An Optimal Expansion Planning of Power Systems considering Cycle-Based AC Optimal Power Flow

Erik F. Alvarez[a], Juan Camilo López[b], Luis Olmos[a], Andres Ramos[a]

[a]*Institute for Research in Technology (IIT), Comillas Pontifical University, Alberto Aguilera 23, Madrid, 28015, Spain*
[b]*School of Electrical and Computer Engineering (FEEC), University of Campinas (UNICAMP), Av. Albert Einstein, Nº 400 - Cidade Universitária, Campinas-SP, 13083-852, Brazil*

Abstract

This paper presents a novel mixed-integer linear optimization formulation of the AC network-constrained, cost-based, integrated expansion planning problem. The formulation is used to determine the investment needs per technology including the location and sizing of new generation, energy storage, and transmission network assets in a future low-carbon power system. To reduce the size of the resulting problem, the AC optimal power flow (AC-OPF) model is represented in a compact way using cycle constraints. A bound tightening procedure is also considered to reduce the search space and improve the solver performance by adjusting the voltage bounds within the AC-OPF. Contrary to typically used formulations of the integrated expansion planning problem, the constraints considered here include all main aspects of system operation, namely unit commitment, energy storage system management, AC-OPF, and reactive power compensation. Thus, in this paper, we examine how both the proposed transmission expansion modeling developments and the interrelation of the integrated planning constraints affect the computation of the solution to the expansion planning problem. The performance of this formulation is assessed on the RTS-GMLC test system by computing the expansion plan and comparing it with the results of three other expansion planning formulations most frequently employed in the recent literature to address the integrated expansion planning problem for medium to large-scale systems. Expansion plans are computed and compared for different case studies and multiple scenarios. According to the comparative analysis, neglecting the AC-OPF or the unit commitment constraints can increase the total system costs by 7.10%–9.57% or 6.29%–8.39%, respectively.

Preprint submitted to SEGAN *2024/05/05*

Unlike other modeling approaches, the proposed approach does not rely on simplifications that impact the quality of the solution. Thanks to the incorporated cycle-based AC-OPF constraints and the consideration of a bound tightening procedure, the computation time is reduced by 17.67%–27.21%.

Keywords: AC optimal power flow, cycle constraints, low-carbon emissions, power system expansion planning, unit commitment

---

Notation

Acronyms:

AC/DC	Alternating/direct current
ESS	Energy storage system
GEP	Generation expansion planning
IEP	Integrated expansion planning
KVL	Kirchhoff's voltage law
MILP	Mixed-integer linear programming
OPF	Optimal power flow
PSH	Pumped-storage hydro
RES	Renewable energy source
RPP	Reactive power planning
SEP	Storage expansion planning
TEP	Transmission expansion planning
UC	Unit commitment
VRE	Variable renewable energy

Sets:

$\mathcal{B}$	Set of nodes
$\mathcal{C}$	Set of circuits
$\mathcal{E}$	Set of ESS units
$\mathcal{E}^e$	Set of existing ESS units, $\mathcal{E}^e \subseteq \mathcal{E}$
$\mathcal{E}^c$	Set of candidate ESS units, $\mathcal{E}^c \subseteq \mathcal{E}$
$\mathcal{G}$	Set of generation units
$\mathcal{G}^e$	Set of existing generation units, $\mathcal{G}^e \subseteq \mathcal{G}$
$\mathcal{G}^c$	Set of candidate generation units, $\mathcal{G}^c \subseteq \mathcal{G}$
$\mathcal{K}$	Set of cycles
$\mathcal{K}^e$	Set of existing cycles, $\mathcal{K}^c \subseteq \mathcal{K}$

$\mathcal{K}^c$	Set of candidate cycles, $\mathcal{K}^e \subseteq \mathcal{K}$
$\mathcal{L}$	Set of lines, $\mathcal{L} \subseteq \mathcal{B} \times \mathcal{B}$
$\mathcal{L}^e$	Set of existing lines, $\mathcal{L}^e \subseteq \mathcal{L}$
$\mathcal{L}^c$	Set of candidate lines, $\mathcal{L}^c \subseteq \mathcal{L}$
$\mathcal{N}$	Set of load levels or time steps
$\mathcal{R}$	Set of capacitor banks
$\mathcal{R}^e$	Set of existing capacitor banks, $\mathcal{R}^e \subseteq \mathcal{R}$
$\mathcal{R}^c$	Set of candidate capacitor banks, $\mathcal{R}^c \subseteq \mathcal{R}$
$\mathcal{S}$	Set of synchronous compensator
$\mathcal{S}^e$	Set of existing synchronous compensator, $\mathcal{S}^e \subseteq \mathcal{S}$
$\mathcal{S}^c$	Set of candidate synchronous compensator, $\mathcal{S}^c \subseteq \mathcal{S}$

Indices:

c	Circuit, $c \in \mathcal{C}$.
e	ESS unit (i.e., battery storage or PSH), $e \in \mathcal{E}$, and $\mathcal{E} \subseteq \mathcal{G}$.
g	Generation unit (i.e., thermal or RES), $g \in \mathcal{G}$.
i	Node, $i \in \mathcal{B}$.
ij	Branch from node i to j.
ijc	Line, $ijc \in \mathcal{L}$.
n	Load level, $n \in \mathcal{N}$.
k	Cycle, $k \in \mathcal{K}$.
r	Capacitor bank, $r \in \mathcal{R}$.
s	Synchronous compensator, $s \in \mathcal{S}$.

System Parameters:

S_B	Base power [MW]
D_n	Duration of the interval between two load levels [h]
C^{shed}	Cost of energy not served [USD/p.u.]
P^{co2}	CO2 emission cost [USD/tCO2]
P^d_{ni}, Q^d_{ni}	Active/reactive power demand [MW, Mvar]

Generation and Storage Parameters:

C^{gen}_g	Annualized fixed installation cost of a candidate generation unit [MUSD]
C^{sto}_e	Annualized fixed installation cost of a candidate ESS unit [MUSD]
$\overline{P}_{ng}, \underline{P}_{ng}$	Maximum/minimum active power generation from thermal units [MW]
$\overline{P}_{ne}, \underline{P}_{ne}$	Maximum/minimum active power generation from storage units [MW]
$\overline{Q}_{ng}, \underline{Q}_{ng}$	Maximum/minimum reactive power generation [Mvar]

$\overline{P}^{c}_{ne}$	Maximum active power consumption of an ESS unit [MW]
CF_g	Fixed cost of a generation unit [MUSD/h]
CV_g	Variable cost of a generation unit (It includes fuel, O&M, and emission costs) [MUSD/MWh]
CV_e	Variable cost of a storage unit [MUSD/MWh]
R^u_g, R^d_g	Ramp-up and ramp-down of a generation unit [MW/h]
T^u_g, T^d_g	Minimum up-time and down-time of a generation unit [h]
C^{su}_g	Startup cost of a generation unit [MUSD]
C^{sd}_g	Shutdown cost of a generation unit [MUSD]
τ_e	Seasonality of the ESS unit (e.g., 24 hours, 168 hours, and 672 hours for daily, weekly, and monthly management) [h]
η_e	Round-trip efficiency of the pump/turbine cycle of a hydropower plant, or charge/discharge of an ESS unit [p.u.]
Ψ_e	Storage capacity of an ESS unit [MWh]
E_e	Energy inflow of an ESS unit [MWh]
pf^{cap}	Capacitive power factor [p.u.]
pf^{ind}	Inductive power factor [p.u.]
E^{co2}_g	CO2 emission rate [tCO2/MWh]

Transmission Parameters:

C^{line}_{ijc}	Annualized fixed cost of a candidate line [MUSD]
C^{sy}_s	Annualized fixed cost of a candidate synchronous compensator[MUSD]
C^{sh}_r	Annualized fixed cost of a capacitor bank [MUSD]
R_{ijc}, X_{ijc}	Resistance and reactance of a line [p.u.]
Z^2_{ijc}, B^l_{ijc}	Squared series impedance and shunt susceptance of a line [p.u.]
B^{sh}_r	Shunt susceptance [p.u.]
$\overline{S}_{ijc}$	Maximum apparent power flow [MVA]
$\overline{\theta}_i$, $\underline{\theta}_i$	Maximum/minimum voltage angle at node i [rad]
$\overline{\theta}_{ijc}$, $\underline{\theta}_{ijc}$	Maximum/minimum voltage angle difference [rad]
$\overline{V}_{ni}, \underline{V}_{ni}$	Maximum/minimum voltage [p.u.]

Continuous Variables:

l^{shed}_{ni}	Load shedding [p.u.]
p_{ng}, q_{ng}	Active and reactive power generation [MW, Mvar]
p^c_{ne}	Active power consumption from a ESS unit [MW]
p^s_{ng}	Active power of the second block of a committed unit [MW]
p^{2c}_{ne}	Active power charged from a ESS unit [MW]

y_{ne}	Stored energy from ESS unit [MWh]
s_{ne}	Spilled energy from ESS unit [MWh]
v_{ni}^2	Square of voltage magnitude [p.u.^{2}]
i_{nijc}^2	Square of current magnitude [A^2]
θ_{ni}	Voltage angle [rad]
f_{nijc}^P	Active power flow [MW]
f_{nijc}^Q	Reactive power flow [Mvar]
q_{nijc}^{shl}	Reactive power injection by an existing/candidate line [Mvar]
q_{nijc}^{shc}	Reactive power injection by an existing/candidate capacitor bank [Mvar]
f_{nijc}^v	Auxiliary variable of voltage drop [p.u.]
f_{nijc}^θ	Auxiliary variable of voltage angle difference [rad]
θ_{nijc}^k	Auxiliary variable which represents the voltage angle difference for the branch between nodes i and j per load level n (i.e., time step or hour) and candidate line ijc considered in constraints for cycle k [rad]

Binary Variables:

i_g^g	Candidate generation unit installed or not
i_e^e	Candidate ESS unit installed or not
i_{ijc}^t	Candidate line installed or not
i_s^q	Candidate synchronous compensator installed or not
i_r^s	Candidate capacitor bank installed or not
uc_{ng}	Commitment of generation unit
su_{ng}, sd_{ng}	Startup/shutdown of generation unit
$i_{ij}^{t,k}$	Auxiliary variable which specifies whether any of the candidate lines for the branch $'ij'$ is built or not when stating the cycle constraints for cycle k

1. Motivation and state of the art

The decommissioning of fossil-fuel-based power plants and the uncertain availability of wind and solar generation are new challenges for power system operation and expansion. The power system requires new sources of flexibility and resilience to deal with these challenges. Thus, classical power system optimization models are being upgraded to include low-carbon emission technologies and novel electricity markets that exploit renewable energy sources. This is also the case for models used to optimize power system expansion planning, which is essential to meet the technical and environmental requirements at a minimum cost.

Traditionally, power system expansion planning has been tackled via decomposition techniques and by defining independent planning subproblems, e.g., [1, 2, 3], because with the currently available computational resources, an integrated approach i.e., addressing the joint expansion planning of energy resources and the network, cannot be used for larger systems due to its poor scalabilitty. However, decomposing the integrated planning problem also has its disadvantages. Given that current European electricity systems operate within an unbundled framework, the solutions resulting from the decomposition can be suboptimal, leading to an increase in the total system cost, as concluded in [4]. Co-optimizing the expansion of generation and transmission, on the other hand, allows network operating utilities to determine the socially optimal development of the system by implementing perfect coordination between both types of investments.

Motivated by the opportunities that coordinated generation and transmission expansion planning can offer to network operators and utilities, many recent studies, e.g. [5, 6], investigate integrated planning approaches. In [5], the authors explore the integrated expansion planning (IEP) of generation, storage and transmission, concluding that the deployment of energy storage systems (ESS) defers the installation of new thermal generation units, minimizes RES curtailment, and reduces transmission congestion. However, to overcome the complexity introduced by the IEP problem, studies employ simplifications that may affect the optimal investment plan. For example, no previous studies addressing the IEP problem include reactive power planning (RPP); see [6, 7]. Although there are works that consider AC-OPF [8], the RPP is not taken into account even though it aims to keep voltage levels within limits and maintain supply capacity and inductive power load in transmission lines. Neglecting the provision of local reactive power support in generation expansion planning can lead to increased investment in generation and, in general, to suboptimal decisions regarding the location of generation assets, as observed in [9]. For example, the authors in [10] propose an optimization model for simultaneous TEP (not including generation and storage expansion planning) and RPP, and conclude that adopting this joint approach can reduce investment costs by up to 18% compared to the case where RPP is not considered. Thus, integrating RPP into the IEP problem could lead to more efficient investment decisions.

Another simplification is the omission of UC constraints, as in the models presented in [4, 5, 11]. Ignoring these constraints significantly affects the efficiency of generation investment decisions, leading to higher operation

costs, higher emissions, and missing climate policy targets, as shown in [12]. Another common practice in power system expansion planning concerns the representation of the power flows. For example, the authors in [5] consider the transportation model. While other studies consider a more accurate network model, such as the DC power flow approximation model in [11]. The DC model is widely used in many works on expansion planning in power systems due to the fact that including the set of constraints related to it does not significantly increase the problem size (number of rows and columns of the optimization problem). However, the expansion plans calculated in works considering a DC model usually need to be adjusted, adding more investments to them, to make the AC operation resulting from their implementation feasible, as shown in [10]. To avoid this, several advanced AC-OPF models have been considered within the expansion planning problem in the literature, as shown in [4, 10, 9, 13, 14, 15]. These advanced models can be categorized according to their complexity into linear programming models like the one proposed in [10], second-order conic programming (SOCP) models [4, 9, 13, 14], and those using semidefinite relaxation (SDR) [15]. Considering SOCP or SDR power flow models, such as the ones within the proposed formulations of the IEP in [8, 16], and 8736 consecutive loadlevels or time steps would result in too large problems (i.e., greater than 20 million rows and columns without counting the binary variables and the problem complexity, either MILP, MISOCP or MISDP) when applied to medium-to-large scale systems due to their large computational burden. Therefore, making use of a linear, compact, AC model could strike the right balance between computing efficient expansion plans resulting in an operation of the system that meets the AC power flow constraints and formulating an IEP problem of a small size for medium-to-large scale systems. Then, there is room for applying appropriate techniques to make the formulation of the linear AC load flow model more compact.

Apart from this, keeping the resulting IEP problem size within reasonable limits has led most authors to apply time-domain reduction techniques to represent system operation. For example, in [6], only one week is considered to represent the operation of the system over each month (84 days/year), while in [5], only 96 hours over a year are considered. These approaches focus on representing short-term operational uncertainties. However, due to the fact that large shares of renewable energy source (RES) based generation are to be expected in future power systems, and the need to deploy relevant amounts of ESS capacity to integrate the output of this generation, the level

of time granularity of the representation made of the system operation may significantly affect the investment decisions computed when solving the IEP problem and the efficiency of the corresponding plan. In fact, according to [17], the level of detail considered when representing the system operation in the temporal and techno-economic domains should have a similar impact on the set of energy system investment decisions made and its efficiency. If the system operation needs to be represented with a high level of detail in IEP problems, in terms of the set of constraints on this operation and its temporal granularity, the development of scalable, flexible, and tractable enough formulations of this problem may require the implementation of advanced search space reduction techniques. Thus, the authors in [18] propose a methodology to meaningfully select the set of candidate transmission lines to consider in TEP problems (thus reducing the number of binary variables in the problem), and the authors in [14, 19] consider cycle constraints to reduce the search space by reformulating the branch flow-based AC-OPF and DC-OPF, respectively. The authors in [14] propose a branch flow-based SOCP-OPF that outperforms the NLP-OPF and SDP-OPF models. Including this formulation in the classical MILP problem (i.e., the expansion planning and unit commitment problems) promises an improvement in its performance. On the other hand, in [19], the authors extend the formulation in [20] to the TEP problem, where they identify two sets of cycles: those that are in the cycle basis and those that are candidate cycles because they comprise at least one branch where only candidate lines are defined. Moreover, the authors in [13] proposed the application of highly effective bound optimization and tightening strategies to strengthen the AC-OPF in network optimization problems, increasing its accuracy while speeding up its computation. Specifically, Coffrin et al. use the minimal network and bounded consistency algorithm to optimize and tighten the bounds on the voltage magnitude and angle variables. It can be concluded that the application of appropriate search space reduction techniques within the IEP problem for large systems seems to be a step in the right direction.

1.1. Contributions

This paper extends and linearizes the cycle-based formulation of the branch flow-based SOCP OPF in [14] so that it can be employed within the integrated, expansion planning problem of medium to large electric systems, considering the deployment of new resources of all major types, including reactive power compensation devices. The formulation of the cycle constraints

in the expansion planning problem that we propose here is based on the formulation proposed in [19]. However, the authors in [19] employed DC OPF, whereas we have adapted their formulation to be employed within an OPF based on an AC branch flow model. In addition, our tight AC-OPF based formulation is comprehensive considering the deployment of a variety of generation and storage resources as well as reactive power compensation devices. Our formulation comprises a detailed representation of the operation of the system, including both all the relevant operation constraints and a temporal granularity that is high enough to be able to represent the economic operation of all the main energy resources, including short-, medium- and long-term storage. In order to make the resulting problem tractable, efficient size reduction techniques are applied, namely the consideration of the cycle-based AC-OPF model and a procedure to obtain tight bounds for the voltage magnitude and angle variables. Adopting a static planning approach is necessary to keep the size of the resulting problem, which considers hourly resolution, within sizable limits for power systems of medium to large size.

The specific contributions of the work discussed here are highlighted next:

- This paper provides a novel, linear, branch flow-based, AC-OPF model using cycle constraints, which have been adapted to be considered within the integrated expansion planning problem.
- The typically employed formulations of the bound-tightening constraints on the voltage magnitude and angle variables are modified to be included in a linearized AC-OPF within the expansion planning problem. To the best of our knowledge, these constraints have not been included in this problem formulation in the literature.
- Using a representative case study, we demonstrate the benefits of considering the bound-tightening and cycle constraints in the formulation of the AC-OPF within the expansion planning problem. The reductions achieved in the size of the problem and its computational burden are highly significant as they considerably improve the solution of the expansion planning problem.
- Lastly, the applicability and efficiency of the proposed formulation are proven by applying it to compute the optimal integrated expansion plan of a realistic power system of medium size based on the deployment of RES-based generation and comparing the performance of this formulation with that of other relevant formulations also applicable to this type of problem.

1.2. Paper organization

The remainder of the paper is organized as follows: Section 2 provides the comprehensive formulation proposed for the IEP problem. In section 3, case studies are defined featuring several scenarios and finally, in Section 4, the case studies results are discussed to analyze the performance of this formulation.

2. Model Formulation

In this section, the mathematical formulation of the proposed cycle constraints within IEP problem and the application of the bound-tightening procedure are discussed. Solving this problem provides an optimal investment plan considering the current and future needs of the power system. The solution computed also includes the optimal operation decisions in each scenario, which can be used in short-term analyses since the model developed considers UC, and hourly operation constraints. Note that most test cases in the literature consider representative hours, days, or weeks, but our test case was defined considering 8736 consecutive hours of system operation within the target year in order to represent system operation accurately enough. Another capability of our formulation is to consider separately each hour (representative time periods), which demonstrates the suitability of the proposed formulation to be applied to large test systems.

Traditionally, these kinds of large-size problems have been addressed by simplifying assumptions or taking representative time periods, which decreases the computational burden of the problem at the expense of accuracy. Here, we provide a formulation that captures as many details of the system's operation as possible. In addition, the investment decisions made are selected within a rich and mixed portfolio of plausible investments comprising generation units, ESSs, transmission lines, synchronous compensators, and capacitor banks. The system operation computed considers cycle-based AC-OPF, UC, storage management, and operating reserve constraints. Expansion and operation results produced include investments in a wide range of system assets, the main hourly system operation variables, the associated costs, and additional results like CO2 emissions, RES energy spillage, the marginal operating costs, or the water value.

The proposed mathematical formulation is divided into three parts: a) the IEP objective function and the impact of the investment decisions made on the system operation; b) the operational constraints concerning the storage

and the generation units, including their UC; and c) the formulation of the cycle-based AC-OPF, based on [10, 14]. The first two parts can be found in Appendix A, and the formulation of the linear cycle-based AC-OPF to be used within the IEP problem is provided next:

2.1. AC power flow: Branch flow model

In this work, the AC-OPF is based on the formulation of the branch flow model presented in [10] and the formulation that uses cycle constraints presented in [14]. The formulation of the cycle constraints was extended t o be embedded within the proposed expansion planning formulation.

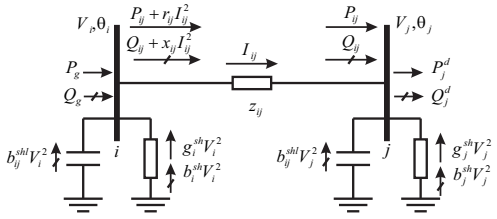

Figure 1: Branch Flow Model.

2.1.1. Active and reactive power balance

The active and reactive power balance constraints are given by (1a) and (1b), respectively. Moreover, when enforcing the reactive power balance, the production, and consumption of reactive power by the generation units, synchronous compensators, transmission lines, and as well as production by capacitor banks are considered.

$$\sum_{g\in\mathcal{G}_i} p_{ng}+\sum_{e\in\mathcal{E}_i} p_{ne}=\sum_{e\in\mathcal{G}_i} p^c_{ne}+P^d_{ni}\left(1-l^{shed}_{ni}\right)+ \sum_{ijc\in\mathcal{L}} f^P_{nijc}-\sum_{jic\in\mathcal{L}} f^P_{njic}-R_{jic}i^2_{njic} : \forall ni, \tag{1a}$$

$$\sum_{g\in\mathcal{G}_i} q_{ng}+\sum_{s\in\mathcal{S}_i} q_{ns}+\sum_{r\in\mathcal{R}_i} q^{shc}_{nr}=Q^d_{ni}\left(1-l^{shed}_{ni}\right) + \sum_{ijc\in\mathcal{L}} f^Q_{nijc}-q^{shl}_{nijc}-\sum_{jic\in\mathcal{L}} f^Q_{njic}+q^{shl}_{njic}-X_{jic}i^2_{njic} : \forall ni, \tag{1b}$$

Note that it is the squared current flow (i^2_{nijc}) that is considered a variable in our formulation instead of the current flow. This is also the case for the square of the voltage magnitude (v^2_{nijc}), which is considered in the following constraints.

2.1.2. Voltage magnitude drop

Equations (2a)-(2c) represent the voltage magnitude drop between nodes, for each branch. In this case, that for existing lines is determined by (2a), while that for candidate lines is determined by (2b)-(2c). Note that these last constraints are disjunctive inequalities to deal with the product of the variables *current flow* and *power flow* of a line with its respective investment (binary) variable in the original formulation.

$$v_{ni}^2 - v_{nj}^2 = Z_{ijc}^2 i_{nijc}^2 + 2\left(R_{ijc} f_{nijc}^P + X_{ijc} f_{nijc}^Q\right) \quad : \forall nijc | ijc \in \mathcal{L}^e, \ (2a)$$

$$v_{ni}^2 - v_{nj}^2 - f_{nijc}^v = Z_{ijc}^2 i_{nijc}^2 + 2\left(R_{ijc} f_{nijc}^P + X_{ijc} f_{nijc}^Q\right) \quad : \forall nijc | ijc \in \mathcal{L}^c, \ (2b)$$

$$\left| f_{nijc}^v \right| \leq \left(\overline{V}_i^2 - \underline{V}_i^2\right)(1 - i_{ijc}^t) \quad : \forall nijc | ijc \in \mathcal{L}^c, \ (2c)$$

2.1.3. Voltage angle differences

Equations (3a)-(3c) represent the voltage angle differences along branches. Those for existing lines are determined in (3a), while those for candidate lines are determined in (3b)-(3c). In the same way as for the constraints applied to candidate lines in Section 2.1.2, here, disjunctive inequalities are also defined. The non-linear equations (3a)-(3b) can be linearized by assuming that $v_{ni}^2 = 1.0$ and $\sin(\theta_{ni} - \theta_{nj}) = (\theta_{ni} - \theta_{nj})$, as done in the initial step in [10].

$$\sqrt{v_i^2}\sqrt{v_j^2}\sin(\theta_i - \theta_j) = X_{ijc} f_{ijc}^P - R_{ijc} f_{ijc}^Q \quad : \forall nijc | ijc \in \mathcal{L}^e, \ (3a)$$

$$\sqrt{v_i^2}\sqrt{v_j^2}\sin(\theta_i - \theta_j) = f_{nijc}^\theta + X_{ijc} f_{ijc}^P - R_{ijc} f_{ijc}^Q \quad : \forall nijc | ijc \in \mathcal{L}^c, \ (3b)$$

$$\left| f_{nijc}^\theta \right| \leq 2\overline{V}_i \overline{V}_j \overline{\theta}_i (1 - i_{ijc}^t) \quad : \forall nijc | ijc \in \mathcal{L}^c, \ (3c)$$

2.1.4. Current flow constraint

The non-linear and non-convex constraint (4) is used to calculate the current flow, based on the concurrent power flows and voltage magnitudes. There are several ways to avoid considering a non-linear constraint. Still, the main ones are a) the use of relaxation techniques to convert it into a rotated second-order conic constraint to bound the square of the active and reactive power flow; and b) using a piecewise linearization process of the square terms. In the proposed model, this last one presented in [10] is considered for (4).

$$(f_{nijc}^P)^2 + (f_{nijc}^Q)^2 = v_{ni}^2 i_{ijc}^2 \quad : \forall nijc, \ (4)$$

2.1.5. Reactive power injection by transmission lines

The injection by existing lines is determined according to (5a), and that by candidates according to the inequalities in (5b)-(5c).

$$q^{shl}_{nijc} = v^2_{ni} B^l_{ijc} \qquad : \forall nijc | ijc \in \mathcal{L}^e, \ (5a)$$

$$-\overline{V}^2_{ni}\left(1-i^t_{ijc}\right) \leq q^{shl}_{nijc} - v^2_{ni} B^l_{ijc} \leq \underline{V}^2_{ni}\left(1-i^t_{ijc}\right) \qquad : \forall nijc | ijc \in \mathcal{L}^c, \ (5b)$$

$$\underline{V}^2_{ni} B^l_{ijc} i^t_{ijc} \leq q^{shl}_{nijc} \leq \overline{V}^2_{ni} B^l_{ijc} i^t_{ijc} \qquad : \forall nijc | ijc \in \mathcal{L}^c, \ (5c)$$

2.1.6. Reactive power injection by capacitor banks

Existing capacitor banks are governed by (6a), and the candidates by the inequalities in (6b)-(6c).

$$q^{shc}_{nr} = v^2_{ni} B^{sh}_r \qquad : \forall nr | r \in \mathcal{R}^e, \ (6a)$$

$$-\overline{V}^2_{ni}(1-i^s_r) \leq q^{shc}_{nr} - v^2_{ni} B^{sh}_r \leq \underline{V}^2_{ni}\,(1-i^s_r) \qquad : \forall nr | r \in \mathcal{R}^c, \ (6b)$$

$$\underline{V}^2_{ni} B^{sh}_r i^s_r \leq q^{shc}_{nr} \leq \overline{V}^2_{ni} B^{sh}_r i^s_r \qquad : \forall nr | r \in \mathcal{R}^c, \ (6c)$$

Where $\mathcal{R}^e$ and $\mathcal{R}^c$ are the subsets of existing and candidate capacitor banks, respectively.

2.1.7. Current flow bounds on existing transmission lines

Bounds to the current flow magnitudes for existing transmission lines are imposed by (7).

$$0 \leq \frac{i^2_{nijc}}{\left(\overline{S}_{ijc}/\overline{V}_i\right)^2} \leq 1 \qquad : \forall nijc | ijc \in \mathcal{L}^e, \ (7)$$

2.2. Cycle constraints

In the branch flow-based AC-OPF model, Kirchhoff's voltage law is represented by the voltage magnitude drop (2), voltage angle difference (3), and current flow (4) equations. Taking as a reference the cycle constraints provided in [14], we propose only replacing voltage angle difference constraints

(3a)-(3c) with cycle constraints to speed up and strengthen the AC-OPF formulation within the expansion planning problem. Then, the formulation of the linear branch flow-based AC-OPF model we employ changes with respect to that proposed in previous works due to the fact that we have considered cycle flow constraints and have reformulated equation (3) making use of equation (8) and the new set of equations (9), which are linear and efficiently formulated, as argued in [14]. Note that, in our work, the only network elements deemed to be part of the cycles are the transmission lines, since the active and/or reactive power compensation devices connected in series with the transmission lines are not represented here.

The cycles considered are defined and classified according to the approach proposed in [19]. Thus, we also define two cycle sets: the existing cycles, which are computed considering the existing transmission network, and the candidate cycles, which comprise at least one branch for which only candidate lines are considered. The sets $\mathcal{K}^e$ and $\mathcal{K}^c$ correspond to the sets of existing and candidate cycles, respectively. The set $\mathcal{K}^e$ corresponds to the cycle basis computed considering the existing network. The set $\mathcal{K}^c$ is the result of taking out of the cycle basis computed considering the complete network (comprising the existing and the candidate lines) all those cycles within set $\mathcal{K}^e$. Then, $\mathcal{K}^c$ is the difference between the cycle basis computed considering all the existing and candidate lines and the set $\mathcal{K}^e$. All the cycles considered that only comprise existing lines are within $\mathcal{K}^e$, and all those considered that at least comprise one candidate line are within $\mathcal{K}^c$. Finally, the voltage angle differences along the cycles $k \in \mathcal{K}$, where $\mathcal{K} = \mathcal{K}^e \cup \mathcal{K}^c$, are expressed as a directed linear combination of those for branches $'ij'$, using the incidence matrix $H_{ijc,k}$ in (8a). Note that, according to the proposed formulation, the KVL is not enforced for those initially defined cycles that include at least one branch for which only candidate lines are considered, if none of these lines is eventually built.

$$H_{ijc,k} = \begin{cases} 1 & \text{if} \quad \text{line } ijc \in k \\ -1 & \text{if} \quad \text{line } jic \in k \\ 0 & \text{if} \quad \text{otherwise} \end{cases} \tag{8a}$$

For those cycles in $\mathcal{K}^e$, the voltage angle differences along them can be represented by the corresponding sum of voltage angle differences along the lines in (8b).

$$\sum_{ijc \in k | k \in \mathcal{K}^e} H_{ijc,k} \theta^k_{nijc} = 0 \qquad \forall n \tag{8b}$$

For every cycle $k \in \mathcal{K}^e$, the equations (8b)-(8e) are defined as proposed in [14]. Where it is considered that the sum of voltage angle differences across all the lines in every cycle k must add up to zero, as imposed by (8b), and (8c) links the auxiliary variable $\theta^k_{nijc}, \forall nijc | ijc \in \mathcal{L}^e$ to its respective voltage angle difference.

$$\theta^k_{nijc} = \theta_{ni} - \theta_{nj} \qquad \forall nijc | ijc \in \mathcal{L}^e, \tag{8c}$$

Finally, (3a) is relaxed into (8d) and (8e).

$$\theta^k_{nijc} \geq \frac{X_{ijc} f^P_{nijc} - R_{ijc} f^Q_{nijc}}{\overline{V}_{ni}\overline{V}_{nj}\cos\frac{\Theta^m_{nijc}}{2}} - \tan\frac{\Theta^m_{nijc}}{2} + \frac{\Theta^m_{nijc}}{2} \qquad \forall nijc|ijc \in \mathcal{L}^e, \text{ (8d)}$$

$$\theta^k_{nijc} \leq \frac{X_{ijc} f^P_{nijc} - R_{ijc} f^Q_{nijc}}{\underline{V}_{ni}\underline{V}_{nj}\cos\frac{\Theta^m_{nijc}}{2}} + \tan\frac{\Theta^m_{nijc}}{2} - \frac{\Theta^m_{nijc}}{2} \qquad \forall nijc|ijc \in \mathcal{L}^e, \text{ (8e)}$$

Equations (8c)-(8e), as defined in [14], do not include cycles with candidate lines. Note that Θ^m_{nijc} is a parameter and is calculated as $\Theta^m_{nijc} = \max[|\underline{\theta}|, |\overline{\theta}|]$, which makes the equations (8d)-(8e) linear.

In our work, we propose the new set of equations (9) for those cycles that include branches only featuring candidate lines, i.e., for candidate cycles. This set of equations is enforced for a certain candidate cycle if, and only if, there is at least one line per branch $'ij'$ within this cycle that is built or already exists. In other words, a candidate cycle is closed, and the cycle constraints are enforced for it, when there is at least one candidate line per branch in the cycle only featuring this type of lines whose investment variable is equal to 1, as stated in equations (9a)-(9c).

For this, we use the auxiliary variables ($\theta^k_{nijc}, \forall nijc|ijc \in \mathcal{L}^c$ and $i^{t,k}_{ij}, \forall ij|ij \in \mathcal{L}^c$) and the set $\mathcal{L}^p$ comprising those candidates lines that belong to the same branch and are in parallel. Note that constraint (8b), which is defined for existing cycles ($k \in \mathcal{K}^e$), is extended to candidate cycles through the constraint (9a) considering the previously defined auxiliary variables θ^k_{nijc}. Within the latter constraint, the sum of the voltage angle differences along the branches is set to zero when the cycle (k in $\mathcal{K}^c$) is closed, i.e., when at least one line is built for each branch without an existing line within the cycle. This sum is zero if all complementary auxiliary investment variables $i^{t,k}_{ij}$ defined for the branches within the cycle are zero. Otherwise, the sum of the voltage angle differences along the branches in the cycle is set to free.

$$\Big|\sum_{ijc\in k|k\in\mathcal{K}^c} H_{ijc,k}\theta^k_{nijc}\Big| \leq \sum_{ijc\in k, ijc\in\mathcal{L}^c|k\in\mathcal{K}^c} i^{t,k}_{ij}\overline{\theta}_{nijc} \qquad \forall n, \text{ (9a)}$$

In the same way, (8c) is extended to (9b) and (9c) by considering the auxiliary variable θ^k_{nijc} equal to its respective voltage angle difference when a candidate line is built; otherwise, that auxiliary variable is equal to zero in order not to be considered in the sum of the left side in (9a).

$$-(1-i^t_{ijc})\overline{\theta}_{nijc} \leq \theta^k_{nijc} - (\theta_{ni}-\theta_{nj}) \leq (1-i^t_{ijc})\overline{\theta}_{nijc} \qquad \forall nijc, \text{ (9b)}$$

$$-i^t_{ijc}\underline{\theta}_{nijc} \leq \theta^k_{nijc} \leq i^t_{ijc}\overline{\theta}_{nijc} \qquad \forall nijc, \text{ (9c)}$$

Note that the value of the auxiliary variable $i^{t,k}_{ij}$, related to the investment for branch $'ij'$, is determined by constraint (9d) for those branches $'ij'$ for which there is only one candidate line, while this is determined by constraints (9e)-(9f) for those branches $'ij'$ for which there are multiple candidate lines in parallel. This auxiliary variable for branch $'ij'$ takes a value of 0 if at least one candidate line is built for branch $'ij'$. Otherwise, it takes a value of 1.

$$i^{t,k}_{ij} = 1 - i^t_{ijc} \qquad \forall nijc | ijc \in \mathcal{L}^c, ijc \notin \mathcal{L}^p, \text{ (9d)}$$

$$0 \leq i^{t,k}_{ij} \leq 1 - i^t_{ijc} \qquad \forall nijc | ijc \in \mathcal{L}^p, \text{ (9e)}$$

$$1 - \sum_{c \in \mathcal{C}} i^t_{ijc} \leq i^{t,k}_{ij} \qquad \forall ij | ij \in \mathcal{L}^p, \text{ (9f)}$$

Finally, as it was defined for cycles comprised of only existing lines, the envelopes in (9g)-(9h) are used to define the bounds for the auxiliary variable θ^k_{nijc} for candidate lines. Note that Θ^m_{nijc} is a parameter that makes the equations linear. And, it is computed as follows $\Theta^m_{nijc} = \max[|\underline{\theta}|, |\overline{\theta}|]$.

$$\theta^k_{nijc} \geq \frac{X_{ijc}f^P_{nijc} - R_{ijc}f^Q_{nijc}}{\overline{V}_{ni}\overline{V}_{nj}\cos\frac{\Theta^m_{nijc}}{2}} - \tan\frac{\Theta^m_{nijc}}{2} + \frac{\Theta^m_{nijc}}{2} \qquad \forall nijc | ijc \in \mathcal{L}^c, \text{ (9g)}$$

$$\theta^k_{nijc} \leq \frac{X_{ijc}f^P_{nijc} - R_{ijc}f^Q_{nijc}}{\underline{V}_{ni}\underline{V}_{nj}\cos\frac{\Theta^m_{nijc}}{2}} + \tan\frac{\Theta^m_{nijc}}{2} - \frac{\Theta^m_{nijc}}{2} \qquad \forall nijc | ijc \in \mathcal{L}^c, \text{ (9h)}$$

An illustrative 3-bus case example where either one single existing or one single candidate cycle is defined is shown in Fig. 2. There, v^2 represents the squared voltage magnitude, θ represents the voltage angle, k_1 corresponds to the cycle defined, and c represents a specific line. This figure shows the 3-bus system in three different cases: a) when there are only existing lines within the system and there is just one existing cycle: 1-2-3-1; b) when there is only one candidate line connecting nodes 1 and 3 and the only cycle defined is candidate cycle is 1-2-3-1; and c) when there are just two candidate lines in parallel connecting nodes 1 and 3, and the only cycle defined is candidate cycle 1-2-3-1.

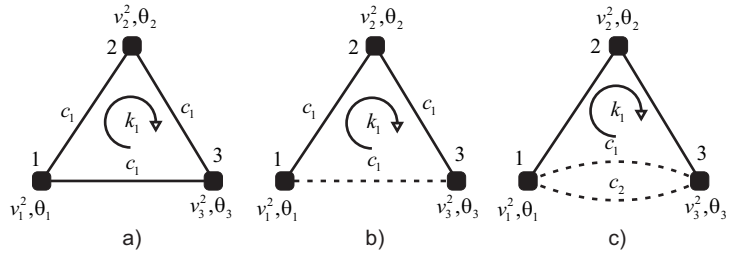

Figure 2: a) 3-bus system where there are only existing lines; b) 3-bus system where one branch within the cycle defined comprises one candidate line; and c) 3-bus system where one branch within the single cycle defined comprises two candidate lines in parallel.

For case a), as shown in Fig. 1a, equation (8b) would read $\theta^k_{12a} + \theta^k_{23a} - \theta^k_{13a} = 0$ if we just consider one hour (time step). Besides, equation set (8c) is a three-equation set that reads as follows: $\theta^k_{12a} = \theta_1 - \theta_2$, $\theta^k_{23a} = \theta_2 - \theta_3$, and $\theta^k_{13a} = \theta_1 - \theta_3$.

For case b), as shown in Fig. 1b, given that there is a branch comprising just one candidate line, equation (9a) applies, and it reads $-i^{t,k}_{13}\overline{\theta}_{13a} \leq \theta^k_{12a} + \theta^k_{23a} - \theta^k_{13a} \leq i^{t,k}_{13}\overline{\theta}_{13a}$. Equation set (8c) is a two-equation set reading: $\theta^k_{12a} = \theta_1 - \theta_2$, and $\theta^k_{23a} = \theta_2 - \theta_3$. For the candidate line for branch $1-3$, equation (9b) applies, and it reads $-(1 - i^t_{13a})\overline{\theta}_{13a} \leq \theta^k_{13a} - (\theta_1 - \theta_3) \leq (1 - i^t_{13a})\overline{\theta}_{13a}$. Together with the former, equation (9c) also applies, reading $-i^t_{13a}\underline{\theta}_{13a} \leq \theta^k_{13a} \leq i^t_{13a}\overline{\theta}_{13a}$. Besides, equation (9d) applies, reading $i^{t,k}_{13} = 1 - i^t_{13a}$.

Finally, for case c), as shown in Fig. 1c, given that there is a branch comprising two candidate lines in parallel, equation (9a) applies, reading $-i^{t,k}_{13}(\overline{\theta}_{13a} + \overline{\theta}_{13b}) \leq \theta^k_{12a} + \theta^k_{23a} - \theta^k_{13a} - \theta^k_{13b} \leq i^{t,k}_{13}(\overline{\theta}_{13a} + \overline{\theta}_{13b})$. Equation set (8c) also applies, and it is a two-equation set reading the same as that in case b). Besides, for candidate lines $'a'$ and $'b'$ in branch $1-3$, equation set (9b) applies, reading $-(1 - i^t_{13a})\overline{\theta}_{13a} \leq \theta^k_{13a} - (\theta_1 - \theta_3) \leq (1 - i^t_{13a})\overline{\theta}_{13a}$, and

$-(1-i^t_{13b})\overline{\theta}_{13b} \leq \theta^k_{13b} - (\theta_1 - \theta_3) \leq (1-i^t_{13b})\overline{\theta}_{13b}$. Also, equation set (9c) applies, reading $-i^t_{13a}\underline{\theta}_{13a} \leq \theta^k_{13a} \leq i^t_{13a}\overline{\theta}_{13a}$ and $-i^t_{13b}\underline{\theta}_{13b} \leq \theta^k_{13b} \leq i^t_{13b}\overline{\theta}_{13b}$. Since there are two parallel lines defined for branch $1-3$, equation set (9e) reads to $0 \leq i^{t,k}_{13} \leq 1 - i^t_{13a}$ and $0 \leq i^{t,k}_{13} \leq 1 - i^t_{13b}$, while equation (9f) reads $1 - i^t_{13a} - i^t_{13b} \leq i^{t,k}_{13}$.

2.3. Bound-tightening procedure

Based on the results computed for relevant case studies so far, we can conclude that a larger reduction of the problem solving time is achieved, without affecting the quality of the computed expansion planning solution, is achieved when, together with the proposed convex relaxations of the AC-OPF formulation, the also proposed tightened bounds of the variables v^2_{ni} and θ^k_{nijc} are implemented. In [13], the minimal network and bound-consistency algorithm is proposed to find tight bounds of these variables and speed up the AC-OPF problem solving process. This tightening process is achieved through propagation techniques that exploit the structure of the problem constraints systematically updating variable bounds based on the information provided by each constraint. Similar to [13], we use the algorithm in a procedure before solving the proposed expansion planning problem as follows:

1. The binary variables defined in the expansion planning problem are relaxed so that they can take any value in [0,1], and the minimal network and bound-consistency algorithm [13] per load level (i.e., time step, here made to coincide with each operation hour) $'n'$ is run.

2. The voltage magnitude bounds $(\overline{V}_{ni}, \underline{V}_{ni})$ and voltage angle bounds $(\overline{\theta}_{ni}, \underline{\theta}_{ni})$ are updated. Additionally, constraint (10) is tightened by defining the convex envelopes of the squared voltages as in [21]

$$v^2_{ni} \in \langle x \rangle \begin{cases} x \geq V^{sqr}_{ni} \\ x \leq (\overline{V}_{ni} + \underline{V}_{ni})\sqrt{V^{sqr}_{ni}} - \overline{V}_{ni}\underline{V}_{ni} \end{cases} : \forall ni, \quad (10)$$

   where V^{sqr}_{ni} is a parameter who takes the value of v^2_{ni} previously computed in Step 1. Moreover, the feasible region for the voltage angle can be tightened according to (11), where Θ^{Δ} is the voltage angle difference, defined as a parameter whose value is made to coincide with the difference of the variables θ_{ni} and θ_{nj} previously computed in Step 1.

$$-\pi/3 \leq -\Theta^{\Delta}_{nijc} \leq \theta_i - \theta_j \leq \Theta^{\Delta}_{nijc} \leq \pi/3 \qquad : \forall nijc, \quad (11)$$

3. The proposed expansion planning problem is solved, including the updated bounds for the voltage angles and magnitudes as well as (10) and (11).

3. Tests, Assumptions, and Parameters

This section presents the test system, the case studies, and the scenarios, as well as the assumptions that were used in the simulations that validated the performance of the proposed model. All the formulated optimization problems were solved using Gurobi 10.0.1 as a commercial MIP solver on a computer with a 3.40 GHz Intel Core i7-10875H processor and 32 GB of RAM. Moreover, the simulations were set up in Julia 1.8.5, where JuMP 1.7.0 was used to develop the optimization models.

The entire year of 2030 has been considered as the time horizon of the study. The system operation is computed for 8736 hours of the target year[1]. Initial bounds have been set for the voltage magnitudes at $\underline{V} = 0.95\ p.u.$ and $\overline{V} = 1.05\ p.u.$

3.1. Test System

The RTS-GMLC test system from [22] is considered to assess the performance of the proposed formulation in each case study. This test system is an updated version of the traditional IEEE Reliability Test System (RTS). It aims to represent current power systems hosting large amounts of RES-based generation but still hosts significant amounts of fossil fuel power plants. The time series considered for the output of the RES generation and the demand, which are the main variable input parameters, are also those provided in [22]. Geographically, the system depicted corresponds to a region in the southwestern part of the United States of America. The network model considered for this region comprises 73 nodes, representing substations distributed within three areas. Each area comprises two sub-areas. The voltage levels considered are 230 kV and 138 kV. The existing network comprises 104 lines (36 of which at 138 kV and 68 at 230kV) and 16 power transformers.

Note that although in our case studies we considered the RTS-GMLC system (73 nodes) and represented the system operation with an hourly resolution, our proposed formulation also supports the consideration of repre-

[1]This unconventional number of hours is considered because it is a multiple of the number of hours a week, which is also related to the storage capacity of PSH units.

sentative periods (representative hours, days, or weeks with an hourly resolution). This would allow us to implement our modeling approach in much larger test systems.

Candidate network investments include the duplication of all existing lines. The admittance, impedance, and apparent power limits of the transmission lines are those in the original RTS test case. Given the set of candidate lines considered, the number of cycles found in the power grid does not change with the set of candidate branches installed. The additional parameters considered in the case studies have been set as follows: the susceptance of the capacitor banks is set at 0.2 p.u., and their cost $\left(C_r^{sh},\ \forall r \in \mathcal{R}^c\right)$ amounts to MUSD 0.05 for each. The costs of the lines depend on the line features. In addition, six candidate synchronous compensators of 200/-150 Mvar are considered, each with an investment cost of 0.25 MUSD/Mvar. They are located in nodes 107, 122, 207, 222, 307, and 322. They were located in these nodes because, by computing a load flow, it was determined in advance that these nodes require voltage regulation. Moreover, a candidate capacitor bank of 50 Mvar is located in every node.

The total projected peak active and reactive load is 12287.8 MW and 9377.5 Mvar, while the total annual electricity demand amounts to 56.2 TWh. Note that the demand considered is 1.5 times the original one. About 46% of the demand is located in nodes at 138 kV. The cost of non-served energy has been set at 10000 USD/MWh.

The power generation mix comprises 372 units, 158 of which are already installed, while the rest are candidate assets. The already installed generation capacities per technology are shown in Table 1. The generation technologies considered include 6 thermal generation technologies and 2 RES-based ones (Solar-PV and Onshore-Wind). Storage is also significant in the system. Candidate investments are only considered for the technologies Gas-CC, Gas-CC-90CCS (90% carbon capture and sequestration), Gas-CT, Solar-CSP with Thermal Energy Storage, Solar-PV, Onshore-Wind, PSH, and BESS. Note that the technology named *Hydro* is conventional hydro-power with reservoirs. Within thermal generation, gas is most flexible in terms of its minimum up-time and down-time, while nuclear is the one whose operation is less flexible. It is used to cover the base demand. Coal and oil provide additional flexibility for the overall amount of power produced to adapt to changes in the net demand.

All the parameters representing the features of the existing generation, storage, and reactive power units have been taken from the RTS-GMLC

Table 1: Installed capacities of the generation, storage, and reactive power compensation technologies within the test system.

	Technology	Existing				Candidate			
		$\overline{P}$	$\underline{P}$	$\overline{Q}$	$\underline{Q}$	$\overline{P}$	$\underline{P}$	$\overline{Q}$	$\underline{Q}$
Thermal	Coal	2317	924	1070	-575	-	-	-	-
	Gas-CC	3550	1700	1500	-250	8400	3780	3600	-1800
	Gas-CC-90CCS	-	-	-	-	12000	5400	4800	-2400
	Gas-CT	1485	594	554	-390	6000	2700	2400	-1200
	Nuclear	400	396	200	-50	-	-	-	-
	Oil	324	131	162	-	-	-	-	-
RES	Solar-PV	2716	-	-	-	4050	-	-	-
	Onshore-Wind	2508	-	-	-	3800	-	-	-
	Hydro	1000	-	320	-200	-	-	-	-
	PSH	-	-	-	-	5900	-	1888	-1180
	Solar-CSP	200	30	-	-	4600	450	-	-
	BESS	50	-	-	-	4380	-	-	-
Reactive Power	Synch. compensator	-	-	600	-150	-	-	1200	-300
	Capacitor bank	-	-	-	-	-	-	3650	-
	Filter	-	-	-	-300	-	-	-	-

system. The features of the candidate units are the same as those of the existing units. The technical characteristics of all the technologies but Gas-CC-90CCS and PSH have been drawn from the technical guide in [23]. Their investment cost has also been drawn from [23], whereas the fuel cost has been drawn from [24]. All the costs are expressed in 2020 USD, using the Consumer Price Index from the U.S. Bureau of Labor Statistics, 2021 [23]. The inflows of the hydro units have been drawn from [22], and those of the PSH are equivalent to those of the hydro. The candidate PSH units are equivalent to the existing hydro units in number and parameters, as well as their location, but with the difference that PSH units can pump water. The output profiles of solar and wind generation, which depend on their geographical location, have been generated using the *System Advisor Model* provided by the National Renewable Energy Laboratory (NREL) [25]. The storage capacity and round-trip efficiency of all the BESS are deemed to be 4 hours and 0.85, respectively, while those of i) Solar-CSP and ii) PSH, and storage hydro in general, are deemed to be 24 hours and 0.8, and 1 week and 0.75, respectively. The deployment of candidate generation & storage units is computed separately for each node, as Fig. 3 shows. Many candidate solar PV units are located in Area 2. Few candidate units are located in Area 1, where there is a prominent presence of natural gas-based technologies since this is an area with few primary renewable resources available. Most of these candidate units are located in the same location as the existing ones. The

complete data set for this system can be found in the repository[2] of the openTEPES model in [26].

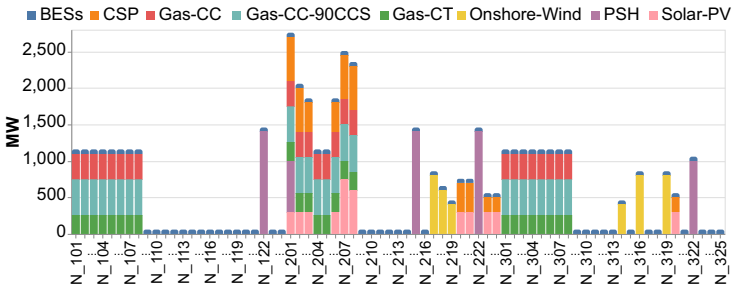

Figure 3: Capacity of candidate generation & storage per technology and node.

3.2. Case studies

In order to validate the proposed model, four cases are defined that correspond to four different configurations of the overall expansion and operation problem, shown in Table 2, considering several scenarios with specific fuel and carbon prices. For each of these scenarios, the optimal development of the test system is computed up to the year 2030. The proposed model M-A is compared with three other expansion planning formulations i.e., M-B, M-C, and M-D, which are also MILP formulations. The three alternatives are the formulations that are typically and most frequently used to solve the IEP problem considering medium and large systems with a highly detailed representation of the system operation. The main features and representative set of equations of each formulation are listed in Table 2. Model M-A includes Unit Commitment (UC) constraints and represents the AC-OPF using a cycle-based Branch Flow Model (BFM). Model M-B is similar to M-A, but the branch flow model it features is equivalent to that in the MILP formulation presented in [10]. M-C and M-D are inspired by classical formulations found in the literature and are widely used to solve power system expansion planning problems. M-C is based on M-B, but it does not consider UC constraints. In model M-D, the DC-OPF and UC constraints are considered, which modifies the objective function ((12) and (13)). Then, the model is set up with the following equations: (12) s.t. (15a), (15c), (16a)–(16c), (13), (17), (21), (22), (23)–(27), (30b), (30c), and the DC-OPF's equations are

[2]Repository

taken from the openTEPES model [26].

$$\min C^{ge}+C_n^{te}+C^{se}+\sum_{n\in\mathcal{N}} C_n^{gen}+C_n^{con}+C_n^{co2}+C_n^{ens} \tag{12}$$

$$C_n^{ens}=\sum_{i\in\mathcal{B}} D_n C^{ens,P} l_{ni}^{shed} P_{ni}^{d} \qquad :\forall n, \tag{13}$$

Table 2: Features of the proposed model (M-A) and alternatives.

Model	Description	Equations
M-A	GEP+TEP+SEP+RPP+AC-OPF+UC+Cycle constraints	(14)-(20), (21)-(30), (1)-(7), (9)
M-B	GEP+TEP+SEP+RPP+AC-OPF+UC	(14)-(20), (21)-(30), (1)-(7)
M-C	GEP+TEP+SEP+RPP+AC-OPF	(14)-(20), (21), (23)-(30), (1)-(7)
M-D	GEP+TEP+SEP+DC-OPF+UC	(14)-(15c), (16a)-(18), (21)-(27), (30b)

3.2.1. Definition of the performance metrics

The proposed model formulation is compared to others previously proposed in the literature according to two main performance metrics: i) the efficiency, or quality, of the planning solution by each formulation, and ii) the number of computational resources, or the CPU time, required by a given computer to solve the expansion planning problem. Given a test system, the efficiency of an expansion plan is primarily assessed in terms of the total expansion and operating costs. However, given the scarcity of the resources to be devoted to the system expansion, one can also assess the efficiency of an expansion plan in per unit terms, corresponding to the reduction in the total system cost per unit of investment cost resulting from the implementation of this plan instead of the plan calculated using the reference formulation. Furthermore, given that in some systems or regions, the minimization of RES energy spillage has been defined as an expansion planning or system operation objective, one could also assess the efficiency of a plan in terms of the RES energy spillage resulting from its deployment. This, if implemented, would be a complementary measure of the efficiency of the plan to be considered in addition to the main ones, which should relate to the system costs corresponding to the situation in which this plan is deployed.

3.2.2. Planning methodology adopted: a 2-stage approach for the computation of the system costs resulting from an expansion solution

This section describes the method used to calculate the total system cost resulting from implementing each expansion planning solution, that is computed with an M-X formulation. Since the system representation differs in each formulation, this method allows for a fair comparison of their system operation cost by providing a common reference for the cost calculation. This means that different expansion planning formulations follow different approaches to the calculation of the operating costs, potentially resulting in significantly different estimates of the operating costs resulting from a given system expansion solution. Therefore, in order to allow a fair comparison, a 2-stage approach to calculating the total costs resulting from implementing an expansion planning solution has been developed.

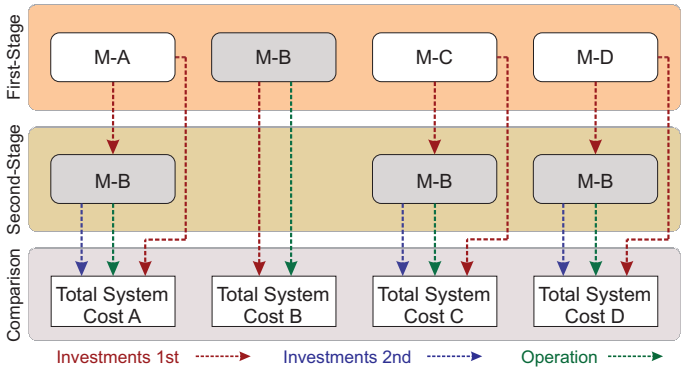

Figure 4: The proposed 2-stage comparative analysis approach.

This approach is shown in Fig. 4, wherein in the first stage, the optimal system expansion solution is computed considering each formulation M-X and scenario. Then, in the second stage, considering the representation of the system operation made in the full-fledged reference formulation, M-B, we compute the total system cost, comprising the investment and the operation ones, resulting from the implementation of the expansion solution produced in the first stage with each formulation, for each scenario. In addition to the investments included in the expansion solution calculated in the first step, this second stage also considers the implementation of additional investments that take up to 2-3 years. The additional investments computed in stage 2 are aimed at avoiding any infeasibility in the system operation, given that the operation resulting from the implementation of the expansion plans computed in stage 1 with some of the formulations compared may be infeasible,

or leading to significant levels of non-served energy, in some of the operation hours. Besides, these additional investments could increase the economic efficiency of the system development.

The types of investments considered in stage 2 of the expansion planning process are limited to certain technologies in order to make this planning process as realistic as possible. The planning of the expansion computed in stage 1 considering each formulation is deemed to take place in the long-term, when there is still time to deploy investments for any technology before the target year for which the impact of investments on the system operation is computed. However, in line with what is common practice in expansion planning analyses, the second stage in this planning process, when the effect of the expansion plans on the system operation is analyzed with a high level of detail, is deemed to take place only 3-4 years before the time when the investments planned should enter into operation (3-4 years before the year for which the impact of these investments on the system operation is considered). By the time stage 2 analyses are carried out, only investments involving certain technologies could be identified and deployed before the start of the representative period of time considered to compute the impact of each plan on the system operation. The generation investment options considered in the second stage (see Fig. 4) are those based on gas, BESS, wind, and solar PV. In addition, grid investments such as transmission lines, capacitor banks, and synchronous compensators are also considered possible in the second stage. These technologies have an average construction/installation time of $\sim$3-4 years [23], plus $\sim$1-2 years for any delays, making them promising reinforcements to consider in the medium term.

Note that the expansion planning problem is solved in each of the two stages (in the second stage, the type of investments to be made is largely restricted) for all the formulations except for formulation M-B. This is because the expansion plan, operating costs, and total system costs computed in the first stage for formulation M-B are directly considered the final ones to be compared with those computed in the second stage for the rest of the formulations.

3.3. Scenario characterization

Within the case studies, several scenarios have been defined to represent different possible future paths for fuel (in this case, gas) and carbon prices due to the relevance of both parameters' values in reaching a low-carbon

emission future. Both prices directly affect the deployment of RES and storage technologies, which may be far more attractive in certain scenarios than in others. The level of fuel and carbon prices in the scenarios has been determined by multiplying the base values of these parameters with the scaling factors 2x, 1x, and 0.5x, as shown in Table 3.

Table 3: Fuel and carbon price levels considered in each of the scenarios considered

		Carbon		
	Times	**0.5x**	**1x**	**2x**
Fuel	**0.5x**	A1B1	A0B1	X
	1x	A1B0	A0B0	A2B0
	2x	X	A0B2	A2B2

The base values considered for the 2030 prices of coal, natural gas, and oil have been taken from the *European Resources Adequacy Assessment* [24] and are 6.62, 9.32, and 12.30 USD/MMBTU, respectively. The base value of the carbon price, which has also been taken from [24], is 110 USD/tCO2. Although there is high uncertainty on the level of fuel prices in the near future, the energy policies applied will, most likely, continue actively promoting LCE technologies. This will be reflected in the future level of prices, as shown in the most updated projections [27]. The scaling factors applied to the base values of these prices to compute the corresponding values in each scenario are given in Table 3. These factors have been determined to consider the following: 1) the crude oil and natural gas prices have historically shown some sudden changes in their generally increasing trends, and 2) the future incremental changes in the carbon price will not be as drastic as they have been in the recent past.

4. Simulation Results

As mentioned earlier, this section presents the performance of the proposed formulation. Table 4 shows the total system cost (expansion + operation) in MUSD for each formulation and scenario discussed in section 3.2. Note that there is no unserved energy in any scenario for any formulation. In addition, there is almost no difference in cost between M-A and M-B, with a maximum difference of 0.0017% for one of the scenarios. In contrast, the percentage difference between M-C and M-D and M-B is in the range [7.10%, 9.57%] and [6.29%, 8.39%], respectively. Ignoring the UC or AC-OPF model constraints in an expansion planning problem can lead to significant cost increases.

Table 4: Total system costs (MUSD) per formulation M-X and scenario. Yearly amortized cost is given in parenthesis next to each total system cost.

	Cycle-based AC+UC	AC+UC	AC	DC+UC
	M-A	M-B	M-C	M-D
A0B0	2119.4 (1081.2)	2119.4 (1081.2)	2315.5 (1277.4)	2291.4 (1253.7)
A0B1	1800.2 (937.0)	1800.2 (936.6)	1972.1 (1124.5)	1951.2 (1090.5)
A0B2	2705.5 (1126.2)	2705.5 (1126.2)	2910.6 (1333.2)	2887.1 (1302.3)
A1B0	1908.6 (998.7)	1908.5 (998.6)	2091.1 (1179.5)	2068.1 (1155.7)
A1B1	1535.7 (682.6)	1535.7 (681.7)	1663.4 (835.2)	1648.2 (826.7)
A2B0	2455.5 (1262.3)	2455.4 (1262.2)	2686.4 (1500.7)	2655.2 (1456.9)
A2B2	3063.8 (1193.4)	3063.8 (1193.2)	3281.5 (1412.4)	3256.5 (1377.9)

Both changes in fuel prices and changes in carbon prices have a significant impact on total system costs, although the former appears to have an even greater impact. The impact of each factor on the total cost depends on several key considerations: the range of analyzed price fluctuations analyzed, the emission intensity of the power output of the studied technologies, and the eventual selection of technologies for deployment, as shown in Fig. 5. This figure shows the capacity of installed generation and storage in gigawatts (GW), the extent of transmission lines in 1000 MW-km, which quantifies construction per unit length and power capacity, and reactive power compensation in gigavolt-amperes reactive (GVAr). Both carbon pricing and fuel costs have a significant impact on the decision process regarding technology deployment and utilization. In particular, new thermal units show significant variability across scenarios, as do the quantity of storage (batteries) installed. In addition, Fig. 6 shows that for the proposed formulation M-A, changes in fuel prices within reasonable limits can produce changes in total system cost in the range [-15%, 28%], while carbon prices produce changes in the range [-28%, 45%]. In addition, Table 5 shows, for each formulation and scenario, the curtailment for solar PV and onshore wind output in absolute terms, as well as the percentage of gross output of each technology that is curtailed. The curtailment ratio for the M-A formulation is a lower percentage of gross than that for M-C and M-D for all scenarios. Then, M-A tends to result in a higher level of utilization of the gross output of the RES-based generation used than M-C or M-D.

The formulations at hand can also be compared in terms of the computational time required to solve them. Table 6 shows, for each formulation and scenario, the CPU time required to compute an expansion planning solution for the first stage problem of the 2-stage process followed. The total

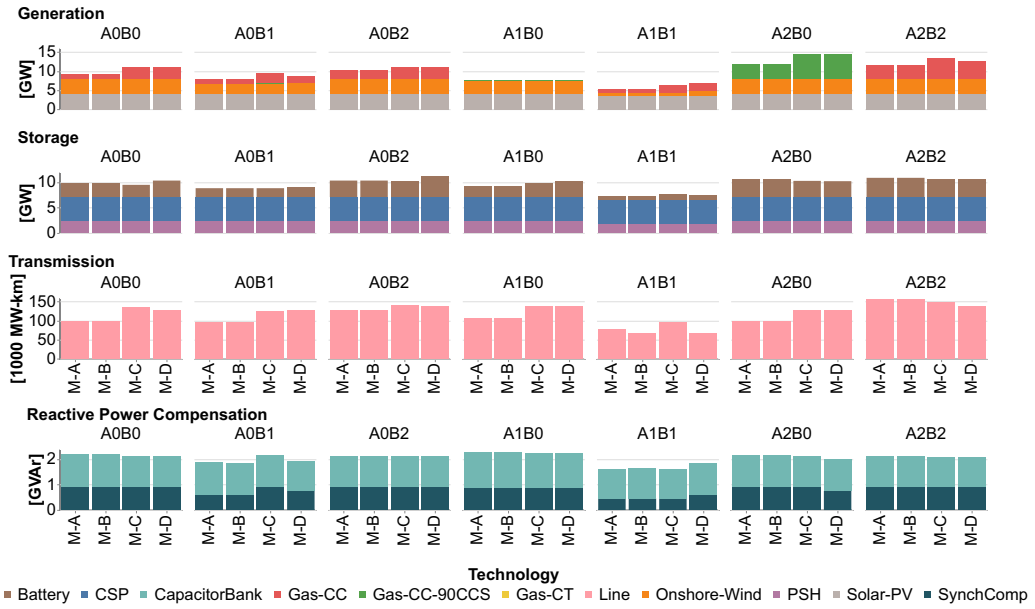

Figure 5: Installed capacity in power generation, energy storage, transmission network and reactive power compensation assets per scenario and formulation.

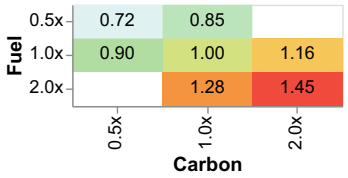

Figure 6: Ratio of total system costs in each scenario to those in the base scenario A0B0 for formulation M-A.

Table 5: Annual curtailment of the Solar-PV and Onshore-Wind output in TWh. Relative curtailment is given in parenthesis next to each annual curtailment.

	Solar-PV			Onshore-Wind		
	Cycle-based AC+UC	AC	DC+UC	Cycle-based AC+UC	AC	DC+UC
	M-A	M-C	M-D	M-A	M-C	M-D
A0B0	0.57 (4.11%)	3.48 (18.08%)	2.24 (13.00%)	0.66 (5.00%)	4.39 (23.21%)	2.31 (16.15%)
A0B1	0.83 (8.20%)	3.91 (27.90%)	2.46 (17.32%)	1.18 (9.82%)	5.22 (28.60%)	3.26 (20.78%)
A0B2	0.41 (2.91%)	2.18 (9.94%)	1.96 (8.96%)	0.47 (3.57%)	1.54 (8.47%)	1.07 (8.47%)
A1B0	0.69 (5.40%)	3.82 (22.88%)	2.34 (16.10%)	0.85 (6.69%)	4.94 (26.44%)	2.67 (17.26%)
A1B1	1.19 (10.27%)	4.17 (30.30%)	2.85 (20.57%)	1.75 (11.30%)	5.38 (29.50%)	4.72 (27.54%)
A2B0	0.46 (3.30%)	2.79 (12.47%)	1.72 (9.19%)	0.42 (3.13%)	2.82 (15.54%)	1.60 (11.00%)
A2B2	0.28 (1.73%)	1.68 (6.57%)	1.35 (5.30%)	0.30 (2.27%)	1.04 (5.23%)	0.84 (4.64%)

CPU time required to solve both stage 1 and stage 2 problems is provided in parenthesis.

Table 6: CPU time (h) required to solve the first stage problem using each formulation, and total CPU time (h) required to solve the problems in both stage 1 and stage 2 in parenthesis.

	Cycle-based AC+UC	AC+UC	AC	DC+UC
	M-A	M-B	M-C	M-D
A0B0	7.41 (14.53)	10.18	6.45 (17.19)	5.71 (15.49)
A0B1	7.27 (15.74)	8.83	6.65 (18.35)	5.70 (16.34)
A0B2	7.44 (14.88)	9.43	6.63 (17.04)	5.65 (15.74)
A1B0	7.40 (14.12)	9.71	6.67 (17.15)	5.63 (15.98)
A1B1	7.27 (13.24)	9.49	7.11 (16.36)	5.67 (14.53)
A2B0	7.73 (13.57)	9.66	7.62 (16.38)	5.82 (14.38)
A2B2	7.25 (13.29)	9.61	6.83 (16.36)	5.61 (14.93)

The 2-stage approach aims at computing an initial, long-term, expansion planning solution while for the M-C and M-D formulations, the aim is to refine the formulations in the medium term within the second stage to make them feasible. The time savings achieved by M-A compared to M-B are in the range of [17.67%, 27.21%], considering their times in the first stage. However, as can be seen, the total CPU time required to solve the problems in both stages is longer than that required to solve the stage 1 problem alone, and it is only provided for M-A, M-C and M-D formulations, since these are the formulations producing expansion plans that may need to be refined in stage 2. As expected, those formulations that consider a lower level of detail in the representation made of the system operation have a lower computational burden than those that consider a more detailed representation of the operation in the first stage. But the merits of each formulation are not the same when considering the total CPU time required to solve the problems in both planning stages as when just considering the time required to solve the problem in stage 1. The total time required to compute the final expansion plan to implement is longest when making use of M-C and M-D formulations, and it is longer for formulation M-C than for formulation M-D because formulation M-C is more expensive computationally speaking. For example, if we compare the total CPU time required to compute the expansion plans to implement based on those computed with formulations M-C and M-A, we can conclude that the time employed using formulation M-C is 3.9% to 23.6% longer, depending on the case study considered. This is probably due to the fact that, within formulation M-C, UC constraints are omitted in stage 1. These are intertemporal constraints that, when considered in stage 2, result

in a significantly different representation of the hourly system operation. UC constraints are already considered in stage 1 when making use of formulation M-A. Considering an AC-OPF model instead of the DC one (this concerns the comparison between M-A and M-D), the CPU time required to compute the first stage problem increases by a range of [27.544%, 34.68%], which is still within reasonable limits. The impact of considering UC constraints is the smallest (this concerns the comparison between M-A and M-C). In fact, there are some scenarios where the computational burden of the problems for M-A and M-C are similar. This is a consequence of the fact that in the time horizon considered, few generators will be subject to UC constraints (typically thermal ones). Then the number of these constraints and the associated binary variables will be small, and consequently they will have a small impact on the problem solution time. The increase in computation time resulting from considering these constraints is greater when cycle constraints are not considered. As expected, formulation M-D has the lowest computational cost. Comparing formulations M-A and M-B in terms of the CPU time required to solve the expansion planning problem in the first stage and the size of this problem, see Table 6 and Table 7. Both tables show that time and problem size are essentially proportional.

Table 7: Sizes of the stage 1 problem after the presolve[3], when considering scenario A0B0. The size of the stage 2 problem is given within parenthesis next to the stage 1 value.

		Cycle-based AC+UC	AC+UC	AC	DC+UC
		M-A	M-B	M-C	M-D
Variables	Continuous	13010112 (9497382)	15336949	10735864 (13803254)	7668475 (13419830)
	Binary	155889 (69215)	155889	539 (140300)	85739 (116917)
Constraints	-	10534324 (7268684)	13875521	9019089 (12141081)	6937761 (12141081)

Considering cycle constraints, the size of this problem is reduced in the first stage as follows: a) the number of continuous variables by 15.17%; b) the number of binary variables by 26.00%; and c) the number of constraints by 24.08%. Although the problem size is smaller for M-A than for M-B, the quality of the solution computed for the former is not inferior to that of the latter. In fact, given that the representation made of the system operation using formulation M-A is more accurate than that for formulations M-C and M-D, the stage 1 problem size for M-C and M-D is smaller than that

[3]The presolve comprises techniques and procedures used by optimization solvers to simplify the problem before attempting to solve it using more complex algorithms.

for M-A, but the stage 2 problem size if smallest for M-A. This is because formulations M-C and M-D neglect the UC constraints and use a DC-OPF instead of an AC-OPF. For example, M-C only considers a small number of binary variables in stage 1, which are only related to the investment decisions. M-D considers DC-OPF and does not consider investments in reactive power compensation devices. Therefore, the number of binary variables in the stage 1 problem is higher for formulations M-A and M-B than for formulations M-C and M-D. However, the system operation computed in stage 1 for formulation M-A is much closer to that considered in the reference formulation M-B than that considered for formulations M-C and M-D. Then, a significantly larger amount of investments need to be computed in stage 2 to make the resulting operation feasible, and increase the efficiency of the system expansion, when taking as a starting point the expansion plans computed using formulations M-C and M-D.

Table 8: Electricity production per technology for the proposed formulation (M-A) in the base scenario (A0B0).

	Technology	[TWh]	[%]
Thermal	Coal	0.0000	0.0
	Gas-CC	12.2322	20.9
	Gas-CC-90CCS	0.0000	0.0
	Gas-CT	0.0000	0.0
	Nuclear	3.0707	5.2
	Oil	0.0000	0.0
RES	Solar-PV	13.2367	22.6
	Onshore-Wind	12.5079	21.3
	Hydro	3.9706	6.8
	PSH	3.8688	6.6
	Solar-CSP	8.5931	14.7
	BESS	1.1525	2.0

The proposed formulation (M-A) supports the design of a power system in a low-carbon future context, as it represents almost 80% of electricity generation based on RES, as shown in Table 8; in contrast to the initial electricity generation of the RTS-GMLC system shown in Table 8 of [22], where there is still a significant amount of electricity generated by coal and significant use of the gas CC units. Table 8 shows that Gas-CC is the only one of the fossil fuel based thermal technologies in operation and has an important share in the energy mix. Nuclear power generation is also significant. Solar PV, onshore wind, solar CSP and, to a lesser extent, hydro dominate electricity generation. There is also significant use of storage in its various forms. Gas-fired power plants then produce all CO_2 emissions. Solar CSP production is significant because its hybridization with thermal storage can

benefit the system.

Regarding the network utilization in the base scenario (A0B) for formulation M-A, the average utilization level of interconnections between areas is 54.31%. At the same time, there are 179 hours when some interconnections are congested. The average network utilization level within areas is much lower, 20.05%, and there are 48 hours when some internal congestion occurs within areas. These are a few hours when the level of flexibility available within each area (mainly provided by storage) is insufficient to avoid network congestion. Since demand response is not considered in this case, the overall flexibility requirements can be measured in terms of the changes in net demand (electricity demand minus solar and wind generation) that the system must accommodate. The largest upward and downward ramps are 8953.57 MW and -1111.55 MW, respectively. The system must be flexible enough to handle these ramps. The ability of the system to do this can be assessed by the ramp margins, which represent the ratio of the maximum ramp the system can handle to the large ramp it faces in calculated system operation. The ramp-up and ramp-down margins in the base scenario A0B0 when using the M-A formulation to calculate system expansion are 1.10% and 1.81%, respectively.

Finally, we assess the level of accuracy of the operation computed using formulation M-A instead of the full original AC power flow equations. A model featuring the non-convex and non-linear fundamental equations of the AC power flow has been employed to separately calculate the system operation in each hour (load level) independently for the base scenario, A0B0. The power flow solution computed this way has been compared to that computed using formulation M-A in stage 1. When calculating the AC power flow, the investment plan (binary variables) and the active and reactive power generated by the units calculated when using the M-A formulation have been fixed to calculate the power flows, voltage magnitudes, and angles. Note that the reference node is treated as the slack bus, where the voltage magnitude and angle are fixed. Note that, for each storage unit, such as the PSH or BESS ones, we consider its net electricity generation in each hour individually. The net electricity generation level of a storage unit is computed as its generation level minus its consumption level in the respective hour. The energy inventory of storage units is managed separately for each unit. Moreover, the procedure used to perform this comparison mimics that used by the authors in [10].

Table 9: Average and maximum computation error per scenario after the first stage of the computation process for the power flows, the voltage magnitudes and the angles, making use of the M-A formulation. The errors made are determined by comparing the values computed for these variables with those resulting from considering the fundamental power flow equations.

Scenario	Feature	V_{ni}, [%]	θ_{ni}, [%]	P_{nijc}, [%]	Q_{nijc}, [%]
A0B0	Max. error	0.925	0.865	0.489	0.557
	Avg. error	0.028	0.032	0.068	0.072
A0B1	Max. error	0.910	0.981	0.572	0.563
	Avg. error	0.043	0.039	0.060	0.084
A0B2	Max. error	0.087	0.072	0.068	0.041
	Avg. error	0.032	0.059	0.070	0.057
A1B0	Max. error	0.916	0.821	0.516	0.683
	Avg. error	0.025	0.049	0.074	0.067
A1B1	Max. error	0.096	0.928	0.341	0.793
	Avg. error	0.041	0.039	0.077	0.076
A2B0	Max. error	0.048	0.062	0.005	0.006
	Avg. error	0.013	0.025	0.032	0.037
A2B2	Max. error	0.032	0.051	0.004	0.008
	Avg. error	0.019	0.026	0.013	0.028

Table 9 shows both the average and the maximum errors that occur in the calculation of power flows, as well as voltage magnitudes and angles. These calculations used the proposed formulation, M-A, across all scenarios and serve as a basis for comparison with the solutions derived from the nonlinear model executed per time step considering the solution provided by M-A. It is important to note that these errors, both average and maximum, are within acceptable limits. The maximum error made over all the considered system states (8736 hours in total per scenario) is less than 1%. In addition, the produced output data reveals the existence of complex relationships among the several variables considered across the different scenarios. In particular, there is a moderate positive correlation between the errors made when computing the active power flows (P_{nijc}) and the reactive power flows (Q_{nijc}). This can be observed across all the scenarios, highlighting the interconnectedness of these critical system variables. The analysis conducted has also allowed us to identify significant correlations existing among the errors made when computing key variables, such as the active and reactive power flows and voltage angles (θ_{ni}), for specific scenarios, such as A0B0, A1B1 and A2B2, highlighting the complex interplay among these variables. Despite the differences between these scenarios, the calculated results show a consistent

upward trend in the errors correlated to the increase in the ratio of fuel and carbon price levels between the scenarios. In terms of the voltage magnitude (V_{ni}), the maximum and average errors show opposite trends when comparing the scenarios A0B1, or A0B2, to scenarios A0B0 or A1B0, i.e. in the latter the maximum error is larger while the average error is lower than in the former. This suggests that the error distribution in the former scenarios is significantly different from that in the later scenarios. The largest deviations from the accurate values of the system operation variables are observed for the voltage magnitudes and angles. However, in average terms, the accuracy of these voltage-related variables exceeds the errors of the active and reactive power flows. This trend can be attributed to the assumptions made when deriving the M-A equations for current flows. The fact that significant maximum errors are made for the voltage variables, particularly those associated with insufficient reactive power generation - a key challenge in power systems with high shares of RES. In this regard, it is noteworthy the fact that the voltage at the nodes is largely affected by the production and consumption of reactive power within each system area.

5. Conclusions

This paper proposed a novel formulation of the IEP (generation, storage, and transmission expansion planning) problem including a linear branch flow-based AC-OPF model using cycle constraints. The deployment of reactive power compensation devices is considered in the proposed formulation while, in addition, the system operation is represented in a high level of detail regarding both the set of constraints and the time granularity. The time step considered is 1 hour and the operation of the system is calculated taking into account 8736 time steps. Although the considered RTS-GLMC comprise 73 nodes, our proposal is adaptable to incorporate representative periods - such as hours, days or weeks - also with hourly resolution. This flexibility enables the implementation of our approach in significantly medium - larger scale systems.

In order to make the resulting problem tractable, the AC-OPF model is represented in a compact way making use of cycle constraints, and the bounds of the voltage magnitude and angle variables are optimized and tightened. This allows one to compute, within reasonable time limits, highly-efficient expansion planning solutions for medium-to-large scale systems, considering the deployment of all types of energy resources and technologies, and

an accurate enough representation of the management of short-, medium-, and long-term storage technologies. As a result, the total system costs savings achieved by the proposed formulation are largely due to the fact that significant savings are achieved in the investments undertaken. Thus, the investments costs for the proposed formulation are 15.33% to 18.27% lower than those computed when considering a DC-OPF and 13.36% to 17.43% lower than those computed when neglecting UC constraints.

Therefore, the results computed for the case study clearly show that the proposed formulation outperforms the alternative ones. This can be attributed to the fact that the proposed formulation is able to more accurately represent the system operation. The robustness and accuracy of the proposed formulation when compared to others is showcased by the results of the analyses conducted for the various scenarios. However, it is important to acknowledge the limitations of the proposed formulation and outline directions for future research. In particular, the reader should be aware of the fact that the AC feasibility assessment we conduct may not fully address the challenges posed by extreme operating conditions or infrequent, but critical, events that could impact the system reliability. In addition, the computational burden of the simulations conducted when implementing our modelling approach could still limit its scalability to very complex and large power systems. However, it is also important to consider that time availability is not an issue in the simulation for expansion planning studies. To address these limitations, future improvements of the formulation proposed here could include the application of adaptive stochastic modeling to more accurately represent the probabilistic aspects of the power system operation. In addition, the use of state-of-the-art computational algorithms and machine learning techniques could significantly improve the computational efficiency of the proposed formulation, thus drastically reducing the computational time while maintaining or improving the accuracy of the representation of the system operation. These improvements will enhance the applicability of the proposed formulation in real-time analyses and ensure its tractability across different system scales, thereby maintaining its value in the dynamically changing energy management landscape.

Declaration of generative AI in scientific writing

During the preparation of this work, the authors used DeepL in order to avoid grammatical errors and improve the readability of the manuscript.

After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A: Mathematical formulation of the IEP problem

5.1. Objective function

The objective function in (14) aims to minimize the total system cost C, comprising the investment cost and the operating cost. The investment cost comprises the terms C^{ge}, C^{te}, C^{se}, and C^{re}, which respectively denote the installation costs of the generation units in (15a), transmission lines in (15b), ESSs in (15c), and reactive power compensation units, that is, synchronous compensators and capacitor banks, in (15d). The operating cost comprises C_n^{gen}, C_n^{con}, C_n^{co2}, and C_n^{ens}, which are summed for each load level $n \in \mathcal{N}$ and, respectively, denote: 1) the generation operating cost in (16a), 2) the cost of the energy consumption by the ESSs when charging in (16b), 3) the cost of emissions in (16c), and 4) the cost of load shedding in (16d).

$$\min C = C^{ge} + C^{te} + C^{se} + C^{re} + \sum_{n \in \mathcal{N}} C_n^{gen} + C_n^{con} + C_n^{co2} + C_n^{ens} \tag{14}$$

$$C^{ge} = \sum_{g \in \mathcal{G}^c} C_g^{gen} i_g^g \tag{15a}$$

$$C^{te} = \sum_{ijc \in \mathcal{L}^c} C_{ijc}^{line} i_{ijc}^t \tag{15b}$$

$$C^{se} = \sum_{e \in \mathcal{E}^c} C_e^{sto} i_e^e \tag{15c}$$

$$C^{re} = \sum_{s \in \mathcal{S}^c} C_s^{sy} i_s^q + \sum_{r \in \mathcal{R}^c} C_r^{sh} i_r^s \tag{15d}$$

$$C_n^{gen} = \sum_{g \in \mathcal{G}} D_n (CV_g p_{ng} + CF_g uc_{ng}) + C_g^{su} su_{ng} + C_g^{sd} sd_{ng} \quad : \forall n, \tag{16a}$$

$$C_n^{con} = \sum_{e \in \mathcal{G}} D_n CV_e p_{ne}^c \quad : \forall n, \tag{16b}$$

$$C_n^{co2} = \sum_{g \in \mathcal{G}} D_n P^{co2} E_g^{co2} p_{ng} \quad : \forall n, \tag{16c}$$

$$C_n^{ens} = \sum_{i \in \mathcal{B}} D_n l_{ni}^{shed} C^{shed} \left(P_{ni}^d + Q_{ni}^d \right) \quad : \forall n, \text{ (16d)}$$

In (15a)–(16d), the sets $\mathcal{G}^c$, $\mathcal{E}^c$, $\mathcal{L}^c$, $\mathcal{S}^c$, and $\mathcal{R}^c$, represent the set of candidate generation units, ESSs, lines, synchronous compensator, and capacitor banks, respectively. Furthermore, the binary variables i_g^g, i_e^e, i_{ijc}^t, i_s^q, and i_r^s denote the decision to install the candidate generator g, the candidate ESS unit e, the candidate line i.e., circuit c between nodes i and j, the candidate synchronous compensator s, and the candidate capacitor bank r. These variables are multiplied by the respective fixed annualized installation cost of each asset. The parameters C_g^{gen}, C_e^{sto}, C_{ijc}^{line}, C_s^{sy}, and C_r^{sh} refer to these costs. The binary variables uc_{ng}, su_{ng}, and sd_{ng} denote the decision for commit, startup, and shutdown of a generation unit, respectively. The variable p_{ng} represents the active power generation of a generator, p_{ne} represents the active power generation of an ESS unit, the variable p_{ne}^c represents the active power consumption of an ESS unit and the variable l_{ni}^{shed} the load shedding in p.u. The parameters CV_g and CF_g respectively denote the variable and fixed operating cost of a generation unit, the parameters C_g^{su}, and C_g^{sd} respectively denote the startup and shutdown cost of a generation unit, the parameter CV_e is the variable cost of an ESS unit during charging, and the parameter C^{shed} denotes the cost of non-served energy. Furthermore, the parameters D_n, P^{co2}, E_g^{co2}, P_{ni}^d, and Q_{ni}^d refer to the duration of a load level, the cost of CO2 emissions, the CO2 emission rate, the active power demand and the reactive power demand, respectively.

5.2. Investment decisions bounds for operational variables

Investment decisions in generation units bound the commitment of non-VRE generation units in (17a). The VRE generation's power output is also bounded by the investment decision of its respective generation unit in (17b). The power output and consumption of ESSs are also bounded by the investment decision in (18).

$$uc_{ng} \leq i_g^g \quad : \forall ng | g \in \mathcal{G}^c, \text{ (17a)}$$

$$\frac{p_{ng}}{\overline{P}_g} \leq i_g^g \quad : \forall ng | g \in \mathcal{G}^c, \text{ (17b)}$$

$$\frac{p_{ne}}{\overline{P}_e} \leq i_e^e; \quad \frac{p_{ne}^c}{\overline{P}_e^c} \leq i_e^e \quad : \forall ne | e \in \mathcal{E}^c, \text{ (18)}$$

$\mathcal{G}^c$ and $\mathcal{E}^c$ are the subsets of candidate generators and ESSs, respectively.

The investment in transmission lines limits the current flow in order to implicitly constrain the flow of power in candidate lines as in (19). In an AC-OPF, the current flow is calculated according to (4). To limit the flow when a DC-OPF is considered, limiting the active power flow is enough, as was made in [26].

$$\frac{i_{nijc}^2}{\overline{S}_{ijc}^2/\overline{V}_{ni}^2} \leq i_{ijc}^t \qquad : \forall nijc | ijc \in \mathcal{L}^c, \ (19)$$

$\mathcal{L}^c$ is the subset of candidate lines.

Finally, the output of candidate synchronous compensators is bounded by its investment decision as in (20). Note that synchronous compensators are modeled as generation units that are unable to produce active power. Investment decisions on capacitor banks set a limit to the reactive power production of the corresponding assets as in (6b) and (6c).

$$-i_s^q \leq \frac{q_{ns}}{-\underline{Q}_s}; \quad \frac{q_{ns}}{\overline{Q}_s} \leq i_s^q \qquad : \forall ns | s \in \mathcal{S}^c, \ (20)$$

$\mathcal{S}^c$ is the subset of candidate synchronous compensators.

5.3. Storage management

The following constraints related to storage management concern hydro reservoirs, pumped-storage hydro (PSH), battery energy storage systems (BESS), or other ESS technology such as compressed air energy storage or hydrogen.

The energy inventory or state of charge of an ESS unit is represented by (21a). This equation relates the state of charge of an ESS unit with the one for previous periods and the charge and discharge operation decisions made as in [28]. The equation considers each device's characteristic management scheme, or charge and discharge cycle, which can be hourly, daily, or weekly. The parameter τ represents the seasonality of the ESS unit.

$$y_{n-\tau_e,e} - y_{ne} + s_{ne} + \sum_{n'=n-\tau_e}^{n} D_{n'} \left(E_{n'e} - p_{n'e} + \eta_e p_{n'e}^c\right) = 0 \qquad : \forall ne, \ (21a)$$

The maximum and minimum capacity of electricity consumption to charge the ESS storage is defined by (21b).

$$\frac{p^{2c}_{ne}}{\overline{P}^{c}_{e} - \underline{P}^{c}_{e}} \geq 0; \quad \frac{p^{2c}_{ne}}{\overline{P}^{c}_{e} - \underline{P}^{c}_{e}} \leq 1 \qquad : \forall ne, \ (21b)$$

We define (21c), inspired in [28], to represent the continuous storage model, which does not employ binary variables to activate the charge and discharge of the ESS unit.

$$\frac{p^{s}_{ne}}{\overline{P}_{e} - \underline{P}_{e}} + \frac{p^{2c}_{ne}}{\overline{P}^{c}_{e} - \underline{P}^{c}_{e}} \leq 1 \qquad : \forall ne, \ (21c)$$

Then, the total charge or electricity consumption of an ESS unit is stated by (21d). Note that we split the electricity consumption into two parts: 1) the part that assures the minimum consumption by $\underline{P}^{c}_{e}$ as a fixed part, and 2) the part that can be managed and can vary between the range $[\underline{P}^{c}_{e}, \overline{P}^{c}_{e}]$. It was inspired by the second block of the power outputs approach presented in [29] as part of a tight formulation of the UC problem.

$$\frac{p^{c}_{ne}}{\underline{P}^{c}_{e}} = 1 + \frac{p^{2c}_{ne}}{\underline{P}^{c}_{e}} \qquad : \forall ne, \ (21d)$$

5.4. Unit commitment

The maximum and minimum output of the second block of a committed unit (all generation technologies except the VRE units) is given by (22a) and (22b). For the UC formulation, the energy-block approach was adopted. In this case, the production of a thermal unit is divided into a first block, related to the minimum power output of the unit, and the second block, related to the rest of the power, supplied or not to the system, and its relationship with multiple electricity products (i.e., reserves, ramps, etc.). The corresponding constraints are based on [30], which proposes a tight formulation of the UC problem considering reserves, ramps, start-up and shut-down constraints.

$$\frac{p^{s}_{ng}}{\overline{P}_{g} - \underline{P}_{g}} \geq 0 \qquad : \forall ng, \ (22a)$$

$$\frac{p^{s}_{ng}}{\overline{P}_{g} - \underline{P}_{g}} \leq uc_{ng} \qquad : \forall ng, \ (22b)$$

The total power output of a committed unit is defined by (22c). It is applied for all the generation units with the exception of the VRE units.

$$\frac{p_{ng}}{\underline{P}_g} = uc_{ng} + \frac{p^s_{ng}}{\underline{P}_g} \qquad : \forall ng, \ (22c)$$

The logical rule that relates the start-up, shut-down and unit commitment variable from the current load level with the previous of it is defined by using (22d) for all the generation units with the exception of the VRE units.

$$uc_{ng} - uc_{n-1,g} = su_{ng} - sd_{ng} \qquad : \forall ng, \ (22d)$$

The initial commitment and unit output are determined based on the merit order given by each unit's fixed and variable costs, including the RES and ESS units.

The maximum ramp-up/down constraints for the second block of energy produced by the generation units [31] with the exception of the VRE units is provided in (27).

$$\frac{-p^s_{n-1,g}+p^s_{ng}}{D_n R^u_{ng}} \leq uc_{ng} - su_{ng} \qquad : \forall ng, \ (22e)$$

$$\frac{-p^s_{n-1,g}+p^s_{ng}}{D_n R^d_{ng}} \geq -uc_{n-1,g} + sd_{ng} \qquad : \forall ng, \ (22f)$$

The variables for minimum up and down time of the thermal unit respectively represented by $su_{n'g}$ and $sd_{n'g}$, are bounded in (22g) and (22h), as stated in [32].

$$\sum_{n'=n+1-T^u_g}^{n} su_{n'g} \leq uc_{ng} \qquad : \forall ng, \ (22g)$$

$$\sum_{n'=n+1-T^d_g}^{n} sd_{n'g} \leq 1 - uc_{ng} \qquad : \forall ng, \ (22h)$$

5.5. Bounds for the OPF

The upper and lower bounds of the variables related to the OPF's variables (i.e., active/reactive power generation, active power consumption, load shedding, voltage magnitude and angle) are described in this section.

5.5.1. Bounds for generation units

All the bounds related to the operation of generation and ESSs are defined by (23)-(25). Bound (26) sets the limits for load shedding, while bounds for the active and reactive power injections are set in (27)-(29b). The reactive power injections are bounded, using (28a)-(29b).

$$0 \leq p^{c}_{ne} \leq \overline{P}^{c}_{e} \qquad : \forall ne, \ (23)$$

$$0 \leq p_{ne} \leq \overline{P}_{e} \qquad : \forall ne, \ (24)$$

$$0 \leq p^{2c}_{ne} \leq \overline{P}^{c}_{e}; \quad 0 \leq y_{ne} \leq \Psi_{e}; \quad 0 \leq s_{ne} \qquad : \forall ne, \ (25)$$

$$0 \leq l^{shed}_{ni} \leq 1 \qquad : \forall ni, \ (26)$$

$$0 \leq p_{ng} \leq \overline{P}_{g}; \quad 0 \leq p^{s}_{ng} \leq \overline{P}_{g} - \underline{P}_{g} \qquad : \forall ng, \ (27)$$

$$\underline{Q}_{g} \leq q_{ng} \leq \overline{Q}_{g} \qquad : \forall ng, \ (28a)$$

$$\underline{Q}_{s} \leq q_{ns} \leq \overline{Q}_{s} \qquad : \forall ns, \ (28b)$$

$$|q_{ng}| \leq p_{ng} \tan(\cos^{-1}(pf)) \qquad : \forall ng, \ (29a)$$

$$|q_{ng}| = \begin{cases} q_{ng} & \text{if} \quad pf = pf^{\text{cap}} \\ -q_{ng} & \text{if} \quad pf = pf^{\text{ind}} \end{cases} \qquad : \forall ng, \ (29b)$$

5.5.2. Bounds for network variables

Bounds for the voltage magnitudes are set in (30a). Equation (30b) sets the voltage angle for the reference node.

$$\underline{V}^{2}_{ni} \leq v^{2}_{ni} \leq \overline{V}^{2}_{ni} \qquad : \forall ni, \ (30a)$$

$$\underline{\theta}_{nijc} \leq \theta_{ni} - \theta_{nj} \leq \overline{\theta}_{nijc} \qquad : \forall nijc \ (30b)$$

$$\theta_{ni} = 0 \qquad : \forall ni | i = \text{reference node}, \ (30c)$$

Acknowledgement

The work leading to this paper has received fundings from the European Union's Horizon 2020 research and innovation programme for the openENTRANCE project (https://openentrance.eu/) under the grant agreement No. 835896.

References

[1] E. Du, N. Zhang, C. Kang, Q. Xia, A High-Efficiency Network-Constrained Clustered Unit Commitment Model for Power System Planning Studies, IEEE Transactions on Power Systems 34 (2019). doi:10.1109/TPWRS.2018.2881512.

[2] P. V. Gomes, J. T. Saraiva, L. Carvalho, B. Dias, L. W. Oliveira, Impact of decision-making models in Transmission Expansion Planning considering large shares of renewable energy sources, Electric Power Systems Research 174 (2019). doi:10.1016/j.epsr.2019.04.030.

[3] N. E. Koltsaklis, M. C. Georgiadis, A multi-period, multi-regional generation expansion planning model incorporating unit commitment constraints, Applied Energy 158 (2015). doi:10.1016/j.apenergy.2015.08.054.

[4] H. Zhang, H. Cheng, L. Liu, S. Zhang, Q. Zhou, L. Jiang, Coordination of generation, transmission and reactive power sources expansion planning with high penetration of wind power, International Journal of Electrical Power and Energy Systems 108 (2019). doi:10.1016/j.ijepes.2019.01.006.

[5] M. Moradi-Sepahvand, T. Amraee, Integrated expansion planning of electric energy generation, transmission, and storage for handling high shares of wind and solar power generation, Applied Energy 298 (2021). doi:10.1016/j.apenergy.2021.117137.

[6] D. V. Pombo, J. Martinez-Rico, H. M. Marczinkowski, Towards 100% renewable islands in 2040 via generation expansion planning: The case of São Vicente, Cape Verde, Applied Energy (2022). doi:10.1016/j.apenergy.2022.118869.

[7] O. H. Abdalla, M. A. A. Adma, A. S. Ahmed, Generation expansion planning considering unit commitment constraints and data-driven robust optimization under uncertainties, International Transactions on Electrical Energy Systems 31 (2021). doi:10.1002/2050-7038.12878.

[8] H. Shaoyun, C. Haozhong, Z. Pingling, Z. Jianping, J. Lu, Composite generation and transmission expansion planning with second order conic relaxation of ac power flow, Asia-Pacific Power and Energy Engineering Conference, APPEEC 2016-December (2016). doi:10.1109/APPEEC.2016.7779838.

[9] S. Wogrin, D. Tejada-Arango, S. Delikaraoglou, A. Botterud, Assessing the impact of inertia and reactive power constraints in generation expansion planning, Applied Energy 280 (2020). doi:10.1016/j.apenergy.2020.115925.

[10] L. H. Macedo, C. V. Montes, J. F. Franco, M. J. Rider, R. Romero, MILP branch flow model for concurrent AC multistage transmission expansion and reactive power planning with security constraints, IET Generation, Transmission and Distribution (2016). doi:10.1049/iet-gtd.2016.0081.

[11] D. Quiroga, E. Sauma, D. Pozo, Power system expansion planning under global and local emission mitigation policies, Applied Energy 239 (2019). doi:10.1016/j.apenergy.2019.02.001.

[12] A. Schwele, J. Kazempour, P. Pinson, Do unit commitment constraints affect generation expansion planning? a scalable stochastic model, Energy Systems 11 (2020). doi:10.1007/s12667-018-00321-z.

[13] C. Coffrin, H. L. Hijazi, P. V. Hentenryck, Strengthening convex relaxations with bound tightening for power network optimization, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9255 (2015). doi:10.1007/978-3-319-23219-5_4.

[14] M. M. U. T. Chowdhury, S. Kamalasadan, S. Paudyal, A Second-Order Cone Programming (SOCP) Based Optimal Power Flow (OPF) Model with Cyclic Constraints for Power Transmission Systems, IEEE Transactions on Power Systems (2023). doi:10.1109/TPWRS.2023.3247891.

[15] E. Alvarez, M. Paredes, M. Rider, Semidefinite relaxation and generalised benders decomposition to solve the transmission expansion network and reactive power planning, IET Generation, Transmission and Distribution 14 (2020). doi:10.1049/iet-gtd.2019.0331.

[16] S. Mahmoudi, B. Alizadeh, Simultaneous generation and network expansion planning in large-scale power systems under exact ac power flow equations, IET Generation, Transmission and Distribution (2022). doi:10.1049/gtd2.12586.

[17] K. Poncelet, E. Delarue, D. Six, J. Duerinck, W. D'haeseleer, Impact of the level of temporal and operational detail in energy-system planning models, Applied Energy 162 (2016) 631–643.

[18] Q. Ploussard, L. Olmos, A. Ramos, A search space reduction method for transmission expansion planning using an iterative refinement of the dc load flow model, IEEE Transactions on Power Systems 35 (2020). doi:10.1109/TPWRS.2019.2930719.

[19] F. Neumann, T. Brown, Transmission Expansion Planning Using Cycle Flows, e-Energy 2020 - Proceedings of the 11th ACM International Conference on Future Energy Systems (2020). doi:10.1145/3396851.3397688.

[20] J. Hörsch, H. Ronellenfitsch, D. Witthaut, T. Brown, Linear optimal power flow using cycle flows, Electric Power Systems Research 158 (2018) 126–135. doi:10.1016/j.epsr.2017.12.034.

[21] B. Kocuk, S. S. Dey, X. A. Sun, Strong SOCP relaxations for the optimal power flow problem, Operations Research (2016). doi:10.1287/opre.2016.1489.

[22] C. Barrows, E. Preston, A. Staid, G. Stephen, J. P. Watson, A. Bloom, A. Ehlen, J. Ikaheimo, J. Jorgenson, D. Krishnamurthy, J. Lau, B. McBennett, M. O'Connell, The IEEE Reliability Test System: A Proposed 2019 Update, IEEE Transactions on Power Systems 35 (2020). doi:10.1109/TPWRS.2019.2925557.

[23] L. Vimmerstedt, T. Stehly, S. Akar, A. Sekar, B. Mirletz, D. Stright, C. Augustine, P. Beiter, P. Bhaskar, N. Blair, S. Cohen, W. Cole,

P. Duffy, D. Feldman, P. Gagnon, P. Kurup, C. Murphy, V. Ramasamy, J. Robins, J. Zuboy, D. Oladosu, J. Hoffmann, 2022 Annual Technology Baseline (ATB) Cost and Performance Data for Electricity Generation Technologies (06 2022). doi:10.25984/1871952.
URL https://data.openei.org/submissions/5716

[24] ENTSO-E, European Resource Adequacy Assessment - ERAA - Downloads, Accessed: 17.03.2023 (2022).
URL https://www.entsoe.eu/outlooks/eraa/2022/eraa-downloads/

[25] NREL: System Advisor Model, Accessed: 17.03.2023 (2022).
URL https://sam.nrel.gov/

[26] A. Ramos, E. F. Alvarez, S. Lumbreras, OpenTEPES: Open-source transmission and generation expansion planning, SoftwareX (2022). doi:10.1016/j.softx.2022.101070.

[27] A. Ari, N. Arregui, S. Black, O. Celasun, D. Iakova, A. Mineshima, V. Mylonas, I. Parry, I. Teodoru, K. Zhunussova, Surging energy prices in europe in the aftermath of the war: How to support the vulnerable and speed up the transition away from fossil fuels, Accessed: 17.03.2023 (2022).
URL https://www.imf.org/-/media/Files/Publications/WP/2022/English/wpiea2022152-print-pdf.ashx

[28] S. Wang, G. Geng, Q. Jiang, R. Bo, Generation Expansion Planning Considering Discrete Storage Model and Renewable Energy Uncertainty: A Bi-Interval Optimization Approach, IEEE Transactions on Industrial Informatics (2022). doi:10.1109/TII.2022.3178997.

[29] G. Morales-Espana, C. M. Correa-Posada, A. Ramos, Tight and Compact MIP Formulation of Configuration-Based Combined-Cycle Units, IEEE Transactions on Power Systems 31 (2016) 1350–1359. doi:10.1109/TPWRS.2015.2425833.

[30] D. A. Tejada-Arango, S. Lumbreras, P. Sanchez-Martin, A. Ramos, Which unit-commitment formulation is best? a comparison framework, IEEE Transactions on Power Systems 35 (2020). doi:10.1109/TPWRS.2019.2962024.

[31] P. Damcı-Kurt, S. Küçükyavuz, D. Rajan, A. Atamtürk, A polyhedral study of production ramping, Mathematical Programming 158 (2016). doi:10.1007/s10107-015-0919-9.

[32] D. Rajan, S. Takriti, Minimum up/down polytopes of the unit commitment problem with start-up costs (2005).
URL https://dominoweb.draco.res.ibm.com/reports/rc23628.pdf