Optimization Methods and Software
Vol. 00, No. 00, January 2008, 1–20

RESEARCH ARTICLE

Optimal Scenario Tree Reduction for Stochastic Streamflows in Power Generation Planning Problems

Welington Luis de Oliveira[1,2]; Claudia Sagastizábal[2,3]; Débora Dias Jardim Penna[2]; Maria Elvira Piñeiro Maceira[2,4]; Jorge Machado Damázio[2,4]

[2] *Electric Power Research Center – CEPEL, CP 68007 - Rio de Janeiro, Brazil.*
3 *IMPA – Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, Brazil.*
Research of this author is supported in part by CNPq Grant 30345-2008/9, by PRONEX-Optimization, and by FAPERJ.
4 *Rio de Janeiro State University – UERJ, São Francisco Xavier 525, Rio de Janeiro Brazil.*

(*October 19, 2009*)

The mid-term operation planning of hydro-thermal power systems needs a large number of synthetic sequences to represent accurately stochastic streamflows. These sequences are generated by a periodic autoregressive model. If the number of synthetic sequences is too big, the optimization planning problem may be too difficult to solve. To select a small set of sequences representing well enough the stochastic process, this work employs two variants of the Scenario Optimal Reduction technique. The first variant applies such technique at the last stage of a tree defined *a priori* for the whole planning horizon while the second variant combines a stage-wise reduction, preserving the periodic autoregressive structure, with re-sampling. Both approaches are assessed numerically on hydrological sequences generated for real configurations of the Brazilian power system.

Keywords: Scenario Reduction, Stochastic Programming, Mid-Term Operation Planning of Hydro-Thermal Systems.

AMS Subject Classification: 65K05; 49J52; 60G07

1. Introduction

The Brazilian power system (BPS) is predominantly hydraulic: hydro-generation represents over 80% of the total electricity produced in the country. Operation of hydro-plants is coupled both in time and in space, due to the presence of many basins with cascaded reservoirs. These features make the BPS operation planning a complex large scale optimization problem, that is very difficult to solve.

In hydro-dominated systems the most important source of uncertainty is the amount of water arriving into the reservoirs at each time of the planning period. For mid-term planning, an accurate representation of such streamflows is crucial. In this context, streamflows are represented by synthetic sequences, generated by a periodic autoregressive model [2]. An accurate representation of the streamflows stochastic process requires a large number of synthetic sequences. However, if this number is too big, the corresponding optimization problem may become too difficult to solve. It is then necessary to have some criterion for choosing, among

1 Email: wlo@cepel.br,gevazp@cepel.br.

ISSN: 1055-6788 print/ISSN 1029-4937 online

DOI: 10.1080/1055678xxxxxxxxxxxxx
http://www.informaworld.com

the many generated sequences, a small subgroup that represents “well enough” the stochastic data. Such a criterion should provide not only optimal decisions (close to the ones that would be obtained when using all the sequences), but also statistical adherence to the underlying stochastic process.

The Scenario Optimal Reduction (SOR) technique, introduced in [4], makes use of the Fortet-Mourier distance and duality theory to compute the distance between two probability measures, corresponding, respectively, to a huge scenario tree with all the sequences and to a smaller scenario tree, of fixed cardinality. The computed distance function defines the objective function for an optimization problem, which selects the best reduced tree by picking the most representative scenarios among all the sets of fixed cardinality. This higher-level optimization problem has a combinatorial nature and is essentially a set-covering problem [3]. Its solution is found by employing a heuristic procedure.

This work presents a methodology along the lines of [4], with one important difference. While in its original variant the SOR technique would destroy insterstage dependence of the sequences when creating a reduced tree, with our modification the reduced tree is composed by **periodic autoregressive sequences**. In this way, not only the reduced tree has statistical properties (means, standard deviations, cross correlations) close to the initial ones, but also at any time step all streamflows can be actually computed, knowing the history of realizations. In addition, the knowledge of the periodic autoregressive structure makes it possible to choose a very specific scaling factor to determine proximity of scenarios. Such scaling factor, depending on the tree autoregressive variance, prevents from eliminating all of the most extreme scenarios. This is an important feature in our application, for the reduced optimization problem to include information on flooding and droughts.

Our multistage problem has T time steps. We propose to build the autoregressive reduced scenario tree either by using the SOR technique on a tree defined *a priori* for the whole planning horizon, or by combining a stage-wise reduction, that preserves the periodic autoregressive structure, with re-sampling techniques. The former strategy, called of *global reduction*, adopts a two-stage approach by applying the original SOR technique only to the last time step (the first stage covers time steps 1 to $T-1$ while the second stage consists of time step T). By contrast, the latter strategy, that we refer to as of *local reduction*, proceeds step by step, from time step 2 to time step T. The local reduction technique is specially suitable for our application because, contrary to the global reduction, it does not need to define an initial (potentially huge) scenario tree. An additional important feature of the local reduction is that (thanks to the facts that no node is aggregated and that the process is autoregressive) no filtration distance is needed to keep close the optimal values of the original and the reduced problem. Indeed, it is shown in Subsection 6.2 that stability follows from the fact that the local reduction makes a continuous selection of the original process.

This work is organized as follows. Section 2 presents the scenario reduction problem, while Section 3 shows some characteristics of the streamflows scenario tree. Section 4 gives two algorithms for constructing reduced scenario trees, and a statistical analysis of the methodology. In Section 5 we validate the approaches on hydrological sequences generated for the whole Brazilian power system. Section 6 discusses the stability properties of the local reduction technique and the paper ends with some concluding remarks.

2. Scenario Optimal Reduction

Let $X \subset \ ^n$ denote the feasible set of operation variables for the mid-term planning, such as thermal generation, energy interchanges between subsystems[1], stored volumes in reservoirs, etc. Let $\Theta \subset \ ^K$ be the sample space of uncertainty, with N scenarios w^i of streamflows, with associated probability distribution, assigning to each scenario w^i a probability p_i, for $i \in I := \{1, \ldots, N\}$. In our multistage setting, the feasible set has the form $X = \prod_{t=1}^{T} X_t$ with $X_t \subset \ ^{n_t}$ and $n = \sum_{t=1}^{T} n_t$. Likewise for the streamflows, $w^i = (w^i_t, t = 1, \ldots, T)$, so $\Theta = \prod_{t=1}^{T} \Theta_t$ with $\Theta_t \subset \ ^{K_t}$ and $K = \sum_{t=1}^{T} K_t$. The notation $w_{[t]}$ and $\Theta_{[t]}$ corresponds, respectively, to sequences $(w_1, \ldots, w_t)$ and $\prod_{s=1}^{t} \Theta_s$, for $t = 1, \ldots, T$. A stochastic process (or simply a tree) is represented by ω, and its $\sigma-$algebra is given by $\mathcal{F}(\omega)$. Given the cost function $f : \Theta \times X \to \ $, the mid-term planning optimization problem can be written as

$$\min_{x \in X} E_P f(x), \text{ where } E_P f(x) := \sum_{i=1}^{N} p_i f(w^i, x) \tag{1}$$

is the expected value of the cost function, taking with respect to probability P.

When the scenario tree ω representing uncertainty is too large, the numerical solution of problem (1) becomes very complex. In this situation, it is convenient to choose a smaller tree, i.e. a subset of representative scenarios $\{w^{j_1}, w^{j_2}, \ldots, w^{j_{Nred}}\}$ with $Nred << N$, and a new probability distribution $Q = \{q(w^{j_1}), \ldots, q(w^{j_{Nred}})\}$. These elements define a new objective function (depending on the new, smaller, tree), denoted by $E_{J,Q} f(x)$, for which the reduced problem

$$\min_{x \in X} E_{J,Q} f(x) \tag{2}$$

will be actually solved. In this expression, $J := I \backslash \{j_1, \ldots, j_{Nred}\}$ is the index subset of discarded scenarios, and

$$E_{J,Q} f(x) := \sum_{l \in I \backslash J} p_l f(w^l, x).$$

When compared to (1), it becomes clear that the choice of J and Q determines the degree of quality of the reduced optimization problem (2). The SOR technique introduced in [4] for two-stage problems provides a tool (efficient both from the theoretical and algorithmical points of view) to keep close both the optimal values and optimal decision variables in (1) and (2). Having the initial tree ω, and denoting by $v^*(\omega)$ the optimal value in (1), consider for $\epsilon > 0$, let the $\epsilon - solution$ set:

$$S_\epsilon(P) := \left\{ x \in X; E_P f(x) \le v^*(\omega) + \epsilon = \min_x E_P f(x) + \epsilon \right\}.$$

To an optimal pair (J^*, Q^*) - the optimal set of discarded scenario indices and the optimal redistribution of preserved scenarios, corresponds a reduced tree $\tilde{\omega}$, whose

[1]A "subsystem" is an electrical geographical system, in our case corresponding to North, Northeast, South and Southeast Brazilian regions.

optimal value $v^*(\tilde{\omega})$ in (2) should satisfy the stability property

$$|v^*(\omega) - v^*(\tilde{\omega})| = |\min_{x\in X} E_P f(x) - \min_{x\in X} E_{J^*,Q^*} f(x)| \le \epsilon \quad \text{with } S_0(Q^*) \subset S_0(P) + B(0,\rho)\,, \tag{3}$$

and where $B(0,\rho)$ is a ball centered in $0 \in \ ^n$ with radius ρ.

Under certain conditions that are satisfied by the cost and constraints functions defining our mid-term planning operation problem, Theorem 1 in [4] ensures that, for any $\epsilon > 0$ there exists a constant $\rho > 0$ such that relation (3) is satisfied in a two-stage setting (the multi-stage case is more subtle, see comments below). Intuitively, this means that it is possible to take almost the same optimal decisions (generation, import-export exchanges, stored volumes) and, consequently, to obtain similar operational cost and deficit using a smaller subset of scenarios, provided this smaller set is properly selected.

In order to suitably select such smaller set, a natural criterion for proximity between solutions of problems (1) and (2) should strive to minimize the gap between the functional values of the respective optimization problems

$$\min_{J,Q} |E_P f(x) - E_{J,Q} f(x)| \ \text{ for each } x \in X\,, \tag{4}$$

or at least keep controlled the difference above. For two-stage problems, the effect of small perturbations of the probability distributions on the optimal value can be measured by using the Fortet-Mourier metric for probability measures, explained below. The multistage case ($T > 2$) is more involved and requires additional considerations, for example appending the Fortet-Mourier metric with filtration structure, or ensuring that the selection is done in a continuous manner; we refer to [9, 10, 15]. In Section 6 below we show that when applied to our power planning problem, the local reduction technique makes a continuous selection of the original stochastic process and, hence, the optimal value function is stable with respect to the "perturbation" yielding the reduced tree.

2.1. *The Fortet-Mourier Metric*

Two given scenarios w and w' are "close" when their distance is small enough. Such distance is measured by a function $d : \Theta \times \Theta \to \ _+$, for example, a norm or a pseudonorm, as in § 4.1 below.

To each tree ω corresponds a probability measure. The Fortet-Mourier metric defines the distance between two probability measures, in our case P and Q. For two-stage problems, note that when the cost function is Lipschitz continuous in variable w and X is a bounded set, we have the following inequality

$$|f(w,\cdot) - f(w',\cdot)| \le L d(w,w'), \tag{5}$$

where L is a Lipschitz constant for f. Therefore, the function $1/Lf(w,\cdot)$ belongs to the functional set

$$\Phi_d := \{\phi : \Theta \times X \to \ \ ; |\phi(w,\cdot) - \phi(w',\cdot)| \le d(w,w')\}. \tag{6}$$

To determine the distance between two probabilities P and Q, the Fortet-Mourier metric, denoted by $D_J(P,Q)$, uses Φ_d as feasible set in the following optimization problem

$$D_J(P,Q) := \sup_{\phi \in \Phi_d} |E_P \phi(x) - E_{J,Q}\phi(x)| \text{ for } x \in X. \tag{7}$$

We see that, for two-stage problems and at least locally, the Fortet-Mourier metric gives an upper bound for the difference between the optimal values (1) and (2).

The SOR problem determines the optimal pair (J^*, Q^*) by minimizing this metric over the variables J and Q:

$$\min_{J,Q} D_J(P,Q). \tag{8}$$

We now recall from [4] how to solve this problem.

2.2. *Bilevel Formulation of Scenarios Problem Reduction*

Problem (8) can be split as follows:

$$\begin{cases} \min c(J) \\ s.t. \;\; J \subset I \\ \qquad |J| = N - Nred, \end{cases}$$

where

$$c(J) := \begin{cases} \min D_J(P,Q) \\ s.t. \;\; Q \geq 0 \\ \qquad \sum_{l \in I \setminus J} q_l = 1. \end{cases} \tag{9}$$

This bilevel formulation, separating optimization of variables J and Q, yields an explicit formulation for the second level problem.

Theorem 2.1 *(Explicit formulation, [4, Theorem 2])*
Given $J \subset I$, the optimal value of the problem (9) is given by the expression

$$c(J) = \sum_{j \in J} p_j \min_{l \in I \setminus J} d(w^j, w^l),$$

where the minimum is attained at

$$q_l = p_l + \sum_{j \in J_l} p_j, \text{ where } J_l := \{j \in J : l \in \arg\min_{l \in I \setminus J} d(w^j, w^l)\}. \square \tag{10}$$

In principle, the SOR problem could be solved in an exhaustive manner by analyzing each possible set J with a fixed cardinality $N - Nred$. But this choice is computationally impracticable, due to the large number of possible combinations of scenario sets. Instead, we use the following heuristic method from [7].

Algorithm of Fast forward selection, [7, Algorithm 2.4].
Given a certain scenario distance matrix $\mathbf{D} \in {}^N \times {}^N$ with entries $\mathbf{D}(\omega^k, \omega^u)$ for $k, u = 1, \ldots, N$, the working index set $J^{[0]} = I = \{1, 2, ..., N\}$, and N_{red}, the desired number of preserved scenarios (or alternatively a tolerance `Tol` > 0), do the following steps.

`Step 0.` Set $J^{[0]} = \{1, 2, \ldots, N\}$ and $d^{[1]}_{ku} = \mathbf{D}(w^k, w^u)$, for k and $u = 1, \ldots, N$.
`Step 1.` Compute

$$z^{[1]}_u := \sum_{k=1, k \neq u}^{N} p_k u^{[1]}_{ku}, \quad u = 1, \ldots, N. \text{ Choose } u_1 \in \arg \min_{u \in \{1, \ldots, N\}} z^{[1]}_u,$$

and set $J^{[1]} = \{1, \ldots, N\} \backslash \{u_1\}$.
`Step i.` Compute

$$d^{[i]}_{ku} = \min\{d^{[i-1]}_{ku}, d^{[i-1]}_{ku_{i-1}}\}, \quad k, u \in J^{[i-1]}, \text{ and } z^{[i]}_u = \sum_{k \in J^{[i-1]} \backslash \{u\}} p_k d^{[i]}_{ku}, \quad u \in J^{[i-1]}.$$

Choose $u_i \in \arg \min_{u \in J^{[i-1]}} z^{[i]}_u$ and set $J^{[i]} = J^{[i-1]} \backslash u_i$.

If $i = N - N_{red}$ (or alternatively $z^{[i]}_{u_i} \leq \mathtt{Tol}$), go to Step $N - N_{red}$. Otherwise, update $i = i + 1$ and go to Step i.
`Step` $N - N_{red}$. The set $J^* = J^{[i]}$ is the index set to be discarded. □

The main idea of this algorithm is to iteratively solve problems of the form

$$\min\{c(J) : J \subset I : \#J = N - i\},$$

where $i = 1, \ldots, Nred$ is the iteration counter.

For more information on the SOR technique, we refer to [4, 5] for two-stage problems and to [8, 9] for the multistage setting. We now present a methodology along the lines of [5], with one important difference, that allows us to use multivariate scenario trees generated by periodic autoregressive models without losing interstage dependence in the reduction process.

3. Streamflows Scenario Tree

Our operation planning problem uses synthetic sequences for representing the stochastic process of streamflows. The model GEVAZP, developed by CEPEL [12], generates multivariate synthetic sequences of monthly streamflows by a periodic autoregressive model, [2].

Figure 1 shows two typical streamflows trees for the Tucuruí hydro-plant (an 8.370 MW hydro-plant in the N subsystem).

To give a hint on the bulk of information that needs to be handled in our planning problem, we mention that there is one tree as in Figure 1 per hydro-plant and that the whole BPS includes more than 100 hydro-plants (111 in our numerical validation in Section 5).

With the information generated by model GEVAZP, the next step is to determine the optimal generation for each power plant, subject to satisfying demand, and taking into account streamflow uncertainty and operational constraints. The corresponding optimization problem is solved by the DECOMP model, [13].

We now detail some important features of the synthetic sequences generation process and the underlying autoregressive structure.

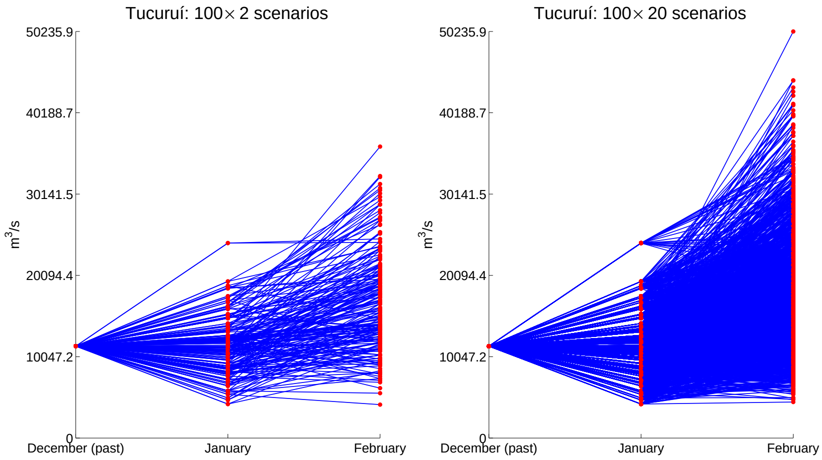

Figure 1. Two streamflow scenario trees.

3.1. *Synthetic Sequences Generation*

For a given hydraulic plant and a given month t, let $\{Z_{t-1}, \ldots, Z_{t-\pi_t}\}$ be known past values of streamflows. Typically, the model order is $\pi_t < 12$. Let μ_t and σ_t^2 be the historical mean and variance, respectively. The periodic autoregressive model for the t^{th}-month streamflow is

$$Z_t = \mu_t + \sum_{i=1}^{\pi_t} \sigma_t \phi_i^t \left(\frac{Z_{t-i} - \mu_{t-i}}{\sigma_{t-i}} \right) + \sigma_t \xi_t, \tag{11}$$

where ξ_t is a correlated and identically distributed noise with zero mean and variance $\sigma_{\xi_t}^2$, and ϕ is the autoregressive coefficients vector. The theoretical expected value of relation (11) conditioned to the past is denoted and defined by

$$E[Z_t|\_] = \mu_t + \sum_{i=1}^{\pi_t} \sigma_t \phi_i^t \left(\frac{Z_{t-i} - \mu_{t-i}}{\sigma_{t-i}} \right),$$

with conditional theoretical variance

$$V[Z_t|\_] = \sigma_t^2 \sigma_{\xi_t}^2.$$

Using (11) written with t replaced by $t+1$ yields the relation

$$\begin{aligned} Z_{t+1} = \mu_{t+1} + \sigma_{t+1}\phi_1^{t+1} \left(\tfrac{E[Z_t|\_] + \sigma_t \xi_t - \mu_t}{\sigma_t} \right) \\ + \textstyle\sum_{i=2}^{\pi_{t+1}} \sigma_{t+1} \phi_i^{t+1} \left(\tfrac{Z_{t+1-i} - \mu_{t+1-i}}{\sigma_{t+1-i}} \right) + \sigma_{t+1}\xi_{t+1}. \end{aligned}$$

As a result, the conditional theoretical values for step $t+1$ are given by

$$E[Z_{t+1}|\_] = \mu_{t+1} + \left(\frac{\sigma_{t+1}}{\sigma_t} \phi_1^{t+1} \right) (E[Z_t|\_] - \mu_t) + \sum_{i=2}^{\pi_{t+1}} \sigma_{t+1} \phi_i^{t+1} \left(\frac{Z_{t+1-i} - \mu_{t+1-i}}{\sigma_{t+1-i}} \right),$$

and

$$V[Z_{t+1}|\_] = \left(\frac{\sigma_{t+1}}{\sigma_t} \phi_1^{t+1} \right)^2 V[Z_t|\_] + \sigma_{t+1}^2 \sigma_{\xi_{t+1}}^2 .$$

The interest of these theoretical values is that they can be used to validate the synthetic sequences statistics.

In [5] the SOR algorithm is applied to a time horizon of T steps, going backwards until time step $t = 1$. This procedure aggregates scenarios and, hence, destroys the autoregressive structure of the stochastic process. In our application, preserving such structure is important, because it ensures that all streamflows at step $t+1$ can be actually computed, given the known realizations until step t. In this way, the temporal dependence is not destroyed.

In order to preserve the periodic autoregressive structure, we modify the method [5] in two, alternative, different ways:

- By performing a "global reduction" that applies [5] only once, at $t = T$.
- By performing a "local reduction" from time step t to time step T.

An important difference between the global and local reduction is that the latter does not need to generate the whole tree. The particular choice of which one of the two alternatives to apply will essentially depend on whether or not it is possible to generate at once the full initial tree, with N multivariate scenarios.

4. Modified SOR Algorithms

Figure 2 shows schematically the global reduction procedure, applied on a full tree.

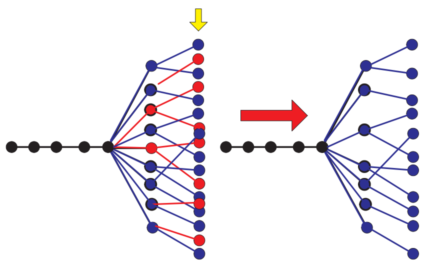

Figure 2. GOR Algorithm.

Algorithm of Global Optimal Reduction – GOR.

Parameters: For $t = T$, let N_{red} be the number of scenarios that will be preserved, with $N_{red} << N$.

`Step 1.` Compute a distance matrix $\mathbf{D} \in {}^{N} \times {}^{N}$ by means some function $d : \Theta \times \Theta \to {}_{+}$.

`Step 2.` Apply the fast forward selection algorithm in Section 2 to find the most representative N_{red} scenarios.

`Step 3.` Redistribute the preserved scenarios by using formula (10).□

Now, suppose we are interested in generating a scenario tree with 12 time steps and $N = 500$ scenarios at each node, where each node corresponds to streamflows arriving into one reservoir. The corresponding total number of scenarios involved is 500^{12} (per hydro-plant), a bulky figure, impossible to be manipulated or even stored (not to mention that there are more than 100 plants!). In this context, the GOR algorithm is computationally infeasible.

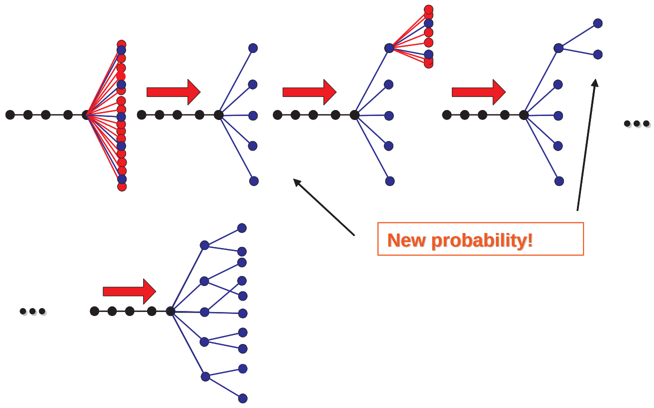

Figure 3. LOR Algorithm.

The LOR algorithm offers a good alternative, because it allows to work with all the information in an implicit way, proceeding step by step. At a given time step, N scenarios are generated, and they are immediately reduced to N_{red} scenarios by applying the SOR strategy. Then, for each scenario preserved, N new scenarios are generated, and the scenario selection strategy is applied to each branch of the new partial tree, covering two times steps. This procedure is repeated until $t = T$. A schematic representation of the algorithm is given in Figure 3.

Algorithm of Local Optimal Reduction – LOR.

Parameters: For $k = t, \ldots, T$, let $N(k)_{red}$ be the number of scenarios that will be preserved, satisfying $N(k)_{red} << N(k)$ and let $J^k \subset I^k := \{1, \ldots, N(k)\}$ be the index subset of discarded scenarios.

`Step 1.` Set $k = t$.
`Step 2.` Generate $N(k)$ scenarios.
`Step 3.` Compute a distance matrix $\mathbf{D} \in {}^{N(k)} \times {}^{N(k)}$ by means some function $d : \Theta \times \Theta \to {}_{+}$.
`Step 4.` Apply some algorithm for selecting scenarios (for example, the fast forward selection) to find $N(k)_{red}$ representative scenarios with index belonging to $I^k \backslash J^k$.
`Step 5.` Redistribute the scenarios by formula (10).
`Step 6.` Set $k = k + 1$. If $k \leq T$ go to Step 2, otherwise stop.$\square$

We see that with this algorithm it is possible to build a reduced scenario tree without having an initial huge tree: at each time step k, $N(k)$ new scenarios are generated, that are immediately reduced to $N(k)_{red}$ scenarios.

Both algorithms depend on the "distance" function d chosen at Step 3. We now comment on how to choose such function in a clever way, that makes use of the autoregressive data.

4.1. *The Pseudonorm Function*

The function d used in [7] is given by

$$d(w, w') = d_r(w, w') := \|w - w'\| \max\{1, \|w\|^r, \|w'\|^r\},$$

where $\|\cdot\|$ is a norm and $r > 1$ a constant.

For the mid-term operation planning problem there are usually more than $Nplant = 100$ hydroelectric plants, so each scenario w_t^i is a vector in ^{Nplant} with coordinates $w_t^i(j)$. The following pseudonorm

$$\breve{d}(w_t, w_t') = \max_{j=1,\ldots,Nplant} \{d^j(w_t, w_t')\} \tag{12}$$

where $d^j(w_t, w_t') = \sqrt{c_t^j}|w_t(j) - w_t'(j)| \max\left\{1, c_t^j|w_t(j)|^2, c_t^j|w_t'(j)|^2\right\}$, measures proximity between scenarios. In this expression, $|\cdot|$ is the absolute value function, and c_t^j is a *well-chosen* scaling factor. For our application, it is convenient to set $c_t^j = \frac{V[Z_t|\_]}{S_t^2}$, the ratio between the theoretical and sample variance (of generated scenarios) for streamflows of the j^{th} hydroelectric plant. Our choice is based on the fact that when scenarios are discarded, variance tends to decrease. But reducing variance is not interesting for the streamflows stochastic process, because it is convenient to keep extreme scenarios, that represent more severe flooding or droughts. The penalty terms

$$\sqrt{c_t^j} \text{ and } \max\left\{1, c_t^j|w_t(j)|^2, c_t^j|w_t'(j)|^2\right\}$$

play a key role in stabilizing the variance values while preventing the elimination of all of the most extreme scenarios.

When the function d is a norm (satisfies the triangular inequality) it is shown in [9] that the function $c(J)$ defined in Theorem 2.1 gives an equivalent representation for the Fortet-Mourier metric. Our function $\breve{d}(\cdot, \cdot)$ from (12) does not satisfy the triangular inequality. In this case, the Monge-Kantorovich functional gives an upper bound for the Fortet-Mourier metric. In our numerical validation, presented below, we observed that the pseudonorm function gives better results than when using the ℓ_1, ℓ_2, or the ℓ_∞ norms.

5. Numerical Results

This section validates, by using statistical tests, the application of the GOR and LOR techniques to hydrological sequences generated by model GEVAZP on a real configuration of the mid-term operation planning problem in Brazil.

Specifically, we consider historical streamflows from years 1931 to 2006, a planning horizon of 2 months (April and May), and a power mix with 111 hydroelectric plants. Results are reported for 4 hydro-plants, named Três Marias, Itá, Tucuruí, and Xingó.

The considered trees are:

- For GOR algorithm: 120 generated scenarios for April, with all of them preserved; and $120 \times 50 \rightarrow 6000$ generated scenarios for May, with 960 preserved (16% of scenarios kept).
- For LOR algorithm: 500 generated scenarios for April, with 120 preserved (24% of scenarios kept); and 60000 generated scenarios for May, with 960 preserved scenarios (1.6% of scenarios kept).

A first assessment of the quality of preserved scenarios is done by comparing the statistics of both trees (initial and reduced) with the statistics predicted by theory.

Figures 4 and 5 report on the mean and standard deviation of the initial and reduced trees, using respectively the GOR and LOR algorithm. Note that the cor-

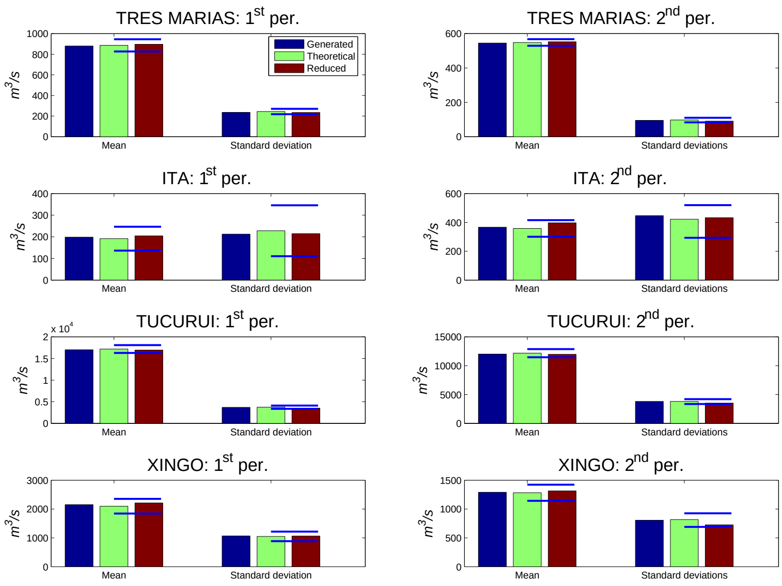

Figure 4. GOR Algorithm: Mean and Standard Deviation

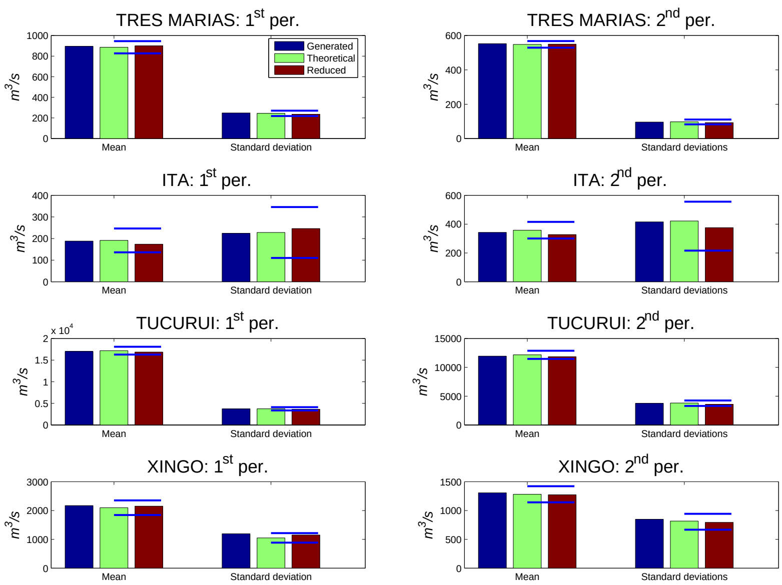

Figure 5. LOR Algorithm: Mean and Standard Deviation

responding statistics are in the acceptability ranges, represented by the error bars in the figures. In relation to the theoretical values, we conclude that the statistics of the reduced trees obtained with both of our techniques are very satisfactory.

Figures 6 and 7 show the linear regression between the means and standard deviations of the initial and reduced trees for all the hydro-plants in the configuration. It is expected that both the slope of each linear regression line, and the determination coefficient R-square, R^2 (measuring the fit of the linear regression line, [6]), take values close to 1.

Since the SOR technique is based on the minimization of the differences between the expected values (recall (4)), the excellent results observed in the figures were

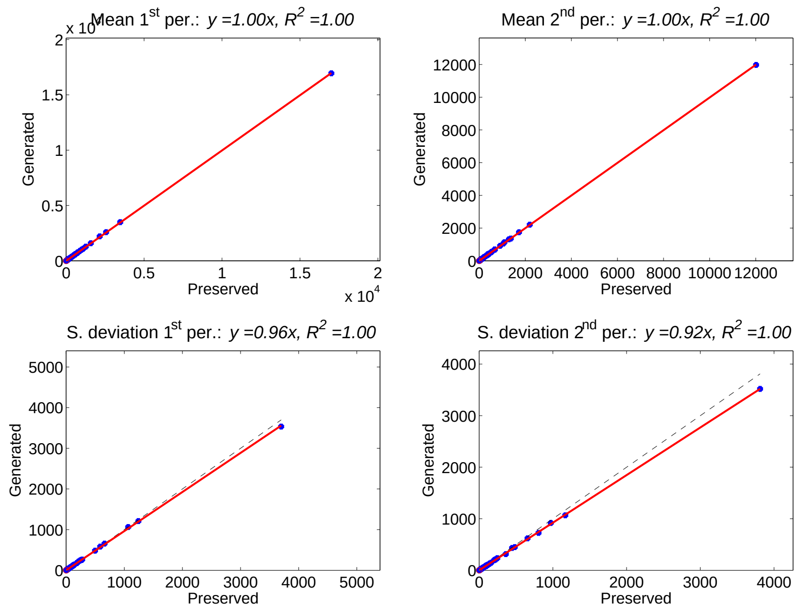

Figure 6. GOR - Linear Regressions: Mean and Standard Deviation

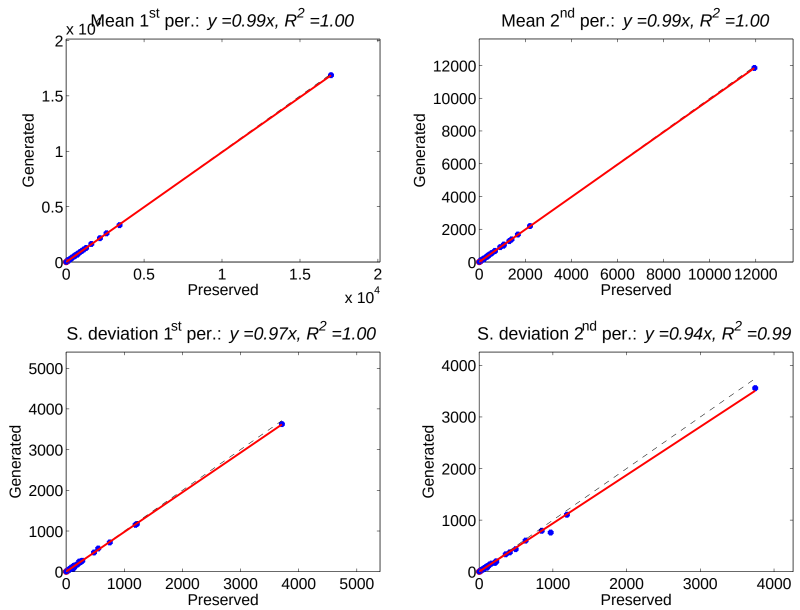

Figure 7. LOR - Linear Regressions: Mean and Standard Deviation

to be expected for the statistical mean. Moreover, the function $\breve{d}$ defined in (12) handles satisfactorily the variance for the preserved scenarios. Indeed, as shown in Figures 6, 7, and the right hand side graphs on Figures 4 and 5, standard deviations are well preserved in both time steps.

In spite of the fact that the function $\breve{d}$ does not satisfy the triangular inequality, our numerical results show a good performance on variance preservation. This fact can be justified by the ratio $\frac{V[Z_t|_-]}{S_t^2}$ scaling the pseudonorm. Indeed, depending on the generated scenarios variance, S_t^2, such ratio can be bigger or smaller than 1. In our multivariate application, it is desirable not to deviate too much from the variance target (theoretical variance $V[Z_t|_-]$). Thus, if $S_t^2 < V[Z_t|_-]$, Algorithm 1 avoids discarding all of the most extreme scenarios, keeping in this way, the initial scenarios variability.

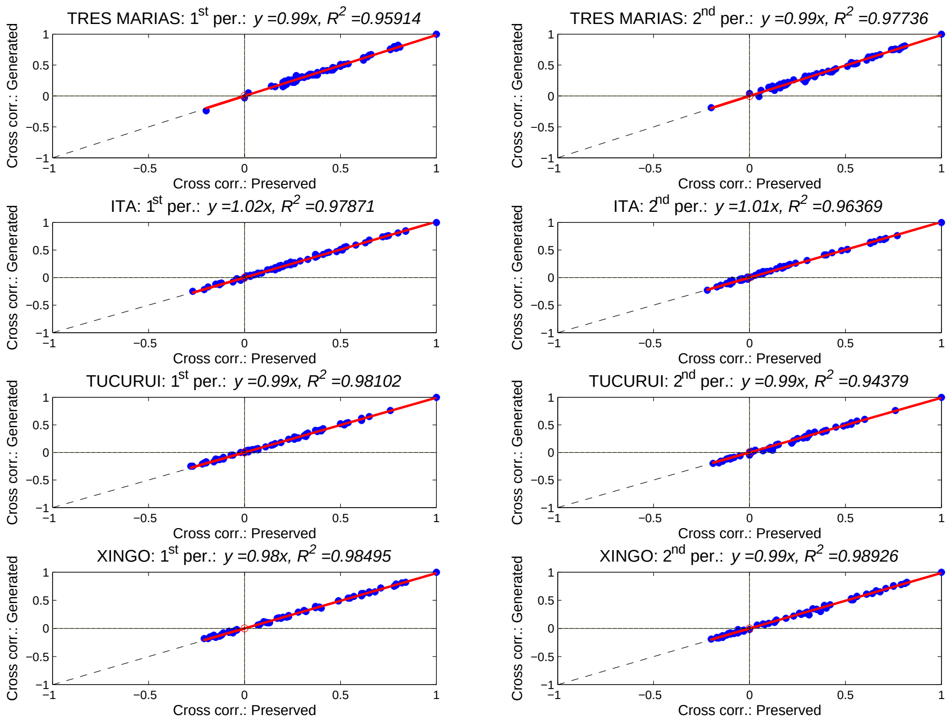

Figure 8. GOR: Cross Correlations

Cross correlations determine the hydrological dependency between different hydroelectric plants. For example, hydro-plants in the same basin are positively correlated.

Figures 8 and 9 show the cross correlations between streamflows of Três Marias, Itá, Tucuruí and Xingó hydroelectric plants, with the others hydroelectric plants in the same configuration. For this study, the number of scenarios preserved by LOR represents only a 24% of the total number of generated scenarios for the first month and 1.6% for the second one. However, as shown in Figure 9, such a small percentage of scenarios defining the reduce tree still manages to reproduce efficiently the desired cross correlations.

Finally, Table 1 reports the values related to critical Kolmogorov-Smirnov [14] and Cramér-von Mises tests [16] for each plant distribution. These tests measure the adhesion between two probability distributions. If the computed value is less than the critical value, the distributions of the big and small trees are considered adherent.

Table 1: Goodness of Fit Test - GOR

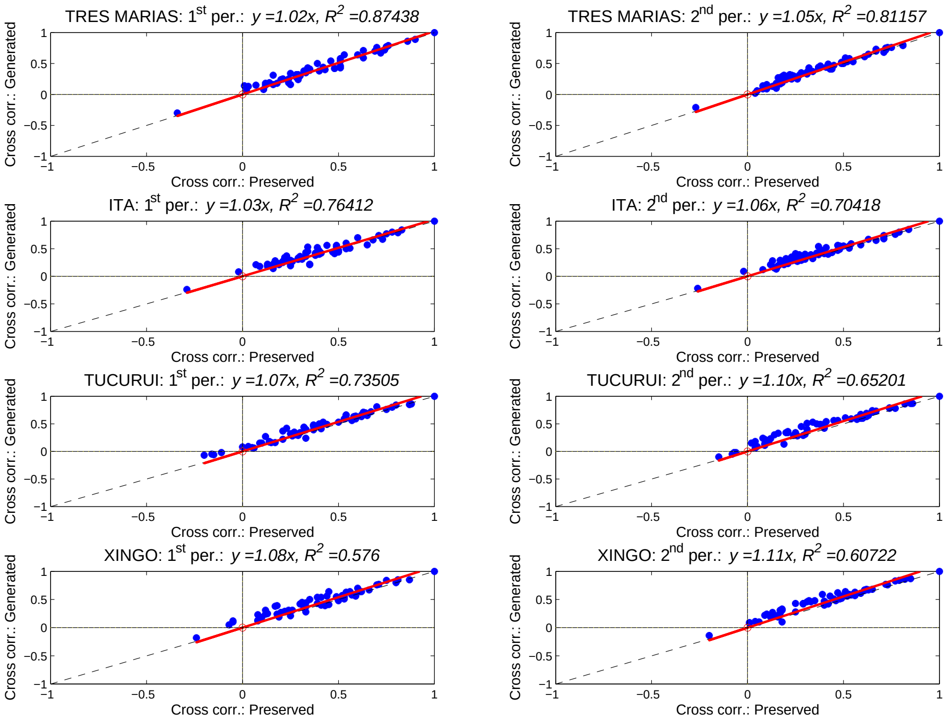

Figure 9. LOR: Cross Correlations

Adhesion Tests	H. Plant	Critical Value	1^{st} per.	2^{nd} per.
Kolmogorov-Smirnov	Três Marias		1.039	1.006
	Itá	95% ⇒ 1.628	1.411	1.517
	Tucuruí	99% ⇒ 1.358	0.921	0.779
	Xingó		0.725	0.964
Cramér-von Mises	Três Marias		0.112	0.148
	Itá	95% ⇒ 0.461	0.483	0.510
	Tucuruí	99% ⇒ 0.743	0.089	0.111
	Xingó		0.117	0.176

Since the LOR algorithm applies the SOR technique at each branch of the tree[1] the Kolmogorov-Smirnov and Cramér-von Mises tests were applied at each branch. Tables 2 and 3 report on the computed values for each time period.

Table 2: Goodness of Fit Test - LOR 1^{st} period

Adhesion Tests	H. Plant	Critical Value	1^{st} per.
Kolmogorov-Smirnov	Três Marias		0.843
	Itá	95% ⇒ 1.628	0.764
	Tucuruí	99% ⇒ 1.358	0.588
	Xingó		0.901
Cramér-von Mises	Três Marias		0.194
	Itá	95% ⇒ 0.461	0.089
	Tucuruí	99% ⇒ 0.743	0.053
	Xingó		0.113

Table 3: Goodness of Fit Test - LOR 2^{nd} per.

[1]Different branches come from different pasts.

	Kolmogorov-Smirnov		Cramér-von Mises	
Critical Value	95%	99%	95%	99%
Três Marias	100%	100%	100%	100%
Itá	94.1%	99.1%	91.6%	95.8%
Tucuruí	98.3%	99.1%	98.3%	99.1%
Xingó	96.6%	100%	96.6%	99.1%

The values reported in Tables 1-3 once more confirm the efficiency of our modified SOR techniques: all computed values are below the acceptance bounds. This result ensures adherence between the probability distributions P and Q.

6. Stability considerations

In the multistage case ($T > 2$), the Lipschitz-like property (5) is harder to be fulfilled and a rather involved analysis needs to be put in place to keep the optimal values and solution of the reduced problem (2) close to the original ones, obtained with (1). For example, stability results in [9] use, in addition to the Fortet-Mourier distance D_J that controls the probability approximation, a so-called *filtration distance* that controls the approximation on the conditional probability distributions.

Our GOR technique essentially works as a 2-stage approach, and as such does not need any additional consideration to ensure stability. By contrast, the LOR technique is a multistage strategy that at first sight only uses the Fortet-Mourier distance without introducing any filtration distance. We now discuss this issue, and show how the autoregressive structure together with the fact that we do not aggregate nodes, helps in ensuring stability. Basically, we apply the results in [10], establishing conditions for the selection mechanism to be continuous and to guarantee stability of the approximation.

We do not enter into too much technicalities here, for a thorough study of stability of multistage stochastic programs, we refer to [9, 10, 15] and references therein.

6.1. *Filtration distance and aggregation of nodes*

Given the initial problem (1), corresponding to a scenario tree ω, and the reduced problem (2), corresponding to a reduced tree $\tilde{\omega}$, their ℓ_∞-*filtration distance* is given by the expression

$$D_{\mathcal{F}}(\omega, \tilde{\omega}) = \sup_{|x|_\infty \le 1} \left\{ \sum_{t=2}^{T} \| E[x_t|\mathcal{F}_t(\omega)] - E[x_t|\mathcal{F}_t(\tilde{\omega})] \| \right\},$$

where $\mathcal{F}_t(\omega)$ is the σ-algebra generated by all the nodes composing the tree ω, from time step 1 to time step t and the expected values $E[x_t|\mathcal{F}_t(\omega)]$ are conditioned to the nodes probabilities at time step t.

Along the lines of (3), for the multistage setting it is shown in [8] that the initial and reduced optimal values, denoted by $v^*(\omega)$ and $v^*(\tilde{\omega})$ respectively, satisfy the relation

$$|v^*(\omega) - v^*(\tilde{\omega})| \le L\Big(D_J(P, \tilde{P}) + D_{\mathcal{F}}(\omega, \tilde{\omega})\Big),$$

where L is a problem-dependent constant, and $D_J(P, \tilde{P})$ is the Fortet-Mourier distance between the corresponding probabilities P and $\tilde{P}$. This result shows that, while proximity between probabilities is controlled by the Fortet-Mourier distance, the filtration distance has the role of keeping close the corresponding expected decisions at each time step t.

Similarly to our LOR approach, the original SOR technique [5] proceeds in a stage-wise manner. An important difference is that in [5] after reduction, nodes are **aggregated**. As shown in the simple example in Figure 10, the initial tree ω, with nodes B and C at time step 2, is reduced with the original SOR technique to the tree $\tilde{\omega}$, with only node C at time step 2.

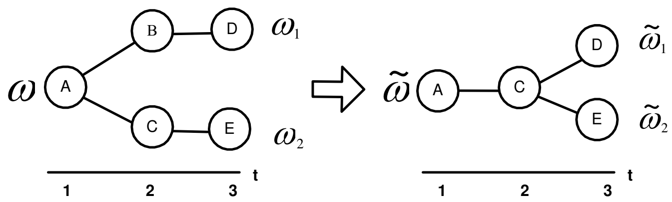

Figure 10. Nodes aggregation in the original SOR technique.

In terms of optimal decisions, the aggregation of nodes has a negative impact for the reduced tree and makes it necessary to introduce the filtration distance. Once more referring to the example in Figure 10, while in the initial tree the decision at node D depends on node B, after aggregation, in the reduced tree node D's decision will depend on node C instead:

$$x_3^D(w^1) \text{ depends on } x_2^B(w^1) \quad \text{but} \quad x_3^D(\tilde{w}^1) \text{ depends on } x_2^C(\tilde{w}^1)\,.$$

Additionally, even when node B is almost identical to node C, there is no reason for the conditional distribution $P_{[\cdot|w_{[2]}]}$ to be close to $\tilde{P}_{[\cdot|\tilde{w}_{[2]}]}$ unless the conditional distributions are continuous. Under such circumstances, using the filtration distance is fundamental to keep stability; see [9, Example 2.6] and [15, Section 2.1].

The whole purpose of the filtration distance is to force the reduction process to identify nodes that are not "too different" in terms of future optimal decisions (noting that, since such optimal decisions are not known, an upper bound needs to be used). Once this identification is done, the original SOR technique, aggregates such nodes. By contrast, neither of our approaches aggregates nodes, because aggregation would destroy the desired stage-wise dependent structure of the periodic autoregressive model.

6.2. *Stability without introducing filtration distance*

Instead of incorporating the filtration distance in the selection process, the work [10] considers a dynamic programming reformulation of multistage stochastic linear programming problems and studies conditions for the recourse functions to depend continuously on the current state. As already mentioned in [11], in this respect it is necessary that conditional distributions themselves are continuous on the state variables. We now detail further the formulation of our problem (1) to show that all relevant conditions in [10], specifically Assumptions 2.1, 2.3, and 2.6 therein are satisfied.

Given some $r \geq 1$, the feasible set in (1) has the form

$$X = X(\omega) := \{x \in L_r(\Theta, \mathcal{F}, P;\ {}^{n}_{+}) : x_1 \in X_1(w_1), x_t \in X_t(x_{t-1}, w_t)\}\,,$$

where each t^{th}-stage feasible set, given by

$$\begin{array}{lll} X_1(w_1) & := \{x_1 \in C_1 \subset {}^{n_1}_{+}; W_1 x_1 = b_1\} & \text{if } t = 1, \text{ and} \\ X_t(x_{t-1}, w_t) & := \{x_t \in C_t \subset {}^{n_t}_{+}; W_t x_t = b_t(w_t) - T_t x_{t-1}\} & \text{if } t = 2, \dots, T, \end{array}$$

is nonempty and compact for each partial scenario $w^i_{[t]} \in \Theta_{[t]}$. All sets C_t are compact and independent of the stochastic process ω.

The cost function $f : \Theta \times {}^n \to {}_+$ is given by

$$f(w^i, x) := \sum_{t=1}^{T} c_t(w^i_t)^\top x_t.$$

In our generation planning application, c_t represents the operational (and deficit) cost at time step $t = 1, \dots, T$. Since $x_t \geq 0$ a.e.w., the function $f(\cdot, \cdot)$ is nonnegative and Assumption 2.1 in [10] is trivially satisfied. In fact, in our problem all feasible sets are compact polyhedrons in some nonnegative orthant and the stochastic process is given by a periodic autoregressive model. Moreover, there is complete fixed recourse, see [1, Chapter 3].

Problem (1) can be rewritten by using a dynamic programming formulation, as follows. Letting $\alpha_t : {}^{n_t}_{+} \times \Theta_{[t]} \to {}_+$ be the recourse function at stage t, the dynamic programming recursion giving $\alpha_1(x_0, w_{[0]})$, the optimal value of (1), starts by setting $\alpha_T \equiv 0$ and defines

$$\alpha_t(x_{t-1}, w_{[t]}) := \min_{x_t \in X_t(x_{t-1}, w_t)} c_t(w^i_t)^\top x_t + E_{[P|w_{[t]}]} \alpha_{t+1}(x_t, w_{[t+1]}) \quad \text{for } t = T, \cdots, 1,$$

where $E_{[P|w_{[t]}]}$ is the conditional expectation relative to $\mathcal{F}(w_{[t]})$. Recourse functions are finite in the presence of complete recourse, and they are well defined and measurable because [10, Assumption 2.1] holds.

Assumption 2.3 in [10] imposes a linear growth condition on the decisions x_t. Since all of our feasible sets are bounded, this condition trivially holds in our application. As a result, by Proposition 2.5 in [10], the mapping $x_{t-1} \mapsto \alpha_t(x_{t-1}, w_{[t]})$ is Lipschitz continuous. Lipschitz continuity of the mapping $w_{[t]} \mapsto \alpha_t(x_{t-1}, w_{[t]})$ is more involved because, unlike the decision variable x_{t-1}, a realization w^i_t impacts the expectation of future realizations of $w_{[\tau]}$, for $\tau > t$. For this reason, the following assumption on continuity of the conditional distributions is required.

Lipschitz Continuity of conditional probability distribution[10, Assumption 2.6]. *There exist constants $W, K > 0$ and $r > 0$, such that with $m_t := 1 + (T-1)(1+r)$ for $t = 1, \dots, T$, the following conditions hold.*

i. For every $t = 1, \dots, T-1$, every Borel set $A_{[t]} \subset \Theta_{[t]}$ with $P_{[|\omega_{[t]}]}[A_{[t]}] = 1$, and P-almost everywhere $w_{[t]}, \hat{w}_{[t]} \in \Theta_{[t]}$

$$D(P_{[|w_{[t]}]}, P_{[|\hat{w}_{[t]}]}) \leq K \max\{1, ||w_{[t]}||, ||\hat{w}_{[t]}||\}^{m_t - 1} ||w_{[t]} - \hat{w}_{[t]}||.$$

ii. For every $t = 1, \dots, T-1$ and P-almost everywhere $w_{[t]} \in \Theta_{[t]}$

$$E_{P_{[|w_{[t]}]}}[\max\{1, ||w_{[T]}||\}^{1+T-t}] \leq W \max\{1, ||w_{[t]}||\}^{m_t}.$$

The function $D(\cdot, \cdot)$ is the Fortet-Mourier metric defined in (7), written with $P = P_{[|w_{[t]}]}$, $Q = P_{[|\hat{w}_{[t]}]}$ and the set J omitted. We mention that the same Lipschitz continuity property, in a slightly different context, is required in [15, Assumption

2] for determining the quality of tree approximations. As discussed in [10, Remark 2.7] and [15], Assumption 2.6.*i* is equivalent to requiring the ergodic coefficient

$$\sup_{w_{[t]},\hat{w}_{[t]}} \frac{D(P_{[|w_{[t]}]}, P_{[|\hat{w}_{[t]}]})}{\max\{1, ||w_{[t]}||, ||\hat{w}_{[t]}||\}^{m_t-1}||w_{[t]} - \hat{w}_{[t]}||}$$

to be bounded above.

By [10, Lemma A.1], see also [10, Example A.2], Assumption 2.6 above holds for autoregressive stochastic processes. Hence, by [10, Theorem 1], in our planning problem, the recourse function α_t is Lipschitz continuous on both arguments. Furthermore, as pointed out in [10, Remark 2.8], our boundedness assumption implies that the mapping $(x_{t-1}, w_{[t]}) \mapsto \alpha_t(x_{t-1}, w_{[t]})$ is uniformly Lipschitz continuous.

The next step to state the stability result in [10, Theorem 3] consists in formalizing the selection and reduction of scenarios as an *approximation* that is nonanticipative:

Definition 6.1 **(Tree approximation)** *A stochastic process $\tilde{\omega} \in (\Theta, \mathcal{F}, P)$ is called an* approximation *of ω, if there exist Borel-measurable mappings such that*

$$s_t : \Theta_{[t]} \to \Theta_t, \; \textit{for } t = 1, \ldots, T,$$

satisfying the following conditions:

i. $\tilde{\omega}_t = s_t(\omega_{[t]})$ *for* $t = 1, \ldots, T$,
ii. $s_{[T]}(\Theta) \subset \Theta$,
iii. $s_1(w_1) = w_1$ *for every* $w_1 \in \Theta_1$, *and*
iv. $s_{[T]}(\omega) \in L^p(\Theta, \mathcal{F}, P)$ *for every* $p \in [1, \infty)$.

In order to see that our LOR technique provides a tree approximation satisfying Definition 6.1, we need to define the selection mappings $s_t(\cdot)$. To this aim, we consider an equivalent formulation of LOR, that starts with an initial huge tree, and moves forward, from stage $t = 1$ up to stage $t = T$, by eliminating from the big tree all descendants of discarded scenarios at stage t. From a theoretical point of view, such rule is equivalent to LOR, because in LOR the fast forward selection of scenarios, yielding a partial reduced tree $\tilde{w}_{[t]}$, is done by using only information available at stage t. The corresponding mapping $\omega_t \mapsto s_t(\omega_t)$ is measurable because the filtration $\mathcal{F}(\tilde{\omega}_{[t]})$ is measurable with respect to $\mathcal{F}(\omega_{[t]})$. Hence, item *i*. above holds, and likewise for *ii*, because all scenarios of a reduced tree $\tilde{\omega}(= \tilde{\omega}_{[T]})$ are extracted from the original huge tree ω. We mention in passing that condition *ii* is not satisfied if there is node aggregation as in the rule proposed in [9]. This phenomenon, that can be seen in Figure 10, makes it necessary to introduce a filtration distance for ensuring stability. Item *iii* holds for LOR because our technique always keeps the root node. Finally, the integrability condition *iv* is also satisfied, because the measurable function s_t is defined by solving a mass transportation problem (9) (cf. Theorem 2.1), depending on the finite distance between scenarios, and considering two measurable probability distributions. Since the same process is applied a finite number of times (for each time stage), the reduced tree obtained by LOR satisfies item *iv*.

As a result, assumptions 2.1, 2.3 and 2.6 in [10] are satisfied, with the LOR tree $\tilde{\omega}$ being an approximation of the tree ω, so [10, Theorem 3] applies:

$$|v^*(\omega) - v^*(\tilde{\omega})| \leq \gamma E_P[\max\{1, ||\omega||, ||\hat{\omega}||\}^{m_1}||\omega - \hat{\omega}||],$$

where $\gamma > 0$ is a problem-depend constant. Closeness of the respective solutions follows from the inequality below, resulting from Assumptions 2.1 and 2.3 in [10]:

$$||x_t(\omega_{[t]})|| \leq L^t \max\{1, ||\omega_{[t]}||\}^{t-1} ,$$

and recalling that w and $\tilde{w}$ are "close".

Concluding remarks

The high cardinality of information needed to represent accurately uncertainty in streamflows for the mid-term planning operation problem makes its solution computationally impracticable for the BPS. In order to preserve the autoregressive structure of the model representing uncertainty, we proposed two modifications of previously defined scenario reduction techniques.

As shown by our results, our modified SOR algorithms are effective tools to build scenario trees that are sufficiently representative, while preserving a good statistical adhesion between the big and small scenario trees. The use of the Fortet-Mourier metric to define the SOR problem and select a subgroup of scenarios from a big tree with many scenarios significantly simplifies the optimization problem, and reduces dramatically CPU times. For example, to build a reduced tree with 960 scenarios, from an initial tree with 60000 scenarios (for each one of the 111 hydro-electric plants), LOR algorithm spent 14 minutes on a Pentium 4 3.00 GHz computer. For comparison, GOR algorithm needed 62 minutes to reduce 6000 scenarios to 960.

Our numerical experience shows satisfactorily low levels for the error in means, standard deviations, and cross correlations.

As a final remark, our discussion in Section 6 shows that for periodic autoregressive stationary processes with a large number of scenarios (as in our application), a stage-wise process combining of a local reduction technique (without aggregating nodes) with re-sampling for the next stage preserves stability.

Acknowledgment The authors wish to thank the reviewers for the careful reading and for offering interesting suggestions and comments. We are particularly grateful for pointing out the continuity property of conditional distributions in Section 6.2, crucial for showing stability for the LOR technique.

References

[1] J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer Science, New York (1997).
[2] G.E.P. Box, G.M. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting & Control*, 3rd ed., Prentice Hall, San Francisco (1994).
[3] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed., The MIT Press, Cambridge (2001), pp. 1033-1038.
[4] J. Dupacova, N. Gröwe-Kuska, and W. Römisch, *Scenario reduction in stochastic programming: An approach using probability metrics*, Mathematical Programming 95 (2003), pp. 493–511.
[5] N. Gröwe-Kuska, H. Heitsch, and W. Römisch, *Scenario reduction and scenario tree construction for power management problems*, IEEE Bologna Power Tech Proceedings (A. Borghetti, C.A. Nucci, M. Paolone eds.) (2003).
[6] D.N. Gujarati, *Basic Econometrics*, 3rd ed., McGraw-Hill Companies (1987).
[7] H. Heitsch and W. Römisch, *Scenario reduction algorithms in stochastic programming*, Computational Optimization and Applications 24 (2003), pp. 187–206.
[8] ———, *Scenario tree reduction for multistage stochastic programs*, Computational Management Science 6 (2009), pp. 117–133.
[9] H. Heitsch, W. Römisch, and C. Strugarek, *Stability of multistage stochastic programs*, SIAM Journal on Optimization 17 (2006), pp. 511–525.
[10] C. Küchler, *On stability of multistage stochastic programs*, SIAM Journal on Optimization 19 (2008), pp. 952–968.

[11] D. Kuhn, *Generalized Bounds for Convex Multistage Stochastic Programs*, 1st ed., Lecture Notes in Economics and Mathematical Systems, Springer, Berlin (2005).
[12] M.E.P. Maceira and C.V. Bezerra, *Stochastic Streamflow model for Hydroelectric Systems*, *In:* Proceedings of 5th International Conference on Probabilistic Methods Applied to Power Systems (1997), Vancouver, Canada, pp. 305-310.
[13] M.E.P. Maceira, L.A. Terry, J.M. Damazio, F.S. Costa, and A.C.G. Melo, *Chain of Models for Setting the Energy Dispatch and Spot Price in the Brazilian System, In:* Power System Computation Conference - PSCC'02 (2002) Sevilla, Spain.
[14] F.J. Massey, *The kolmogorov-smirnov test for goodness of fit*, Journal of the American Statistical Association 46 (1956), pp. 68–77.
[15] R. Mirkov and G.C. Pflug, *Tree approximations of dynamic stochastic programs*, SIAM Journal on Optimization 18 (2007), pp. 1082–1105.
[16] S. Seigel, *Nonparametric statistics for the behavioral sciences*, McGraw-Hill (New York) (1956).