Multi-Plant Hydro Power Production Function Modeling: Application to the Stochastic Mid-Term Hydrothermal Planning Problem

Pedro T. M. Lira, André Luiz Diniz
Electric Energy Research Center - CEPEL
Rio de Janeiro, Brazil
pedrolira@poli.ufrj.br, diniz@cepel.br

Carmen L. T. Borges
Federal University of Rio de Janeiro - UFRJ
Rio de Janeiro, Brazil
carmen@nacad.ufrj.br

***Abstract*—This work proposes the modeling and application of the so-called "multi-hydro plants" for the stochastic mid-term hydrothermal planning problem. The approach consist in aggregating sub-cascades of hydroelectric plants based on the fact that the operation of the cascades is dictated only by decisions on the release of the reservoirs with regularization capacity, and the operation of run-of-the-river plants is a consequence of such decisions. As a consequence, we obtain a formulation of the problem with a smaller number of variables and constraints, without any loss of representation of the physical and operation constraints of the hydro plants. This yields a reduction in the computational time to solve typical stochastic hydrothermal coordination problems of more than 20%, and results are presented for a real case of the Brazilian system. Furthermore, there is even a gain in accuracy of about 6% in the approximation of the non-concave hydro production function, and an analytical investigation regarding this aspect is conducted.**

***Index Terms*—Hydro production function, mid-term planning, stochastic programming, piecewise-linear models**

I. Introduction

The operation planning of hydrothermal systems aims to coordinate the use of water and thermal resources to obtain a less costly operation, while ensuring the security of supply. In such problem, it is crucial to accurately model the hydro generation, which is a nonlinear function of storage, turbine discharge, and, in some plants, spillage [1]. Given its complexity and associated challenges, operation planning is traditionally split into long, mid, and short-term [2] problems, with different levels of detail among them. In long-term planning (1-5 years, monthly time steps), aggregation methods for representing hydro plants still prove to be advantageous and, since the seminal work [3], several approaches have still been proposed [4]-[6]. Although this type of modeling allows representing the variation of generation of hydro plants with the water head, the water balance equations along the cascade – which is essential for mid-term (up to 1 year, weekly steps) [7] and short-term (hourly steps, 1 week horizon) [8] planning - is usualy neglected.

Submitted to the 23rd Power Systems Computation Conference (PSCC 2024).

An important aspect of hydrothermal coordination is that the operation of hydro cascades is dictated by plants with reservoirs, since run-of-the-river plants must release its incoming water flow, preferably by the turbines. Based on this, the work [9] proposed the aggregation of each reservoir with its run-of-the-river plants immediately downstream in the cascade, to form a so-called "multi-plant", for which a single hydro production function – which is able to represent generation as accurately as in the full individualized modelling – is built.

In this context, the main objective of this work is to enhance the modeling of the previously proposed multi hydro production function (MHPF) [9] and to improve its performance, in order to turn it more useful for application in larger and more general systems. The contribution of this paper are as follows: (i) extension of the MHPF model to more general configurations of cascaded hydro plants; (ii) application of this concept to stochastic hydrothermal coordination problems; (iii) a more rigorous evaluation of the accuracy and computational efficiency of the methodology for very large systems.

II. Individual Hydro Production Function (HPF)

The generation gh' of a hydro unit depends on its turbined outflow q', the turbine (η_t) and generator (η_g) efficiencies and the net head h, as shown in (1), where k is a constant ($9.81 \times 10^{-3}(\mathrm{kg/m^2s^2})$) comprising the acceleration due to gravity, the density of water, and a unit energy conversion factor.

$$gh' = k\eta_t\eta_g hq' \tag{1}$$

The net head h is the difference between the forebay level $h_{up}(v)$ – which depends on the storage v of the reservoir – and the tailrace level $h_{dw}(q,s)$ – which depends on the total turbined (q) and spillage (s) outflows of the hydro plant – and later subtracting an average value h_{loss} for penstock losses:

$$h = h_{up}(v) - h_{dw}(q,s) - h_{loss} \tag{2}$$

For run-of-the-river hydro plants, the value of h_{up} is constant since v does not vary. Assuming an average efficiency value η, the generation gh of the whole plant, given by the hydro production function (HPF) (3), is the sum of the generations of its NU units:

$$gh = \sum_{u=1}^{NU} k\eta h q'_u = \\ k\eta(h_{up}(v) - h_{dw}(q,s) - h_{loss}) \sum_{u=1}^{NU} q'_u = \mathrm{HPF}(v,q,s) \quad (3)$$

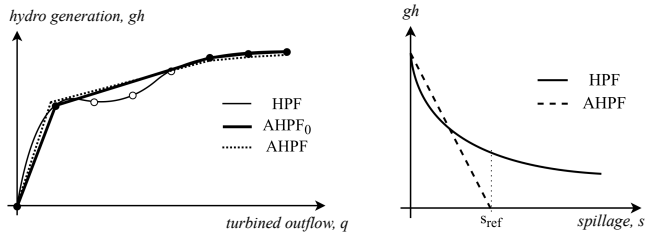
In short-term planning, the output of the hydro plant is highly dependent on the distribution of discharge among its units, as well as efficiency of turbines/generators, leading to non-concave functions. However, for mid-term planning, where weekly/monthly operation is usually considered, the function is better behaved, showing a reasonable concave shape [10], which allows the use of concave approximations [1], [11].

In the sequel, all subscripts refer to variables for the whole hydro plant, rather than to each individual generating unit.

A. Approximate Hydro Production Function (AHPF)

For the cases where the HPF is nearly concave, an approximate piecewise linear model can be computed, by the following steps [1]: (i) first, a grid of N points in the $v \times q$ plane is defined, yielding a value of gh for each point (v_i, q_i), for $s = 0$; (ii) the convex hull of the points $gh(v_i, q_i), i = 1, \ldots, N$, is calculated [12], providing an initial model AHPF_0; (iii) using a linear regression technique, a scaling factor α is calculated for the AHPF_0 model, aiming to minimize the average difference from the exact HPF, leading to the model $\mathrm{AHPF}(v,q) = \alpha \mathrm{AHPF}_0(v,q)$; (iv) finally, the AHPF model is extended to include spillage effects through a secant approximation, due to the positive curvature of the function along this axis. The formulation of the final model $\mathrm{AHPF}(v,q,s)$ is given by (4), where γ are the coefficients for each variable and K_i is the number of hyperplanes for i. The procedure is illustrated in Fig. 1, for the case of a run-of-the-river hydro plant. For details we refer to [1].

$$gh_i \leq \alpha \left(\gamma^k_{0,i} + \gamma^k_{v,i} v_i + \gamma^k_{q,i} q_i\right) + \gamma^k_{s,i} s_i,\ k = 1, \ldots, K_i \quad (4)$$



(a) (b)

Fig. 1. (a) Illustration of the construction of the AHPF model, for a given value v'; (b) secant approximation for spillage as an additional axis.

We note that the secant approximation for spillage is a drawback of the AHPF, as significant deviations from the exact HPF are expected along this axis. As described later, the multi-plant model proposed in this paper mitigates this limitation.

Despite providing adequate accuracy of hydro generation, the AHPF model requires an individual modeling of all plants within a hydrographic basin, which motivated the introduction of the concept of "multi-hydro production function (MHPF)", proposed in [9], which is further improved in this work. In the sequel we briefly describe the MHPF with the enhancements proposed in this work, which are detailed in [13].

III. Enhanced Multi-Hydro Production Function

The MHPF model consists of aggregating, within a single function, the generation of a subset of cascade plants, comprised of a reservoir and several run-of-the-river plants [9]. Since it does not require the conversion of storage or inflow/discharges into energy, the MHPF is a good alternative to the use of energy equivalent reservoirs [3]. Moreover, it reduces the dimensionality of the problem while maintaining the explicit representation of the hydro cascade, i.e., modeling the individual operation of each hydroelectric plant.

The sole assumption made in the construction of the MHPF is that spillage occurs only when the plant is at its maximum turbine discharge. Thus, a variable $d = q + s$ is created to represent the total release of the plant. Mathematically, this means that there is one less variable in the modeling of the MHPF, and therefore, a lower computational effort is expected in constructing and using the function in the optimization problem. It is important to emphasize that this assumption is acceptable for long/mid-term problems [14], while the approach in [1] remains useful when using a more detailed temporal granularity, as in the daily scheduling problem.

A. Identification and Construction of Multi-Plant Topology

The identification of multi-plants in a topology with reservoirs (triangles) and run-of-the-river (circles) plants is shown in Fig. 2, in a broader context where each plant may have more than one upstream plant, which is an enhancement as compared to the topology in [9]. To simplify the explanation, we initially assume that all hydro plants are located in the same subsystem. Since the operation of the hydro plants in each set (rectangles) depends only on the operation of the upstream reservoirs, it is possible to create an MHPF for each set.

In tree-shaped configurations, where a plant may have more than one upstream plant (e.g., plants 5, 6, 12, and 17), which is common in large river basins, special attention is required. In [9], the construction of a multi-plant with more than one reservoir in parallel was mentioned, which would lead to a convex envelope in the $\mathbb{R}^{2n+1}$ space, where n is the number of reservoirs. However, computing the convex hull for the MHPF in this case becomes prohibitive: if only 3 points out of 16 discretization points are used for variables v and d, up to 2.1×10^{16} hyperplanes can be evaluated for $n \geq 3$ [9].

For this reason, going from the downstream to the upstream plants, whenever there is a bifurcation in the topology with at least one reservoir in each branch, a different multi-plant is constructed for each branch (e.g., multiple hydroelectric plants E and F). Therefore, the piecewise linear approximation of the MHPF (see section III.C) remains in the $\mathbb{R}^3$ space.

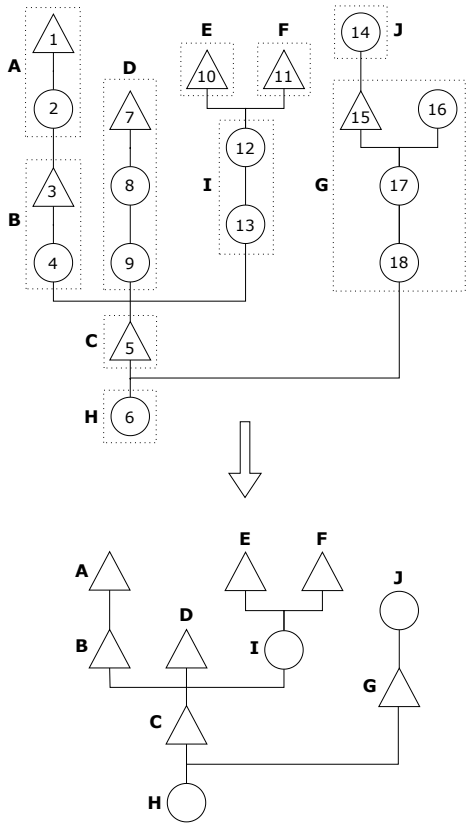


Fig. 2. Example of the construction of multi-plants, yielding a reduction from 18 to 10 hydro plants.

B. Calculation of the Exact MHPF for Each Multi-Plant

The assumption that spillage only occurs when the plant reaches its maximum discharge ($\overline{q}$) leads to (5a) and (5b).

$$q(d) = \min\{d, \overline{q}\} \tag{5a}$$
$$s(d) = d - q(d) \tag{5b}$$

When using (3) for the generation gh_i of each multi-plant i, two situations are considered: a multi-plant with reservoir and a multi-plant with only run-of-the-river individual plants.

1) Multi-plant with a reservoir r: The generation is given by $gh_r = \text{HPF}_r(v_r, q_r(d_r), s_r(d_r))$ for the reservoir r and $gh_j = \text{HPF}_j(q_j(d_r), s_j(d_r))$, for each remaining run-of-the-river plant, with the operation of each plant j given by:

$$d_j = d_r + \sum_{k \in \Omega_j^{up}} (I_k - R_k) \tag{6a}$$
$$q_j(d_r) = \min\{d_j, \overline{q_j}\} \tag{6b}$$
$$s_j(d_r) = d_j - q_j(d_r) \tag{6c}$$

where Ω_j^{up} is the set of upstream run-of-the-river plants belonging to the same multiple plant of the individual plant j (including j itself), and I_k and R_k are the incremental inflows and water withdrawals for other uses of water (input data) of each individual plant k. The MHPF, which is the sum of individual HPFs for the entire set Ω_i of plants belonging to multi-plant i, can be expressed as a function of storage v_r and release d_r of the reservoir r, as shown in (7), where $\Omega_i \setminus r$ is the set of individual plants in multi-plant i excluding plant r:

$$\text{MHPF}_i(v_r, d_r) = \text{HPF}_r(v_r, q_r(d_r), s_r(d_r)) + \sum_{j \in \Omega_i \setminus r} \text{HPF}_j(q_j(d_r), s_j(d_r)) \tag{7}$$

2) Run-of-the-river multi-plant: The generation equation is given by $gh_j = \text{HPF}_j(q_j(ud_i), s_j(ud_i))$ for all plants, where ud_i is the sum of releases from all reservoirs of the upstream multi-plants of plant i (such as multi-plants E and F in the case of multi-plant I in Fig. 2). Therefore, we have:

$$d_j = ud_i + \sum_{k \in \Omega_j^{up} \cup j} (I_k - R_k) \tag{8a}$$
$$q_j(ud_i) = \min\{d_j, \overline{q_j}\} \tag{8b}$$
$$s_j(ud_i) = d_j - q_j(ud_i) \tag{8c}$$

The MHPF is again given by the sum of individual HPFs for the entire set Ω_i of plants. However, in this case, since there is no reservoir, it is expressed solely as a function of ud_i, as shown in (9):

$$\text{MHPF}_\text{i}(ud_i) = \sum_{j \in \Omega_i} \text{HPF}_\text{j}(q_j(ud_i), s_j(ud_i)) \tag{9}$$

We note that: (i) in both types of multi-plants, the MHPF is uniquely determined based on the operation of the reservoirs; (ii) the set of equations (6a)-(9) are not included in the stochastic mid-term hydrothermal planning problem described in section IV. Instead, they are used to make an a priori computation of the points of the exact MHPF, which are used to build the approximate AMHPF model, as detailed below.

C. Approximate Multi-Hydro Production Function (AMHPF)

The same algorithm proposed in [1] for the individual HPF is applied to build the piecewise linear model for the MHPF, as detailed in [13], with some differences described below.

1) Spillage: To model spillage, a single point is added to the outflow window, with 110% of the maximum turbine discharge (q_{max}) of the multi-plant, which is the highest value of maximum turbine discharge among all individual plants within the multi-plant. This point was chosen because it is desirable to have a better representation of the function for lower spillage values, since for larger values of spillage the water value tends to be very low, and larger deviations between the exact and approximate functions have less impact. Fig. 3 illustrates the linear constraint on the spillage axis for a constant value v, passing through points q_{max} and $1.1q_{max}$.

2) Adding "non-differentiable" points: The MHPF is non-differentiable at points where the maximum turbined discharge of each plant within a multi-plant is reached (see Fig. 4). Exactly at these points, the multi-plant has a decrease in its overall efficiency since one plant starts to spill instead of increasing generation. Since the optimization process tends to find a solution at these points, it is interesting to explicitly include them in the piecewise linear approximation (AMHPF)

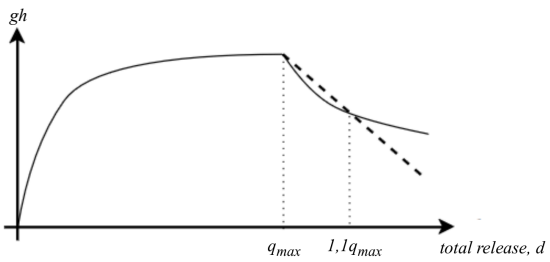


Fig. 3. Additional point in the d-grid for modeling spillage and how the linear constraint would be in this region in case of a run-of-the-river multi-plant.

of the MHPF . Therefore, for N_i being the number of plants within the multi-plant, at most $N_i - 1$ points are added to the set of discretization points used to build the AMHPF (note: since the point corresponding to the plant with the highest maximum turbine discharge is already in the window, there is no need to add it twice). However, not all $N_i - 1$ points are necessary: if the sum of known inflows to an individual plant is already greater than its maximum turbine discharge in a given scenario, that point does not need to be included.

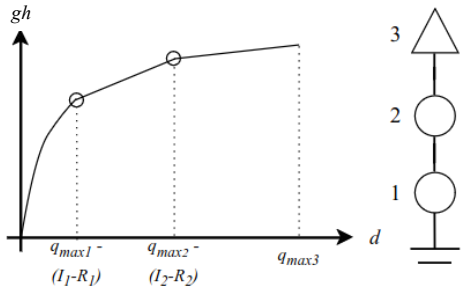


Fig. 4. Example of a MHPF for a multi-plant composed of three hydro plants ($q_{max3} > q_{max2} > q_{max1}$) and the additional points that would be included (the point q_{max3} would be already in the window).

3) Piecewise linear concave approximation for MHPF: After applying all these procedures, the approximate multiple hydroelectric production function (AMHPF), composed of a set of K_i linear inequalities for the generation ghm_i of each multi-plant i can be obtained, for each type:

- for a multi-plant i with a reservoir r:

$$ghm_i \leq \gamma^k_{0,i} + \gamma^k_{v,i} v_r + \gamma^k_{d,i} d_r, \quad k = 1, \ldots, K_i \quad (10)$$

- for a run-of-the-river multi-plant i:

$$ghm_i \leq \gamma^k_{0,i} + \gamma^k_{d,i} \sum_{u \in \Omega^{UR}_i} d_u, \quad k = 1, \ldots, K_i \quad (11)$$

where Ω^{UR}_i is the set of reservoirs upstream of multi-plant i, and $\sum_{u \in \Omega^{UR}_i} d_u = s d_u$. If $\Omega^{UR}_i = \emptyset$, $ghm_i = \gamma^k_{0,i}$, meaning it is a constant value that depends only on the natural inflows for the hydro plants. In principle, this generation value could be subtracted from the demand. However, to avoid infeasibilities in the optimization problem, this type of run-of-the-river multi-plant is also included in the set of problem constraints, and the solution to (11) allows a value in the range $[0, \gamma^k_{0,i}]$

IV. Formulation of the Hydrothermal Coordination Problem with Multi-Plants

In theory, the application of the multi-hydro plant concept leads to a reduction in the size of the problem, for two main reasons: (i) since there are fewer multi-plants than individual plants (see Fig. 2), the number of water balance equations is much lower, as explained later in this section. (ii) only one variable d is needed to represent water release (see section III-B) for each multi-plant, time step, scenario and load level, while in the AHPF model two variables (q and s) are necessary. Consequently, a decrease in computational time is also expected. However, since the AMHPF model requires more hyperplanes in its modeling, this premise must be carefully evaluated.

In addition to the proposed modifications in the MHPF, other constraints of the mid-term hydrothermal coordination problem (see more in [7]) may also be modified, allowing for an even greater reduction in problem size. We note that the hydrothermal coordination problem with a representation of multi-hydro plants is still capable of preserving the individual characteristics of the hydro plants, as described below.

A. Objective Function

The traditional objective function used in hydrothermal planning is cost minimization. After a time decomposition of the problem in order to apply Dual Dynamic Programming (DDP) [15], the subproblem of each time step t and inflow scenario a considers the sum of present costs and the expected value of future costs, as shown in (12), where p indexes the NL load levels in which time step t is divided:

$$Z = \min \sum_{l=1}^{NT} \sum_{p=1}^{NL} (ct_l gt_l^{p,t,a} + cd^t def^{p,t,a}) + E[Q^{t,a}(v^t)] \quad (12)$$

where NT is the number of thermal power plants, ct_l is the thermal incremental cost, and $gt_l^{p,t,a}$ is the generation of thermal power plant l; cd^t is the unit deficit cost, and $def^{p,t,a}$ is the deficit value. The term $E[Q^a_t(v^t)]$ is the expected value of the recourse function for time step t and scenario a, evaluated at the storage vector v^t. Such function is composed by the Benders cuts generated by the DDP algorithm and, for the leaf nodes of the scenario tree, represents the future cost function (FCF) provided by the long-term planning model [16]. Slack variables with artificial costs are included for all constraints, in order to guarante relatively complete recourse.

B. Load Supply

The satisfaction of demand $D_e^{p,t}$ of subsystem e, period t, and load level p is given by:

$$\sum_{l \in \Omega^T_e} gt_l^{p,t,a} + \sum_{i \in \Omega^{MH}_e} ghm_i^{p,t,a} + \sum_{f \in \Omega^{Int}_e} (int^{p,t,a}_{f \to e} - int^{p,t,a}_{e \to f}) def_e^{p,t,a} = D_e^{p,t} \quad (13)$$

where Ω_i^T / Ω_i^{MH} are the sets of thermal / multi-hydro plants in subsystem e; $ghm_i^{p,t,a}$ is the generation of multi-plant i; Ω_e^{Int} is the set of subsystems directly connected to subsystem e and $int_{f\to e}^{p,t,a}$ is the interchange from subsystem f to e.

C. Reservoir Constraints

The reservoir constraints are composed by water balance equations and evaporation. The number of water balance equations is also reduced, as only one equation is included for each multi-plant. The modified water balance equation for each multi-plant i is given by (14), where ζ is a conversion factor from m^3/s to hm^3 and a' is the ancestor node of a. In addition to the substitution $d = q + s$, other modifications are required in the formulation with individual hydro plants:

$$v_r^{t,a} - v_r^{t-1,a'} + \sum_{p=1}^{NL} \zeta^{p,t} \left[d_r^{p,t,a} - \sum_{j \in \Omega_i^{UR}} d_j^{p,t,a} \right] + qevap_r^{t,a} = \zeta^t \left[(I_r^{t,a} - R_r^t) + \sum_{j \in \Omega_i^{UF}} (I_j^{t,a} - R_j^t - qevap_j^{t,a}) \right] \tag{14}$$

- on the right-hand side, the water inflow is the sum of the inflow into the reservoir (if any) $I_r^{t,a}$, and the sum of the inflows $I_j^{t,a}$ for all the plants j in the set Ω_i^{UF} of upstream run-of-the-river plants for multi-plant i up to the first upstream reservoir;
- on the left-hand side, only the contributions of the outflows $d_j^{p,t,a}$ from the multi-plants j in the set Ω_i^{UR} of the first upstream reservoirs of multi-plant i are considered;
- since evaporation $qevap$ is a function of storage [7], it is constant for run-of-the-river plants (thus, included on the right-hand side of the equation) and must be decided only for the reservoir of each multi-plant i.

Multi-plants composed solely of run-of-the-river plants can also have their water balance equation represented for each load level, as shown in (15). The subscript ref refers to the most upstream plant of the run-of-the-river multi-plant, chosen as a reference and whose variables d_{ref} and $qevap_{ref}$ are included in the optimization problem.

$$\zeta^{p,t} \left[d_{ref}^{p,t,a} - \sum_{j \in \Omega_i^{UR}} d_j^{p,t,a} \right] = \frac{\delta^{p,t}}{\Delta t} \zeta^t \left[(I_{ref}^{t,a} - R_{ref}^t) + \sum_{j \in \Omega_i^{UF}} (I_j^{t,a} - R_j^t - qevap_j^{t,a}) - qevap_{ref}^{t,a} \right] \tag{15}$$

where $\delta^{p,t}$ is the duration of each load level p in period t and Δt is the duration of period t. One key aspect should be highlighted: all run-of-the-river hydro plants in a multi-plant have an implicit water balance equation per load level, as if total release d_r of upstream reservoir r was evenly distributed in the water balance equations of downstream plants by load level, according to their duration. This means that run-of-the-river plant cannot modulate water from one load level to another, and evaporation / incremental inflows are evenly distributed along the period. This is an important difference from water balance constraints of run-of-the-river plants in [7], which are defined per period, and not per load level.

The evaporation of each reservoir r is modeled as a linear function of the average storage in each period [7]:

$$qevap_r^{t,a} = a_{ev_r}^t + b_{ev_r}^t \left(\frac{v_r^{t-1,a'} + v_r^{t,a}}{2} \right) \tag{16}$$

where the terms $a_{ev_r}^t$ and $b_{ev_r}^t$ are 1^{st} order Taylor approximations of the nonlinear evaporation function, computed using initial storage as a reference. Since the right hand side of (16) contains the state variable $v_r^{t-1,a'}$, the multiplier of this expression is also used when building the coefficient related to storages in the Benders cuts built by the DDP algorithm. This is also the case of constraints (10), where we consider the average storage of the plant and then also use their multipliers in the computation of Benders cuts.

We note that there are less water balance equations in the multi-plant modeling than in the individual representation, since the first model does not require water balance equations for run-of-river plants, except for equations (15), which are applied only for the most upstream run-of-the-river plant of a run-of-the-river multi-plant.

D. Hydroelectric Generation Limits

The maximum generation of a multi-plant i is given by the sum of the maximum generations of each individual hydroelectric plant in the set Ω_i. On the other hand, the minimum generation needs to be zero since our model does not allow a minimum generation for an individual plant within a multi-plant. These statements are described in (17).

$$0 \leq ghm_i^{p,t,a} \leq \sum_{j \in \Omega_i} \overline{gh_j^t} \tag{17}$$

E. Recovery of Individual Hydroelectric Generation

Once the solution of the optimization problem is obtained, it is necessary to calculate the individual generation for each plant that composes each multi-plant. With the values of v^* and d^* in the optimal solution, it is possible to calculate the exact individual generation gh_j of each hydro plant j using (3), and thus, the exact generation ghm_i of the multi-plant. With these values, the $\alpha_{j,i}$ ratio of each plant j belonging to the multi-plant i is calculated:

$$\alpha_{j,i} = \frac{gh_j}{ghm_i} = \frac{gh_j}{\sum_{k \in \Omega_i} gh_k}. \tag{18}$$

These ratios are multiplied by the optimal solution ghm_i^* for the generation of the multi-plant to obtain each individual generation gh_j^*, as described in (19):

$$gh_j^* = \alpha_{j,i} ghm_i^* \tag{19}$$

Since there are no minimum generation constraints, this distribution is always feasible. However, if, after an initial redistribution, any individual plant has a generation greater than its maximum limit, the surplus value is redistributed to the other plants within the mult-plant, following the same process.

V. Numerical Results

To assess the performance of applying the multi-plant concept for the hydro plants, a typical mid-term hydrothermal coordination problem was constructed using real hydroelectric and thermal power plants from the Brazilian interconnected system and solved using the DDP technique. We represented all constraints of the previous section with the exception of the future cost function (FCF) at the leaf nodes. We note that there is no loss of generality in the comparison, since the use of the FCF does not favor any of the two modeling approaches. Actually, the absence of the FCF makes the analysis even more robust because the reservoirs will tend to use their entire volume down to the minimum storage level, allowing for the evaluation of the performance of both representations over a wider range of reservoir operating conditions.

In order to properly compare the two methodologies, the individual representation will have both types of water balance equations: only by period (AHPF*), as currently used in Brazil) and also for each load level for the run-of-the-river plants (AHPF), which is closer to the multi-plant modeling, as highlighted in section IV-C. In the individual modeling, the AHPF for each plant was constructed with 5 discretization points for both the turbined outflow and volume discretization windows (see [1], [13] for details). For the multi-plant modeling, the AMHPF was initially constructed using 4 points in the total release window, with the eventual addition of the breakpoint points described in section III-C, and 5 points for the volume window. The case was run on a computer cluster with the following configuration: CentOS Linux release 7.1.1503 operating system, 12 cores, 24 threads - AMD Opteron(TM) 6238 1.530 GHz processor, and 98 GB of RAM. The implementation was developed in C++ language, with linear problems solved using CPLEX v22.1 [17].

A. Case Description

The case was based on data from the official Monthly Operation Program of January 2022 developed by the Brazilian Independent System Operator (ISO) [18]. A total of 150 hydroelectric plants were considered, with the same characteristics data as in the official studies. The 102 thermal power plants had their variable costs per unit ($/MWh), minimum and maximum generation adapted for the case, as well as the load demand for each period. There were 12 periods with monthly discretization, 3 load levels, and 2 scenario branches at each node, with the same probability of occurrence, leading to a scenario tree with a total of 4,095 nodes. The tolerance gap for DDP convergence was set to 0.001%.

B. Solution Time and Deviation Analysis

Table I shows the major data and optimization results for the test case. The first information that can be extracted is that the total operating costs with all modeling approaches for the hydro production function are very close, with a difference of less than 1.9%. Such difference is expected since the problems are not identical, and its small magnitude indicates that the multi-plant concept provides a close approximation to the original problem. The AHPF variant shows a higher total operating cost than the AHPF* variant, which is expected since the AHPF problem is more restricted than the AHPF* problem. We note that the main purpose of the AMHPF is not to have a lower optimal value then the AHPF problem, but to obtain a more accurate representation of the exact HPF.

TABLE I
Comparison between total costs, time (average for three runs), and problem size between the representations.

	AHPF*	AHPF	**AMHPF**
Total cost (10^6 $)	8082.87	8106.61	**7954.55**
Number of iterations	77	73	**51**
Solution time (min)	461.29	742.31	**242.56**
Time/iteration (min)	5.99	10.17	**4.76**
Number of variables	19,271,070	20,642,895	**16,208,010**
Number of constraints	11,339,055	12,047,490	**11,666,655**

The number of variables in the multi-plant model is significantly lower, especially because the turbined outflow q and spillage s are replaced by the total release d. However, the number of constraints is (at first sight surprisingly) almost the same. This occurs because, although there are fewer water balance equations for the multi-plant model, more linear constraints are needed to model the AMHPF of each multi-plant, as compared to the sum of constraints needed to model the AHPF of each individual plant belong to that multi-plant. Also, it is likely that more hyperplanes are generated in the construction of AMHPF, since there are breakpoint points being added in the discretization window for the total release axis of this function.

In terms of CPU time, the performance of the multi-plant representation was significantly better: reductions of over 20% per iteration as compared to the AHPF* variant and over 50% as compared to the AHPF model, which is the problem that comes closest to the multi-plant representation. Although the number of constraints is similar, their "complexity" appear to be lower, due to the significant reduction in the number of non-zero elements in the constraints matrix.

Table II presents the average deviations between the generation provided by the approximate AMHPF model (Equations (10), (11)) and the corresponding exact MHPF functions (Equations (7), (9)) for the same values of storage, turbine discharge, and spillage obtained as outputs of the hydrothermal planning problem for each time step and scenario. The same procedure applies to the exact and approximate models for individual plants (Equations (4) and (3)), to which our proposed approach is compared. We observe that the AMHPF model exhibits lower deviations, especially for plants with reservoirs and downward deviations.

Fig. 5 shows the average spillage at each month in our test case, ranging between 3 and 11%, which is consistent with the spillage levels observed in the Brazilian system. Therefore, the absence of the FCF does not favor, in principle, any of the compared approaches.

TABLE II
AVERAGE DEVIATION CONSIDERING ALL PLANTS / LOAD LEVELS / NODES.

		AHPF*	AHPF	**AMHPF**
Per plant	MW	2.41	2.65	**2.27**
	%	0.70	0.76	**0.65**
Per reservoir	MW	2.16	2.15	**1.65**
	%	0.63	0.63	**0.48**
Per run-of-the-river	MW	0.25	0.51	**0.62**
	%	0.07	0.15	**0.18**
Upwards	MW	0.98	0.97	**0.96**
	%	0.29	0.28	**0.28**
Downwards	MW	1.43	1.68	**1.30**
	%	0.42	0.49	**0.37**

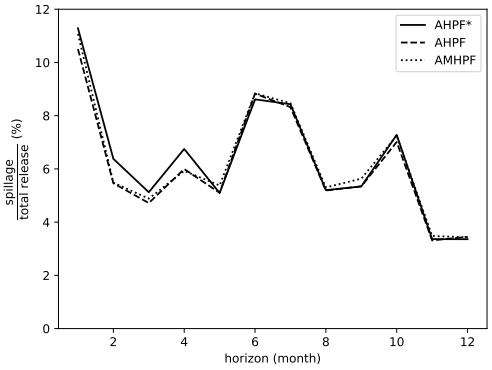


Fig. 5. Ratio of spillage to total release for each month of the case study.

Fig. 6 illustrates the average deviation for each period, and we highlight that even without the FCF, no significant advantage is being given to any representation, and the behavior is similar across all periods. Fig. 7 shows the empirical cumulative distribution function for each node in the scenario tree. It is possible to observe that the curve for AMHPF is more to the left, indicating smaller deviations, with about 90% of them being lower than 3 MW. Finally, Fig. 8 and Fig. 9 present the empirical cumulative distribution functions of the average deviations for each hydro plant. These figures show that up to the point with a value of 1 (both in MW and percent) on the x-axis, there are more deviation values for both AHPF* and AHPF approaches as compared to the proposed AMHPF model. This means that for around 60% of hydro plants with the smallest average deviations, these deviations are smaller in the individual approach. However, the remaining 40% of deviations larger than 1 (in MW and percent) have a greater impact on the calculation of the overall average, which is why AHMPF has a smaller overall average deviation.

It is worth noting that there is no evidence that the AMHPF modeling can reduce deviations that are already small in the individual modeling, as illustrated in Figs. 8 and 9. Accuracy gains in the hydro production function are achieved when individual plants are spilling, as the AMHPF is capable

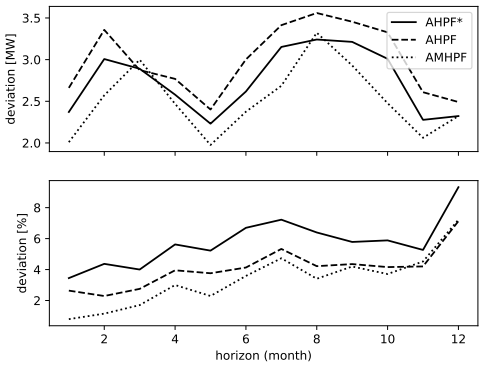


Fig. 6. Average deviation considering all hydro plants / load levels per period.

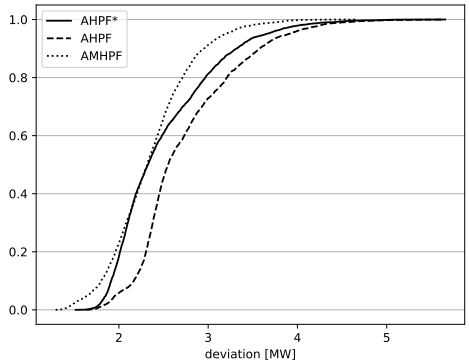


Fig. 7. Empirical cumulative distribution function of the average deviation per plant and per load level for each node.

of better representing the hydro production function in this situation: as discussed in [9], the HPF of the multi-plant only presents a non-concave shape (which is much more difficult to approximate) when all plants are spilling, while the AHPF of the individual plant will face the non-concave region when any of the plants in the cascade is spilling.

C. Sensitivity analysis on the size of the problem

Table III presents an analysis on the computation effort with the complexity of the problem. Since no paralellization scheme was applied in our implementation, one should look only on relative CPU times. We note that, even though the AMHPF model scales similar to AHPF model (and worse than the AHPF* model) with the number of hydro plants, it scales better with the number of load levels. However, extensive numerical experiments are necessary to obtain more solid conclusions.

TABLE III
CPU TIMES (MIN) VS. NUMBER OF HYDRO PLANTS/LOAD LEVELS.

Case	AHPF*	AHPF	AMHPF
13 hydro, 3 load levels	11.57	15.13	7.34
50 hydro, 3 load levels	92.89	79.31	26.73
150 hydro, 3 load levels	**461.29**	**742.31**	**242.56**
150 hydro, 1 load level	342.69	347.77	205.13

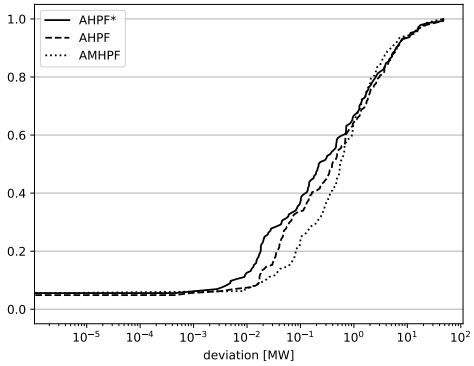


Fig. 8. Empirical cumulative distribution function of the average deviation per node/load level for each hydro plant, in MW (log-scale for the x-axis).

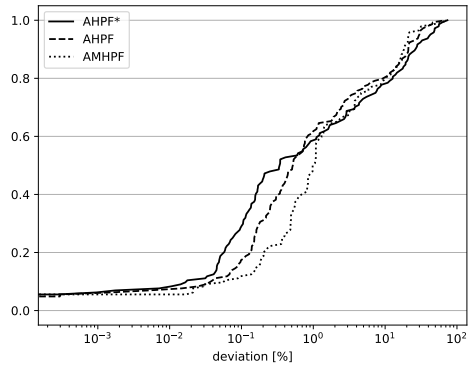


Fig. 9. Empirical cumulative distribution function of the average deviation per node/load level for each hydro plant, in % (log-scale for the x-axis).

VI. Conclusions

This work presented a formulation for the mid-term operation planning problem using the concept of multi-plants, which is an intermediate modeling approach to represent the hydro plants, where reservoirs are aggregated with their immediate downstream run-of-the-river plants in the same cascade, since hydraulic resources are primarily managed by the reservoirs. Despite this aggregated modeling, the individual characteristics of each plant are still represented, which is a great advantage to the use of equivalent energy reservoirs. With the reduction in the number of non-zero elements in the constraints matrix, resulting from the decrease in the number of variables and the number of water balance constraints, there was a reduction in the computational time in approximately 20% as compared to the individual modeling. In addition, there was also a reduction in deviations between the actual hydro production function and the proposed AMHPF, as compared to the same deviations using the current AHPF. Gains were mainly achieved when there is spilling in the operation, which is a non-convex region that needs to be modeled using only a single point in the window and always causes larger deviations due to the inevitable loss of accuracy.

The results presented in this work show that it is possible to apply the multi-plant concept to the stochastic mid-term hydrothermal coordination problem, as this modeling approach brought improvements as compared to the individual model. We list as possible future works: (i) use of AMHPF with electrical network and individual generation constraints for the hydro plants; (ii) consideration of water delay times between consecutive plants within a multi-plant, which is much more challenging; (iii) extension of the multi-plant concept to long-term operation planning model, solved by SDDP, where the main challenge is the use of autoregressive models for water inflows [16], which become not explicitly known a priori, but obtained as a result of forward scenarios sampled during the iterations of the algorithm. Finally, the AMHPF model could also be applied in other solving strategies, e. g., metaheuristics, once it becomes important to reduce the size of the problem.

References

[1] A. L. Diniz and M. E. P. Maceira, "A four-dimensional model of hydro generation for the short-term hydrothermal dispatch problem considering head and spillage effects," *IEEE Trans.Power Syst.*, v.23, n.3, pp. 1298–1308, 2008.

[2] E. Gil and J. Bustos and H. Rudnick, "Short-term hydrothermal generation scheduling model using a genetic algorithm," *IEEE Trans. Power Syst.*, v. 18, n. 4, pp. 1256–1264, 2003.

[3] N. V. Arvanitidis and J. Rosing, "Composite representation of a multireservoir hydroelectric power system," *IEEE Trans. Power Appar.and Syst.*, v. 89, n. 2, pp. 319-326, 1970.

[4] A. Helseth, M. Bo. "Hydropower aggregation by spatial decomposition—an SDDP approach," *IEEE Trans. Sust. Energy*, v.14, n.1, pp. 381-392, 2023.

[5] E. Blom, L. Soder, D. Risberg, "Performance of multi-scenario equivalent hydropower models," *Electric Power Syst. Res.*, v. 187, 106486, 2020.

[6] J.L.E. Brandão, L. Soder, D. Risberg, "Performance of the Equivalent Reservoir Modelling Technique for Multi-Reservoir Hydropower Systems," *Water Resources Management*, v. 24, n.12, 3101-3114, 2010.

[7] A. L. Diniz, F. S. Costa, M. E. Maceira, T. N. Santos, L. C. Brandão, and R. N. Cabral, "Short/mid-term hydrothermal dispatch and spot pricing for large-scale systems - the case of Brazil," in *20th Power Systems Computation Conference*, Dublin, 2018.

[8] J. I. Pérez-Díaz, J. R. Wilhelmi, L. A. Arévalo, "Optimal short-term operation schedule of a hydropower plant in a competitive electricity market," *Energy Conv. Manag.*, v. 51, n. 12, p. 2955-2966, 2010.

[9] A. L. Diniz, A. L. Saboia, R. M. Andrade, "An exact multi-plant hydro power production function for mid/long term hydrothermal coordination," in *19th PSCC – Power Systems Computation Conference*, Genoa, 2016.

[10] V. Bissonnette, L. Lafond, G. Cote, "A Hydro-Thermal Scheduling Model for the Hydro-Québec Production System," *IEEE Trans. on Power Syst.*, v.1, n.2, pp.204-210, 1986.

[11] Q. Goor, R. Kelman, A. Tilmant, "Optimal Multipurpose-Multireservoir Operation Model with Variable Productivity of Hydropower Plants," *J. Water Res. Plan. Manag.*, v.137, n.3, pp. 258-267, 2011.

[12] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls," *ACM Trans. Math. Softw.*, v. 22, n. 4, pp. 469–483, dec 1996.

[13] P. T. M. Lira, *Multi-plant hydro power production function modeling: application to the stochastic problem of mid-term operation planning*, Master Thesis, COPPE/UFRJ, Rio de Janeiro, 2023.

[14] S. H. F. Cunha, S. Prado, J. P. Costa, "Modelagem da produtividade variável de usinas hidrelétricas com base na construção de uma função de produção energética," in *XII Simp. Bras. Rec. Hídr.*, Vitória, 1997.

[15] J. R. Birge, "Decomposition and partitioning methods for multistage stochastic linear programs," *Oper. Res.*, v.33, n.5, pp. 989-1007, 1985.

[16] M. E. Maceira, D. D. J. Penna, A. L. Diniz, R. J. Pinto, A. C. G. Melo, C. V. Vasconcellos, C. B. Cruz, "Twenty years of application of stochastic dual dynamic programming in official and agent studies in Brazil – main features and improvements on the NEWAVE model," in *20th PSCC - Power Systems Computation Conference*, Dublin, 2018.

[17] CPLEX, I. I., "V12.1: User's Manual for CPLEX," *International Business Machines Corporation*, v. 46, no. 53, p. 157, 2009.

[18] ONS, "Acervo Digital," Available in: https://www.ons.org.br/paginas/conhecimento/acervo-digital/documentos-e-publicacoes, 15/06/2022.