsilon \delta_{s,i,t} \leq u_{s,i,t} \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (3f)$$

$$\sum_{s \in \mathcal{S}} p_s \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} \delta_{s,i,t} \leq \alpha \quad (3g)$$

$$\delta_{s,i,t} \in \{0, 1\} \quad \forall s \in \mathcal{S}, \forall i \in \mathcal{N}, \forall t \in \mathcal{T} \quad (3h)$$

Constraints (3d) ensure that the operational decisions in every scenario are optimal and that they only depend on the investment decisions. Constraints (3e) and (3f) couple the $u_{s,i,t}$ and $\delta_{s,i,t}$ variables. Constraint (3g) ensures the satisfaction of the “ α -time units criterion”.

Due to constraints (3d), we can replace $\phi_s(\mathbf{P}, \mathbf{F})$ in the objective function by $\sum_{i \in \mathcal{N}^*} \sum_{t \in \mathcal{T}} g_{s,i,t}^{\text{prod}} h_{s,i,t} + \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} g_{s,a,t}^{\text{flow}} \bar{f}_{s,a,t} + M \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} u_{s,i,t}$.

The formulation (3) defines an optimistic bilevel problem with one leader and $|\mathcal{S}|$ followers. We therefore refer to our two-stage stochastic generation and transmission expansion planning problem with the “ α -time units criterion” as the bilevel stochastic generation and transmission expansion planning problem that we abbreviate as Bilevel- α -Stoch-GTEPP.

Remark 1. *As the investment costs 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.*

The problem obtained by replacing constraints (3d) by $(\mathbf{h}_s, \bar{\mathbf{f}}_s, \mathbf{f}_s, \mathbf{u}_s) \in \mathcal{R}_s(\mathbf{P}, \mathbf{F})$ in formulation (3) is referred to as the *high-point relaxation* of the Bilevel- α -Stoch-GTEPP.

We show in Proposition 1 that there exist instances of the Bilevel- α -Stoch-GTEPP for which the ratio between the value of any optimal solution to its high-point relaxation, denoted sol , and the optimal value of the lower-level problem obtained by fixing the upper-level variables to their value in sol is not bounded by a constant.

Proposition 1. *For any $A > 0$, there exists an instance of Bilevel- α -Stoch-GTEPP such that for any optimal solution $(\mathbf{P}^*, \mathbf{F}^*, (\mathbf{h}_s^*, \bar{\mathbf{f}}_s^*, \mathbf{f}_s^*, \mathbf{u}_s^*)_{s \in \mathcal{S}})$ to its high-point relaxation, we have $\sum_{i \in \mathcal{N}^*} \sum_{t \in \mathcal{T}} g_{s,i,t}^{\text{prod}} h_{s,i,t}^* + \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} g_{s,a,t}^{\text{flow}} \bar{f}_{s,a,t}^* + M \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} u_{s,i,t}^* > A \cdot \phi_s(\mathbf{P}^*, \mathbf{F}^*)$ for a scenario s .*

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

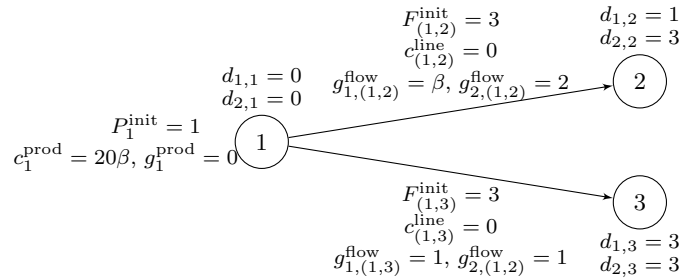


Figure 2: An instance with 3 nodes, 2 arcs, 2 scenarios, 1 time unit, and such that $M = 10\beta$, $\epsilon = 0.1$, $\alpha = 1.5$, $p_1 = 0.9$ and $p_2 = 0.1$.

The optimal solution to the high-point relaxation 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 $\bar{f}_{1,(1,2)}^* = 1 - \epsilon$ ($1 - \epsilon$ unit of flow is sent to node 2 during time unit 1 in scenario 1), $f_{1,(1,3)}^* = \epsilon$, $\bar{f}_{2,(1,3)}^* = 0$, and $\bar{f}_{2,(1,2)}^* = 1$. We show this solution in Figure 3. Note that the expected value of the sum over the nodes of the time units with ENS is equal to $0.9 + 0.1 \times 2 = 1.1 < 1.5 = \alpha$.

The operational cost associated with scenario 1 for the optimal solution to the high-point relaxation is $z_1^*(\beta) = (1 - \epsilon)\beta + \epsilon + 3M$. We also have $\phi_1(\mathbf{P}^*, \mathbf{F}^*) = 1 + 3M$, as it is cheaper to send all the energy to node 2. We finally obtain:

$$\lim_{\beta \rightarrow \infty} \frac{z_1^*(\beta)}{\phi_1(\mathbf{P}^*, \mathbf{F}^*)} = +\infty$$

□

Proposition 1 means that we can create an instance of Bilevel- α -Stoch-GTEPP, for which in an optimal solution to its high-point relaxation, there exists a scenario for which we pay A times more than an optimal

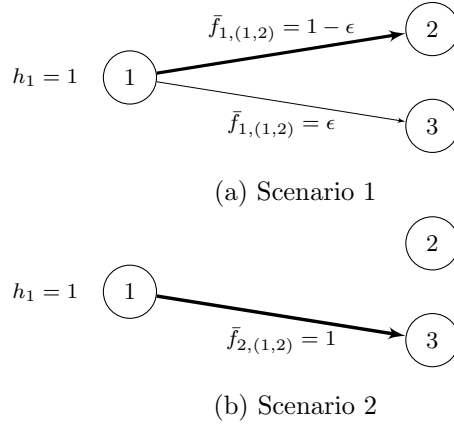


Figure 3: Optimal solution to the high-point relaxation

recourse solution in order to satisfy the “ α -time units criterion” and this for any value $A > 0$. This does not make sense as after the uncertainty is revealed (i.e., we know which scenario has been realized), there would be no reason to pay more than the cost of an optimal recourse solution to satisfy a constraint involving other scenarios that we already know will not realize. This shows the fundamental difference between Bilevel- α -Stoch-GTEPP and its high-point relaxation. We formalize this observation in the following remark.

Remark 2. *The high-point relaxation violates the **time-consistency principle**: “At every state of the system our optimal decisions should not depend on scenarios which we already know cannot happen in the future” (Shapiro, 2009). In the case of two-stage stochastic programming, after the first-stage decision is taken, a realization of the randomness is observed and we know that all but one scenarios will not happen. For the sake of consistency, the set of optimal solutions to the second-stage problem for a given scenario must only depend on first-stage decisions, without any regard for the other scenarios.*

We conclude with a more general remark that contradicts a result stated in Grimm et al. (2016)⁸.

Remark 3. *An optimal solution of the high point relaxation 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 (3) is a special case of a MIP-LP bilevel problem and is not equivalent to its high-point relaxation. 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 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

⁸We 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.

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 integrity 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 markets 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. They also show that their problem can be reformulated as an equivalent bilevel problem using the special structure 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). Using KKT conditions, a large equivalent single-level problem can be obtained.

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. Solving the high-point relaxation of the problem is already challenging as the constraint imposing the “ α -time units criterion” makes the operational sub-problem no longer decomposable by scenarios.

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 defines a sequence of single-level problems for which we expect feasible or optimal solutions to be bilevel feasible as the number of iterations increases. Single-level problems are obtained by relaxing the “ α -time units criterion” (see formulation (1a)-(1q)). We nevertheless impose a lower bound on the total investment cost to increase the likelihood of satisfying the “ α -time units criterion”. We implement a heuristic binary search strategy for determining the value of this lower bound. We show that under some mild assumptions, our procedure effectively computes a bilevel feasible solution in a finite number of iterations.

For readability purposes, we introduce a more concise representation of the problem: vector of variables \mathbf{x} is the concatenation of the vector of variables \mathbf{P} and \mathbf{F} (i.e., $\mathbf{x} = (\mathbf{P}, \mathbf{F})$) and the vector \mathbf{c} is the concatenation of the vectors \mathbf{c}^{prod} and \mathbf{c}^{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 $\bar{\mathbf{f}}_s$ (i.e., $\mathbf{y}_s = (\mathbf{h}_s, \bar{\mathbf{f}}_s, \mathbf{f}_s)$) and \mathbf{g}_s is the associated vector of operational costs. Hereafter we denote by $\mathbf{1}$ a vector of ones of any appropriate dimension. A more concise form of problem (2) reads as follows (where W , W' and T are matrices, $\ell_s = (|\mathcal{N}^*| + |\mathcal{A}|)|\mathcal{T}|$, and $\ell'_s = |\mathcal{A}||\mathcal{T}|$):

$$\begin{aligned} \phi_s(\mathbf{x}) = & \begin{cases} \min & \mathbf{g}_s^\top \mathbf{y}_s + M \mathbf{1}^\top \mathbf{u}_s \\ \text{s.t.} & W \mathbf{y}_s + W' \mathbf{u}_s \geq \mathbf{b}_s - T \mathbf{x} \\ & \mathbf{y}_s \in \mathbb{R}_+^{\ell_s} \times \mathbb{R}^{\ell'_s}, \mathbf{u}_s \in \mathbb{R}_+^{|\mathcal{N}| \times |\mathcal{T}|} \end{cases} \end{aligned} \tag{4a}$$

$$\tag{4b}$$

$$\tag{4c}$$

We refer to problem (4) as $(SP_s(\mathbf{x}))$. This leads to a concise form of problem (3), where matrices D and E and

vector \mathbf{q}_s are appropriately defined to represent constraints (3e)-(3f) (these elements depend on ϵ and C_s).

$$\min \quad \mathbf{c}^\top \mathbf{x} + \sum_{s \in \mathcal{S}} p_s (\mathbf{g}_s^\top \mathbf{y}_s + M \mathbf{1}^\top \mathbf{u}_s) \quad (5a)$$

$$s.t. \quad \mathbf{x} \in \mathbb{R}_+^{|\mathcal{N}^*|+|\mathcal{A}|} \quad (5b)$$

$$(\mathbf{y}_s, \mathbf{u}_s) \in \arg \min_{\substack{(\mathbf{y}'_s, \mathbf{u}'_s) \in \{\mathbf{y}''_s \in \mathbb{R}_+^{\ell_s} \times \mathbb{R}^{\ell'_s}, \mathbf{u}''_s \in \mathbb{R}_+^{|\mathcal{N}| \times |\mathcal{T}|} \\ : W \mathbf{y}''_s + W' \mathbf{u}''_s \geq \mathbf{b}_s - T \mathbf{x}\}}} \{\mathbf{g}_s^\top \mathbf{y}'_s + M \mathbf{1}^\top \mathbf{u}'_s\} \quad \forall s \in \mathcal{S} \quad (5c)$$

$$D \boldsymbol{\delta}_s + E \mathbf{u}_s \geq \mathbf{q}_s \quad \forall s \in \mathcal{S} \quad (5d)$$

$$\sum_{s \in \mathcal{S}} p_s (\mathbf{1}^\top \boldsymbol{\delta}_s) \leq \alpha \quad (5e)$$

$$\boldsymbol{\delta}_s \in \{0, 1\}^{|\mathcal{N}| \times |\mathcal{T}|} \quad \forall s \in \mathcal{S} \quad (5f)$$

Constraints (5c) 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 (5) as $(\mathbf{x}, \mathbf{y}, \mathbf{u})$.

We now introduce the *investment free expansion planning problem*, in which investment decisions have no cost.

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

$$\begin{cases} \min \sum_{s \in \mathcal{S}} p_s \phi_s(\mathbf{x}) \\ s.t. \quad (5b) \end{cases} \quad (6)$$

In the remainder of this paper, we assume that the VOLL M is large enough to prefer satisfying a demand whenever it is possible if only operations costs are considered. This hypothesis is realistic at a strategical level. We formalize this in Assumption 1 which says that if investing in generation and transmission capacity is free, then there does not exist any optimal solution with ENS.

Assumption 1. *We have $\mathbf{u} = \mathbf{0}$ in any optimal solution to the investment free expansion planning problem (6).*

Let \mathbf{x}^* be an optimal solution to the investment free problem and 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 (5).

Proposition 2. *Let \mathbf{x}^* be an optimal solution to the investment free problem. Every bilevel feasible solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{u}})$ to problem (5) 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. As \mathbf{x}^* is an optimal solution to the investment free 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 deduced from the fact that $\mathbf{c}^\top \hat{\mathbf{x}} \geq \mathbf{c}^\top \mathbf{x}^*$. \square

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. Note that if Assumption 1 is not verified, then our method can still be used but without guaranteeing that the restriction of the search space has not excluded all optimal solutions to Bilevel- α -Stoch-GTEPP (unlikely in practice). In the following, we use Iverson brackets⁹ for computing the number of time units with ENS from variables \mathbf{u} .

⁹ $[P] = 1$ if a statement P is true, and $[P] = 0$ 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 “ α -time units criterion” is not taken into account. Observe that $(\text{Inv}(\Lambda))$ is a two-stage stochastic single-level linear program.

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

$$\begin{cases} \min \mathbf{c}^\top \mathbf{x} + \sum_{s \in \mathcal{S}} p_s \phi_s(\mathbf{x}) \\ s.t. \text{ (5b)} \\ \mathbf{c}^\top \mathbf{x} \geq \Lambda \end{cases} \quad (7)$$

The core idea of our heuristic is to solve $(\text{Inv}(\Lambda))$ for every value of Λ selected during a binary search and to check if the computed solutions are bilevel feasible solutions to the Bilevel- α -Stoch-GTEPP (or if it is possible to derive such feasible solutions from the solutions computed during our method). Algorithm 1 outlines the general structure of our heuristic. We start by solving the investment free problem (6) and by initializing our binary search strategy (lines 2 to 4 in Algorithm 1). Then, given a lower bound Λ_{\min} (initially 0) and an upper bound Λ_{\max} (initially Λ_{free}) on the total investment cost, we solve $(\text{Inv}(\Lambda))$ 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 31), with $\tau \in (0, 1)$ the binary search parameter. If, when solving $(\text{Inv}(\Lambda))$, 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 Λ_{\min} is set to Λ (the motivation for this operation is that solutions with a larger total investment cost are more likely to satisfy the “ α -time units criterion”). Otherwise Λ_{\max} is set to the lowest total investment cost associated to a bilevel feasible solution (the rationale behind this operation is that an optimal bilevel feasible solution is more likely to have a lower total investment 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 (1). 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 29). 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 (7), 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:

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

This auxiliary subproblem searches, among all optimal solutions to $(SP_s(\bar{\mathbf{x}}))$, one solution that minimizes the sum over the nodes of the expected number of time units with ENS. 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 $\bar{\mathbf{x}} \in \mathbb{R}_+^{|\mathcal{N}^*| + |\mathcal{A}|}$. There exists $(\mathbf{y}_s, \mathbf{u}_s)$ optimal solution to $(SP_s(\bar{\mathbf{x}}))$ for every $s \in \mathcal{S}$ such that $(\bar{\mathbf{x}}, \mathbf{y}, \mathbf{u})$ is bilevel feasible if and only if $\sum_{s \in \mathcal{S}} p_s \varphi_s(\bar{\mathbf{x}}) \leq \alpha$.

Proof. (\Rightarrow) Let $(\mathbf{y}_s, \mathbf{u}_s)$ be an optimal solution to $(SP_s(\bar{\mathbf{x}}))$ for every $s \in \mathcal{S}$ such that $(\bar{\mathbf{x}}, \mathbf{y}, \mathbf{u})$ is bilevel feasible. 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 $\varphi_s(\bar{\mathbf{x}}) \leq \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} [u_{s,i,t} > \epsilon]$. Because $(\bar{\mathbf{x}}, \mathbf{y}, \mathbf{u})$ is bilevel feasible, we have $\sum_{s \in \mathcal{S}} p_s \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} [u_{s,i,t} > \epsilon] \leq \alpha$ 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 $\mathbf{g}_s^\top \mathbf{y}_s^* + M \mathbf{1}^\top \mathbf{u}_s^* \leq \phi_s(\bar{\mathbf{x}})$ 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. \square

Solving $|\mathcal{S}|$ mixed integer linear programs, with $|\mathcal{N}| \times |\mathcal{T}|$ binary variables in each 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.

Checking whether a given first-stage solution can produce a bilevel solution can be performed at different places in our method. In Algorithm 1, we define a parameter named **freq** as the *frequency* of the bilevel feasibility checking. This parameter is either set to **All**, if the algorithm checks the bilevel feasibility for all the first-stage solutions computed in the Benders decomposition algorithm during the whole optimization process, or set to **Opt** if the algorithm checks only the bilevel feasibility of optimal solutions to the current minimum investment expansion planning problem.

In total, our heuristic has the four following configurations, obtained by combining the values of the two parameters of accuracy and frequency described above: **All-Rand**, **All-MinENS**, **Opt-Rand**, **Opt-MinENS**.

Let $\Phi(\Lambda)$ be the minimum value of the sum over the nodes of the expected number of time units with ENS in an optimal solution to $(\text{Inv}(\Lambda))$. It is important to note that function Φ may not be monotonous in general (see B).

4.2 A Benders decomposition algorithm for solving $(\text{Inv}(\Lambda))$

As problem $(\text{Inv}(\Lambda))$ decomposes into as many independent subproblems as there are scenarios when the investments 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 $\mathbf{x} \in \mathbb{R}_+^{|\mathcal{N}^*| + |\mathcal{A}|}$ and we denote by $\text{Vert}(\Pi_s)$ the set of its extreme points. The multicut Benders reformulation of problem $(\text{Inv}(\Lambda))$, often referred to as the master program, is defined as follows:

$$\min \mathbf{c}^\top \mathbf{x} + \sum_{s \in \mathcal{S}} p_s \theta_s \quad (8a)$$

$$s.t. \text{ (5b)} \quad (8b)$$

$$\mathbf{c}^\top \mathbf{x} \geq \Lambda \quad (8c)$$

$$\theta_s \geq \boldsymbol{\pi}_s^\top (\mathbf{b}_s - T\mathbf{x}), \forall s \in \mathcal{S}, \forall \boldsymbol{\pi}_s \in \text{Vert}(\Pi_s) \quad (8d)$$

Constraints (8d) are Benders (optimality) cuts. The relaxation of (8) defined with a subset of the Benders cuts is called the relaxed master program and we denote this program at the 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 (i.e., the lower bound resulting from solving the relaxed master program is equal to the value of a feasible solution to $(\text{Inv}(\Lambda))$).

Algorithm 1: The Benders-based heuristic solution method

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

- 1 $k \leftarrow 0$, $\text{x_best_updated} \leftarrow \text{False}$, $\Lambda \leftarrow 0$, $\Lambda_{\min} \leftarrow 0$
- 2 Solve the investment free expansion planning problem (6) and retrieve \mathbf{x}_{best} an optimal first-stage solution, and the values $(\phi_s(\mathbf{x}_{\text{best}}))_{s \in \mathcal{S}}$
- 3 $z_{\text{best}} \leftarrow \mathbf{c}^\top \mathbf{x}_{\text{best}} + \sum_{s \in \mathcal{S}} p_s \phi_s(\mathbf{x}_{\text{best}})$
- 4 $\Lambda_{\max} \leftarrow \mathbf{c}^\top \mathbf{x}_{\text{best}}$
- 5 **while** $\Lambda_{\max} - \Lambda_{\min} > \epsilon_{\text{Inv}}$ **do**
- 6 Set the total minimum investment cost to Λ in $(\text{RMP})^{(k)}$ \triangleright Selecting a new bound on the total investment cost
- 7 $\text{UB}^{(k)} \leftarrow +\infty$, $\text{LB}^{(k)} \leftarrow -\infty$
- 8 $\text{x_best_updated} \leftarrow \text{False}$
- 9 **while** $\text{UB}^{(k)} > \text{LB}^{(k)} + \epsilon_{\text{opt}}$ **do** \triangleright Solving $(\text{Inv}(\lambda))$ with a Benders decomposition algorithm
- 10 $k \leftarrow k + 1$
- 11 Solve $(\text{RMP})^{(k)}$ and retrieve an optimal solution $(\check{\mathbf{x}}^{(k)}, \check{\boldsymbol{\theta}}^{(k)})$
- 12 $\text{LB}^{(k)} \leftarrow \mathbf{c}^\top \check{\mathbf{x}}^{(k)} + \sum_{s \in \mathcal{S}} p_s \check{\theta}_s^{(k)}$
- 13 **for** $s \in \mathcal{S}$ **do**
- 14 Solve $(SP_s(\check{\mathbf{x}}^{(k)}))$ and retrieve an optimal solution $(\mathbf{y}_s, \mathbf{u}_s)$, its value $\phi_s(\check{\mathbf{x}}^{(k)})$, and $\boldsymbol{\pi}_s \in \text{Vert}(\Pi_s)$
- 15 Add $\theta_s \geq \boldsymbol{\pi}_s^\top (\mathbf{b}_s - T\mathbf{x})$ to $(\text{RMP})^{(k)}$
- 16 $\text{UB}^{(k)} \leftarrow \min \{ \text{UB}^{(k-1)}, \mathbf{c}^\top \check{\mathbf{x}}^{(k)} + \sum_{s \in \mathcal{S}} p_s \phi_s(\check{\mathbf{x}}^{(k)}) \}$
- 17 $(\text{RMP})^{(k+1)} \leftarrow (\text{RMP})^{(k)}$
- 18 **if** $\text{freq} = \text{All}$ **and** $\text{updateBestBilevelSolution}(\text{acc}, \mathbf{x}^{(k)}, \mathbf{u}, z_{\text{best}})$ **then**
- 19 $\mathbf{x}_{\text{best}} \leftarrow \check{\mathbf{x}}^{(k)}$, $z_{\text{best}} \leftarrow \mathbf{c}^\top \check{\mathbf{x}}^{(k)} + \sum_{s \in \mathcal{S}} p_s \phi_s(\check{\mathbf{x}}^{(k)})$
- 20 $\text{x_best_updated} \leftarrow \text{True}$
- 21 $\Lambda_{\max} \leftarrow \min \{ \Lambda_{\max}, \mathbf{c}^\top \check{\mathbf{x}}^{(k)} \}$
- 22 **if** $\text{freq} = \text{Opt}$ **and** $\text{updateBestBilevelSolution}(\text{acc}, \mathbf{x}^{(k)}, \mathbf{u}, z_{\text{best}})$ **then**
- 23 $\mathbf{x}_{\text{best}} \leftarrow \check{\mathbf{x}}^{(k)}$, $z_{\text{best}} \leftarrow \mathbf{c}^\top \check{\mathbf{x}}^{(k)} + \sum_{s \in \mathcal{S}} p_s \phi_s(\check{\mathbf{x}}^{(k)})$
- 24 $\text{x_best_updated} \leftarrow \text{True}$
- 25 $\Lambda_{\max} \leftarrow \min \{ \Lambda_{\max}, \mathbf{c}^\top \check{\mathbf{x}}^{(k)} \}$
- 26 **if** $\neg \text{x_best_updated}$ **then**
- 27 $\Lambda_{\min} \leftarrow \Lambda$
- 28 **else**
- 29 **if** $\Lambda = 0$ **then return** \mathbf{x}_{best} $\triangleright \mathbf{x}_{\text{best}}$ is an optimal solution
- 30 **if** $\Lambda_{\max} > 100 * \Lambda_{\min}$ **then**
- 31 **if** $\Lambda_{\min} \neq 0$ **then** $\Lambda \leftarrow 10 * \Lambda_{\min}$ **else** $\Lambda \leftarrow 0.1 * \Lambda_{\max}$
- 32 **else**
- 33 $\Lambda \leftarrow \tau \Lambda_{\min} + (1 - \tau) \Lambda_{\max}$
- 34 **return** \mathbf{x}_{best} $\triangleright \mathbf{x}_{\text{best}}$ is a heuristic solution

Function $\text{updateBestBilevelSolution}(\text{acc}, \mathbf{x}, \mathbf{u}, z)$:

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

return **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** \triangleright Check if the given solution is bilevel feasible

return **True**

else

if $\text{acc} = \text{minENS}$ **then** \triangleright Check if there exists a bilevel feasible solution with the same lower-level cost

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**

return **True**

return **False**

It is important to observe that the Benders cuts (8d) computed when solving $(\text{Inv}(\Lambda))$ for a given $\Lambda \geq 0$ are still valid when solving $(\text{Inv}(\Lambda'))$ 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.

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 $(\text{Inv}(\Lambda))$, 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 $\hat{\mathbf{x}}^{(k)}$ to $(\text{RMP}^{(k)})$. Specifically, we have $\mathbf{x}^{(k)} = \kappa \hat{\mathbf{x}}^{(k)} + (1 - \kappa) \hat{\mathbf{x}}^{(k)}$ 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 $\hat{\mathbf{x}}^{(k)}$ by $\mathbf{x}^{(k)}$ from line 12 to line 16, and to update κ when computing $\text{UB}^{(k)}$ in line 16.

5 Experimental design and numerical results

We present in this section the results obtained by the four versions of the matheuristic presented in Section 4 on randomly generated instances. We compare these results to those obtained by solving the KKT-reformulation of the bilevel program using SOS1 to enforce complementarity constraints. We first present our instance generation procedure and the parameters we used to generate our instances, then we present the numerical results.

5.1 Instance generation

We describe how we generate instances of Bilevel- α -Stoch-GTEPP. 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.

We first present the graph generation algorithm we use. 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 nodes 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 a desired graph density is reached. We report in C the detailed procedure to produce our instances, and in Algorithm 2 for the graph generation procedure.

$ \mathcal{N} $	$ \mathcal{S} $	$ \mathcal{T} $	g^{flow}
10	5	5	$\{0.1, 1, 5\}$
20	$\{50, 250\}$	$\{24, 168\}$	$\{0.1, 1, 5\}$
30	$\{50, 250\}$	$\{24, 168\}$	$\{0.1, 1, 5\}$

Table 1: Parameters of the random generated data

We set $g_{s,a,t}^{\text{flow}} = g^{\text{flow}}$ for every $s \in \mathcal{S}, a \in \mathcal{A}, t \in \mathcal{T}$. Second, we generate random data with the parameters of Table 1. For each set of parameters, we generate three random instances with seeds in $\{|\mathcal{S}| + |\mathcal{N}| + |\mathcal{T}| + i : i = 0, 1, 2\}$. We denote an instance with $|\mathcal{N}|$ nodes, $|\mathcal{S}|$ scenarios, and $|\mathcal{T}|$ time units and a transmission operational cost g^{flow} as $\text{N}|\mathcal{N}|\text{-S}|\mathcal{S}|\text{-T}|\mathcal{T}|\text{-}g^{\text{flow}}\text{-}i$ and as $\text{N}|\mathcal{N}|\text{-S}|\mathcal{S}|\text{-T}|\mathcal{T}|\text{-}g^{\text{flow}}$ the set of three instances that share the same set of parameters. The other parameters are listed in Table 2.

Parameter	Value
c_i^{prod}	$U(1000, 8000)$
c_a^{line}	$U(1000, 8000)$
ϵ	10^{-3}
α	3
C	800
M	1000
$g_{s,i,t}^{\text{prod}}$	$U(20, 80)$
F_a^{init}	$U(0, 100)$
Δ_i	0.3
$d_{s,i,t}$	$U(50, 400)$

Table 2: Parameters for the data generation

This leads to a dataset of 81 instances. The minimum and maximum number of variables and constraints of instances for each number of nodes, time steps and scenarios are given in Table 3.

5.2 Numerical results

In order to evaluate the proposed heuristic, we run the four strategies, namely **All-Rand**, **All-MinENS**, **Opt-Rand**, **Opt-MinENS** on one core (sequential mode), on an Intel® Xeon® Gold SKL-6130 processor at 2,1 GHz with 96 GB of RAM with the TURBO boost (up to 3.7 GHz). The time limit is fixed to six hours for every algorithm. Julia JuMP is used as a modeler, and the mixed integer linear programming solver is CPLEX 12.10 (IBM, 2019). We set $\tau = 0.5$ for the binary search parameter, $\epsilon_{\text{opt}} = 10^{-6}$ and $\epsilon_{\text{Inv}} = 10^{-6}$. We use the Julia Package BilevelJuMP (Garcia et al., 2022) with CPLEX 12.10 to solve the single level KKT-reformulation of the bilevel problem (3) using SOS1. We hereafter refer to this method as **KKT-SOS1**.

We first show in Figure 4 the evolution of the sum over the nodes of the expected total number of time units with ENS according to the minimum investment value Λ on an instance with 10 nodes, 5 time steps and 5 scenarios. We sample 500 different values of Λ according to a normal law where the mean is set equal to the best feasible solution value obtained with the **All-Rand** heuristic and the standard deviation is such that the probability to sample a value 20% higher than the mean is 0.1. For every value of Λ , we show the value of $\sum_{s \in \mathcal{S}} p_s(\mathbf{1}^\top \delta_s)$ given by the solver in an optimal solution \mathbf{x}^* to the minimum investment problem (“Random” curve). We also report $\sum_{s \in \mathcal{S}} \varphi_s(\mathbf{x}^*)$ (“Minimum” curve), and the optimal value of $(\text{Aux}_s(\mathbf{x}^*))$ if we change the optimization direction (“Maximum” curve). We observe that the three curves are not perfectly decreasing (this confirms the non-monotonicity of function Φ discussed in B), although there is a clear correlation between a larger value of Λ with a smaller number of time units with ENS. The curves are decreasing almost everywhere, which justifies the use of the proposed heuristic in order to find a good feasible solution. 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. This shows that solving the auxiliary subproblems may help to find better solutions and motivates the introduction of the configuration $\text{acc} = \text{minENS}$ in our heuristic.

We now present in Table 4 the results we obtained for the Bilevel- α -Stoch-GTEPP on the 81 instances we generated. On each line, we show the mean time and mean gap to the best bilevel feasible solution found on the three instances with the same parameters. Let \mathbf{M} be the set of methods we compare. We denote by $z_{\mathbf{p},\mathbf{m}}$

$ \mathcal{N} $	$ \mathcal{T} $	$ \mathcal{S} $	# Variables	# Constraints
			Min - Max	Min - Max
10	5	5	1267 - 1319	1526 - 1676
20	24	50	128836 - 127238	153601 - 160801
30	24	50	190856 - 194459	237601 - 248401
20	24	250	624036 - 642039	768001 - 822001
30	24	250	936053 - 984061	1134001 - 1278001
20	168	50	873636 - 890438	1075201 - 1125601
30	168	50	1352458 - 1402864	1713601 - 1864801
20	168	250	4410037 - 4494039	5502001 - 5754001
30	168	250	6594054 - 6720057	8064001 - 8442001

Table 3: Minimum and maximum number of variables and constraints of instances, for each number of nodes, time steps and scenarios.

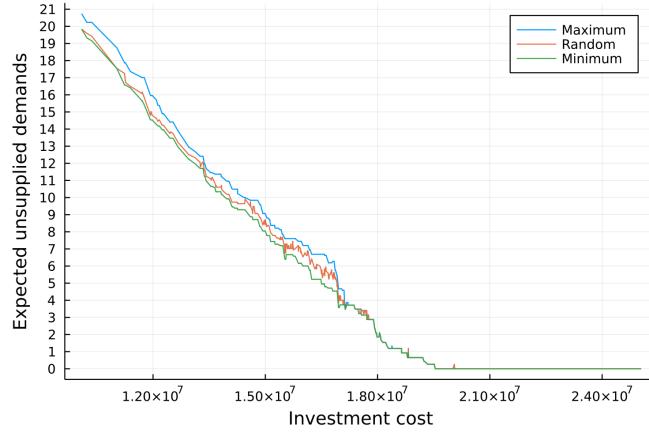


Figure 4: Evolution of $\sum_{s \in \mathcal{S}} p_s(\mathbf{1}^\top \delta_s)$ in an optimal solution to $(\text{Inv}(\Lambda))$ according to the minimum investment cost Λ , for an instance with 10 nodes, 5 time steps and 5 scenarios.

the best value of a solution computed with method \mathbf{m} on instance \mathbf{p} , and by $\bar{z}_{\mathbf{p}} = \min_{\mathbf{m} \in \mathbf{M}} \{z_{\mathbf{p}, \mathbf{m}}\}$ the best value obtained with all the methods on instance \mathbf{p} . The mean gap shown in Table 4 for a method $\mathbf{m} \in \mathbf{M}$ is computed as $\frac{1}{|(\mathcal{N}|\mathcal{N}|-\mathcal{S}|\mathcal{S}|-\mathcal{T}|\mathcal{T}|-\mathbf{g}^{\text{flow}})|} \sum_{\mathbf{p} \in \mathcal{N}|\mathcal{N}|-\mathcal{S}|\mathcal{S}|-\mathcal{T}|\mathcal{T}|-\mathbf{g}^{\text{flow}}} \frac{z_{\mathbf{p}, \mathbf{m}} - \bar{z}_{\mathbf{p}}}{\bar{z}_{\mathbf{p}}}$.

We first remark that, even on the small instances with 10 nodes, 5 time steps and 5 scenarios, **KKT-SOS1** is not able to find any feasible solution in six hours. This result encourages the use of a heuristic solution method in an operational setting to find feasible solutions. We also see, as we could expect, that the heuristic with the **All-MinENS** configuration is not able to complete the binary search in six hours on the largest instances. Indeed, every auxiliary subproblem is a large-scale binary program and solving these subproblems for every first-stage solution evaluated during the algorithm is very time-consuming. We can see that, even with the **Opt-MinENS** configuration, in which the auxiliary subproblems are only solved for an optimal solution to the minimum investment problem, our heuristic terminates almost at the time limit.

The second remark is that the heuristic with both the **All-Rand** and the **All-MinENS** configurations leads on average to the best results, mostly for the large instances with 168 time steps. One can find a solution up to 16% better than the solution computed with an **Opt** configuration. This shows that the use of a Benders decomposition algorithm as an exploratory algorithm to select first-stage solutions allows to improve the results. For a fixed value of the accuracy parameter acc , the results of the heuristic with the **All** configuration are always better than the ones obtained with the **Opt** configuration (**All-Rand** compared to **Opt-Rand** and **All-**

Instance	Opt-Rand		All-Rand		Opt-MinENS		All-MinENS		KKT-SOS1	
	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap
N10-S5-T5-0.1	11.56	0.29%	11.67	0.29%	12.59	0.00%	14.45	0.00%	21600.00	—
N10-S5-T5-1	11.76	0.47%	11.86	0.12%	12.58	0.35%	14.36	0.00%	21600.00	—
N10-S5-T5-5	11.57	0.57%	11.62	0.25%	12.57	0.00%	14.82	0.00%	21600.00	—
N20-S50-T24-0.1	66.45	0.28%	69.65	0.18%	145.80	0.00%	403.40	0.00%	21600.00	—
N20-S50-T24-1	70.95	0.29%	74.24	0.22%	135.28	0.04%	411.81	0.00%	21600.00	—
N20-S50-T24-5	73.05	0.38%	76.41	0.18%	156.13	0.05%	451.74	0.00%	21600.00	—
N30-S50-T24-0.1	171.32	2.11%	178.72	0.70%	343.67	0.37%	1217.57	0.00%	21600.00	—
N30-S50-T24-1	154.42	0.78%	162.17	0.42%	341.48	0.10%	1168.56	0.00%	21600.00	—
N30-S50-T24-5	156.59	1.81%	164.05	0.79%	317.82	1.23%	1174.49	0.00%	21600.00	—
N20-S250-T24-0.1	422.90	0.27%	436.77	0.16%	847.50	0.00%	1729.02	0.00%	21600.00	—
N20-S250-T24-1	465.59	0.40%	486.82	0.15%	860.54	0.16%	1901.98	0.00%	21600.00	—
N20-S250-T24-5	455.92	0.38%	476.75	0.21%	861.77	0.10%	2036.31	0.00%	21600.00	—
N30-S250-T24-0.1	1203.59	0.25%	1232.15	0.19%	2025.75	0.02%	4456.90	0.00%	21600.00	—
N30-S250-T24-1	1147.62	0.29%	1179.92	0.25%	1979.65	0.17%	4367.37	0.00%	21600.00	—
N30-S250-T24-5	1226.13	0.29%	1255.79	0.21%	2049.97	0.09%	4822.33	0.00%	21600.00	—
N20-S50-T168-0.1	406.94	14.55%	472.96	0.00%	2439.40	13.90%	19943.22	0.00%	21600.00	—
N20-S50-T168-1	412.26	10.25%	470.42	0.08%	2078.35	10.25%	20080.08	0.00%	21600.00	—
N20-S50-T168-5	380.15	5.40%	437.25	0.11%	1611.97	5.27%	18766.62	0.00%	21600.00	—
N20-S250-T168-0.1	2355.89	6.20%	2966.55	0.00%	13030.07	5.85%	21600.00	1.74%	21600.00	—
N20-S250-T168-1	2468.53	3.96%	3050.63	0.00%	13780.36	3.96%	21600.00	0.47%	21600.00	—
N20-S250-T168-5	2643.71	2.02%	3097.23	0.00%	17224.53	1.97%	21600.00	34.22%	21600.00	—
N30-S50-T168-0.1	1092.90	9.80%	1204.49	0.00%	7928.70	9.13%	21600.00	2.07%	21600.00	—
N30-S50-T168-1	1019.33	6.28%	1130.36	0.00%	7869.48	5.50%	21600.00	0.00%	21600.00	—
N30-S50-T168-5	975.12	3.73%	1084.62	0.00%	5210.29	3.49%	21600.00	3.25%	21600.00	—
N30-S250-T168-0.1	4947.79	15.35%	6430.68	0.00%	21600.00	14.84%	21600.00	63.89%	21600.00	—
N30-S250-T168-1	5324.71	16.67%	6669.99	0.00%	16568.41	16.67%	21600.00	65.80%	21600.00	—
N30-S250-T168-5	6763.42	10.89%	8364.26	0.00%	19465.20	10.89%	21600.00	26.51%	21600.00	—

Table 4: Solving time and gap to \bar{z}_p for the four configurations of the heuristic and for **KKT-SOS1**

MinENS compared to **Opt-MinENS**), except when the **All-MinENS** heuristic does not finish within the time limit. We see, for example for the N30-S250-T168-1 instances, that the **All-MinENS** heuristic finds the best solutions although it reaches the time limit. Finally, according to the results we obtained on the tested instances, it seems that, depending on the size of the problem, either the **All-MinENS** or the **All-Rand** are the best configurations for our heuristic.

We now analyze the sensitivity of the heuristic to the optimality gap ϵ_{opt} used in the Benders decomposition algorithm performed at each iteration of the heuristic. As we can see in Figure 5, when the optimality gap is larger, the evolution of the expected number of time units with ENS tends to be fuzzier. Indeed, with a larger optimality gap, the Benders decomposition algorithm may return investment solutions in which the expected total ENS is larger than the same value in an optimal solution to the minimum investment expansion planning problem, and these solutions are likely to present a larger expected number of time units with ENS. A lower optimality gap tends to produce better results, but at the expense of solving the minimum investment problem, which takes time. We present in Table 5 the results on our 81 instances obtained by the **All-Rand** heuristic for two values of ϵ_{opt} . As expected, we see that the total computing time for the **All-Rand** heuristic is significantly increased when the optimality gap of the Benders decomposition is smaller. However, the **All-Rand** heuristic with an optimality gap of 10^{-10} in the Benders decomposition gives better results than with an optimality gap of 10^{-6} , as we could have expected from the results presented in Figure 5. It is competitive with the **All-MinENS** heuristic on middle-sized instances, and outperforms the other configurations on large-sized instances. It seems that, in order to get the best possible results with our heuristic, one should try to reach the best possible optimality gap in the Benders decomposition algorithm. However, a small optimality gap might be hard to reach for large-sized instance as well as very expensive in terms of computing time, and one should find a compromise between the quality of the solution and the computing time.

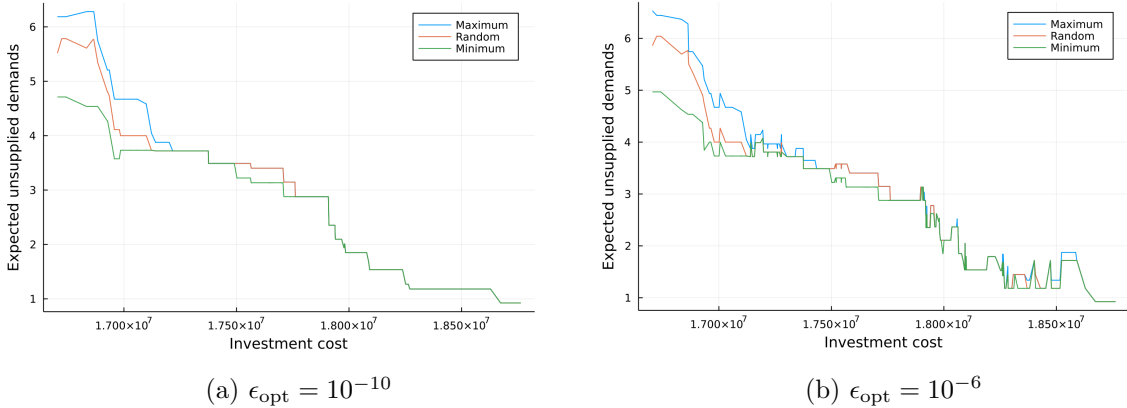


Figure 5: Evolution of $\sum_{s \in \mathcal{S}} p_s(\mathbf{1}^\top \delta_s)$ in an optimal solution to $(\text{Inv}(\Lambda))$ according to the minimum investment cost Λ for an instance with 10 nodes, 5 time steps and 5 scenarios.

We finally present in Table 6 a comparison between the best feasible solutions found with the heuristic compared with an optimal solution to the two-stage stochastic expansion planning problem (1) that ignores the “ α -time units criterion”. This latter problem is a relaxation of the Bilevel- α -Stoch-GTEPP ((3) or equivalently (5)). We observe that optimal solutions to this relaxation have a very large value for the sum over the nodes of the number of time units with ENS. To assess the quality of the solutions computed by the heuristic, we compute a lower bound on the optimal value of our bilevel problem as the maximum between the optimal value of the relaxation (1), and the best lower bound obtained from the solver with **KKT-SOS1**. Although we do not expect this bound to be strong, it allows a first evaluation of the quality of the solutions returned by our heuristic. The average gap is approximately 15%, but answering the question of whether this is more related to the weakness of this bound or to the quality of the solutions of our heuristic is out of our reach at this moment.

Instance	All-Rand $\epsilon_{\text{opt}} = 10^{-6}$		All-Rand $\epsilon_{\text{opt}} = 10^{-10}$		All-MinENS $\epsilon_{\text{opt}} = 10^{-6}$	
	Time	Gap	Time	Gap	Time	Gap
N10-S5-T5-0.1	11.67	0.29%	11.72	0.21%	14.45	0.00%
N10-S5-T5-1	11.86	0.12%	11.93	0.11%	14.36	0.00%
N10-S5-T5-5	11.62	0.32%	11.91	0.26%	14.82	0.07%
N20-S50-T24-0.1	69.65	0.42%	95.20	0.00%	403.40	0.24%
N20-S50-T24-1	74.24	0.44%	100.20	0.00%	411.81	0.22%
N20-S50-T24-5	76.41	0.28%	102.71	0.00%	451.74	0.10%
N30-S50-T24-0.1	178.72	1.96%	305.37	0.00%	1217.57	1.25%
N30-S50-T24-1	162.17	1.41%	288.43	0.01%	1168.56	0.99%
N30-S50-T24-5	164.05	0.79%	281.32	0.44%	1174.49	0.00%
N20-S250-T24-0.1	436.77	0.22%	706.59	0.14%	1729.02	0.06%
N20-S250-T24-1	486.82	0.19%	737.62	0.22%	1901.98	0.04%
N20-S250-T24-5	476.75	0.21%	788.50	0.18%	2036.31	0.00%
N30-S250-T24-0.1	1232.15	0.45%	2321.90	0.00%	4456.90	0.26%
N30-S250-T24-1	1179.92	0.35%	2544.04	0.00%	4367.37	0.10%
N30-S250-T24-5	1255.79	0.24%	2612.04	0.05%	4822.33	0.03%
N20-S50-T168-0.1	472.96	2.10%	696.77	0.01%	19943.22	2.10%
N20-S50-T168-1	470.42	1.13%	680.51	0.00%	20080.08	1.05%
N20-S50-T168-5	437.25	0.60%	668.13	0.71%	18766.62	0.49%
N20-S250-T168-0.1	2966.55	0.70%	7003.94	0.34%	21600.00	2.44%
N20-S250-T168-1	3050.63	1.86%	5100.85	0.02%	21600.00	2.34%
N20-S250-T168-5	3097.23	2.29%	5104.82	0.00%	21600.00	35.35%
N30-S50-T168-0.1	1204.49	1.35%	2178.98	0.00%	21600.00	3.46%
N30-S50-T168-1	1130.36	1.26%	2082.04	0.33%	21600.00	1.26%
N30-S50-T168-5	1084.62	3.69%	2032.77	0.00%	21600.00	7.06%
N30-S250-T168-0.1	6430.68	6.16%	11648.15	0.00%	21600.00	73.79%
N30-S250-T168-1	6669.99	3.30%	12146.08	2.37%	21600.00	70.47%
N30-S250-T168-5	8364.26	1.75%	14958.71	33.63%	21600.00	28.79%

Table 5: Solving time and gap to the best feasible solution (computed by our heuristic in all our tests) for the best solution computed by the heuristic with the **All-Rand** configuration with an optimality gap $\epsilon_{\text{opt}} = 10^{-10}$, and with the configurations **All-Rand** and **All-MinENS** with $\epsilon_{\text{opt}} = 10^{-6}$.

Instance	Problem (1)		LB = $\max\{\mathbf{z}_{(1)}^*, \text{LB}_{\text{KKT-SOS1}}\}$	Gap= $(\mathbf{z}_{\text{Heur}} - \text{LB})/\text{LB}$
	$\sum_{s \in \mathcal{S}} p_s(\mathbf{1}^\top \boldsymbol{\delta}_s)$	$\mathbf{z}_{(1)}^*$		
N10-S5-T5-0.1	32.85	$9.256 \cdot 10^6$	$1.286 \cdot 10^7$	21.54%
N10-S5-T5-1	28.61	$9.258 \cdot 10^6$	$1.291 \cdot 10^7$	21.10%
N10-S5-T5-5	30.11	$9.266 \cdot 10^6$	$1.302 \cdot 10^7$	20.21%
N20-S50-T24-0.1	54.15	$4.184 \cdot 10^7$	$4.222 \cdot 10^7$	11.80%
N20-S50-T24-1	56.54	$4.193 \cdot 10^7$	$4.232 \cdot 10^7$	11.73%
N20-S50-T24-5	56.27	$4.231 \cdot 10^7$	$4.273 \cdot 10^7$	11.85%
N30-S50-T24-0.1	83.24	$5.998 \cdot 10^7$	$6.014 \cdot 10^7$	18.81%
N30-S50-T24-1	83.92	$6.011 \cdot 10^7$	$6.035 \cdot 10^7$	18.20%
N30-S50-T24-5	87.03	$6.065 \cdot 10^7$	$6.090 \cdot 10^7$	16.81%
N20-S250-T24-0.1	55.93	$4.087 \cdot 10^7$	$4.087 \cdot 10^7$	13.53%
N20-S250-T24-1	54.47	$4.095 \cdot 10^7$	$4.095 \cdot 10^7$	13.41%
N20-S250-T24-5	56.97	$4.130 \cdot 10^7$	$4.130 \cdot 10^7$	13.35%
N30-S250-T24-0.1	65.98	$5.801 \cdot 10^7$	$5.801 \cdot 10^7$	14.85%
N30-S250-T24-1	66.45	$5.815 \cdot 10^7$	$5.815 \cdot 10^7$	14.73%
N30-S250-T24-5	65.79	$5.875 \cdot 10^7$	$5.875 \cdot 10^7$	14.61%
N20-S50-T168-0.1	28.22	$7.638 \cdot 10^7$	$7.638 \cdot 10^7$	8.84%
N20-S50-T168-1	30.74	$7.695 \cdot 10^7$	$7.695 \cdot 10^7$	11.22%
N20-S50-T168-5	30.81	$7.941 \cdot 10^7$	$7.941 \cdot 10^7$	7.02%
N30-S50-T168-0.1	46.97	$1.129 \cdot 10^8$	$1.129 \cdot 10^8$	15.34%
N30-S50-T168-1	42.99	$1.138 \cdot 10^8$	$1.138 \cdot 10^8$	13.89%
N30-S50-T168-5	45.93	$1.173 \cdot 10^8$	$1.173 \cdot 10^8$	9.64%
N20-S250-T168-0.1	26.74	$7.456 \cdot 10^7$	$7.456 \cdot 10^7$	8.90%
N20-S250-T168-1	25.26	$7.518 \cdot 10^7$	$7.518 \cdot 10^7$	9.08%
N20-S250-T168-5	26.29	$7.779 \cdot 10^7$	$7.779 \cdot 10^7$	6.73%
N30-S250-T168-0.1	53.26	$1.212 \cdot 10^8$	$1.212 \cdot 10^8$	16.20%
N30-S250-T168-1	45.64	$1.223 \cdot 10^8$	$1.223 \cdot 10^8$	19.37%
N30-S250-T168-5	49.27	$1.268 \cdot 10^8$	$1.268 \cdot 10^8$	13.88%

Table 6: Comparison between the solution to the two-stage stochastic expansion planning problem (1) (that ignores the “ α -time units criterion”) and the best bilevel solution (with value \mathbf{z}_{Heur}) found by our heuristic.

6 Conclusion

We presented in this paper how we can embed a reliability standard called the “ α -time units criterion” into a stochastic transmission and generation investment planning problem defined with an integrated approach. Modeling an upper bound on the sum over the regions of the expected number of time units with ENS induces a bilevel formulation of the problem where the lower level problem decomposes into an operational subproblem for every scenario. We proved that relaxing the optimality conditions on the solutions to the lower-level (i.e., working with the high-point relaxation) leads to invalid solutions, and showed that the ratio between the value of an optimal solution to its 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 proposed a heuristic solution method based on iteratively solving stochastic expansion planning problems with a minimum investment constraint. This solution method computes feasible solutions even on instances with several millions variables and constraints in less than six hours.

The numerical complexity of the studied problem makes the production of good lower bounds on the optimal value a very difficult challenge. Future research on this problem should focus on producing lower-bounds to better assess the quality of the solutions computed by our method. It would prove useful if solving exactly the high-point relaxation can be done in a time efficient way. Designing advanced decomposition-based methods that solve exactly the problem would be the next step.

Another important path to study is the relevance of the “ α -time units criterion” to guarantee the reliability of the electricity network. As it is formulated, it 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). This calls for a redefinition of the RS to improve the resilience of the electrical network. Chance constrained optimization or robust optimization are avenues to explore for defining a more risk-averse criterion.

Acknowledgments

This project has been funded by RTE (Réseau de Transport d’Electricité), French company in charge of the electricity network management, through the projects Antares and Antares Xpansion: <https://github.com/AntaresSimulatorTeam/antares-xpansion>, which are used for long-term adequacy studies. Computer time for this study was provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l’Adour.

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.
- Ben-Ameur, W. and Neto, J. (2007). Acceleration of cutting-plane and column generation algorithms: Applications to network design. *Networks*, 49(1):3–17.
- Braga, A. and Saraiva, J. (2005). A multiyear dynamic approach for transmission expansion planning and long-term marginal costs computation. *IEEE Transactions on Power Systems*, 20(3):1631–1639.
- 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.
- 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.
- IBM (2019). *IBM (2019) IBM ILOG CPLEX 12.10 User’s Manual (IBM ILOG CPLEX Division, Incline Village, NV)*.
- 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., Sirola, 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.
- 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.
- Palintier, 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%A9veloppement%20de%20r%C3%A9seau%202019%20-%20Synth%C3%A8se%20E2%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.
- 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.
- 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.

A An illustrative example of the need to explicitly address the “ α -time units criterion”

We consider the example presented in Figure 6.

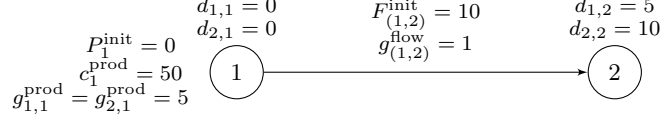


Figure 6: An instance with 2 nodes, 1 arc, 2 scenarios, 1 time unit, and such that $M = 100$ and $\epsilon = 0$.

Using notation introduced in Section 3 but we omit the time unit index:

$$\left\{ \begin{array}{l} \min \quad 50P_1 + 0.9(h_{1,1} + 100u_{1,2}) + 0.1(h_{2,1} + 100u_{2,2}) \\ s.t. \quad h_{s,1} \leq P_1 \quad \forall s \in \{1, 2\} \\ \quad \quad h_{s,1} - f_{s,(1,2)} + u_{s,1} \geq 0 \quad \forall s \in \{1, 2\} \\ \quad \quad f_{1,(1,2)} + u_{1,2} \geq 5 \\ \quad \quad f_{2,(1,2)} + u_{2,2} \geq 10 \\ \quad \quad \bar{f}_{s,(1,2)} \geq f_{s,(1,2)} \quad \forall s \in \{1, 2\} \\ \quad \quad \bar{f}_{s,(1,2)} \geq -f_{s,(1,2)} \quad \forall s \in \{1, 2\} \\ \quad \quad h_{s,1} \in \mathbb{R}_+ \quad \forall s \in \{1, 2\} \\ \quad \quad u_{s,i} \in \mathbb{R}_+ \quad \forall s \in \{1, 2\}, i \in \{1, 2\} \\ \quad \quad f_{s,(1,2)} \in \mathbb{R}_+ \quad \forall s \in \{1, 2\} \\ \quad \quad P_1 \in \mathbb{R}_+ \end{array} \right.$$

We show in the following table the distribution of the cost for different investment solutions.

Value of P_1	Investment cost	Operational cost	Total cost	$(u_{1,2}, u_{2,2})$
0	0	550	550	(5, 10)
5	250	55	305	(0, 5)
10	500	5.5	505.5	(0, 0)

The optimal solution to the problem is obtained for $P_1 = 5$. We observe that in this solution, the demand is not fully satisfied in scenario 2. Investing more in order to satisfy all the demands in every scenario is more expensive, and is not economically viable as the probability associated with this demand is low (0.1 in this example). The example shows that if a given level of reliability of the system must be ensured, it is necessary to explicitly add a constraint to the problem.

B The non-monotonicity of function Φ

Figure 7 presents an instance such that the optimal solution to $(\text{Inv}(0))$ contains 1 node with ENS (i.e., $\Phi(0) = 1$), while the optimal solution to $(\text{Inv}(10))$ contains 2 nodes with ENS (i.e., $\Phi(10) = 2$) and the optimal solution to $(\text{Inv}(520))$ contains 2 nodes with ENS (i.e., $\Phi(520) = 1$) (see Table 7). Note that the instance satisfies Assumption 1.

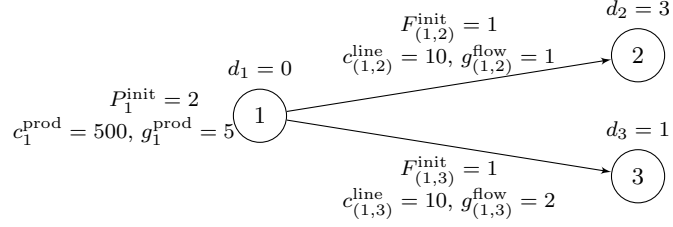


Figure 7: An instance with 3 nodes, 2 arcs, 1 scenario, 1 time unit, and such that $M = 50$ and $\epsilon = 0.05$.

Variable	Λ		
	0	10	520
P_1	0	0	1
h_1	2	2	3
$F_{(1,2)}$	0	1	2
$\bar{f}_{(1,2)}$	1	2	3
$F_{(1,3)}$	0	0	0
$\bar{f}_{(1,3)}$	1	0	0
u_1	0	0	0
u_2	2	1	0
u_3	0	1	1

Table 7: The optimal solution to $(\text{Inv}(\Lambda))$

C Graph generation procedure

Algorithm 2: Undirected graph generation procedure

Parameters: Number of nodes to generate N , Grid ratio $r > 0$, Proportion of nodes $p \in (0, 1)$, desired graph density $d \in (0, 1)$, A maximum number of tries $k_{\max} \in \mathbb{N}^*$ to achieve the desired graph density

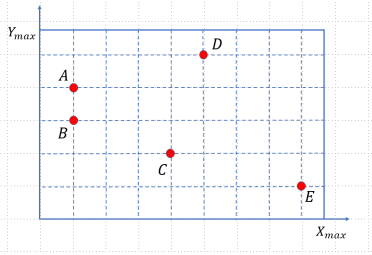
- 1 **Initialization:** $\mathcal{N} \leftarrow \emptyset, \mathcal{A} \leftarrow \emptyset$
- 2 Compute $X_{\max} = \sqrt{\frac{N \cdot r}{p}}$ and $Y_{\max} = \sqrt{\frac{N}{p \cdot r}}$
- 3 Sample N integer points in $[1, X_{\max}] \times [1, Y_{\max}]$ and add them to \mathcal{N}
- 4 $D_{\max} \leftarrow 0$
- 5 **while** G is not connected **do** ▷ Adding edges until connectivity
 - 6 $D_{\max} \leftarrow D_{\max} + 1$
 - 7 $\mathcal{A} \leftarrow \{(i, j), 1 \leq i \leq N, i + 1 \leq j \leq N \mid \text{EuclDist}(i, j) \leq D_{\max}\}$
- 8 **while** $\frac{2|\mathcal{A}|}{N(N-1)} > d$ **do** ▷ Removing edges to get required density
 - 9 $k \leftarrow 0$
 - 10 Choose (i, j) in \mathcal{A}
 - 11 **while** graph $(\mathcal{N}, \mathcal{A} \setminus \{(i, j)\})$ is not connected **do**
 - 12 **if** $k = k_{\max}$ **then return** $G = (\mathcal{N}, \mathcal{A})$ ▷ Unable to reach density
 - 13 $k \leftarrow k + 1$
 - 14 Choose (i, j) in \mathcal{A}
 - 15 $\mathcal{A} \leftarrow \mathcal{A} \setminus \{(i, j)\}$
- 16 **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 $[1, X_{\max}] \times [1, Y_{\max}]$, where $X_{\max} = \sqrt{\frac{N \cdot r}{p}}$ and $Y_{\max} = \sqrt{\frac{N}{p \cdot r}}$. Figure 8a shows a possible result of such a sampling procedure.

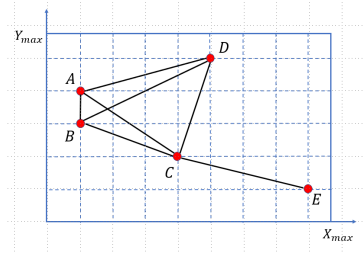
The second step of the graph generation procedure consists of adding edges to the graph so that it becomes connected. Let $D_{\max} = 1$, we add all the edges between nodes which are at an Euclidean distance of at most D_{\max} . If the resulting graph is connected, we stop, otherwise D_{\max} increases by one and we restart a similar process. In the example presented in Figure 8a, this leads to the connected graph of Figure 8b.

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 keeping the graph connected after a given number of trials for edge removal. For example, in the graph of Figure 8b, edge (C, E) cannot be removed as it would lead to a disconnected graph.

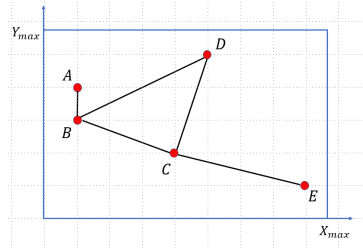
In our experiments, we use $r = 1.5$, $p = 0.005$, $d = 0.3$, and $k_{\max} = 100$.



(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 8: A possible result of Algorithm 2 for $N = 5$, $r = 1.5$, $p = 0.1$ and $d = 0.5$.