Quantifying the impact of scenario tree generation and reduction methods on the solution of the short-term hydroscheduling problem

Maissa Daadaa[1,2*], Sara Séguin[1,2], Miguel F. Anjos[2,3] and Kenjy Demeester[4]

[1*] Université de Québec à Chicoutimi, G7H 2B1, Canada.
[2]GERAD, Montréal (Qc), H3T 1J4, Canada.
[3] University of Edinburgh, EH8 9YL,United Kingdom.
[4]Rio Tinto, Saguenay (Qc), G7S 4R5, Canada.

*Corresponding author(s). E-mail(s): maissa.daadaa1@uqac.ca;
Contributing authors: sara.seguin@uqac.ca;
anjos@stanfordalumni.org; kenjy.demeester@riotinto.com;

Abstract

This paper studies the properties of a stochastic optimization model for the short-term hydropower generation and reduction problem with uncertain inflows. The price of energy is not considered. The uncertainty of the inflows is represented using scenario trees. Backward reduction and neural gas methods are used to generate and reduce a full scenario tree. The objective of this work is to evaluate the impact of scenario tree generation and reduction methods on the solution of the optimization. First, statistical tests are done where the expected volume, the variance and the standard deviation of each scenario tree are calculated and compared. Second, operational tests are realized, where the scenario trees are used as input to the stochastic programming model and the value of the objective function and solution are evaluated and compared. The model are tested on a 14 forecasted days and for a 10 days rolling-horizon for two powerhouses with five turbines each located in the Saguenay-Lac-St-Jean region of the province of Québec in Canada.

Keywords: Stochastic optimization, scenario tree generation, backward reduction, neural gas.

Acknowledgement

This work was supported by a Mitacs-Rio Tinto Accelerate Internship Award.

Nomenclature

Sets	
$i \in \{1, 2, .., S\}$	indexes of the set of scenarios.
$n \in \{1, 2, .., N_i\}$	indexes of the set of nodes for each scenario i.
$c \in \{1, 2, .., C\}$	indexes of the set of powerhouses.
$l \in \{1, 2, .., U^c\}$	indexes of the set of powerhouses upstream of each powerhouse c.
$j \in \{1, 2, .., J_n^c\}$	indexes of the set of turbines associated to the node n and powerhouse c.
$b \in \{1, 2, .., B^c\}$	indexes of the set of combinations of each powerhouse c.
$k \in \{1, 2, .., K_c^b\}$	indexes of the set of efficiency points associated to powerhouse c and combination b.
Parameters	
$P_{k,n}^c$	power output of powerhouse c at node n and point k (MW).
$q_{k,n}^c$	water discharge of powerhouse c at node n and point k (m^3/s).
π_i^c	probability of scenario i for powerhouse c.
ξ_n^c	inflow of powerhouse c at node n (m^3/s).
β	conversion factor from (m^3/s) to (hm^3/h).
θ^c	estimated energy losses from maximum storage (MW) at powerhouse c.
ϵ^c	start-up penalty of turbine (MW) at powerhouse c.
N_{max}^c	maximum number of start-ups for powerhouse c.
V_{max}^c	maximum volume of reservoir c (hm^3).
v_{ini}^c	initial volume of reservoir c (hm^3).
v_{final}^c	final volume of reservoir c (hm^3).
Δt	the duration of the stage (h).
$A_{n,k,j}^c = \begin{cases} 1 \text{ if the turbine } j \text{ of powerhouse } c \text{ at the point } k \text{ is activated at node } n. \\ 0 \text{ otherwise.} \end{cases}$	
Decision variables	
$y_{k,n}^c = \begin{cases} 1 \text{ if the point } k \text{ is chosen at node } n \text{ for powerhouse } c. \\ 0 \text{ otherwise.} \end{cases}$	
$z_{j,n}^c = \begin{cases} 1 \text{ if the turbine } j \text{ of powerhouse } c \text{ is started at node } n. \\ 0 \text{ otherwise.} \end{cases}$	
v_n^c	volume of the reservoir of powerhouse c at node n (hm^3).
d_n^c	water spillage at powerhouse c and node n (m^3/s).

1 Introduction

In hydropower management, the uncertainty of the inflows is a major issue in the decision-making process. The producers are required to define a schedule that allows them to produce more energy under uncertainty. Stochastic short-term hydropower models are used to solve this problem. The aim of these models is to optimally dispatch the amount of water available in the reservoirs between the turbines by considering uncertain inflows. Therefore, the amount of the water discharge, the turbines in operation and the volume of the reservoirs are defined. Different research papers have looked into stochastic short-term hydropower models. Usually, the short-term optimization model is considered as deterministic [8, 11, 21]. An overview on mathematical programming approaches for the deterministic short-term hydropower problem

is presented in [23]. In [7], the authors assume that the inflows and the prices are known as previously forecasted. All the cited works are based on the fact that there is no new information arriving over time, and decisions are made in advance for the whole planning horizon. In contrast, in stochastic models, new information about uncertain data arrives as time evolves along the planning horizon which allows the producers to plan and operate hydropower in an optimal way [24]. In this regard, the purpose of this work is to develop a stochastic short-term hydropower optimization model giving more realistic production plan.

To solve stochastic short-term hydropower problems, a multistage stochastic programming model can be used [22, 27]. For such models, the decisions are taken at the beginning of the horizon before knowing the realization of uncertainty. These decisions are adjusted once the uncertainties are known [16]. Usually, uncertainty is represented by scenario trees. Each path of the tree represent a scenario and each scenario has a probability. The nodes in the scenarios represent the values of inflows into the reservoir system at each stage. The scenario tree serve as input to the stochastic programming model. Implementing a scenario tree with a large number of scenarios requires many resources (memory capacity, computational time). Therefore, several approaches for generating and reducing scenario trees are proposed. These approaches define different methods that allow to determine the fewest number of scenarios that can be used in the stochastic programming model and preserved the most information from the initial distribution.

An overview of methods for scenario tree generation is presented in [13]. The moment matching methods are used in [6, 18]. This method aims to match some specific statistical properties of the scenario tree and historical data series. Statistical properties are chosen by the users according to their needs. In [26], the first four moments (variance, mean, kurtosis and skewness) are matched in order to create a discrete distribution of the uncertain parameter. By contrast, in [12], the results show that the moment matching may lead to strange approximations and may not be able to match the target distribution. The Monte Carlo method can also be used to generate the scenario trees [15, 20].

Clustering methods are proposed in several studies in order to recover the structure of the scenario tree by applying a cluster analysis. k-means clustering is used in [5] where the sum of distances from all objects in that cluster is minimized. In [22], the k-means method is used to minimize the nested distance between the stochastic process of historical inflow data and the multistage stochastic process represented in the scenario tree. Four clustering methods (conditional clustering, progressive clustering, node clustering and neural gas) are analyzed and compared in [14]. Conditional clustering is used to build the tree by sampling scenarios from the distribution probability, and

fitting them at the best position in the tree. Node clustering is used to reduce the size of the tree by joining the closest node available and the algorithm stops where the maximum number of the node is achieved. Progressive clustering aims to cluster the series and to take the centroids as the values of the scenarios to represent the series. Neural gas is a soft competitive learning method used to update the value of the nodes in the scenario tree in such a manner that the distance between the scenario tree and the observed series is reduced gradually. The comparison between these methods has shown that the neural gas outperforms other clustering methods when applied to the case of hydro inflows. Neural gas is used in several papers to generate a scenario tree for hydro inflows data [10, 17, 27].

Another method proposed in the literature to generate and to reduce the size of the scenario tree by preserving as much information as possible is the backward reduction method [24]. This method aims at iteratively selecting scenarios to delete from a full scenario tree in a way that the probability distribution distance between the reduced and the full scenario trees is minimized. In [9], backward reduction is used to reduce the size of the scenario tree in order to reduce the computation time in the case of hydropower. This method is considered the state-of-the-art and is implemented in Scenred/GAMS [3]. Many other approaches have used backward reduction on the basis that the statistical information is maintained in the best possible way [10, 25].

All the cited works make progress towards solving the problem. Since several methods are proposed in this paper, we carry out a study of the impact of the choice of the generation and reduction methods on the solution. Two different methods are chosen: neural gas and backward reduction. The choice of the methods is based on two criteria: the performance of the method based on the literature and the ease of implementation. The reduced scenario trees obtained by each method are used as input to the stochastic model. The results are compared to determine if the choice of the generation and reduction method has an impact on the solution of the optimization problem. A multistage stochastic model is developed in order to maximize the energy production. Since hydropower production functions are nonconvex and nonlinear, different approximation and linearization techniques are proposed, and the production function is modelled in many different ways. Therefore, instead of approximating the hydropower production functions and discretizing the water discharge in order to find the best value that maximizes the energy produced, the best values are already known in operational reality. These values can be determined using the efficiency curves of water discharge for each possible combination of active turbines and it is at these points that the maximum of power produced is reached. The model selects one of these points to maximize the energy production. Since, different combinations can be selected from one period to another, a maximum number of turbine changes is imposed and the turbine start-ups are penalized with a fixed cost.

This new approach was proposed and tested on a deterministic model in [8]. The results show that the proposed model allows to produce more energy than the energy obtained from the real operational decisions. Moreover, using the efficiency points decreases the number of parameters and variables and makes the problem easier to solve. The use of the efficiency curves is an innovative method that allows the producer to implement directly the optimized solution since it is obtained on the efficiency points and therefore agrees with how the engineers want to operate the powerhouses. Since our objective is to find a solution that reflects the operational reality, this deterministic model is adopted and updated to consider uncertain inflows.

The proposed model is tested on a real hydroelectric system owned by Rio Tinto in the Saguenay region of the province of Québec in Canada. In this province, the market is regulated by Hydro-Québec, a public company responsible for the generation, transmission and distribution of electricity. As a result, all producers must buy and sell to them and negotiate fixed price contracts every year. This means that Rio Tinto canno't bid on the spot market and must transact through Hydro-Québec.
Rio Tinto is a company that produces aluminium in the Saguenay region. They own a hydroelectric system that provides 90% of its energy needs. The remaining energy needs are purchased at a known price from Hydro-Québec. Therefore, the primary objective of the optimization is to maximize energy production with the available water in the reservoirs. Hence, it is crucial to consider the uncertainty of inflows in the development of the model. The energy prices are not considered, but the proposed formulation could be extended to consider uncertain prices.

The paper is organized as follows. Section 2 presents the methods used to generate and to reduce a full scenario tree. The methodology and the hydropower optimization model are presented in Section 3. Numerical results are discussed in Section 4 and concluding remarks are presented in Section 5.

2 Scenario tree generation and reduction

In this Section,we present two the methods of scenario tree generation and reduction to study the impact of the choice of the tree reduction method on the solution of the optimization problem.

2.1 Backward reduction

Backward reduction aims to define a sub-scenario tree ξ_i for $i \in \{1,\ldots,S^*\}\backslash I$ with probability π by deleting scenarios from a full scenario tree ξ_s for $s \in \{1,\ldots,S^*\}$ with probability P in a way that the probability distance between the reduced and the full trees is minimized [9]. The problem that has to be solved is to find the set I of deleted scenarios in terms of a given probability distance between P and π. In the context of stochastic power management models [10], the Kantorovich distance $D(P,\pi)$ is used and calculated by Eq.(1)

and Eq.(2):

$$D(P,\pi) = \sum_{s\in I} p_s \min_{i\notin I} C_T(\xi_s,\xi_i) \tag{1}$$

$$C_T(\xi_s,\xi_i) = \sum_{t\in T} \parallel \xi_{s_t} - \xi_{i_t} \parallel \tag{2}$$

where p_s is the probability of the assessed scenario, $I \subset \{1,\ldots,S^*\}$ is the set of deleted scenarios and C_T is the function that measures the distance between two scenarios ξ_s, $s \in \{1,\ldots,S^*\}$ and ξ_i, $i \in \{1,\ldots,S^*\}\backslash I$ on the time horizon $t \in T$. In order for the reader to grasp the link between the tree methodology and the notation for the mathematical model. Each scenario is represented by a node, leading to multiple nodes per period as presented in Fig 1 .
The idea is to compare the Kantorovich distance of the reduced tree and the full one on $\{1,..,t\}$, $t = T, T-1,\ldots,2,1$, and to delete scenarios if the reduced tree is still close enough to the full one with a given accuracy ε_τ as shown in Eq.(3). This accuracy is defined as the reduction percentage. It specifies the desired reduction in terms of the suitably measured distance between the full and the reduced scenario trees. For example, if the reduction percentage is 10%, it means that the distance between the reduced and the full trees is less than 10%, and that, the reduced tree retains 90% of the information contained in the full tree. The probability π_i of the preserved scenario ξ_i, $i \notin I$ is equal to the sum of its former probability p_i and of the probabilities p_i of deleted scenarios that are closest to it, as shown in Eq.(4) and Eq.(5):

$$\sum_{s\in I} p_s \min_{i\notin I} C_T(\xi_s,\xi_i) < \varepsilon_\tau \tag{3}$$

$$\pi_i = p_i + \sum_{s\in I(i)} p_s \;\; \forall i \notin I \tag{4}$$

where

$$I(i) = \{s \in I \; : i = i(s)\}, \; i(s) \in \operatorname*{argmin}_{i\notin I} C_T(\xi_s,\xi_i) \;\; \forall s \in I \tag{5}$$

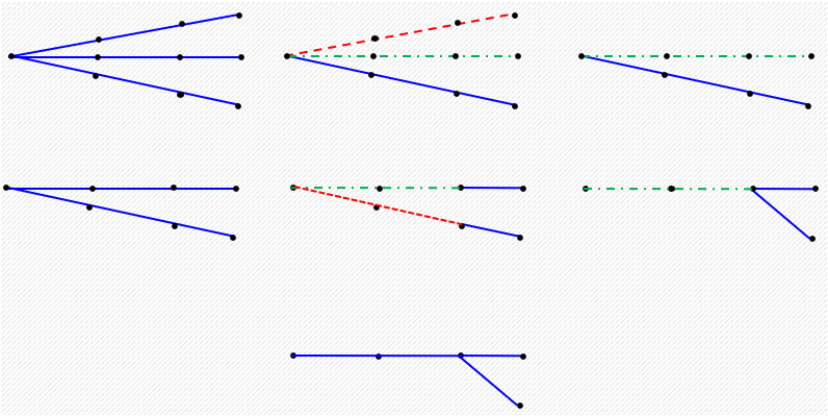

Fig. 1 Backward reduction method [10]

Fig. 1 shows an example of the strategy used to reduce the scenario tree with backward reduction method. At each time horizon t, the distance is calculated (solid line in blue color), one scenario is deleted (dashed line - - in red color) and the probability of this scenario will be added to the preserved one (dashed-dot line -..- in green color).

The advantage of the backward reduction method is that the implementation can be done directly using Scenred2/GAMS [3].

2.2 Neural Gas

Neural gas is a soft competitive learning method used to update the value of the nodes in the scenario tree in such a manner that the distance between the scenario tree and the observed series is reduced gradually [17]. The desired number of scenarios S is defined as a parameter. Neural gas extracts several representative sequences from forecasted streamflow series H as different scenarios to form a streamflow scenario tree ξ_i, $i = 1,\ldots,S$. The neural gas algorithm works as follows:

(1) The nodal value of each scenario in the scenario tree is randomly selected from historical streamflow series H at each stage t, $t \in T$. The selection is done by Eq.(6):

$$\xi_{i_t} = H_{rand(),t} \; \forall i \in S, \; \forall t \in T \tag{6}$$

(2) A new entire series H_w from forecasted streamflow series is randomly selected and the Euclidean distance $d_{i,w}$ between this series and all the scenarios in the tree is calculated by Eq.(7):

$$d_{i,w} = \sum_{t \in T} \parallel H_{w_t} - \xi_{i_t} \parallel \tag{7}$$

(3) The distances are sorted in array D in an ascending order and an array O is generated to record the distance rank of each scenario Eq.(8):

$$O = order\ (D) \tag{8}$$

(4) The values of each node in the scenario tree is updated at each iteration r according to their order in the array O using Eq.(9),(10),(11), and (12):

$$\xi_{i_t}^{r+1} = \xi_{i_t}^{r} + \epsilon(r).h_\lambda(O_i).(H_{w_t} - \xi_{i_t}) \tag{9}$$

$$\epsilon(r) = \epsilon_0.(\epsilon_f/\epsilon_0)^{r/r_{max}} \tag{10}$$

$$h_\lambda(O_i) = e^{-(O_i/\lambda)} \tag{11}$$

$$\lambda = \lambda_0.(\lambda_f/\lambda_0)^{r/r_{max}} \tag{12}$$

where ϵ_0 and ϵ_f are step size parameters. λ_0 and λ_f are the adaptation parameters. O_i is the distance rank of the scenario i. r_{max} is the maximum number of iterations. The value of these parameters can be determined by fine-tuning for each case. In this study, the choice of the values of these parameters is based on the most commonly used values in the literature [14][17][27][19]. Accordingly, we set step size parameters ϵ_0=0.5 and ϵ_f=0.05 and adaptation parameters λ_0=10 and λ_f=0.01. The maximum number of iterations r_{max}=1000.

(5) Steps 2 to 4 are repeated until r_{max} is reached.
(6) The probability of the scenarios π_i is calculated as the proportion of the series randomly chosen whose closest scenario is the scenario i as for Eq.(13):

$$\pi_i = Count\{l' \in [1, L], l' \mid d_{i,l'} = \min_{i' \in [1,S]} (d_{i',l'})\}/L \tag{13}$$

where π_i is the probability of the scenario i and $Count\{.\}$ is a counting function.

2.3 Implementation

2.3.1 Construction

We implemented the Backward reduction and neural gas methods to reduce a full scenario tree (initial tree) with S^* scenarios. The probabilities of the scenarios of the full tree are equal because our aim is to analyze the impact of the choice of the scenario trees generation methods on the decision-making. The backward reduction is used with different reduction percentages. In this work, the reduction percentage are 10%, 20% and 30%. With more than 30% of reduction, the number of scenarios is very small (less than 5 scenario). For this reason, in order to obtain a meaningful results, the maximum reduction percentage is 30%. For each reduction percentage, a reduced scenario tree is obtained using the backward reduction method, as explained in Section 2.1. The number of scenarios of each reduced tree is used to define the desired number of scenarios in the neural gas method as explained in Section 2.2. For example, if the full tree has 10 scenarios, and backward reduction with 10% leads a reduced tree with 7 scenarios, then, the desired number of scenarios in neural gas is set to 7. We recall that the percentage of reduction specifies the desired reduction in terms of the relative distance between the initial and reduced scenario trees.

2.3.2 Tests

The reduced scenario trees obtained from both methods have undergone two tests: statistical and operational tests. In **statistical tests**, the expected volume (E), the variance (V) and the standard deviation (SD) of each scenario

tree are calculated as follows:

$$E = \sum_{i \in S} \pi_i \times \sum_{n \in N_i} \xi_n \tag{14}$$

$$V = \sum_{i \in S} \pi_i \times (\sum_{n \in N_i} \xi_n - E)^2 \tag{15}$$

$$SD = \sqrt{V} \tag{16}$$

where S are the number of the scenarios of each reduced tree, π_i is the probability of each scenario $i \in S$ and ξ_n are the values of the inflows at each node $n \in N_i$.

In **operational tests**, the reduced scenario trees obtained from both backward reduction and neural gas are used as input to the stochastic programming model and the total energy produced is calculated. Theses tests are used to evaluate the impact of each method on the solution.

3 Stochastic multistage mixed integer linear model

In this Section, a stochastic model is developed based on a new formulation that is tested on a deterministic model in [8]. This formulation uses the efficiency points of water discharge at the maximum storage, and it is at these points that the maximum of power produced is reached. Using the efficiency points decreases the number of parameters and variables and makes the problem easier to solve. Therefore, a pair of points of maximum efficiency for water discharge and the power produced are determined as shown in Fig 2. These points corresponding to the maximum efficiency of water discharge and the adjacent points of this maximum with $\pm\ \zeta\ (m^3/s)$ where ζ are integer parameters. The choice of these parameters depends on the management of the powerhouses. A pair of points of maximum efficiency for water discharge and the power produced are determined for each combination of active turbines. Table 1 gives an example of the sixteen possible combinations with 5 available turbines. The minimum number of the active turbines (here 3 turbines) is required due to the physical constraints of the powerhouse.

Table 1 Combinations of 5 available turbines

3 active turbines	4 active turbines	5 active turbines
123 124	1234	12345
125 134	1235	
135 145	1245	
234 235	1345	
245 345	2345	

The objective of the problem is to select a pair of points of maximum efficiency for water discharge and power produced and find the best combinations of active turbines in order to maximize the energy produced. The model can select different combinations from one period to another. However, in practice, it is recommended to have a limited number of start-ups. For that, a maximum number of turbine changes is imposed and the turbine startups are penalized with a fixed cost.

Since the efficiency points are determined at the maximum storage and the reservoir is not always full, a correction of produced power is done. The problem is formulated as a Mixed Integer Linear Programming problem (MILP) to find the best operation point that maximizes total energy production and the turbines in operation while penalizing start-ups. The objective function of the MILP is composed of three terms. The first term computes the power output at each efficiency point for each combination and at each stage. The second term makes a correction between the power produced at the current volume and the maximum storage since the efficiency points are determined at the maximum storage. The third term of the objective function allows to reduce the number of changes by penalizing unit start-ups. This formulation is updated and reformulated as a stochastic multistage mixed integer linear model to consider uncertain inflows.

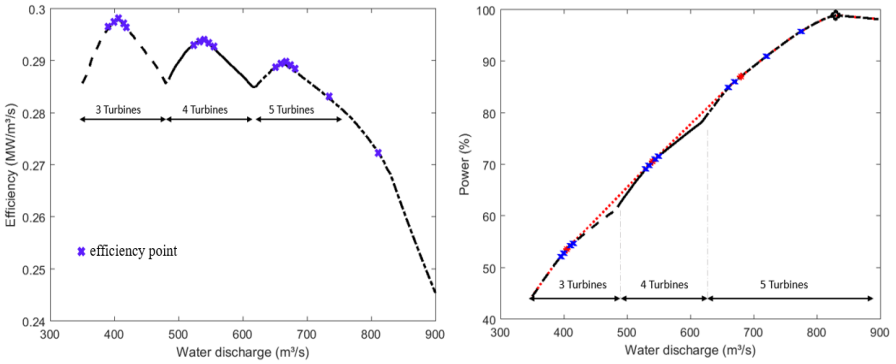

Fig. 2 The efficiency curves

3.1 Mathematical model

In this Section, we present the proposed stochastic multistage mixed integer model. The uncertain inflows are provided from the scenario tree. The variables are $y^c_{k,n}$, which allow the model to select the efficiency point k for a given water discharge and power produced, the volume and the turbines in operation, for each node and powerhouse in the scenario tree. The decision variables are partitioned into stages in a way that decisions made for one stage are not affected by the information for the following stages. The objective is to

maximize energy production in stage 0 and expected energy production in future stages:

$$\max_{y,v,z}[\sum_{c\in C}\sum_{b\in B^c}\sum_{k\in K_b^c} P_{k,0}^c \times \ y_{k,0}^c - \sum_{c\in C}\theta^c \times (V_{max}^c - v_0^c) - \sum_{c\in C}\sum_{j\in J}\varepsilon^c \times z_{j,0}^c +$$

$$\sum_{i\in S}\pi_i^c \times \Delta_t \times [\sum_{c\in C}\sum_{n\in N_i}\sum_{b\in B^c}\sum_{k\in K_b^c} P_{k,n}^c \times \ y_{k,n}^c - \sum_{c\in C}\sum_{n\in N_i}\theta^c \times (V_{max}^c - v_n^c)] -$$

$$\sum_{c\in C}\sum_{n\in N_i}\sum_{j\in J}\varepsilon^c \times z_{j,n}^c] \tag{17}$$

Subject to:

$$v_{n+1}^c = v_n^c + \Delta_t \times [(\xi_n^c \times \beta) - \sum_{b\in B}\sum_{k\in K_b^c}(q_{n,k}^c \times y_{k,n}^c \times \beta) - (d_n^c \times \beta) +$$

$$\sum_{l\in U^c}\sum_{b\in B^c}\sum_{k\in k_b^l}(q_{n,k}^l \times \ y_{k,n}^l \times \beta) + (d_n^l \times \beta)] \quad \forall c \in C, \forall n \in N_i, \forall i \in S \tag{18}$$

$$\sum_{b\in B^c}\sum_{k\in K_b^c} y_{k,n+1}^c \times A_{n+1,k,j}^c - \sum_{b\in B^c}\sum_{k\in K_b^c} y_{k,n}^c \times A_{n,k,j}^c \leq z_{j,n}^c$$

$$\forall c \in C \ , \forall n \in N_i \ , \forall i \in S \ , \forall j \in J \tag{19}$$

$$\sum_{b\in B^c}\sum_{k\in K_b^c} y_{k,n}^c \quad = 1 \qquad \forall c \in C, \forall n \in N_i, \forall i \in S \tag{20}$$

$$\sum_{i\in S}\sum_{n\in N_i}\sum_{j\in J} z_{j,n}^c \leq N_{max}^c \quad \forall c \in C \tag{21}$$

$$v_{min}^c \leq v_n^c \ \leq V_{max}^c \quad \forall c \in C, \forall n \in N_i, \forall i \in S \tag{22}$$

$$v_0^c \ = v_{ini}^c \quad \forall c \in C \tag{23}$$

$$v_{N_i}^c \geq v_{final}^c \quad \forall c \in C, \forall i \in S \tag{24}$$

$$y_{k,n}^c, y_{k,n}^l, z_{j,n}^c \ \in \ \mathcal{B} \quad \forall c \in C, \forall n \in N_i, \forall i \in S, \forall b \in B^c, \forall k \in K_b^c, \forall l \in U^c \tag{25}$$

$$d_n^c, d_n^l, v_n^c \ \in \ \mathcal{R}^+ \quad \forall c \in C, \forall n \in N_i, \forall i \in S, \forall l \in U^c. \tag{26}$$

Constraints (18) ensure water balance at the plants. Constraints (19) are the link between start-up variables and the chosen combination considering the set of points. Constraints (20) force the model to choose only one operating efficiency point at each node for each powerhouse. A maximum number of start-ups N_{max} is imposed with constraints (21). Constraints (22) are the bounds on reservoir volumes, and constraints (23)-(24) specify initial and final volumes. Finally, constraints (25) define the binary variables and (26) the real variables.

3.2 Case study

The proposed model is tested with real data from the Saguenay-Lac-St-Jean hydroelectric system owned by Rio Tinto. For the purpose of this paper, two powerhouses Chute-Du-Diable (CD) and Chute-Savane (CS) are considered. These powerhouses are in series and both have 5 turbines. The planning horizon of the rolling-horizon is about 10 days. For every day of the rolling-horizon, the forecasts of inflows are provided. The distribution of these inflows allows to build a full scenario tree, the forecast is for 14 days. For day 1 of the rolling-horizon, predictions are for days 1 to 14, for day 2 of the rolling-horizon, predictions are for days 2 to 15, and so on. The full scenario trees are reduced using backward reduction and neural gas . The model is updated according to the management of the powerhouses at Rio Tinto. To schedule hydropower production, Rio Tinto aims to obtain hourly decisions. For this reason, the stochastic multistage mixed integer model is updated as follows:

- Stage 0 : the decisions are taken hourly (decisions for hour 1 and decisions for the next hours).
- Following stage : the decisions are taken daily (decisions for everyday).

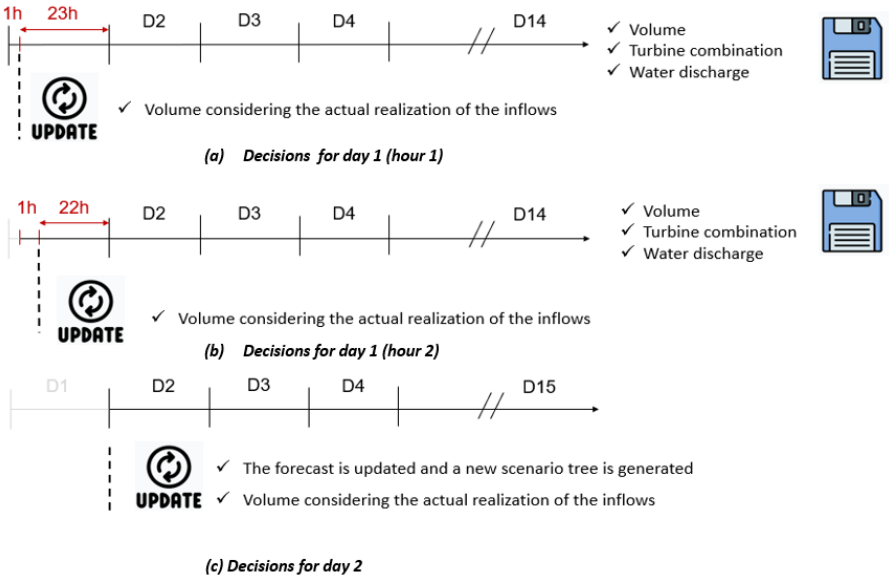

Fig. 3 The methodology according to the management of the powerhouses

For example, for day 1 of the rolling-horizon:

- Stage 0 : the decisions are taken for hour 1 and decisions for the next 23 hours as shown in Fig 3-a.
- Following stage : the decisions are taken for day 2 to 14, one stage per day.

Only the solutions for the volume, the water discharge, the produced power and the combination of active turbines for the first hour of the first-stage are retained. After that, the volume of the reservoir is updated considering the actual realization of the inflow as shown in Fig 3-a. Thereafter, the decisions are taken as follows:

- Stage 0 : the decisions are taken for hour 1 and decisions for the next 22 hours as shown in Fig 3-b
- Following stage : the decisions are taken for day 2 to 14, one stage per day.

The same process is repeated until the end of the day. At the beginning of the day 2, the forecast is updated. The scenario tree is generated for the corresponding day as shown in Fig 3-c and new decisions are taken:

- Stage 0 : the decisions are taken for hour 1 and decisions for the next 23 hours.
- Following stage : the decisions are taken for day 3 to 15, one stage per day.

The process is repeated during the planning horizon. Therefore, the objective is to maximize the total energy production over the whole rolling-horizon, by evaluating the energy production for the first hour of the first stage of each day in the rolling-horizon and penalizing turbine startups.

4 Numerical results

This Section details the results on which two methods of generation and reduction are compared. First, the construction of the reduced scenario trees for each method is done. For backward reduction method, the scenario trees are obtained using Scenred2/GAMS [3], and for the neural gas, the scenario trees are obtaining by solving the algorithm using Python [4]. Second, two tests are made, statistical and operational tests. For statistical tests, the variance, the expected volume and the standard deviation of each scenario tree are calculated to determine the effect of the choice of the reduction method on the preservation of the mathematical aspects. For the operational tests the produced energy obtained from each scenario tree is calculated by the stochastic multistage mixed integer linear model. The formulation is solved using the servers of compute Canada [2] and Xpress solver accessed via Python [1]. The model is tested using 4 data sets (July, September, October and December) from the year 2021, for two powerhouses from a real world system. Finally the results obtained from backward and gas neural methods are compared in order to define the impact of the choice of the method on the objective function and solution.

4.1 Reduction of the scenario tree

For every day of the rolling-horizon, the forecasts of inflows are provided. The distribution of these inflows allows to build a full scenario tree with N scenarios

and the reduced scenario trees are determined on a daily basis, the forecast is for 14 days.

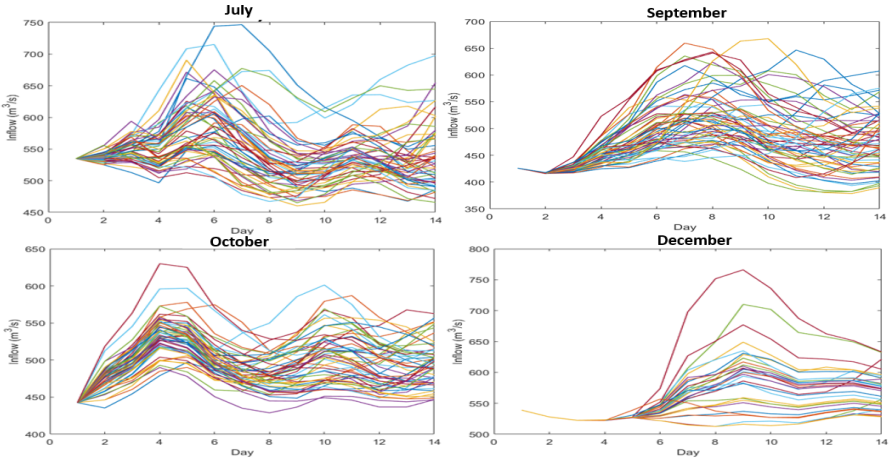

Fig. 4 The full scenario trees for the first day in the rolling horizon for July, September, October and December

Fig 4 shows an example of the full scenario trees obtained on the first day for the 4 test cases.For example, for July, a full scenario tree with 57 scenarios is obtained, for September 57 scenarios, for October 56 scenarios and for December 38 scenarios for the powerhouse CD. These full scenario trees are reduced using backward and neural gas. For example, for October, with 10% of reduction, the number of scenarios is 31, with 20% of reduction the number of scenarios is 17 and with 30% of reduction the number of scenarios is 10. Recall that the percentage of reduction specifies the desired reduction in terms of the distance between the initial and reduced scenario trees.

Fig 5 and Fig 6 illustrate the difference between the full scenario tree and the reduced scenario trees with 10%, 20% and 30% of reduction for October with the two methods. The final number of the scenarios of both methods is the same, but the structure of the scenario tree is different. As explained in Section 2.1, backward reduction aims to delete scenarios from a full scenario tree in a way that the probability distribution distance between the reduced and the full scenario trees is minimized. Neural gas aim to update the value of the nodes in the scenario tree in such a manner that the distance between the scenario tree and the observed series is reduced gradually. For this reason, the structure of the scenario trees are different. After determining the reduced scenario trees for the 4 test cases (July, September, October and December), statistical tests are made for each scenario tree to determine if the choice of the method has an impact on the characteristic of the scenario tree (the expected volume, the variance and the standard deviation). Moreover, operational tests are made to define the impact of each method on the solution.

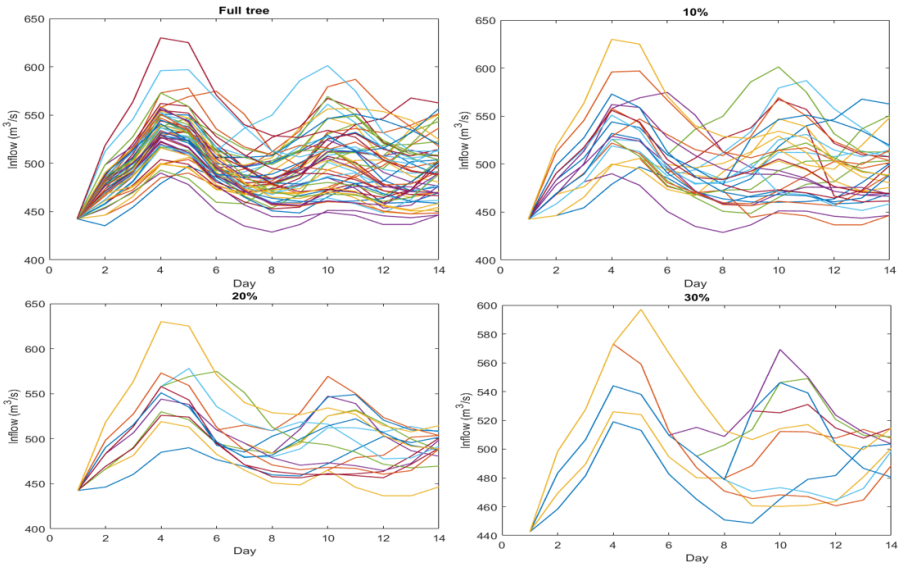

Fig. 5 An example of the full and the reduced scenario trees with 10%, 20% and 30% of reduction with backward reduction

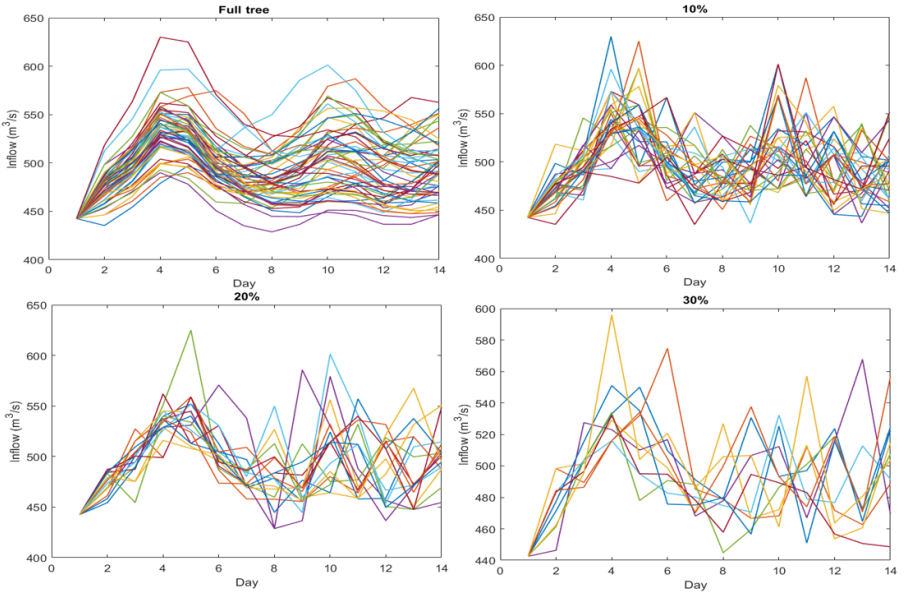

Fig. 6 An example of the full and the reduced scenario trees with 10%, 20% and 30% of reduction with neural gas

4.2 Statistical tests

In this Section, the expected volume, the variance and standard deviation of each scenario tree are calculated during the 10 days of the rolling-horizon for the 4 test cases as explained in Section 2.3.2 in order to determine the effect of the choice of the reduction method on the preservation of the mathematical aspects. The results obtained from the full scenario tree and the reduced scenario trees of each method are compared for the 4 test cases for the powerhouses CD and CS. The results show that for all the test cases:
(1) For the backward reduction, the expected volume is preserved when the scenario tree is reduced since the difference presents slight variations. The variance and standard deviation are preserved until 20% of reduction. With 30% of reduction, the variance and standard deviation are not preserved. For example, Table A1 illustrates the results obtained in detail for July for the powerhouse CD. For the expected volume, the mean difference between the full scenario tree and the reduced scenario tree is 0.132% with 10% of reduction and 0.683% with 30% of reduction. For the variance, the mean difference between the full scenario tree and the reduced scenario tree is 0.676% with 10% of reduction and 32.117 % with 30% of reduction as shown in Table A1 and Fig 7.

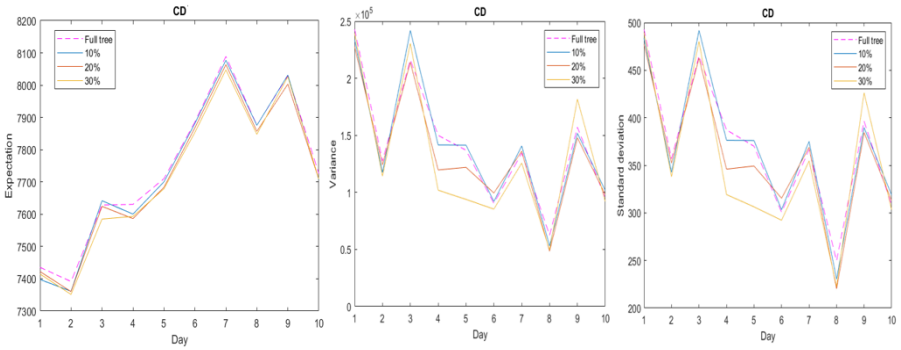

Fig. 7 An example of the statistical tests for backward reduction for powerhouse CD

(2) For the neural gas, the results show that the expected volume is preserved, but not the variance nor the standard deviation when the scenario tree is reduced. For example, for July, the mean difference between the variance of the full tree and all the reduced scenario tree is important from 10% of reduction as shown Fig 8 and Table A1. This can cause loss in the diversity. All clustering and moment matching methods suffer from the same problem. [27]. For September, October and December, the same observations hold. For backward reduction, the expected volume is preserved, the variance and the standard deviation are preserved until 20% of reduction. For neural gas, the expected volume is preserved, but not the variance nor the standard deviation. Due to the large size of data, only the detailed results for July are reported in this Section.

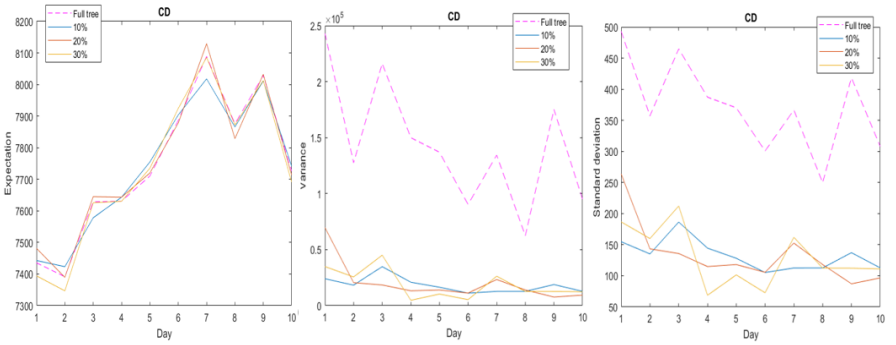

Fig. 8 An example of the statistical tests for neural gas for powerhouse CD

4.3 Operational tests

The present Section describes in more details the impact of the choice of the reduction method on the objective function (total energy produced) and on the solution of the optimization problem.

4.3.1 Impact on the total energy production

We calculated the energy produced throughout the 10 days rolling-horizon for each scenario tree for the two methods. The decisions are taken each hour, so 240 decisions are available for the 10 days.

Table 2 Total energy production for 4 test cases

Reduction	Full tree	Backward	Neural gas	Diff-B -NG(%)
		July 2021 (GWh)		
10%	82.6001	82.6986	82.5570	+0.17
20%	82.6001	82.6805	82.3650	+0.38
30%	82.6001	82.6917	82.3263	+0.44
		September 2021 (GWh)		
10%	78.2898	78.4234	78.2783	+0.185
20%	78.2898	78.6051	78.6000	+0.006
30%	78.2898	76.3866	78.4427	-2.692
		October 2021 (GWh)		
10%	72.6595	72.6545	72.5864	+0.094
20%	72.6595	72.6125	72.5867	+0.036
30%	72.6595	72.5515	72.5299	+0.030
		December 2021 (GWh)		
10%	91.8711	92.045	92.045	0.000
20%	91.8711	86.0004	91.9312	-6.896
30%	91.8711	92.0818	91.7845	+0.323

Table 2 illustrates the difference of the total energy produced between the two methods (Diff-B-NG) for the 4 test cases. A positive value indicates that backward reduction produces more energy than the neural gas and a negative value indicates the opposite. The results show that, in most cases, the backward reduction method produces more energy than neural gas. The highest differences are observed in December with 20% and in September with 30%. In these cases, the neural gas scenario trees produce more energy than those of the backward method. This is due to the fact that for backward reduction, the maximal number of start-ups is quickly reached compared to the neural gas.

For example, Fig 9 indicates the produced power and the turbines in operation for the powerhouse (CS) for both methods. The backward reduction solutions are presented in dashed line and the neural gas solutions in solid line. For backward reduction, the maximum number of the startups (2 start-ups) is reached at hour 65 (solid point). The model requires the activation of 4 turbines during the rest of the planning horizon. However, until this hour, the number of start-ups proposed by neural gas is less than 2 (dashed point), so the model is allowed to propose another start-up. For this reason, at hour 135, the neural gas solution proposes to switch from 4 to 5 active turbines and therefor in these cases neural gas produced more energy than backward reduction. However, in practice, it is recommended to have a limited number of start-ups since the frequent start-ups cause maintenance costs and decreasing the life time of the turbines.

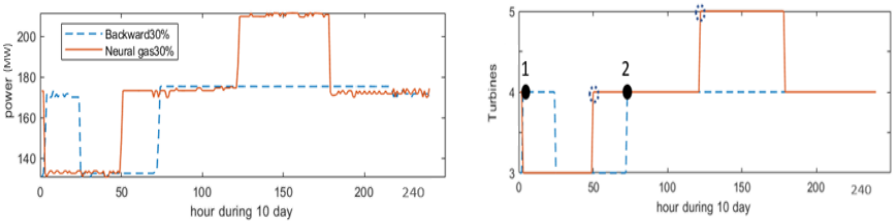

Fig. 9 Maximum number of start-ups

Let us compare the total energy produced from the full tree and the reduced scenario trees of each reduction method in order to define their impact on the objective function. Table 2 reports the values of the total energy produced obtained from the full scenario tree and the reduced scenario trees of each method for the 4 test cases. The results show that, except for the cases where the maximum number of the start-ups is reached, the total energy produced from the full tree and the reduced trees are close. In September and December the results obtained with neural gas are closer to the total energy obtained from the full tree. For the months of July and October, the total energy produced with backward reduction is closer to the total energy from the full scenario

tree. To get further insights, a study of the impact of each method on the solution of the optimization problem is done in the next Section.

4.3.2 Impact on the solutions of the optimization problem

In this Section, the solutions obtained from the full and the reduced scenario trees derived by both methods are compared. The cases of December with 20% of reduction and September with 30% (where the maximum number of the start-ups is reached) are not considered in the analysis of the results because the differences are very high compared to the other cases. Therefore, they are considered as extreme cases.
The results show that for all the test cases: (1) the solutions proposed from the different methods are broadly similar when the reduction is 10%. (2) For the months of July and October, beyond 20% of reduction, the solutions are different and (3) for the month of December and September the differences in the solutions are huge with 30% of reduction. According to the Table 2 and except for the cases where the maximum number of the start-ups is reached, the highest difference (+0.44%) is observed in July and the lowest one (0.00%) is observed in December. In this Section, the results from December and July are detailed.

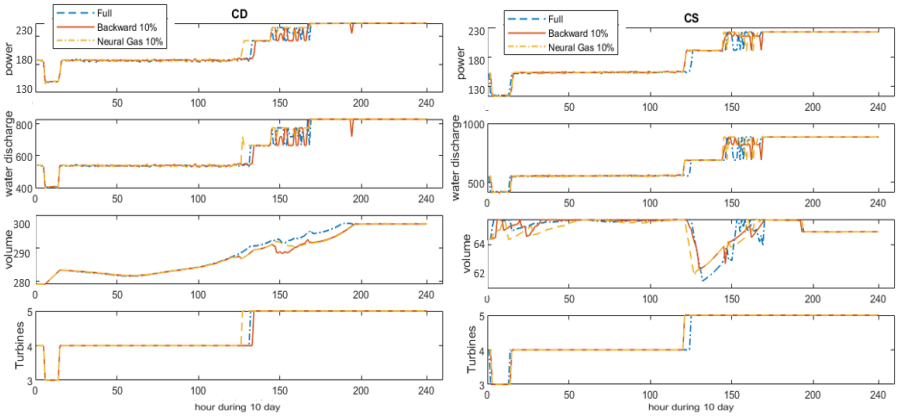

Fig. 10 Comparison of the solutions for backward reduction, neural gas and the full tree for December with 10% of reduction

Let us start with December. Fig 10 and 11 illustrate the proposed solutions: the power produced in (MW), the amount of water discharge in (m^3/s), the volume of the reservoir in (hm^3) and the number of turbines in operation for backward reduction (solid line), neural gas (dashed line) and for the full tree (dotted line) for two powerhouses CD and CS. As shown in Fig 10, the solutions proposed with 10% of reduction from two methods and from the full tree are broadly similar. This similarity is due to the fact that, for December,

the expected volume of the reservoir of the full and the reduced scenario trees during the rolling-horizon are quite similar, as shown in Fig 12 for both powerhouses CD and CS. On the other hand, the results show that from 30% of reduction, the solutions exhibit a small difference. There is a time lag between the proposed solutions of each method as can be seen in Fig 11. This is due to the loss of the information with 30% of reduction.

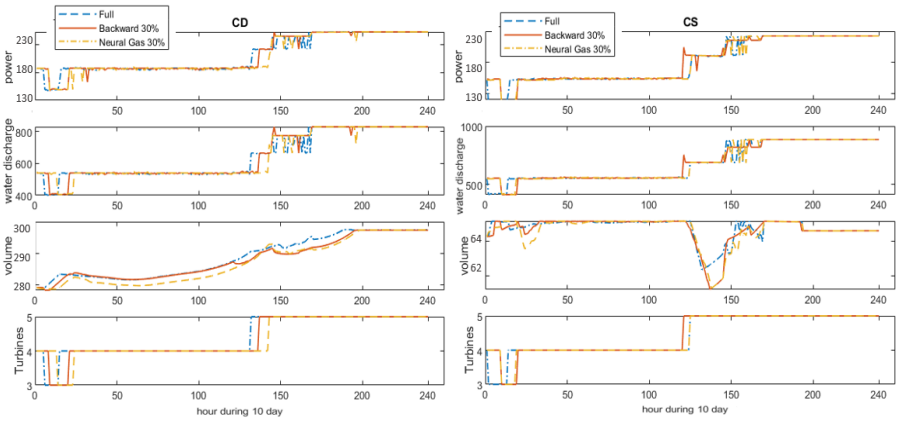

Fig. 11 Comparison of the solutions for backward reduction, neural gas and the full tree for December with 30% of reduction

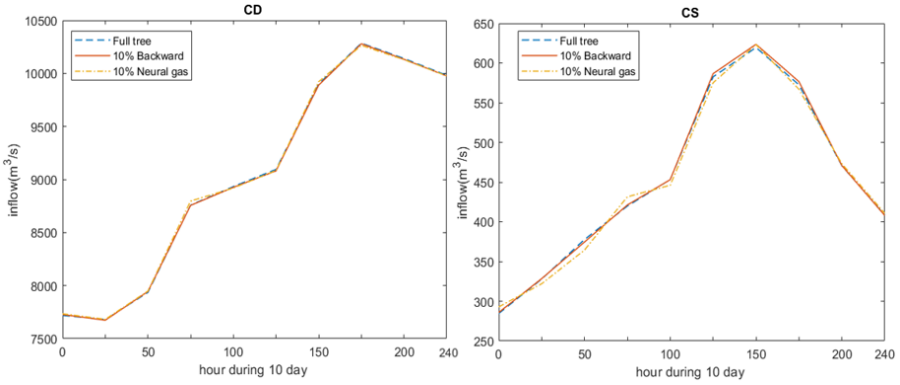

Fig. 12 The expected volume of the reservoir for the full tree, backward reduction and neural gas with 10% of reduction

For the month of July, an analysis of the results for the powerhouses CD and CS shows that the total energy produced by backward reduction with all the reduction percentage is higher than energy produced by neural gas and the full tree. Let us analyse the solution obtained with 20% of reduction since in this case the highest difference between the two methods is detected. Fig 13 and

14 illustrate the proposed solutions for the two powerhouses. The left figures indicate that the produced power and the amount of the water discharge of the full tree and the reduced tree are different. This difference is due to the fact that the expected volume of the reservoir presents slight variations compared to December as shown in the right of Fig 13 and Fig 14. This is due to the variability of inflow scenarios in this month. For example, Fig 4 indicates that the variation in July is greater than in December.

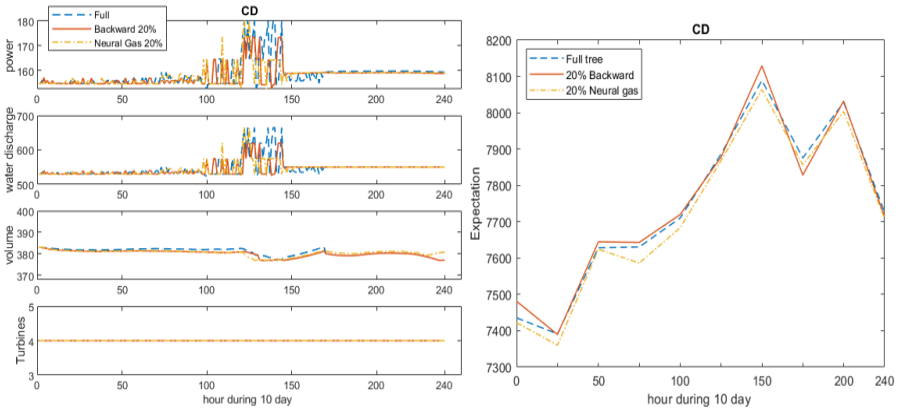

Fig. 13 Comparison of the solutions for backward reduction, neural gas and the full tree for the powerhouse CD for July with 20% of reduction (left figure) and the expected volume of the reservoir for each method (right figure)

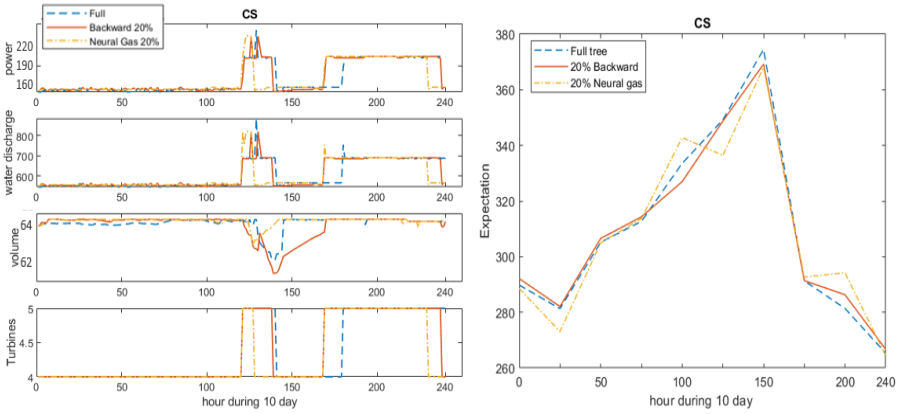

Fig. 14 Comparison of the solutions for backward reduction, neural gas and the full tree for the powerhouse CS for July with 20% of reduction (left figure) and the expected volume of the reservoir for each method (right figure)

4.4 Computational time

The average time to determine the reduced scenario trees are quite similar for both methods. The obtained reduced scenario tree for each method for the powerhouse CD and CS is used as an input to the stochastic programming model.

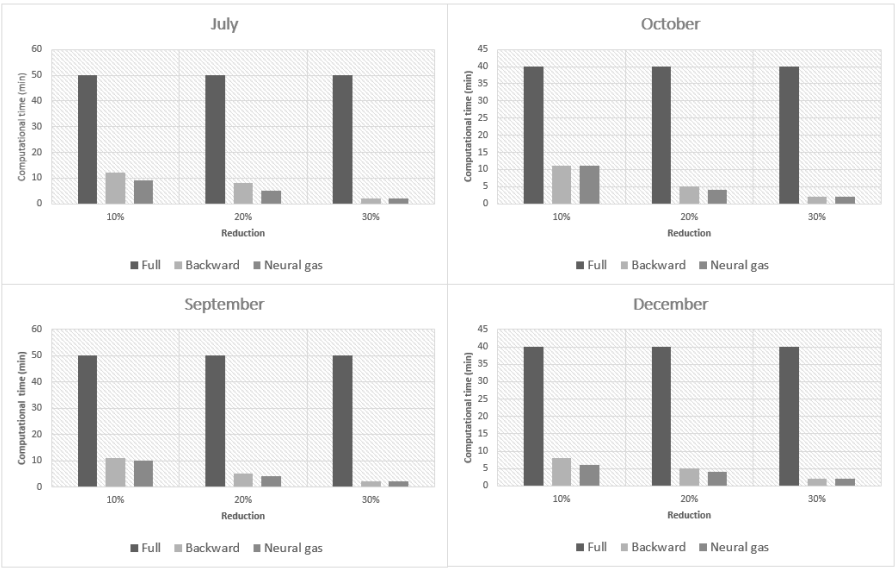

Fig. 15 Comparison of the average computational time for the full scenario tree and the reduced scenario trees

In order to define the impact of the choice of the generation and the reduced method on the solution, the average time to optimize a single day in the rolling-horizon procedure for each scenario tree for each method is calculated and illustrated in Fig 15. The results show that the computational time decreases with the number of scenarios. Since, a decision is made at each node of the scenario tree, the number of variables is reduced with the reduction of the number of the scenarios and as a consequence the computing time is reduced. However, the results show that optimizing with neural gas is faster than backward reduction and the full scenario tree. For example, in July, optimizing with a full scenario tree takes 50 min, with a reduced scenario tree with 20% of reduction, by backward reduction takes 8 min, and by the neural gas method takes 5 min.

4.5 Solution Quality Assessment

In this Section, the solutions obtained from the scenario tree reduction from each method are compared to the solution obtained from the median scenario of the inflows from the full tree in order to define the interest of using a

stochastic model. In addition, a comparison between the stochastic and the deterministic models with the real realizations of the inflows is done in order to assess the quality of the solution.

4.5.1 Stochastic vs median scenario

In order to assess the interest of using a stochastic model, we compare the solutions obtained from the scenario tree generation methods and the median scenario of the inflows. Every day, the median scenario is determined and the problem is solved in a deterministic way. The solutions obtained by backward reduction, neural gas and by the full scenario tree are compared to a rolling median.

As shown in Fig 16, the scenario tree methods produce more energy than the median scenario for all the test cases (except for the cases where the maximum number of the start-ups is reached). However, for December and September, the energy produced with stochastic methods is close to the energy produced with median scenario compared to July and October. In December and September, the expected volume of the reservoir of the median scenario is close to the expected volume of the stochastic method compared to the July and October.

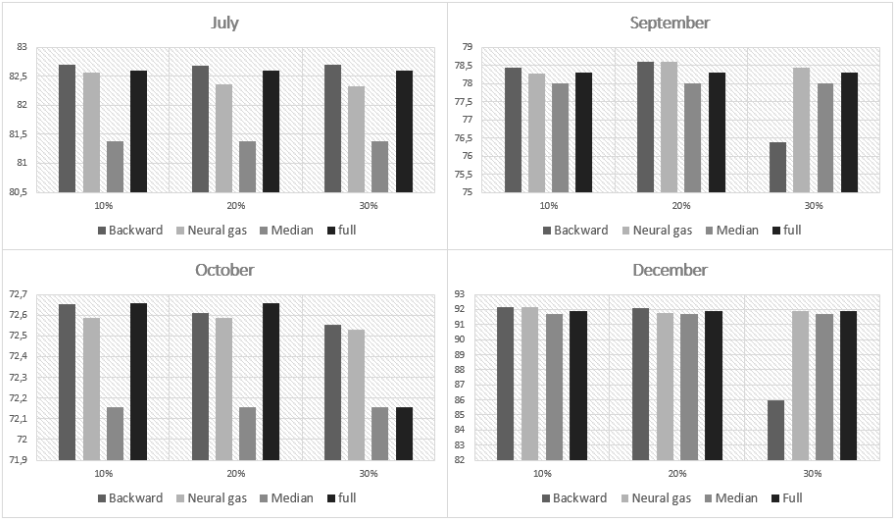

Fig. 16 Total energy produced for stochastic methods vs median method

For example, Fig 17 shows a comparison between the expected volume of the reservoir of the median scenario (with a solid line) and the stochastic method (with a dash line). The results show that the expected volume in December is closer to the expected volume for the median scenario compared to July for the powerhouse CS.

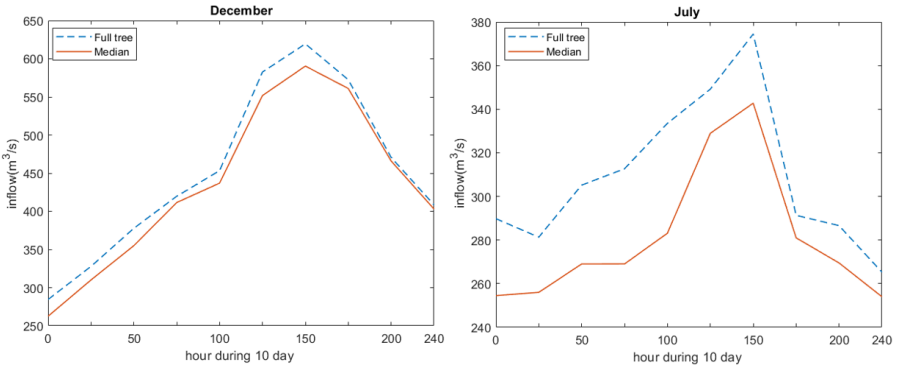

Fig. 17 A comparison between the expected volume of the reservoir of the median scenario and the stochastic method

For the computational time, optimizing with the median scenario is faster than the stochastic methods. The average time to optimize a single day in the rolling-horizon procedure is less than 1 min.

4.5.2 Stochastic vs Deterministic methods

In this Section, we compare the stochastic model and the deterministic one with the real realizations of the inflows in order to assess the quality of the solutions. Table 3 shows the difference between the deterministic and the stochastic models using the full tree (Diff-F-D), backward reduction (Diff-B-D) and neural gas (Diff-NG-D). The results show that the scenario tree generation method is consistent, as the difference between the energy produced of the stochastic and the deterministic models present slight variations. For example, for the December test case, with 10% of reduction, the difference between the deterministic model and both backward reduction and neural gas is only 0.01%.

Table 3 The difference of the total energy production between the stochastic methods and the deterministic one

	July 2021 (%)			**September 2021 (%)**		
Reduction	Diff-B-D	Diff-NG-D	Diff-F-D	Diff-B-D	Diff-NG-D	Diff-F-D
10%	0.61	0.44	0.56	0.62	0.44	0.61
20%	0.84	0.46	0.56	0.21	0.21	0.61
30%	0.89	0.45	0.56	0.41	3.02	0.61
	October 2021 (%)			**December 2021 (%)**		
10%	0.62	0.53	0.52	0.01	0.01	0.31
20%	0.62	0.57	0.52	0.24	6.68	0.31
30%	0.70	0.67	0.52	0.40	0.08	0.31

Fig 18 illustrates the proposed decisions from the deterministic model and the stochastic methods (the full tree, backward reduction and neural gas) with

10% of reduction for the powerhouses CD and CS more precisely : the produced power, the water discharge, the volume of the reservoir and the number of turbines in operation.
As shown in Fig 18, the solution provided by the stochastic and deterministic methods have different strategies, which is not surprising. For example, for the deterministic solutions, the reservoir of the powerhouse CD is filled then lowered and filled again. This is due to the fact that, the amount of the inflows that will occur at the next period is known exactly, so the deterministic model has more liberty to vary the reservoir volumes and that is why a micro-cycle trend can be observed.

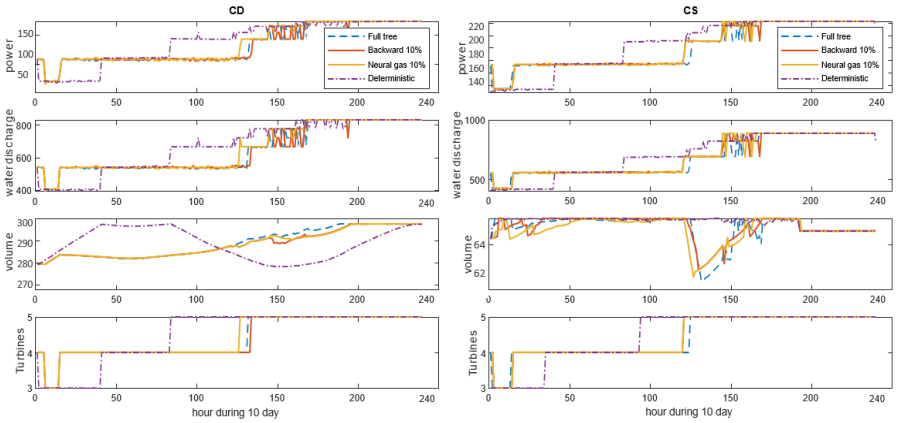

Fig. 18 A comparison between the deterministic and the stochastic solution

5 Discussions and conclusion

This paper presents a stochastic short-term hydropower model for optimizing energy production. The uncertainty of the inflows is represented using the scenario trees. The backward reduction and neural gas methods are used to generate and reduce a full scenario tree. The objective is to determine the impact of the choice of method on the results of the objective function, the solution and the computational time. For this purpose, we carry out statistical tests to determine the impact of the choice of the generation methods on the preservation of the expected volume, the variance and the standard deviation when the scenario tree is reduced. We also carry out operational tests, where the scenario trees are used as input to the stochastic programming model. The total energy produced and the proposed solution with different reduced trees of each method are evaluated and compared to determine the impact of the choice of the methods on the solution. In addition, the proposed model is compared to the median scenario in order to define the interest of the use of the stochastic model and compared to the deterministic one with the real realizations of the inflows in order to assess the quality of the solution. The

results show that, ***for statistical tests,*** neural gas preserves the expected volume, but not the variance nor the standard deviation when the scenario tree is reduced. By contrast, the backward reduction preserves the expected volume, the variance and the standard deviation until 20% of reduction. ***For operational tests,*** the results show that, (1) the solutions proposed from the different methods are broadly similar when the reduction is 10%. (2) For December and September, where the variability of inflow scenarios is not very high and the expected volume provided from the different scenario tree is close, the solutions of both reductions and from the median scenario are close. (3) For July and October, where the variability of inflow scenarios is very high, backward reduction produces more energy than neural gas and the solution proposed with backward reduction is close to the solution of the full scenario tree. (4) Optimizing with the neural gas method is faster than backward reduction and the median scenario is faster than both.

In conclusion, if the variability of inflow scenarios is high and the expected volume between the scenario trees is different, it is recommended to use a method that preserves the variation when the scenario tree is reduced, even though the computational time can be a little longer. Otherwise, any of the methods can be used, but the reduction should not be very important in order to be close to reality. If the variability of inflow scenarios is very low, the median scenario can be used since it is faster and the results obtained from the stochastic method and from the median scenario are close.

Appendix A

Table A1 Comparison of the expected volume, the variance and standard deviation with 10%, 20% and 30% of reduction for July for powerhouse CD

July 2021 10%

Day	EF	EB	Diff (%)	VF	VB	Diff (%)	SDF	SDB	Diff(%)	EF	ENG	Diff(%)	VF	VNG	Diff (%)	SDF	SDNG	Diff (%)
1	7435,49	7397,85	0,51	243067,41	233808,56	3,81	493,02	483,54	1,92	7435,49	7442,78	-0,10	243067,41	23978,69	90,13	493,02	154,85	68,59
2	7391,63	7361,38	0,41	127544,41	117477,33	7,89	357,13	342,75	4,03	7391,63	7423,41	-0,43	127544,41	18265,21	85,68	357,13	135,15	62,16
3	7628,60	7642,00	-0,18	216103,92	241895,01	-11,93	464,87	491,83	-5,80	7628,60	7578,19	0,66	216103,92	34769,09	83,91	464,87	186,46	59,89
4	7630,72	7600,54	0,40	149947,13	141616,96	5,56	387,23	376,32	2,82	7630,72	7643,25	-0,16	149947,13	20878,82	86,08	387,23	144,50	62,68
5	7710,40	7700,51	0,13	136953,40	141411,37	-3,26	370,07	376,05	-1,61	7710,40	7754,87	-0,58	136953,40	16419,40	88,01	370,07	128,14	65,37
6	7886,47	7881,67	0,06	90374,16	92025,81	-1,83	300,62	303,36	-0,91	7886,47	7903,61	-0,22	90374,16	11082,97	87,74	300,62	105,28	64,98
7	8089,08	8077,29	0,15	134380,03	140716,29	-4,72	366,58	375,12	-2,33	8089,08	8017,94	0,88	134380,03	12670,86	90,57	366,58	112,56	69,29
8	7875,87	7875,56	0,00	62375,45	53052,87	14,95	249,75	230,33	7,78	7875,87	7865,92	0,13	62375,45	12708,30	79,63	249,75	112,73	54,86
9	8028,96	8031,42	-0,03	157060,14	51737,66	3,39	396,31	389,54	1,71	7882,98	8011,08	-1,63	175127,52	18805,68	89,26	418,48	137,13	67,23
10	7726,19	7706,90	0,25	94947,50	101688,65	-7,10	308,14	318,89	-3,49	7726,19	7741,52	-0,20	94947,50	12729,59	86,59	308,14	112,83	63,38
mean			0,132			0,33			0,24			-0,17			86,39			63,32

July 2021 20%

Day	EF	EB	Diff (%)	VF	VB	Diff (%)	SDF	SDB	Diff(%)	EF	ENG	Diff(%)	VF	VNG	Diff (%)	SDF	SDNG	Diff (%)
1	7435,49	7414,65	0,28	243067,41	239504,82	1,47	493,02	489,39	0,74	7435,49	7481,23	-0,62	243067,41	69592,72	71,37	493,02	263,80	46,49
2	7391,63	7351,46	0,54	127544,41	114316,10	10,37	357,13	338,11	5,33	7389,69	7391,63	0,03	20527,79	127544,41	83,91	143,28	357,13	59,88
3	7628,60	7585,14	0,57	216103,92	230530,91	-6,68	464,87	480,14	-3,28	7644,93	7628,60	-0,21	18429,37	216103,92	91,47	135,75	464,87	70,80
4	7630,72	7593,33	0,49	149947,13	101945,25	32,01	387,23	319,29	17,55	7643,33	7630,72	-0,17	13212,26	149947,13	91,19	114,94	387,23	70,32
5	7710,40	7678,99	0,41	136953,40	93846,32	31,48	370,07	306,34	17,22	7719,95	7710,40	-0,12	13950,13	136953,40	89,81	118,11	370,07	68,08
6	7886,47	7853,72	0,42	90374,16	85394,16	5,51	300,62	292,22	2,79	7877,95	7886,47	0,11	11154,92	90374,16	87,66	105,62	300,62	64,87
7	8089,08	8047,52	0,51	134380,03	125734,62	6,43	366,58	354,59	3,27	8129,40	8089,08	-0,50	23299,82	134380,03	82,66	152,64	366,58	58,36
8	7875,87	7847,71	0,36	62375,45	49584,05	20,51	249,75	222,67	10,84	7828,68	7875,87	0,60	13946,87	62375,45	77,64	118,10	249,75	52,71
9	8028,96	8026,52	0,03	157060,14	181649,88	-15,66	396,31	426,20	-7,54	8032,49	7882,98	-1,90	7603,86	175127,52	95,66	87,20	418,48	79,16
10	7726,19	7708,88	0,22	94947,50	92026,49	3,08	308,14	303,36	1,55	7713,71	7726,19	0,16	9318,13	94947,50	90,19	96,53	308,14	68,67
mean			0,38			8,85			4,85			-0,26			86,16			63,93

July 2021 30%

Day	EF	EB	Diff (%)	VF	VB	Diff (%)	SDF	SDB	Diff(%)	EF	ENG	Diff(%)	VF	VNG	Diff (%)	SDF	SDNG	Diff (%)
1	7435,49	7321,53	1,53	243067,41	111229,19	54,24	493,02	333,51	32,35	7435,49	7394,35	0,55	243067,41	34862,29	85,66	493,02	186,71	62,13
2	7391,63	7353,87	0,51	127544,41	86643,40	32,07	357,13	294,35	17,58	7394,35	7391,63	0,61	25551,58	127544,41	79,97	159,85	357,13	55,24
3	7628,60	7595,46	0,43	216103,92	239737,08	-10,94	464,87	489,63	-5,33	7625,75	7628,60	0,04	45012,52	216103,92	79,17	212,16	464,87	54,36
4	7630,72	7597,43	0,44	149947,13	30062,18	79,95	387,23	173,38	55,22	7630,48	7630,72	0,00	4724,10	149947,13	96,85	68,73	387,23	82,25
5	7710,40	7704,58	0,08	136953,40	70989,37	48,17	370,07	266,44	28,00	7735,00	7710,40	-0,32	10284,51	136953,40	92,49	101,41	370,0	72,60
6	7886,47	7840,87	0,58	90374,16	104838,18	-16,00	300,62	323,79	-7,71	7924,37	7886,47	-0,48	5289,75	90374,16	94,15	72,73	300,62	75,81
7	8089,08	8020,77	0,84	134380,03	89043,77	33,74	366,58	298,40	18,60	8085,67	8089,08	0,04	26219,98	134380,03	80,49	161,93	366,58	55,83
8	7875,87	7798,05	0,99	62375,45	27427,85	56,03	249,75	165,61	33,69	7872,32	7875,87	0,05	12637,68	62375,45	79,74	112,42	249,75	54,99
9	8028,96	7989,20	0,49	157060,14	131513,25	16,26	396,31	362,65	8,49	8012,80	7882,98	-1,65	12593,25	175127,52	92,81	112,22	418,48	73,18
10	7726,19	7653,23	0,94	94947,50	68694,13	27,65	308,14	262,10	14,94	7691,43	7726,19	0,45	12330,22	94947,50	87,01	111,04	308,14	63,96
mean			0,68			32,12			19,58			-0,07			86,83			65,04

EF : the expected volume of the full tree
EB : the expected volume of backward reduction
Diff : the difference
VF : the variance of the full tree
VB : the variance of backward reduction
SDF : the standard deviation of the full tree
SDB : the standard deviation of backward reduction
ENG : the expected volume of neural gas
SDNG : the standard deviation of neural gas
VNG : the variance of neural gas

No conflicts of interest to declare.

References

[1] FICO Xpress Optimization Suite. Available at: https://www.msi-jp.com/xpress/learning/square/optimizer-2015.pdf, 2014.

[2] Digital Research Alliance of Canada. Available at: https://docs.alliancecan.ca/wiki/Technical_documentation, 2022.

[3] GAMS SCENRED2. Available at: https://www.gams.com/latest/docs/T_SCENRED2.html, 2022.

[4] Python. Available at: https://www.python.org/, 2022.

[5] SP Adhau, RM Moharil, and PG Adhau. K-means clustering technique applied to availability of micro hydro power. *Sustainable Energy Technologies and Assessments*, 8:191–201, 2014.

[6] Ignacio Aravena and Esteban Gil. Hydrological scenario reduction for stochastic optimization in hydrothermal power systems. *Applied Stochastic Models in Business and Industry*, 31(2):231–240, 2015.

[7] Alberto Borghetti, Claudia D'Ambrosio, Andrea Lodi, and Silvano Martello. An milp approach for short-term hydro scheduling and unit commitment with head-dependent reservoir. *IEEE Transactions on power systems*, 23(3):1115–1124, 2008.

[8] Maissa Daadaa, Sara Séguin, Kenjy Demeester, and Miguel F Anjos. An optimization model to maximize energy generation in short-term hydropower unit commitment using efficiency points. *International Journal of Electrical Power & Energy Systems*, 125:106419, 2021.

[9] Turid Follestad, Ove Wolfgang, and Michael M Belsnes. An approach for assessing the effect of scenario tree approximations in stochastic hydropower scheduling models. In *Proc. of the 17th Power System Computation Conference*, pages 271–277, 2011.

[10] Nicole Growe-Kuska, Holger Heitsch, and Werner Romisch. Scenario reduction and scenario tree construction for power management problems. In *2003 IEEE Bologna Power Tech Conference Proceedings,*, volume 3, pages 7–pp. IEEE, 2003.

[11] Lucas SM Guedes, Pedro de Mendonça Maia, Adriano Chaves Lisboa, Douglas Alexandre Gomes Vieira, and Rodney Rezende Saldanha. A unit commitment algorithm and a compact milp model for short-term hydropower generation scheduling. *IEEE Transactions on Power Systems*,

32(5):3381–3390, 2016.

[12] Ronald Hochreiter and Georg Ch Pflug. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Annals of Operations Research*, 152(1):257–272, 2007.

[13] Michal Kaut and W Stein. *Evaluation of scenario-generation methods for stochastic programming*. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät ..., 2003.

[14] Jesús Ma Latorre, Santiago Cerisola, and Andres Ramos. Clustering algorithms for scenario tree generation: Application to natural hydro inflows. *European Journal of Operational Research*, 181(3):1339–1353, 2007.

[15] Hernan Leövey and Werner Römisch. Quasi-monte carlo methods for linear two-stage stochastic programming problems. *Mathematical Programming*, 151(1):315–345, 2015.

[16] Can Li and Ignacio E Grossmann. A review of stochastic programming methods for optimization of process systems under uncertainty. *Frontiers in Chemical Engineering*, 2:622241, 2021.

[17] Jinshu Li, Feilin Zhu, Bin Xu, and William W-G Yeh. Streamflow scenario tree reduction based on conditional monte carlo sampling and regularized optimization. *Journal of Hydrology*, 577:123943, 2019.

[18] Ekaterina Moiseeva and Mohammad Reza Hesamzadeh. Strategic bidding of a hydropower producer under uncertainty: Modified benders approach. *IEEE Transactions on Power Systems*, 33(1):861–873, 2017.

[19] Frederik Questier, Qian Guo, Beata Walczak, DL Massart, C Boucon, and S De Jong. The neural-gas network for classifying analytical data. *Chemometrics and Intelligent Laboratory Systems*, 61(1-2):105–121, 2002.

[20] Ignacio Rios, Roger JB Wets, and David L Woodruff. Multi-period forecasting and scenario generation with limited data. *Computational Management Science*, 12(2):267–295, 2015.

[21] Sara Séguin, Pascal Côté, and Charles Audet. Self-scheduling short-term unit commitment and loading problem. *IEEE Transactions on Power Systems*, 31(1):133–142, 2015.

[22] Sara Séguin, Stein-Erik Fleten, Pascal Côté, Alois Pichler, and Charles Audet. Stochastic short-term hydropower planning with inflow scenario trees. *European Journal of Operational Research*, 259(3):1156–1168, 2017.

[23] Raouia Taktak and Claudia D'Ambrosio. An overview on mathematical programming approaches for the deterministic unit commitment problem in hydro valleys. *Energy Systems*, 8(1):57–79, 2017.

[24] Yelena Vardanyan. *On stochastic optimization for short-term hydropower planning*. PhD thesis, KTH Royal Institute of Technology, 2012.

[25] Yelena Vardanyan and Mohammad Reza Hesamzadeh. The coordinated bidding of a hydropower producer in three-settlement markets with time-dependent risk measure. *Electric Power Systems Research*, 151:40–58, 2017.

[26] Begoña Vitoriano, S Cerisol, and Andrés Ramos. Generating scenario trees for hydro inflows. In *Proceedings of the 6th International Conference Probabilistic Methods Applied to Power Systems PMAPS*, volume 2, 2000.

[27] Bin Xu, Ping-An Zhong, Renato C Zambon, Yunfa Zhao, and William W-G Yeh. Scenario tree reduction in stochastic programming with recourse for hydropower operations. *Water Resources Research*, 51(8):6359–6380, 2015.