UQAC
Université du Québec
à Chicoutimi

OPTIMISATION DE LA PRODUCTION HYDROÉLECTRIQUE À COURT TERME

PAR MAISSA DAADAA

THÈSE PRÉSENTÉE À L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI COMME EXIGENCE PARTIELLE EN VUE DE L'OBTENTION DU GRADE DE PHILOSOPHIÆ DOCTOR (PH.D.) EN INFORMATIQUE

QUÉBEC, CANADA

RÉSUMÉ

L'hydroélectricité est la principale source d'énergie utilisée au Québec grâce à son vaste réseau de cours d'eau et de rivières. Les gestionnaires des centrales hydroélectriques jouent un rôle important dans la production d'énergie en gérant efficacement l'eau stockée dans les réservoirs, tout en répondant aux contraintes opérationnelles telles que les contraintes hydriques, les arrêts et démarrages des turbines, etc. La résolution des problèmes de gestion des centrales repose sur la détermination des turbines en opération dans les centrales, du débit turbiné par chacune d'elles et des volumes des réservoirs pour chaque période de planification. Cette résolution s'avère complexe en raison de la non-linéarité et de la non-convexité des fonctions de production, qui dépendent de plusieurs facteurs tels que l'efficacité des turbines qui diffère d'une turbine à une autre, les démarrages des turbines qui doivent être minimisés et de l'incertitude des apports naturels. Le défi de cette thèse est non seulement de résoudre ce type de problème, mais d'obtenir des solutions qui peuvent être appliquées directement dans les centrales. Pour réussir ce défi, une technique novatrice basée sur les courbes d'efficacité des débits turbinés et des puissances produites pour toutes les combinaisons possibles de turbines est utilisée. Cette technique permet de déterminer un ensemble optimal de paires de points de débit turbiné et puissance produite ayant des efficacités maximales. Ces points seront utilisés dans les modèles d'optimisations proposés durant cette thèse.

Tout d'abord, une modélisation déterministe du problème à court terme pour le réseau hydroélectrique du Saguenay Lac-St-Jean est présentée. Ce modèle linéaire mixte en nombres entiers exploite les couples de points de débit turbiné et de puissance produite pour déterminer le débit optimal qui maximise l'énergie produite tout en pénalisant les démarrages des turbines. Des tests sont effectués pour démontrer l'efficacité de la méthode proposée.

Ensuite, le modèle déterministe est étendu afin de prendre en compte l'incertitude des apports. Ces incertitudes sont représentées en utilisant des arbres de scénarios qui sont générés et réduits en utilisant deux méthodes de génération et de réduction : le gaz neuronal et le backward reduction. Les arbres de scénarios obtenus par chaque méthode sont utilisés comme des paramètres pour résoudre un modèle de programmation linéaire en nombres entiers stochastique à plusieurs étapes. Les résultats obtenus par ces deux méthodes sont comparés afin d'évaluer impact du choix de la méthode sur la solution. Cette étude démontre que la méthode backward reduction surpasse la méthode du gaz neuronal en termes de préservation des informations statistiques. En outre, une comparaison entre le modèle stochastique et le modèle déterministe est réalisée afin d'évaluer la qualité de la solution. Cette comparaison a ouvert de nouvelles perspectives en conduisant une étude sur le choix des modèles d'optimisation pour le problème d'optimisation hydroélectrique à court terme.

Étant donné qu'il est possible d'utiliser à la fois des modèles déterministes et stochastiques, une approche qui vise à déterminer le meilleur modèle d'optimisation pendant l'horizon de planification en se basant sur les prévisions des apports afin de maximiser la production

totale d'énergie est présentée. Pour ce faire, une formulation de programmation linéaire mixte en nombres entiers où les apports sont représentés par un scénario médian de tous les scénarios possibles est utilisée en tant qu'un modèle déterministe, et une formulation de programmation linéaire en nombres entiers stochastique à plusieurs étapes où l'incertitude des apports est représentée en utilisant un arbre de scénario est utilisée en tant qu'un modèle stochastique. L'approche proposée permet d'identifier le modèle d'optimisation qui s'adaptera le mieux aux conditions variables du système hydroélectrique. Le problème est formulé comme un problème d'optimisation boîte noire, dans lequel les solutions sont examinées, y compris une évaluation des volumes d'apport attendus. Les résultats obtenus montrent que le choix du modèle d'optimisation est influencé par la variabilité observée des apports naturels.

ABSTRACT

Hydropower is the main source of energy in Québec, thanks to the extensive network of streams and rivers. Powerhouse managers play an important role in energy production by efficiently managing the water stored in reservoirs, subject to operational constraints such as water balance, turbine start-ups, etc. Solving a hydropower management problem determines the turbines in operation in the powerhouses, the water discharged by each turbine, and the volume of the reservoirs for each planning period. The resolution proves to be complex due to the nonlinearity and nonconvexity of the production functions, which depends on several factors, such as turbine efficiency, which varies from turbine to turbine, turbine start-ups that need to be minimized, and inflow uncertainty. The challenge of this thesis is not only to solve these types of problems, but also to find solutions that can be directly applied in the powerhouses. To overcome this challenge, an innovative technique based on the efficiency curves of water discharge and power produced for all possible turbine combinations is used. This technique allows to find of an optimal set of pairs of points of water discharge and power produced with maximum efficiencies. These points are used in the optimization models proposed in this thesis.

First, a deterministic modeling of the short-term problem for the Saguenay Lac-St-Jean hydropower system is presented. This mixed-integer linear programming model uses the pair of points of water discharge and power produced to determine the optimal water discharge that maximizes energy produced while penalizing turbine start-ups. Tests are conducted to prove the effectiveness of the proposed method.

Second, the deterministic model is extended to account for inflow uncertainty. These uncertainties are represented using scenario trees generated and reduced using two generation and reduction methods : neural gas and backward reduction. The scenario trees obtained by each method are used as parameters to solve a multi-stage mixed-integer linear stochastic programming model. The results obtained with these two methods are compared to evaluate the impact of the choice of the method on the solution. This study shows that the backward reduction method outperforms the neural gas method in terms of preserving statistical information. Moreover, a comparison is made between the stochastic model and the deterministic model to evaluate the quality of the solution. This comparison opens new perspectives by conducting a study on the choice of optimization models for the short-term hydroelectric optimization problem.

Since it is possible to use both deterministic and stochastic models, an approach is presented that aims to determine the best optimization model during the planning horizon based on inflow forecasts in order to maximize total energy production. For this purpose, a mixed-integer linear programming formulation, where the inflows are represented by a median scenario of all possible scenarios, is used as a deterministic model, and, a multi-stage mixed-integer linear stochastic programming formulation, where the uncertainty of the inflows

is represented by a scenario tree, is used as a stochastic model. The proposed approach allows to find the optimal optimization model that best fit the varying conditions of the hydropower system. The problem is formulated as a blackbox optimization problem where solutions are examined, including an evaluation of the expected volumes of the inflows. Results show that the choice of the optimization model is influenced by the observed variability of inflows.

TABLE DES MATIÈRES

RÉSUMÉ ii

ABSTRACT iv

LISTE DES TABLEAUX ix

LISTE DES FIGURES x

REMERCIEMENTS xiii

CHAPITRE I – INTRODUCTION 1

1.1 CONTEXTE 2

1.2 OBJECTIFS DE LA THÈSE 4

1.3 TRAVAUX DÉCOULANT DE LA THÈSE 5

1.4 STRUCTURE DE LA THÈSE 6

CHAPITRE II – REVUE DE LITTÉRATURE 8

2.1 LES MÉTHODES DÉTERMINISTES 8

2.1.1 LA PROGRAMMATION DYNAMIQUE 9

2.1.2 PROGRAMMATION MIXTE EN NOMBRES ENTIERS 10

2.1.3 RELAXATION LAGRANGIENNE 13

2.2 LES MÉTHODES STOCHASTIQUES 15

2.2.1 LA PROGRAMMATION STOCHASTIQUE : MODÈLES DE RECOURS 16

2.2.2 OPTIMISATION ROBUSTE 23

2.2.3 AUTRES MÉTHODES 23

2.3 L'OPTIMISATION BOÎTE NOIRE 24

CHAPITRE III – ORGANISATION DE LA THÈSE 28

CHAPITRE IV – ARTICLE 1 : AN OPTIMIZATION MODEL TO MAXIMIZE ENERGY GENERATION IN SHORT-TERM HYDROPOWER UNIT COMMITMENT USING EFFICIENCY POINTS 30

4.1 INTRODUCTION 32

4.1.1 MOTIVATION 32

4.1.2 LITERATURE REVIEW 32
4.1.3 CONTRIBUTIONS 34
4.1.4 ORGANIZATION OF THE PAPER 35
4.2 SHORT-TERM HYDRO-POWER SCHEDULING PROBLEM 35
4.2.1 POWER PRODUCTION 35
4.2.2 TURBINE EFFICIENCY 36
4.2.3 COMBINATIONS OF THE TURBINES 36
4.2.4 START-UPS 38
4.2.5 PROBLEM DESCRIPTION 38
4.2.6 MATHEMATICAL MODEL 41
4.3 CASE STUDY 42
4.3.1 EXPERIMENTAL SETUP 43
4.3.2 COMPUTATIONAL RESULTS 45
4.3.3 VALIDATION OF THE MODEL 47
4.3.4 INTERPRETATION OF THE RESULTS 48
4.3.5 DISCUSSIONS 51
4.4 CONCLUSION 52
CHAPITRE V – ARTICLE 2 : QUANTIFYING THE IMPACT OF SCENARIO TREE GENERATION AND REDUCTION METHODS ON THE SOLUTION OF THE SHORT-TERM HYDROSCHEDULING PROBLEM 53
5.1 INTRODUCTION 55
5.2 SCENARIO TREE GENERATION AND REDUCTION 59
5.2.1 BACKWARD REDUCTION 59
5.2.2 NEURAL GAS 61
5.2.3 IMPLEMENTATION 63
5.3 STOCHASTIC MULTISTAGE MIXED INTEGER LINEAR MODEL 65
5.3.1 MATHEMATICAL MODEL 67
5.3.2 CASE STUDY 68

5.4 NUMERICAL RESULTS 70
5.4.1 REDUCTION OF THE SCENARIO TREE 71
5.4.2 STATISTICAL TESTS 74
5.4.3 OPERATIONAL TESTS 76
5.4.4 COMPUTATIONAL TIME 83
5.4.5 SOLUTION QUALITY ASSESSMENT 85
5.5 DISCUSSIONS AND CONCLUSION 89
CHAPITRE VI – ARTICLE 3 : OPTIMIZING STRATEGIES FOR SHORT-TERM HYDROPOWER SCHEDULING USING A BLACKBOX OPTIMIZATION FRAMEWORK 91
6.1 INTRODUCTION 93
6.2 SHORT TERM HYDROPOWER SCHEDULING 99
6.2.1 DETERMINISTIC MODEL 100
6.2.2 STOCHASTIC MODEL 102
6.2.3 OPTIMIZATION PROBLEM 105
6.3 BLACKBOX OPTIMIZATION 106
6.3.1 BLACKBOX DESCRIPTION 107
6.3.2 BLACKBOX FORMULATION 108
6.4 COMPUTATIONAL RESULTS 111
6.4.1 CASE STUDY 111
6.4.2 AGGREGATED STOCHASTIC MODEL 113
6.4.3 BLACKBOX SOLUTION 116
6.5 CONCLUSION 126
CONCLUSION 128
BIBLIOGRAPHIE 132
APPENDICE A – COMPARAISON STATISTIQUE ENTRE LES MÉTHODES BACKWARD REDUCTION ET GAZ NEURONAL 146

LISTE DES TABLEAUX

TABLEAU 4.2 : COMBINATIONS OF 4 AVAILABLE TURBINES 36

TABLEAU 4.3 : COMBINATION OF 5 AVAILABLE TURBINES 44

TABLEAU 4.4 : TOTAL ENERGY PRODUCTION AND AVERAGE IMPROVEMENTS 46

TABLEAU 4.5 : COMPARISON BETWEEN THE ENERGY APPROXIMATED AND THE REAL VALUES 47

TABLEAU 5.2 : COMBINATIONS OF 5 AVAILABLE TURBINES 65

TABLEAU 5.3 : TOTAL ENERGY PRODUCTION FOR 4 TEST CASES. 77

TABLEAU 5.4 : THE DIFFERENCE OF THE TOTAL ENERGY PRODUCTION BETWEEN THE STOCHASTIC METHODS AND THE DETERMINISTIC ONE 88

TABLEAU 6.2 : THE DIFFERENCE OF THE TOTAL ENERGY PRODUCTION (GWH) BETWEEN NON-AGGREGATED AND AGGREGATED MODELS 114

TABLEAU 6.3 : THE DIFFERENCE OF THE COMPUTATIONAL TIME BETWEEN NON-AGGREGATED AND AGGREGATED MODELS . . .114

TABLEAU 6.4 : TOTAL ENERGY PRODUCTION (GWH) FROM THE OPTIMIZATION MODELS 117

TABLEAU 6.5 : PROPOSED SOLUTIONS FROM BLACKBOX OPTIMIZATION ©MAISSA DAADAA 2023 118

TABLEAU 6.6 : PROPOSED SOLUTIONS FROM BLACKBOX MODEL USING THE AGGREGATED AND NON-AGGREGATED MODEL 118

TABLEAU 6.7 : ENERGY PRODUCED IN GWH USING BBO METHOD AND THE REAL REALIZATION 119

TABLEAU A1 : COMPARISON OF THE EXPECTED VOLUME, THE VARIANCE AND STANDARD DEVIATION WITH 10%, 20% AND 30% OF REDUCTION FOR JULY FOR POWERHOUSE CD 147

LISTE DES FIGURES

FIGURE 2.1 – EXEMPLE D'UN ARBRE DE SCÉNARIOS 17

FIGURE 2.2 – ALGORITHME MADS 27

FIGURE 4.1 – POWER OUTPUT FOR ALL COMBINATIONS 37

FIGURE 4.2 – ZOOM-IN OF THE RECTANGLE IN FIG.1 37

FIGURE 4.3 – PAIRS OF EFFICIENCY POINTS (P,q) 39

FIGURE 4.4 – THE SET OF EFFICIENCY POINTS OF WATER DISCHARGE. . . . 39

FIGURE 4.5 – HISTOGRAM OF AVERAGE IMPROVEMENTS OF ENERGY 47

FIGURE 4.6 – INSTANCE 32 FOR CD 48

FIGURE 4.7 – INSTANCE 32 FOR CS 48

FIGURE 4.8 – INSTANCE 36 FOR CD 49

FIGURE 4.9 – INSTANCE 36 FOR CS 49

FIGURE 4.10 – INSTANCE 22 FOR CD 50

FIGURE 4.11 – INSTANCE 22 FOR CS 50

FIGURE 4.12 – INSTANCE 7 FOR CD 51

FIGURE 4.13 – INSTANCE 7 FOR CS 51

FIGURE 5.1 – BACKWARD REDUCTION METHOD 61

FIGURE 5.2 – THE EFFICIENCY CURVES 66

FIGURE 5.3 – THE METHODOLOGY ACCORDING TO THE MANAGEMENT OF THE POWERHOUSES 69

FIGURE 5.4 – THE FULL SCENARIO TREES FOR THE FIRST DAY IN THE ROLLING HORIZON FOR JULY, SEPTEMBER, OCTOBER AND DECEMBER 72

FIGURE 5.5 – AN EXAMPLE OF THE FULL AND THE REDUCED SCENARIO TREES WITH 10%, 20% AND 30% OF REDUCTION WITH BACKWARD REDUCTION 73

FIGURE 5.6 – AN EXAMPLE OF THE FULL AND THE REDUCED SCENARIO TREES WITH 10%, 20% AND 30% OF REDUCTION WITH NEURAL GAS 74

FIGURE 5.7 – AN EXAMPLE OF THE STATISTICAL TESTS FOR BACKWARD REDUCTION FOR POWERHOUSE CD 75

FIGURE 5.8 – AN EXAMPLE OF THE STATISTICAL TESTS FOR NEURAL GAS FOR POWERHOUSE CD. 76

FIGURE 5.9 – MAXIMUM NUMBER OF START-UPS 78

FIGURE 5.10 – COMPARISON OF THE SOLUTIONS FOR BACKWARD REDUCTION, NEURAL GAS AND THE FULL TREE FOR DECEMBER WITH 10% OF REDUCTION 80

FIGURE 5.11 – COMPARISON OF THE SOLUTIONS FOR BACKWARD REDUCTION, NEURAL GAS AND THE FULL TREE FOR DECEMBER WITH 30% OF REDUCTION 81

FIGURE 5.12 – THE EXPECTED VOLUME OF THE RESERVOIR FOR THE FULL TREE, BACKWARD REDUCTION AND NEURAL GAS WITH 10% OF REDUCTION 81

FIGURE 5.13 – 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). 82

FIGURE 5.14 – 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) 83

FIGURE 5.15 – COMPARISON OF THE AVERAGE COMPUTATIONAL TIME FOR THE FULL SCENARIO TREE AND THE REDUCED SCENARIO TREES 84

FIGURE 5.16 – TOTAL ENERGY PRODUCED FOR STOCHASTIC METHODS VS MEDIAN METHOD 86

FIGURE 5.17 – A COMPARISON BETWEEN THE EXPECTED VOLUME OF THE RESERVOIR OF THE MEDIAN SCENARIO AND THE STOCHASTIC METHOD 87

FIGURE 5.18 – A COMPARISON BETWEEN THE DETERMINISTIC AND THE STOCHASTIC SOLUTION 89

FIGURE 6.1 – THE EFFICIENCY CURVES 99

FIGURE 6.2 – AN EXAMPLE OF THE FULL AND THE REDUCED SCENARIO TREES 103

FIGURE 6.3 – OPTIMIZATION MODELS 106

FIGURE 6.4 – BLACKBOX PROCESS. 110

FIGURE 6.5 – COMPARISON OF THE SOLUTIONS OBTAINED FROM THE AGGREGATED AND NON-AGGREGATED MODELS 116

FIGURE 6.6 – SOLUTIONS FOR STOCHASTIC, DETERMINISTIC, BBO1 AND BBO2 MODELS FOR THE DECEMBER 2021 CASE 120

FIGURE 6.7 – THE EXPECTED VOLUME OF INFLOWS FOR THE STOCHASTIC AND DETERMINISTIC MODELS FOR THE DECEMBER 2021 CASE 121

FIGURE 6.8 – THE EXPECTED VOLUME OF INFLOWS FOR THE STOCHASTIC AND DETERMINISTIC MODELS FOR THE SEPTEMBER 2021 CASE. 122

FIGURE 6.9 – THE EXPECTED VOLUME OF INFLOWS FOR THE STOCHASTIC AND DETERMINISTIC MODELS FOR THE OCTOBER 2021 CASE 123

FIGURE 6.10 – A BOXPLOT 124

FIGURE 6.11 – THE BOXPLOT FOR THE MEAN INFLOW FOR SEPTEMBER AND OCTOBER 2021 CASES FOR THE POWERHOUSE CS 125

REMERCIEMENTS

Je profite de ce rapport pour exprimer mes vifs remerciements à toutes les personnes qui ont contribué de différentes manières à la réussite de cette thèse.

Je tiens à remercier mes directeurs, Sara Séguin et Miguel F. Anjos, pour leur bonne humeur constante, leur expertise, leurs précieux conseils et leur confiance qui m'ont permis de mener à bien ce travail. Je ne saurais trouver les mots justes pour exprimer à quel point j'ai apprécié travailler avec vous. Je ne me suis jamais senti seul, vous avez toujours été disponibles, même pendant les moments les plus difficiles que j'ai traversés. Votre soutien moral restera gravé dans ma mémoire. J'espère que tous les encadrants seront aussi exemplaires que vous. Vous êtes un modèle pour moi.

Je souhaite remercier Kenjy Demenster, analyste en recherche opérationnelle chez Rio Tinto à Saguenay, pour son accueil chaleureux et son implication dans le projet. J'ai énormément appris de son expérience et de ses compétences.

Je tiens à exprimer ma gratitude envers mes chers parents pour leur soutien moral, ainsi que pour leurs sacrifices. Merci d'avoir fait de moi la personne que je suis aujourd'hui. Je remercie également ma sœur Nana et mes frères. Je tiens à exprimer une reconnaissance particulière et éblouissante envers mon cher mari. Ce simple remerciement ne suffira pas à exprimer toute ma gratitude. Si j'écris ma thèse aujourd'hui, c'est grâce à toi. Merci pour ton soutien, tes encouragements, ta motivation, ton aide et ton affection. Merci d'être présent dans ma vie. Je remercie également ma coquette "Nanou" d'avoir apporté de la joie à ma vie. Tes câlins le matin et ton sourire m'ont donné la force de devenir meilleure chaque jour.

Une pensé pour mes amis rencontrés à Chicoutimi, Asma, Rania, Imène, Souhail, Belgacem, Yessine, Amine et Rihana. Merci d'avoir animé mes années d'études. Un grand merci à mes plus chères amies, Wissal et Abir, d'être à mes côtés malgré la distance.

Enfin, je souhaite remercier toutes les personnes, de près ou de loin, qui ont contribué à cette réussite. Sans oublier à remercier moi-même d'être aussi forte au cours de ces années.

CHAPITRE I

INTRODUCTION

Au cours des dernières décennies, des efforts substantiels ont été faits afin d'orienter la production d'énergie vers des sources plus propres et renouvelables, en vue d'assurer une production d'énergie durable. L'hydroélectricité est l'une des principales énergies produites et consommées dans le monde en 2022 selon l'agence internationale d'énergie (IEA) [1], l'année la plus récente pour laquelle des données de production mondiale sont disponibles. Selon des statistiques, un taux de croissance annuel moyen de la production atteindra près de 4% entre 2023 et 2030. Le Canada est classé au troisième rang des producteurs d'hydroélectricité après la chine et le Brésil en 2022, avec une part mondiale de 8.9%. L'hydroélectricité représente plus de 60% de la production canadienne d'électricité grâce à ces vastes ressources hydroélectriques. Dans la province du Québec, 99% de ses besoins en électricité sont couverts par l'hydroélectricité [2]. Cette énergie est générée par des centrales hydroélectriques qui peuvent être de deux types : des centrales au fil de l'eau, qui ne disposent d'aucune réserve d'eau, et des centrales à réservoir, qui retiennent l'écoulement naturel de l'eau en créant un lac de retenue.

Pour les centrales à réservoir, lorsque l'eau est stockée, et en fonction des besoins, des vannes sont ouvertes au niveau du réservoir pour que l'eau achemine dans des tuyaux appelés conduite forcée. Ces tuyaux conduisent l'eau vers les centrales hydroélectriques situées en bas du réservoir. À la sortie de la conduite, la force d'eau fait tourner une turbine qui fait à son tour tourner un alternateur. L'interaction entre la partie stator, la partie rotor et les aimants de l'alternateur produit un courant électrique. Finalement, l'eau turbinée rejoint la ligne naturelle de la rivière par un canal appelé canal de fuite.

1.1 CONTEXTE

Les producteurs d'électricité cherchent souvent à améliorer la gestion de leur production en maximisant l'énergie produite ou en minimisant les coûts. Ces problèmes de gestion sont très complexes et dépendent de plusieurs facteurs avec plusieurs sources d'incertitudes. Afin d'assurer cette gestion, différents modèles d'optimisation sur différents horizons sont utilisés : des modèles à long terme, à moyen terme et à court terme, chacun étant résolu par des méthodes d'optimisation et des techniques dédiées.

L'optimisation à long terme permet de définir le potentiel de production future sur un horizon de plusieurs années en prenant en compte l'incertitude des apports. L'optimisation à moyen terme permet de gérer les réservoirs de manière hebdomadaire ou mensuelle sur une période annuelle, en prenant en compte plusieurs facteurs tels que les volumes des réservoirs, l'incertitude des apports, les demandes et les prix de l'électricité. L'optimisation à court terme permet de définir la stratégie optimale de production journalière ou horaire sur une période allant de quelques jours à quelques semaines, en indiquant la répartition de la quantité d'eau disponible dans les réservoirs entre les différentes centrales ainsi qu'entre les turbines de chaque centrale, dans le but de maximiser l'énergie produite ou de minimiser les coûts de production.

Cette thèse est dédiée aux problèmes d'optimisation à court terme en particulier chez Rio Tinto, un fabricant d'aluminium situé dans la région du Saguenay-Lac-St-Jean. La production d'aluminium étant une activité énergivore, Rio Tinto s'appuie sur un système hydroélectrique composé de six centrales électriques avec quatre réservoirs et 44 turbines, permettant de répondre à 90% de ses besoins énergétiques. La résolution des problèmes d'optimisation à court terme s'avère complexe en raison de la présence de multiples facteurs interdépendants, tels que la fonction de production, l'efficacité des turbines, les démarrages et les incertitudes.

Fonction de production

La fonction de production est une fonction non-lineaire et non-convexe qui dépend du débit

turbiné q (m^3/s), de l'efficacité de la turbine η, de l'accélération gravitationnelle G (m/s^2), de la densité ρ (kg/m^3) et de la hauteur de chute nette h_n (m) qui dépend du débit total turbiné et déversé Q (m^3/s) ainsi que du volume du réservoir v (hm^3) [3]. La puissance p (W) générée par une turbine peut être donnée par l'équation :

$$p = G \times \eta \times h_n(Q, v) \times q \times \rho. \tag{1.1}$$

La hauteur de chute nette h_n est la déférence entre l'élévation amont (h_f), l'élévation aval (h_t) et les pertes de charge associées au frottement d'eau dans la conduite forcée (h_p). La hauteur de chute nette est calculée par :

$$h_n(Q, v) = h_f(v) - h_t(Q) - h_p(Q, q). \tag{1.2}$$

La représentation de la production des centrales hydroélectriques dans un modèle mathématique est difficile, puisqu'elle dépend de fonctions non-linéaires et non-convexes.

L'efficacité des turbines

L'efficacité des turbines est le facteur le plus important en ce qui concerne la production de l'unité. C'est la capacité de la turbine à transformer la puissance mécanique de l'eau en une puissance électrique. L'efficacité dépend de la hauteur de chute nette et du débit turbiné. Chaque turbine a sa propre courbe d'efficacité, et pour la même valeur du débit turbiné, les turbines produisent une puissance différente.

Les démarrages des turbines

Les démarrages fréquents des turbines réduisent leur durée du vie et entraînent des coûts de maintenance. En hydroélectricité, à chaque démarrage, la turbine nécessite un certain temps afin d'atteindre son efficacité optimale, ce qui entraîne des pertes en termes d'énergie. Ces pertes ont été estimées par Rio Tinto et seront utilisées en tant que paramètre dans les modèles d'optimisation dans cette thèse.

Les incertitudes

Les incertitudes sont généralement liées aux apports naturels, qui comprennent les précipitations sous forme de pluies et de neige dans les réservoirs, au prix de l'énergie lorsqu'il s'agit d'un marché déréglementé (les marchés déréglementés sont des marchés ouverts à la concurrence où le prix d'énergie est fournis par la concurrence entre les fournisseurs), aux défaillances des machines et aux demandes. Considéré l'incertitude dans le développement des modèles d'optimisation permet d'avoir les solutions plus réalistes. Dans cette thèse, l'incertitude liée aux apports naturels est considérée, puisqu'au Québec où les modèles d'optimisations sont testés, le marché est réglementé par Hydro-Québec. Par conséquent, tous les producteurs doivent acheter, vendre et négocier des contrats à prix fixe chaque année.

Chargement des unités (Unit Commitment)

Le chargement des unités (Unit Commitment) est l'un des grands problèmes d'optimisation pour les systèmes hydroélectriques. Il permet de déterminer le calendrier de production pour chaque unité dans le système afin de maximiser l'énergie ou de minimiser les coûts [4].

1.2 OBJECTIFS DE LA THÈSE

Cette thèse, rédigée par articles, a pour objectif de développer des modèles d'optimisation à court terme pour le réseau hydroélectrique du Saguenay-Lac-St-Jean afin de maximiser l'énergie produite en pénalisant les démarrages des turbines et en prenant en compte l'incertitude des apports naturels.

Le réseau hydroélectrique du Saguenay-Lac-St-Jean est constitué de six centrales : Chute-Savane, Chute-du-Diable, Chute-des-Passes, Isle-Maligne, Chute-à-Caron et Shipshaw avec un total de 44 turbines. Quatre réservoirs sont disponibles avec des capacités différentes. Dans cette thèse, les deux centrales Chute-Savane et Chute-du-Diable sont considerées. L'objectif consiste à déterminer la meilleure stratégie de production d'électricité en proposant des solu-

tions qui reflètent plus la réalité opérationnelle et qui peuvent être appliquées directement en pratique. Ces solutions sont : le débit turbiné, le volume du réservoir, la puissance produite, et les turbines en marche pour chaque période de l'horizon de planification et pour chaque centrale.

Pour réussir cet objectif, un modèle d'optimisation déterministe à court terme est développé, en utilisant une nouvelle technique, dans le but de maximiser l'énergie totale produite tout en pénalisant les démarrages des turbines. Ce modèle est testé et comparé aux décisions opérationnelles historiques des centrales Chute-Savane et Chute-du-Diable. Ensuite un modèle d'optimisation stochastique est développé en utilisant la même technique novatrice, dans le but de prendre en compte l'incertitude des apports. Deux méthodes de génération et de réduction d'arbres de scénarios, permettant de représenter cette incertitude, sont mises en oeuvre et comparées pour évaluer l'impact du choix de la méthode de génération sur la solution. Enfin, une étude sur le choix du modèle d'optimisation à utiliser en se basant les previsions des apports afin de maximiser la production totale d'énergie est effectuée. Étant donné qu'il est possible d'utiliser à la fois le modèle déterministe, le modèle stochastique ou les deux, cette approche vise à déterminer le modèle d'optimisation le plus adapté tout au long de l'horizon de planification, en se basant sur les prévisions des apports naturels. Le problème est formulé en tant que problème d'optimisation en boîte noire.

1.3 TRAVAUX DÉCOULANT DE LA THÈSE

Au cours de cette thèse et de la réalisation des objectifs fixés, un article de conférence a été présenté lors de la conférence *International conference on optimization and learning (OLA2020)* qui s'était déroulé du 17 au 19 février 2020, à Cadix, Espagne. Le rapport technique de cet article est publié au cahier du GERAD. La référence est donnée ci-dessous :

- Daadaa, M., Séguin, S., Anjos, M., et Demeester, K. A linear mixed-integer formulation of the short-term hydropower problem. Les Cahiers du GERAD, G-2019-63, 2019.

Deux autres rapports techniques ont été publiés au cahier du GERAD. Les références complètes sont données ci-dessous :

- Daadaa, M., Séguin, S., Anjos, M., et Demeester, K. Quantifying the impact of scenario tree generation methods on the solution of the short-term hydroscheduling problem. Les Cahiers du GERAD, G-2022-47, 2022.
- Daadaa, M., Séguin, S., Anjos, M. F., et Demeester, K. Optimizing strategies for short-term hydropower scheduling using a blackbox optimization framework. Les Cahiers du GERAD, G-2023-38, 2023.

Trois articles scientifiques découlent de la thèse, deux ont été publiés dans les revues *International Journal of Electrical Power & Energy Systems* et *Energy systems*, et un soumis à la révue *IEEE Transactions on Power Systems*. Les références des articles publiés sont données ci-dessous :

- Daadaa, M., Séguin, S., Anjos, M., et Demeester, K. An optimization model to maximize energy generation in short-term hydropower unit commitment using efficiency points. International Journal of Electrical Power & Energy Systems, vol.125, p.106419, 2021.
- Daadaa, M., Séguin, S., Anjos, M. F., et Demeester, K. Quantifying the impact of scenario tree generation and reduction methods on the solution of the short-term hydro-scheduling problem. Energy Systems, 2023. Repéré à : https ://doi.org/10.1007/s12667-023-00623-x.

1.4 STRUCTURE DE LA THÈSE

Le deuxième chapitre présente une revue de la littérature concernant les différentes méthodes et techniques utilisées pour résoudre les problèmes d'optimisation à court terme.

Le troisième chapitre propose une présentation détaillée de l'organisation de la thèse. Les chapitres quatre, cinq et six sont dédiés à la présentation des articles de recherche. Enfin, une conclusion générale vient synthétiser les résultats obtenus, discuter des limitations de la solution proposée, et esquisser les perspectives de recherche futures.

CHAPITRE II
REVUE DE LITTÉRATURE

Plusieurs méthodes et modèles ont été développés afin de résoudre les problèmes d'optimisation hydroélectrique à court terme. Ces approches peuvent être de nature déterministe ou stochastique, et elles sont applicables aux marchés réglementés et déréglementés. La première partie de ce chapitre se concentre sur la présentation des principales méthodes utilisées pour résoudre les problèmes déterministes. La deuxième partie aborde les principales méthodes utilisées pour modéliser les problèmes stochastiques, afin de prendre en compte l'incertitude. Enfin, la dernière partie introduit le concept d'optimisation boîte noire, qui sera utilisée dans le cadre du dernier objectif de cette thèse. Les notions présentées dans ce chapitre sont essentielles à une meilleure compréhension des objectifs de cette thèse.

2.1 LES MÉTHODES DÉTERMINISTES

Plusieurs travaux de recherche ont été consacrés à la résolution des problèmes d'optimisation des systèmes hydroélectriques en utilisant des méthodes déterministes, où tous les paramètres sont considérés comme étant connus. Les fonctions objectifs varient en fonction de la problématique, mais elles sont généralement définies pour maximiser la production d'énergie, minimiser les coûts, ou encore maximiser les revenus dans le cas d'un marché déréglementé.

Dans le domaine de l'hydroélectricité, les méthodes les plus couramment utilisées pour résoudre les problèmes à court terme comprennent la programmation dynamique [3], la programmation linéaire en nombres entiers [5, 6], la relaxation lagrangienne [7], ainsi que d'autres approches telles que les algorithmes génétiques [8, 9].

2.1.1 LA PROGRAMMATION DYNAMIQUE

La programmation dynamique (PD) est une méthode algorithmique introduite par Richard Bellman en 1950 [10]. Elle consiste à décomposer le problème en sous-problèmes appelés *étapes*. Chaque étape n est caractérisée par différents *états* (S_n) du système. À chaque étape, une décision (X_n) est prise, et cette décision affecte l'état de la prochaine étape. Le principe fondamental de la PD consiste à relier les étapes les unes aux autres de manière récursive afin d'obtenir une solution optimale. Autrement dit, commencer à résoudre la dernière étape en reculant d'une étape à chaque fois, tout en stockant les résultats intermédiaires $f_n^*(S_n)$ jusqu'à ce que tous les sous-problèmes soient résolus. L'équation récursive de Bellman est la suivante [11] :

$$f_n(S_n, X_n) = f_{n+1}^*(S_{n+1}) + g_n(S_n, X_n) \ , \forall n \in N. \tag{2.1}$$

où $f_{n+1}^*(S_{n+1})$ représente le coût final qui dépend de l'état précédent et $g_n(S_n, X_n)$ représente le coût associé à chaque étape du système. Pour éviter une surcharge de calcul, la PD nécessite une discrétisation des espaces d'état. La manière de modéliser ces espaces peut varier d'un problème à l'autre.

Appliquée à l'optimisation en hydroélectricité, la PD est l'une des méthodes les plus utilisées pour résoudre les problèmes de chargement des unités. Dans l'article [3], cette technique est développée pour déterminer le débit turbiné à chaque période et pour chaque turbine selon un débit total Q et un volume du réservoir donné v. L'objectif principal est de maximiser la puissance produite P. Dans ce modèle, les étapes sont les turbines $j \in \{n, n-1, ..., 1\}$ où n est le nombre total des turbines, les états sont les débits restant à turbiner $S_j \in \{1, 2, ..., r\}$ et les variables sont les débits turbinés $q_j \in \{1,2,...,min\{\bar{q}_j, Q\}\}$ avec $\bar{q}_j$ le débit maximal turbiné par la turbine. La puissance optimale est calculée par $f_j^*(S_j) = Max_{q_j} f_{j+1}^*(S_j - q_j) + P(S_j, v)$.

Dans l'article [12], la PD est utilisée pour déterminer le nombre des turbines en marche à chaque période, dans le but de minimiser les coûts de production. Un compromis est établi entre l'efficacité de la production et les pertes occasionnées par les arrêts et les démarrages des turbines. Dans ce modèle, les étapes sont les heures, les états sont le nombre de turbines en marche à chaque période et les variables sont le nombre d'arrêts et de démarrages des turbines à chaque étape. Une autre approche [13] aborde le problème de chargement des unités dans un marché déréglementé en utilisant la PD. L'objectif est d'augmenter les revenus en prenant en compte le nombre de démarrages des turbines, l'efficacité des turbines et la variation de la hauteur de chute nette. Le modèle est testé sur un cas réel avec un seul réservoir. Les résultats obtenus montrent une augmentation des revenus variant entre 0,9% et 3,3%. En revanche, l'application du modèle proposé dans les systèmes avec des multi-réservoirs complexifie le problème, rendant la résolution difficile. De nombreux autres articles [14, 15] ont également utilisé la PD pour résoudre leurs modèles de planification à court terme.

La PD fonctionne bien dans les problèmes de chargement des unités et permet de déterminer des solutions optimales. Cependant, cette technique souffre d'une augmentation exponentielle du temps de calcul, rendant parfois la résolution impossible [16]. Cette augmentation du temps de calcul est principalement due à l'augmentation du nombre de variables, ce qui exige des ressources de calcul importantes. Afin de surmonter cette limitation, plusieurs travaux ont proposé de combiner la PD avec des heuristiques ou des méthodes de décomposition [17] afin de réduire les nombres d'états.

2.1.2 PROGRAMMATION MIXTE EN NOMBRES ENTIERS

Les problèmes d'optimisation hydroélectrique contiennent des variables continues telles que le volume (v) et le débit turbiné (q), ainsi que des variables entières (binaires) telles que l'arrêt/démarrage des turbines ou les zones d'opération interdites. Ces zones correspondent

aux périodes pendant lesquelles certaines unités ne peuvent pas fonctionner en raison de phénomènes de cavitation, de vibrations mécaniques ou de perte d'efficacité. La programmation mixte en nombres entiers est l'une des méthodes les plus efficaces pour résoudre ces problèmes d'optimisation à court terme [18], et qui peut être appliquée à la fois aux problèmes linéaires et non-linéaires.

LES MODÈLES MIXTES NON-LINÉAIRES EN NOMBRES ENTIERS

Les modèles mixtes non-linéaires ont été le sujet de plusieurs travaux de recherche. Les objectifs varient d'une étude à l'autre, mais dans la majorité des articles, ils visent à maximiser les revenus ou la production d'énergie [19, 20, 21]. Dans l'article [22], une approche non-linéaire mixte en nombres entiers est développée dans le but de maximiser les revenus tout en prenant en compte la hauteur de chute, les arrêts/démarrages des unités et les limites des débits turbinés. Le modèle est testé sur un système hydroélectrique au Portugal, et les résultats obtenus sont comparés à un modèle linéaire en nombres entiers dont la fonction de production varie linéairement seulement avec le débit. Les résultats montrent que l'approche proposée permet d'augmenter les revenus, mais le temps de calcul est relativement élevé par rapport au modèle linéaire en nombres entiers.

Dans l'article [23], une formulation quadratique en nombres entiers est proposée pour résoudre un problème de chargement des unités. Les contraintes des arrêts/démarrages, le bilan de la puissance et les contraintes quadratiques liées au carburant sont pris en compte. Le modèle est testé sur un cas réel au Mexique. Les résultats montrent l'efficacité de la méthode avec un temps du calcul assez acceptable avec 29933 contraintes et 31152 variables dont 26760 sont des variables binaires.

Un autre modèle non-linéaire mixte en nombres entiers est utilisé afin de maximiser la puissance produite [24]. Pour ce faire, les pertes dues aux arrêts/démarrages des unités et les

pertes de puissance dues au frottement d'eau dans les conduites forcées sont minimisées. Dans cette approche, les contraintes hydriques sont relaxées pour réduire le temps de calcul.

Les modèles d'optimisation non-linéaire en nombres entiers sont efficaces et permettent d'obtenir des solutions plus réalistes. Cependant, traiter les non-linéarités est une tâche très difficile et comporte des limitations, notamment en ce qui concerne la taille des instances et le temps de calcul. Par conséquent, plusieurs chercheurs ont opté pour l'utilisation de la programmation linéaire mixte en nombres entiers afin de faciliter la résolution des problèmes d'optimisation hydroélectrique [25, 26].

LES MODÈLES MIXTES LINÉAIRES EN NOMBRES ENTIERS

Les modèles mixtes linéaires en nombres entiers sont utilisés dans plusieurs travaux de recherche pour approximer la non-linéarité de la fonction de production. Dans [26], un modèle mixte linéaire en nombres entiers est utilisé afin de maximiser l'énergie produite en prenant en compte les zones d'opération interdites. La fonction de production est approximée à l'aide d'un modèle d'approximation linéaire par morceaux. Le modèle est testé sur des centrales hydroélectriques brésiliennes. Les résultats ont montré que la formulation proposée permet d'obtenir de bons résultats, mais pour des solutions plus précises, il est nécessaire d'augmenter le nombre de morceaux linéaires dans le modèle, ce qui entraîne une augmentation du temps de calcul. Dans [27], un sous-modèle linéaire par morceaux plus précis est proposé pour approximer la relation non linéaire entre la puissance produite, le débit turbiné et la hauteur de chute dans la fonction de production hydroélectrique. Le problème est résolu à l'aide d'un modèle mixte linéaire en nombres entiers afin de maximiser le profil tout en tenant compte des démarrages des turbines.

Dans [28], une technique d'interpolation est utilisée pour linéariser les fonctions de production

hydroélectrique en prenant en compte l'effet de la hauteur de chute. Le problème est formulé comme un modèle linéaire mixte en nombres entiers visant à minimiser les coûts de production. Les variables de décision sont les arrêts/démarrages des unités et le débit total turbiné. L'auteur a affirmé que lorsque le nombre de réservoirs augmente, ce modèle devient rapidement difficile à résoudre en raison du nombre accru de variables et de contraintes. Dans [29], un modèle linéaire mixte en nombres entiers est utilisé pour maximiser les revenus en minimisant des coûts des arrêts/démarrages et en prenant en compte l'effet de la hauteur de chute. Les variables continues sont le débit turbiné, le volume du réservoir, la puissance produite, le déversement d'eau, tandis que les variables binaires sont les arrêts/démarrages des unités. Une technique de linéarisation est utilisée pour résoudre la non-linéarité de la relation entre la puissance et le débit turbiné. Les résultats ont montré que cette technique permet d'obtenir une bonne représentation de la fonction de production, cependant, elle est difficilement applicable en pratique suite à la taille du problème à résoudre et dans certains cas la solution proposée peut être non réalisable en pratique comme par exemple, le non-respect des contraintes liées aux zones interdites. Plusieurs, autres travaux ont utilisé des modèles linéaires mixtes en nombres entiers pour résoudre les problèmes d'optimisation à court terme [30, 31, 32].

Les modèles d'optimisation linéaire en nombre entier simplifient la résolution du problème et permettent d'obtenir de bonnes solutions avec des temps de calcul raisonnables. Cependant la modélisation et la technique d'approximation revêtent une grande importance dans les modèles linéarisés car la solution peut être influencée par la précision, et dans certains cas, l'application des solutions en pratique devient difficile.

2.1.3 RELAXATION LAGRANGIENNE

La Relaxation Lagrangienne (RL) est une technique qui permet de relaxer les contraintes strictes qui rendent le problème compliqué, voire les supprimer. Les contraintes relaxées sont

intégrées dans la fonction objective sous la forme d'une pénalité avec un multiplicateur de Lagrange associé à chacune d'elles. La résolution repose sur la théorie de la dualité. À chaque itération, un sous-gradient du dual et une borne inférieure sont trouvés. La résolution du problème dual permet de trouver les valeurs des multiplicateurs, ce qui permet de trouver une solution optimale qui respecte toutes les contraintes.

Les problèmes d'optimisation hydroélectrique sont complexes à résoudre, car les fonctions de production dépendent de plusieurs variables : la hauteur de chute nette, l'efficacité des turbines, le débit turbiné pour chaque turbine et le débit total turbiné. De plus, l'inclusion de certaines contraintes telles que les démarrages, les *ramping* et les $down/up\text{-}time$ rendent le problème encore plus difficile à résoudre. Les contraintes du ramping assurent que la puissance produite par l'unité de production ne peut pas changer arbitrairement : l'augmentation de la puissance lors du démarrage est limitée par une contrainte de *ramp-up* et la diminution de la puissance lors de l'arrêt est limitée par une contrainte de *ramp-down* [4]. Les contraintes $down\text{-}time$ et $up\text{-}time$ expliquent le fait qu'une unité ne peut pas être allumée ou arrêtée de manière arbitraire. L'état de l'unité doit être maintenu pendant une certaine période avant de pouvoir être changé, car les variables sont interdépendantes. Plusieurs travaux ont utilisé une combinaison de la RL et d'autres techniques afin de résoudre ces problèmes. Dans l'article [33], une combinaison de la RL, de la programmation évolutionnaire et quadratique est effectuée afin de trouver une solution fiable. La résolution du problème est effectuée en deux phases. Dans la première phase, la RL est utilisée afin de déterminer un calendrier de production réalisable. Ce calendrier est ensuite amélioré par un algorithme de programmation évolutionnaire. Dans la deuxième phase, un algorithme de programmation quadratique est appliqué afin de respecter toutes les contraintes.

Dans d'autres travaux, des méthodes heuristiques et méta-heuristiques sont combinées à la RL pour trouver une solution réalisable et fiable. Dans l'article [34], la RL est utilisée pour déterminer un calendrier de production en respectant les contraintes de demande, mais

en négligeant les contraintes *down*/*up-time*. Ensuite, une heuristique est appliquée afin de raffiner la solution obtenue. Une autre approche [35] a fait appel à une méta-heuristique suite à l'application de la RL afin de trouver une solution réalisable proche de l'optimum en prenant en compte les contraintes de démarrage et de *up/down time*. Plusieurs autres travaux de recherches [4, 36, 37, 38, 39] ont utilisé la RL pour résoudre les problèmes d'optimisation hydroélectrique à court terme.

La RL est une technique qui permet de trouver rapidement de très bonnes limites inférieures et permet de résoudre des instances de très grandes tailles. Cependant, dans certains cas, la solution trouvée est sous-optimale [40], et dans d'autres cas, elle nécessite l'utilisation d'autres techniques ou d'heuristiques pour obtenir des solutions réellement réalisables. Parfois, ces heuristiques doivent être changées à chaque modification du modèle mathématique pour garantir la réalisabilité de la solution [41].

2.2 LES MÉTHODES STOCHASTIQUES

Les problèmes d'optimisation hydroélectrique sont très complexes et dépendent de plusieurs facteurs avec plusieurs sources d'incertitudes. Ces incertitudes sont généralement liées aux apports naturels qui sont les précipitations sous forme de pluies et de neige dans les réservoirs, au prix de l'énergie dans le contexte d'un marché déréglementé, aux demandes et aux défaillances des turbines. Les incertitudes doivent être prises en compte lors du développement des modèles d'optimisation afin d'obtenir des solutions plus pertinentes. Dans cette optique, plusieurs méthodes sont utilisées afin de résoudre ces problèmes. Puisque le deuxième objectif de cette thèse consiste à développer un modèle stochastique qui prend en considération l'incertitude des apports naturels, cette section est dédiée à présenter les différentes méthodes utilisées pour répondre à l'aspect stochastique comme la programmation stochastique, l'optimisation robuste ainsi d'autres techniques.

2.2.1 LA PROGRAMMATION STOCHASTIQUE : MODÈLES DE RECOURS

La programmation stochastique est la méthode la plus utilisée pour la prise de décision sous incertitude. Plusieurs modèles sont proposés dans la littérature afin de résoudre les problèmes d'optimisation à court terme. Les modèles à recours sont parmi les méthodes les plus répandues. Ces modèles sont généralement résolus en deux étapes ou en plusieurs étapes. La première étape consiste à prendre une décision immédiate avant de connaître la réalisation des incertitudes, en se basant sur les ressources actuelles. Cette décision sera ajustée une fois que les incertitudes sont connues [42]. Par exemple, en hydroélectricité, les quantités des apports sont incertaines. Les gestionnaires doivent prendre une décision sur la quantité des débits à turbiner sans connaître les valeurs des apports. Une fois que les vraies valeurs des apports sont disponibles, la décision prise dans l'étape précédente est ajustée, et un nouveau problème est résolu afin de trouver la prochaine décision, et ainsi de suite. Ou par exemple, dans les marchés dérèglementés, les producteurs d'électricité doivent fournir leurs propositions de prix et les volumes de leurs réservoirs avant midi de chaque journée sans connaître les prix et les quantités des apports futurs.

Dans l'article [43], un programme stochastique à plusieurs étapes est développé pour obtenir un plan de production efficace sur un horizon de 7 jours. Chaque étape correspond à une journée. La première étape consiste à définir un plan de production pour la première journée en appliquant un modèle linéaire mixte en nombres entiers. Ensuite, un programme stochastique est utilisé pour les 6 jours suivants. L'objectif est de trouver un plan de production pour la première journée qui permet d'avoir un équilibre entre les bénéfices actuels et les bénéfices futurs attendus. Les incertitudes des prix sont gérées par un arbre de scénarios. Dans l'article [44], un modèle stochastique à deux étapes est développé afin de maximiser les revenus en considérant l'incertitude des apports naturels et des prix. Dans la plupart des travaux de recherche, les incertitudes sont représentées par des arbres de scénarios.

L'ARBRE DE SCÉNARIOS

L'arbre de scénario est un outil utilisé pour gérer l'incertitude dans de nombreuses situations. La Fig.2.1 montre un exemple d'une représentation graphique d'un arbre avec 4 scénarios. À chaque étape, une information est disponible, et une décision est prise. Les étapes ne sont pas nécessairement des périodes de temps.

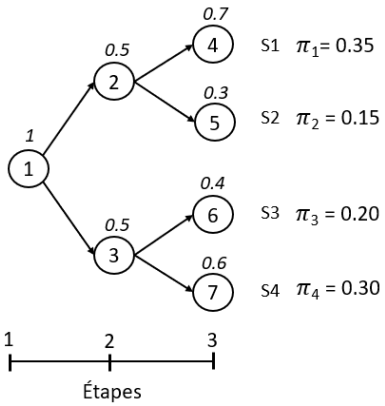


FIGURE 2.1 : Exemple d'un arbre de scénarios

Un scénario $s \in \{1,2,3,..,S\}$ est constitué du plusieurs arcs et noeuds [45]. Chaque noeud $n \in \{1,2,3,..,N_s\}$ du scénario s représente un état possible résultant d'un ensemble de données. Les arcs ou les branches représentent les scénarios pour les prochaines étapes. Par exemple, dans la Fig.2.1, chaque noeud a deux branches ce qu'indique l'existence de deux scénarios possibles pour l'étape suivante. Les noeuds finaux $l \in \{1,2,3,..,L_s\}$ (les noeuds $n = 4$ et $n = 5$ dans la Fig.2.1) sont appelés noeuds feuilles pour chaque scénario s. Une probabilité, qui définit la possibilité de réalisation est attribuée à chaque noeud. Étant donné qu'un scénario est constitué d'un ensemble de noeuds, une probabilité π_s est calculée pour chaque scénario. Cette probabilité est égale aux produits des probabilités des noeuds qui constituent ce scénario. Par

exemple, dans la Fig.2.1, la probabilité du scénario $s1$ est $\pi_1 = 0.35$, calculée comme étant le produit de 0.5*0.7. La somme des probabilités de tous les scénarios est égale à 1 ($\sum_s^S \pi_s = 1$).

GÉNÉRATION ET RÉDUCTION D'ARBRE DE SCÉNARIOS

Trouver une représentation précise pour l'incertitude est une étape cruciale dans l'optimisation stochastique en raison de son influence sur les résultats du problème. Différentes méthodes de génération sont utilisées comme les méthodes de partitionnement de données, les méthodes des moments, et plusieurs autres techniques.

Le partitionnement de données (*Clustering*) : est une technique d'analyse des données qui consiste à organiser les données en des paquets homogènes (*clusters*). Chaque paquet regroupe les données ayant des caractéristiques communes. Le principe de cette méthode est utilisé dans la génération de l'arbre afin de représenter l'incertitude des apports naturels ou/et des prix. Dans l'article [46], un arbre de 3038 scénarios sur 30 jours de prévision est défini en se basant sur les données historiques et les prévisions météorologiques. Afin de générer cet arbre, une méthode de partitionnement des données appelé les k-moyennes ($k-means$) a été développée. Les données sont regroupées en des *clusters* de façon que la distance entre chaque point de données et la moyenne du *cluster* auquel il appartient soit minimisée. Ces clusters sont améliorés par une méthode d'approximation stochastique et un critère de convergence basé sur la distance de *Wasserstein*. Un test de stabilité a montré que la méthode de génération proposée est cohérente. Dans [47], quatre autres méthodes de génération : le partitionnement de données conditionnel, le partitionnement de noeuds, le partitionnement progressif et la méthode du gaz neuronal sont présentés. Ces quatre méthodes sont comparées pour déterminer la meilleure méthode permettant d'obtenir un arbre de scénarios correspondant le plus à la distribution initiale et les résultats obtenus ont montré que la méthode de gaz neuronal performe les autres méthodes. La méthode du gaz neuronal est une méthode basée sur les réseaux

de neurones. Cette méthode consiste à définir la structure d'arbre de scénarios à l'entrée et attribuer une valeur aléatoire pour chaque noeud en se basant sur les données historiques. À chaque itération, une nouvelle série est choisie aléatoirement à partir des données et une évaluation entre la distance de l'arbre et la série choisie est effectuée. Les scénarios sont triés en ordre croissant selon leur distance. Par la suite, les valeurs des noeuds de chaque scénario seront modifiées selon l'ordre et l'itération en utilisation une équation d'adaptation jusqu'à ce que la distance globale soit minimisée. Les probabilités sont attribuées à chaque scénario comme la proportion des séries choisies aléatoirement et qui sont les plus proches au scénario. Dans [48], la méthode du gaz neuronal est utilisée pour représenter l'incertitude des apports. Pour ce faire, un arbre de scénarios des apports $\{w^i\}$,$\forall\ i = \{1,2,3..I\}$ est défini où I est le nombre total des scénarios. Chaque scénario contient une séquence des noeuds $\{w_t^i\}$, $\forall\ t = \{1,2,3..T\}$ où T est le nombre total des périodes. Les valeurs des noeuds sont initialisées en sélectionnant des valeurs aléatoires à partir d'un ensemble des données historiques IN : $w_t^i = IN_{rand(),t}, \forall i \in \{1,..,I\}, \forall t \in \{1,..,T\}$ avec $rand()$ un nombre aléatoire uniformément distribué dans la plage $[1,K]$, et K est le nombre total de séries des apports historiques annuels. À chaque itération, une nouvelle série d'apports IN^k est sélectionnée aléatoirement à partir de IN. La distance entre le scénario i à la série k est calculée par la distance euclidienne :

$$d_{i,k} = \sum_{t \in T} ||IN_t^k - w_t^i||\ \forall i \in \{1,..,I\},\ \forall k \in \{1,..,K\}. \tag{2.2}$$

Par la suite, ces scénarios sont triés et enregistrés dans un tableau de distance O. À chaque itération j et selon l'ordre du scénario dans le tableau O, les valeurs des scénarios sont mises à jour par une valeur d'adaptation Δw_t^i :

$$\Delta w_t^i = \varepsilon(j).h_\lambda(O_i).(IN_t^k - w_t^i) \tag{2.3}$$

$$\varepsilon(r) = \varepsilon_0.(\varepsilon_f/\varepsilon_0)^{j/j_{max}} \tag{2.4}$$

$$h_\lambda(O_i) = e^{-(O_i/\lambda)} \tag{2.5}$$

$$\lambda = \lambda_0.(\lambda_f/\lambda_0)^{j/j_{max}} \tag{2.6}$$

$$w_t^{i,j+1} \leftarrow w_t^{i,j} + \Delta w_t^i. \tag{2.7}$$

L'algorithme s'arrête lorsque le temps d'itération atteint un seuil prédéterminé j_{max}. La probabilité de chaque scénario est calculée en fonction du nombre de séries qui ont une distance minimale par rapport à ce scénario :

$$\pi_i = Count\{k' \in [\,1,K\,]\,, k' | d_{i,k'} = \min_{i' \in [\,1,k]} \{d_{i',k'}\}\}\backslash K. \tag{2.8}$$

Où π_i est la probabilité du scénario i et $Count$ est une fonction de comptage. Les résultats obtenus montrent que l'arbre de scénarios convergera progressivement vers le centroïde des échantillons choisis. Plusieurs autres travaux ont utilisé la méthode du gaz neuronal générer l'incertitude [48, 49].

Une autre technique intitulée **la méthode des moments** est utilisée pour générer l'arbre de scénarios en utilisant des propriétés statistiques (espérance, variance, etc.). Cette technique vise à approximer la distribution de l'incertitude afin que les moments statistiques d'arbre de scénario correspondent aux moments statistiques des échantillons observés [50]. Les propriétés statistiques sont choisies par les utilisateurs selon leurs jugements. Dans l'article [51] les apports naturels sont considérés comme des variables aléatoires suivant une distribution continue inconnue. L'objectif consiste à trouver une représentation discrète de la distribution de ces variables de sorte que les quatre propriétés statistiques (espérance, variance, asymétrie, kurtosis) des approximations discrètes et les distributions continues sont similaires. Pour ce faire, une méthode qui minimise la distance entre les moments statistiques de l'approximation discrète et les moments statiques de la distribution sous-jacente est effectuée. La même

méthode a été présentée dans l'article [52], cependant, les auteurs ont indiqué que dans certains cas il est extrêmement difficile à faire correspondre les propriétés statistiques suite à l'incohérente ou à la non-convexité du problème. D'autres approches ont traité le problème de génération de l'arbre de scénarios en utilisant un modèle auto-régressif ARIMA (Auto-Regressive Integrated Moving Average) afin de représenter l'incertitude des prix [53, 54], tandis que d'autres utilisent la méthode de Monte-Carlo [55, 56].

Dans la programmation stochastique avec les modèles à recours, le temps de calcul augmente avec l'augmentation du nombre de scénarios et de nombre d'étapes. Par conséquent, la réduction de l'arbre de scénario peut être un moyen efficace pour réduire la charge de calcul. La réduction de la taille de l'arbre de scénarios entraîne une perte d'informations, car les scénarios réduits peuvent modifier les caractéristiques des échantillons observés. Pour cela, différentes techniques de réduction sont utilisées afin de trouver un compromis entre le temps de calcul et la précision de la solution. La méthode du gaz neuronal est utilisée aussi pour réduire la taille de l'arbre. Dans [57, 48] cette méthode est utilisée pour générer et réduit la taille de l'arbre en utilisant différents niveaux de réduction (10%,20%,..,90%). Une comparaison entre les différents arbres réduits (comparaison de la dimensionnalité du problème et les valeurs de la fonction objective) est faite. Les résultats ont montré qu'à partir de 40% de réduction la précision de solution se dégrade.

Une autre méthode est utilisée dans la littérature pour réduire la taille de l'arbre de scénario en préservant le plus d'informations possible est la méthode *backward reduction*. Cette méthode vise à définir un sous-arbre ξ_i pour $i \in \{1,\ldots,S^*\}\backslash I$ avec une probabilité π en supprimant des scénarios d'un arbre de scénarios initial ξ_s pour $s \in \{1,\ldots,S^*\}$ avec une probabilité P de manière à minimiser la distance de distribution de probabilité entre l'arbre réduit et initial. Le problème à résoudre est de trouver l'ensemble I de scénarios à supprimer en fonction d'une distance de probabilité entre P et π. Dans le contexte des modèles de gestion de l'énergie [58],

la distance de Kantorovich $D(P,\pi)$ est utilisée et calculée par Eq.(2.9) et Eq. (2.10) :

$$D(P,\pi) = \sum_{s\in I} p_s \min_{i\notin I} C_T(\xi_s,\xi_i) \tag{2.9}$$

$$C_T(\xi_s,\xi_i) = \sum_{t\in T} \parallel \xi_{s_t} - \xi_{i_t} \parallel . \tag{2.10}$$

où p_s est la probabilité du scénario évalué, $I \subset \{1,\ldots,S^*\}$ est l'ensemble des scénarios supprimés et C_T est la fonction qui mesure la distance entre deux scénarios ξ_s, $s \in \{1,\ldots,S^*\}$ et ξ_i, $i \in \{1,\ldots,S^*\}\backslash I$ sur l'horizon temporel $t \in T$.

L'idée est de comparer la distance de Kantorovich de l'arbre réduit et de l'arbre initial sur l'horizon $\{1,..,t\}$, $t = T$, $T-1$,...,2, 1, et de supprimer des scénarios si l'arbre réduit est encore assez proche de l'arbre initial avec une certaine précision donnée ε_τ comme indiqué dans Eq.(2.11). Cette précision est définie comme le pourcentage de réduction. Ce pourcentage spécifie la réduction souhaitée en termes de distance entre les arbres de scénarios initial et réduit. Par exemple, si le pourcentage de réduction est de 10%, cela signifie que la distance entre l'arbre réduit et l'arbre complet est inférieure à 10%, en revanche, l'arbre réduit conserve 90% des informations contenues dans l'arbre complet. La probabilité π_i du scénario préservé ξ_i, $i \notin I$ est égale à la somme de son ancienne probabilité p_i et des probabilités p_i des scénarios supprimés qui lui sont les plus proches, comme montré dans Eq.(2.12) et Eq.(2.13) :

$$\sum_{s\in I} p_s \min_{i\notin I} C_T(\xi_s,\xi_i) < \varepsilon_\tau \tag{2.11}$$

$$\pi_i = p_i + \sum_{s\in I(i)} p_s \;\forall i \notin I \tag{2.12}$$

$$I(i) = \{s \in I : i = i(s)\},\ i(s) \in \operatorname*{argmin}_{i\notin I} C_T(\xi_s,\xi_i)\ \forall s \in I. \tag{2.13}$$

L'avantage de cette méthode est que son implémentation se fait facilement en utilisant l'outil *SCENRED*. Cet outil fait partie de la bibliothèque *GAMS* [59]. Plusieurs travaux de recherche [60, 61] ont utilisé cet outil.

2.2.2 OPTIMISATION ROBUSTE

L'optimisation robuste évite de définir des scénarios et de faire des hypothèses sur leurs probabilités. Les incertitudes sont représentées dans un ensemble déterministe borné en utilisant certaines informations comme la variance et l'espérance [4]. L'objectif est de trouver une solution réalisable, peu importe la réalisation, en particulier, la pire des cas.

En hydroélectricité, l'optimisation robuste a été un sujet d'intérêt de plusieurs travaux de recherche pour résoudre des problèmes de chargement des unités avec des paramètres incertains : les demandes [62], les prix [63, 64] et les pannes des unités [65]. Dans [66], un modèle robuste pour minimiser les coûts en prenant en compte l'incertitude de l'élasticité de la demande au prix (c'est une mesure qui montre l'élasticité de la quantité demandée lorsque les prix sont changés) a été développé. Le modèle proposé est comparé avec un modèle déterministe et les résultats obtenus ont montré l'efficacité de la méthode. Dans [67], une comparaison entre un modèle stochastique basé sur les scénarios et un modèle robuste est effectuée. Les résultats montrent que les solutions obtenues par le modèle stochastique sont plus stables, cependant, le temps de calcul est plus élevé comparant au modèle robuste. De plus, les solutions obtenues par le modèle robuste sont sensibles et dépendent de l'intervalle d'incertitude.

2.2.3 AUTRES MÉTHODES

Pour résoudre les problèmes d'optimisation avec incertitude, certaines contraintes peuvent être tellement contraignantes que le modèle devient non-réalisable. Dans ces cas, les modèles avec contraintes probabilistes peuvent être utilisés en assignant une probabilité à certaines ou à toutes les contraintes. Ces modèles sont peu utilisés dans les problèmes d'optimisation hydroélectrique à court terme [68, 69]. Il existe d'autres méthodes comme la PD

stochastique [70], la PD duale stochastique [60] et la programmation stochastique dynamique par échantillonnage [71] qui traitent l'incertitude, mais ces méthodes sont plutôt appliquées dans l'optimisation à moyen terme.

2.3 L'OPTIMISATION BOÎTE NOIRE

L'optimisation boîte noire est un processus qui permet de trouver un optimum d'une fonction qui ne peut être calculée qu'à partir d'un code informatique, cependant, la structure interne de la boîte noire n'est pas disponible analytiquement. Un modèle d'optimisation boîte noire peut s'écrie :

$$min_{x\in\Omega}\ f(x). \tag{2.14}$$

où f est la fonction qui doit être minimisée. Les méthodes de recherche directes sont souvent utilisées pour traiter les problèmes de boîtes noires. Ces méthodes permettent de guider la recherche en se basant sur la valeur de f et l'évaluation des contraintes. Pour ce faire, plusieurs algorithmes sont développés.

Le premier algorithme est inventé en 1952 par Fermi et Metropolis sous le nom la recherche par coordonnées (CS) [72]. C'est une méthode itérative qui consiste à discrétiser l'espace d'états en un treillis et de générer à chaque itération $k \in \{0,1,2,..\}$ une liste de points P_k aux alentours des itérés dans un rayon prédéfini Δ_k afin de trouver un point qui améliore la fonction objective. Cette étape est appelée *la sonde locale*. La méthode consiste à définir, initialement, un paramètre de taille du treillis Δ_0 et un point x_0 et d'évaluer la fonction objective de ce point. À chaque itération k, la fonction f est évaluée sur $2n$ points de sonde qui se situent dans les directions de base de $\mathbb{R}^n$ avec une distance Δ_k de l'itéré courant x_k appelé centre de la sonde. L'ensemble des points de sonde est défini alors de l'équation suivante :

$$P_k = \{x_k \pm \Delta_k e_i : i = 1,2,..,n\}. \tag{2.15}$$

Où e_i est i-ème colonne de la matrice identité I_n.

Le principe de la méthode CS est d'évaluer tous les points de l'ensemble de sonde P_k, si l'un de ces points permet d'améliorer la valeur de la fonction objective, il sera défini comme meilleure solution. Dans le cas où aucun point n'améliore la fonction objective, le paramètre de taille du treillis Δ_{k+1} sera la moitié de Δ_k et une nouvelle itération est effectuée. L'algorithme s'arrête selon une condition d'arrêt prédéfinie. Cette condition peut être : une valeur limite de la taille du treillis, un temps d'exécution bien défini, un certain nombre d'itérations, etc.

Cette méthode est simple à utiliser et peut-être appliquer sur les problèmes d'optimisation sans contraintes ou sans aucune hypothèse sur la fonction objective, mais les directions sont limitées, et dans certains cas l'algorithme CS peut rester bloqué au point du départ [73].

En 1997, Torczon [74] développe une méthode de recherche généralisée par motifs (Gps : Generalized Pattern Search). Cet algorithme permet d'augmenter la flexibilité de l'algorithme (CS) en explorant davantage l'espace des solutions. Une étape de recherche globale flexible est ajoutée avant la sonde locale. Au débit le treillis est défini par :

$$M_k = \{x + \Delta_k D_z : x \in v_k,\ z \in \mathbb{N}^{\mathrm{l}}\}. \tag{2.16}$$

où v_k est l'ensemble des points déjà évalués au début de l'itération k, Δ_k est le paramètre de taille de treillis et D est l'ensemble des directions possibles. Tous les points évalués qui que ce soit évalués en phase de recherche globale ou pendant la sonde locale sont évalués et sont situés sur le treillis. Ensuite, le paramètre du treillis est mis à jour suivant l'équation Eq.2.17 :

$$\Delta_{k+1} = \tau^{w_k} \Delta_k. \tag{2.17}$$

Où $\tau \in \mathbb{Q}$ et w_k est un entier positif si une nouvelle solution est trouvée à l'itération k. Sinon

w_k est négatif. Contrairement à l'algorithme (CS), la recherche par motif permet d'augmenter la taille du treillis Δ_{k+1} lors d'un succès. Finalement, un sous-ensemble du treillis est défini par :

$$P_k = \{x_k + \Delta_k d_k : \ d \in D_k\}. \tag{2.18}$$

La méthode est relancée jusqu'au atteindre un critère d'arrêt prédéfini. L'inconvénient de cette méthode et que l'algorithme GPS utilise un ensemble fini de direction dans sa recherche et dans certains cas, la solution obtenue peut être un optimum local.

Une autre méthode de recherche directe plus riche que l'algorithme GPS est la recherche directe par treillis adaptatifs (MADS) [75]. La résolution du cet algorithme est effectuée en deux étapes : la première étape consiste à explorer l'espace des solutions afin de trouver la meilleure solution. Si la recherche est réussie, la taille du treillis est augmentée. Sinon, une deuxième étape (sonde locale) est effectuée. Pour ce faire, le treillis est défini par :

$$M_k = \{x + \Delta_k^m D_z : x \in v_k, \ z \in \mathbb{N}^{n_D}\}. \tag{2.19}$$

où $\Delta_k^m \ \forall \ k \in \mathbb{R}^+$ est le paramètre de taille du treillis à l'itération k, v_k est l'ensemble des points déjà évalués au début de l'itération k et $D \in \mathbb{R}^{n \times n_D}$ est la matrice composée des directions de recherche. D est souvent prise comme la matrice [I_n - I_n] où I_n est la matrice identité de taille n. Ensuite, une sonde qui utilise un sous-ensemble de directions est définie afin de sélectionner des points à une distance inférieure à Δ_k^p qui est la taille du cadre de sonde. Si la recherche globale ou la sonde locale réussissent d'avoir une solution meilleure que la solution courante, la taille du treillis Δ_k^m et la taille du cadre de la sonde Δ_k^p sont augmentées. Sinon, si la sonde locale échoue à trouver une meilleure solution, les tailles sont diminuées (voir Fig.2.2).

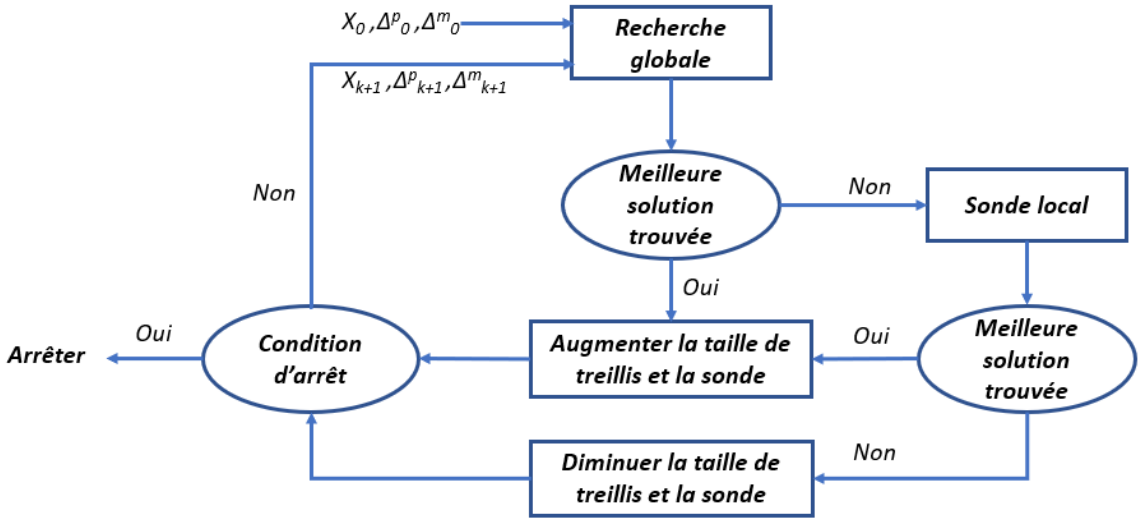


FIGURE 2.2 : Algorithme MADS

L'implémentation de l'algorithme est effectuée sur le logiciel *NOMAD* [76]. L'algorithme *MADS* permet de converger plus rapidement avec une solution plus performante puisqu'il y a deux paramètres la taille du treillis Δ_k et la taille de la sonde δ_k contrairement à l'algorithme GPS.

En hydrologie, l'optimisation boîte noire est utilisée pour trouver l'emplacement optimal des capteurs *GMON* afin d'estimer l'équivalent en eau de la neige au sol pour une gestion efficace des barrages hydroélectriques [77]. Une autre approche utilise la boîte noire pour trouver les valeurs des apports naturels présentés dans l'arbre qui maximisent l'énergie produite. Pour ce faire, l'algorithme *MADS* a été appliqué et l'implémentation a été effectuée par *NOMAD* [78]. La même technique est utilisée dans l'approche [79] afin de calibrer le modèle hydrologique pour six bassins canadiens utilisés par Hydro-Québec. Deux critères d'arrêt ont été utilisés dans cette étude : le nombre maximal d'évaluations (20000) et la limite de la taille de la sonde.

CHAPITRE III
ORGANISATION DE LA THÈSE

Cette thèse, rédigée par articles, a pour objectif de développer des modèles d'optimisation à court terme pour le réseau hydroélectrique du Saguenay-Lac-St-Jean chez Rio Tinto. Comme présenté dans le chapitre précédent, différentes méthodes sont utilisées pour résoudre les modèles d'optimisation déterministe et stochastique. Cependant, il a été observé que dans certains travaux, les solutions obtenues peuvent être difficiles à appliquer sur le terrain. Dans ce contexte, Rio Tinto souhaite disposer d'un modèle qui lui permette d'obtenir des solutions directement applicables dans ses centrales, en tenant compte des arrêts et des démarrages des turbines, ainsi que de l'incertitude liée aux apports.

Dans un premier temps, un modèle d'optimisation déterministe à court terme est développé. Ce modèle est formulé comme un problème linéaire mixte en nombres entiers. Son objectif principal est de maximiser la production totale d'énergie en utilisant une nouvelle technique. Des contraintes sur le nombre maximal de démarrages sont imposées afin de garantir la faisabilité des solutions. Ce modèle est ensuite testé et comparé aux décisions opérationnelles historiques des centrales Chute-Savane et Chute-du-Diable. Les résultats de ces travaux ont été publiés dans la revue *International Journal of Electrical Power & Energy Systems* et sont présentés dans le chapitre 4 de cette thèse.

En deuxième lieu, le modèle d'optimisation déterministe est adapté afin de considérer l'incertitude liée aux apports naturels. Un modèle d'optimisation stochastique à court terme est développé. Pour représenter l'incertitude des apports, deux méthodes de génération et réduction de l'arbre de scénarios sont utilisées. Les deux méthodes sont comparées afin d'étudier l'impact du choix de la méthode sur la solution. Pour ce faire, les prévisions d'apports sont mise à

jour quotidiennement et les arbres de scénarios sont générés et réduit. Le modèle est résolu et seule la solution du premier noeud de l'arbre de scénarios est considérée. Une fois les vrais apports sont connus, les volumes des réservoirs sont mis à jour et le processus est répété durant l'horizon de planification. Les résultats obtenus des deux méthodes sont comparés, et en outre, une comparaison est effectuée entre le modèle stochastique et le modèle déterministe en utilisant le médian des apports. Les résultats sont présentés dans la revue *Energy Systems, Special Issue on Hydropower Scheduling* et se trouvent au chapitre 5 de cette thèse.

En troisième lieu, une étude sur le choix du modèle d'optimisation à court terme est effectuée. L'objectif est de déterminer le meilleur modèle d'optimisation à utiliser (déterministe, stochastique ou les deux) sur un l'horizon de planification, dans le but de maximiser la production totale d'énergie tout en réduisant le temps de simulation. Le problème est formulé comme un problème d'optimisation boîte noire. Dans un premier temps, une agrégation du modèle stochastique est effectuée afin de réduire le temps de calcul. Ensuite, une comparaison est réalisée entre le modèle agrégé et le modèle non agrégé afin d'évaluer la qualité de la solution. La représentation des apports est effectuée en utilisant une méthode de génération backward reduction pour le modèle stochastique, tandis que le scénario médian de tous les scénarios possibles est utilisé pour le modèle déterministe. Seulement la solution du premier noeud de l'arbre est conservée. Les résultats fournis par la boite noire sont analysés pour déterminer comment choisir le meilleur modèle d'optimisation tout au long de l'horizon de planification. Les résultats sont présentés dans un article soumis à *IEEE Transactions on Power Systems* et se trouvent au chapitre 6 de cette thèse.

Finalement, une conclusion générale, accompagnée d'une analyse des limites et des perspectives pour améliorer les travaux futurs sont présentées.

CHAPITRE IV

ARTICLE 1 : AN OPTIMIZATION MODEL TO MAXIMIZE ENERGY GENERATION IN SHORT-TERM HYDROPOWER UNIT COMMITMENT USING EFFICIENCY POINTS

Cet article est publié :

Daadaa, M., Séguin, S., Demeester, K., & Anjos, M. F. (2021). 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.

Abstract : This paper presents a linear mixed-integer formulation to solve the short-term hydropower unit commitment problem. It uses the pair of maximum efficiency points of water discharge and the power produced at the maximum storage for all possible combinations of turbines. The goal is to maximize total energy production for all periods. The objective function is calculated using the correction between the power produced at the current volume and the maximum storage and penalizes unit start-ups. Constraints on the maximum number of turbine changes are imposed to find a viable solution. Computational results are reported for 65 instances with two powerhouses of five turbines each located in the Saguenay-Lac-St-Jean region of the province of Quebec in Canada.

Key words : Hydro unit commitment, short-term hydropower optimization, mixed integer linear programming.

NOMENCLATURE

Sets

$t \in \{1,2,3,..,T\}$: set of periods.

$c \in \{1,2,3,..,C\}$: set of powerhouses.

$l \in \{1,2,3,..,U^c\}$: set of powerhouses upstream of each powerhouse c.

$j \in \{1,2,3,..,J_t^c\}$: set of turbines associated to period t and powerhouse c.

$b \in \{1,2,3,..,B^c\}$: set of combinations of each powerhouse c.

$k \in \{1,2,3,..,K_c^b\}$: set of points (the maximum and the adjacent points) associated to powerhouse c and combination b.

Parameters

$P_{k,t}^c$: power output of powerhouse c at period t and point k (MW)

$q_{k,t}^c$: water discharge of powerhouse c at period t and point k (m^3/s).

δ_t^c : inflow of powerhouse c at period t (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).

Δ_t : the duration of period t (h).

$$A_{t,k,j}^c = \begin{cases} 1 & \text{if the turbine } j \text{ of powerhouse } c \text{ at the point } k \text{ is activated at period } t. \\ 0 & \text{otherwise.} \end{cases}$$

Variables

$$y_{k,t}^c = \begin{cases} 1 & \text{if the point } k \text{ is chosen at period } t \text{ for powerhouse } c. \\ 0 & \text{otherwise.} \end{cases}$$

$$z_{j,t}^c = \begin{cases} 1 & \text{if the turbine } j \text{ of powerhouse } c \text{ is started at period } t. \\ 0 & \text{otherwise.} \end{cases}$$

v_t^c : volume of the reservoir of powerhouse c at period t (hm^3).

d_t^c : water spillage at powerhouse c and period t (m^3/s).

4.1 INTRODUCTION

4.1.1 MOTIVATION

Hydropower is the third source of renewable energy in the world according to the International Energy Agency (IEA) in 2017 [80]. Electricity producers seek to manage their production either by maximizing the energy production or minimizing the operational cost. Managing hydroelectric systems is complex and requires different optimization processes. Long-term optimization models are used to determine the future production potential for a few years of the planning horizon and take into account the uncertainty of the inflows [81, 82]. Medium-term models are used to define the quantity of water available in the reservoir for hydropower production for a weekly scheduling horizon [83, 84]. Short-term models aim at defining the optimal strategy of the daily production by dispatching the quantity of water between the turbines in order to maximize the energy produced using the power production function or to minimize costs [3].

This paper focuses on the short-term model which determines for each powerhouse and for each time stage the optimal water discharge, the volume of the reservoirs, and the state of each unit (on or off) while respecting some constraints. For the short-term problem, these constraints usually involve water balance constraints, energy demand, water discharge constraints, reservoir limits and start-ups of the units [85].

4.1.2 LITERATURE REVIEW

The study of the unit commitment problem has taken place during the past few decades. Several methods and algorithms can be used to solve this problem : dynamic programming

[86, 3], Benders decomposition [87], heuristic methods like genetic algorithms and particle swarm optimization [9, 88] and sequential quadratic programming [89] among others. Many unit commitment problems are formulated by Mixed Integer Linear Programming (MILP). The advantage of MILP is that the discrete nature of the state of the units can be modelled using integer variables [90]. In addition, MILP can allows large size scheduling problems in energy power systems [91].

Because the hydroelectric production function are non-linear and non-convex, different techniques are used to handle this non-linearity. In [28] a MILP was proposed to solve the hydropower unit commitment problem in order to minimize the operational costs. In this model, the decision variables were the start-up and shutdown of the units and the water releases. The formulation of the problem considers the variation of the net head water and the non-linearity of the production function. This non-linearity was accounted for with a three-dimensional interpolation technique. The model was tested on a real case in China with one reservoir and 32 heterogeneous generating units. Another formulation in [26] uses MILP to maximize power efficiency. The problem was split into two phases. The first phase was a preprocessing based on unit commitment and a piecewise linear generation function to define the water discharge and the total powerhouse generation, considering the maximum discharge bound, the efficiency curves and restricted operating zones for each unit. In the second phase, a MILP formulation with fewer binary variables was built based on the preprocessing phase in order to maximize the final storage energy.

In [29] a unit commitment problem with head dependent reservoir was developed to find the optimal scheduling of a multiunit pump-storage hydropower system. A mixed integer linear model was formulated to solve the problem. The continuous variables are the water discharge, the volume of the reservoir, the produced power, the water spillage and the binary variables are the start-ups and shutdowns of the units. An enhanced linearization technique was used to solve the non-linearity of the relationship between power and water flow. However, this

technique was difficult to apply in practice, given the size of the problem.

Previous research used MILP to facilitate the introduction of the integer variables and the linearization of non-linear functions. In [92] a MILP was developed to maximize the profits by estimating the income of produced power and the start-up/shut-down costs. An analysis of the linearization effects of non-linear functions and related constraints on solution feasibility was conducted. It was found that the linearization may result in solutions that cannot be used in practice, for example, because they do not respect the restricted operational zone constraint. These zones are the periods in which some units cannot operate due to cavitation phenomena, mechanical vibration or loss in efficiency. To obtain a feasible solutions, an approximation of the error caused by MILP approximation formulation was done.

4.1.3 CONTRIBUTIONS

All of the cited works make progress towards solving the problem by using different approximation and linearization techniques. These give good results. 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 efficiency curves of water discharge and power produced for each possible combination of turbines can be used. These curves allow to determine a set of efficiency points of water discharge and it is at these points that the maximum of power produced is reached.

This paper proposes a mathematical formulation of the unit commitment problem where efficiency points of power produced and water discharge for each combination of active turbines are derived and the model selects one of these points to maximize the energy produced. Energy losses caused by the unit start-ups are taken into account. Transmission constraints, the reserve and load constraints are not considered in this paper. The solution provided by the model can directly be implemented in practice since it is obtained on the efficiency points, as

the engineers want to operate the power plants. Moreover, using the efficiency points decreases the number of parameters and variables and makes the problem easier to solve.

4.1.4 ORGANIZATION OF THE PAPER

This paper is organized as follows. Section 4.2 presents the characteristics of the problem and the mathematical model developed. The results are discussed in Section 4.3 and concluding remarks are presented in Section 4.4.

4.2 SHORT-TERM HYDRO-POWER SCHEDULING PROBLEM

The unit commitment problem is used to determine the optimal production plan up to one week. The purpose is to maximize the energy production and penalize unit start-ups. To maximize energy produced, several factors are considered. This section defines the different factors and describes the short-term problem.

4.2.1 POWER PRODUCTION

The production function depends on the water discharge q in m^3/s, the efficiency of the unit η, the gravitational acceleration G in m/s^2, the density ρ in kg/m^3 and the net water head h_n in m which in turn depends on the total water discharge Q (sum of the water discharge and water spillage) in m^3/s and the volume of the reservoir v in hm^3 [3]. The power output p in W from a hydro generating unit is given by the equation :

$$p(q,h_n) = G \times \eta \times h_n(Q,v) \times q \times \rho. \tag{4.1}$$

The net water head is the difference between the forebay elevation h_f, tailrace elevation h_t and the losses caused by the friction in the penstock h_p. The net water head is calculated by :

$$h_n(Q,v) = h_f(v) - h_t(Q) - h_p(Q,q). \tag{4.2}$$

4.2.2 TURBINE EFFICIENCY

Turbine efficiency is the most important factor in the output of the unit. It represents the capability of the turbine to transform the mechanical power of the water into the most electrical power as possible. The turbine efficiency depends on the water discharge and the net water head. Each turbine has its own efficiency curve and for the same value of water discharge and volume, different turbines produce different quantities of power.

4.2.3 COMBINATIONS OF THE TURBINES

The total power output depends on the total water discharge and the number of active turbines. In the operational reality, the active turbines are grouped into combinations. Table 4.2 shows an example of the possible combinations of a powerhouse that has 4 available turbines. The operator can use up to 4 turbines. In the operation of a powerhouse, a minimum number of active turbines (for example 3 turbines) is required due to the physical constraints of the powerhouse.

TABLEAU 4.2 : Combinations of 4 available turbines ©Maissa daadaa 2023

3 active turbines	4 active turbines
123 124 134 234	1234

As explained in Section 4.2.2, each turbine has its own efficiency. The combination of active turbines has also its own efficiency. Fig.4.1 shows the power output depending on the water discharge for all possible combinations and this for a given forebay elevation.

As shown in Fig.4.1 , the power output decreases when the maximum water discharge of the turbine combination has been reached. For example, in the instance where the active turbines are 234 (curve with 'square' marker) the power output decreases once the maximum water discharge is reached (510 m^3/s). It is useless to increase the quantity of water discharge because it will be spilled, therefore the tailrace elevation will increase and consequently the net water head will be reduced. For this reason the power output is decreased. Fig.4.2 shows that for the same number of active turbines and the same value of water discharge, there is a small variation in the production which confirms that the power produced differs from one combination to another.

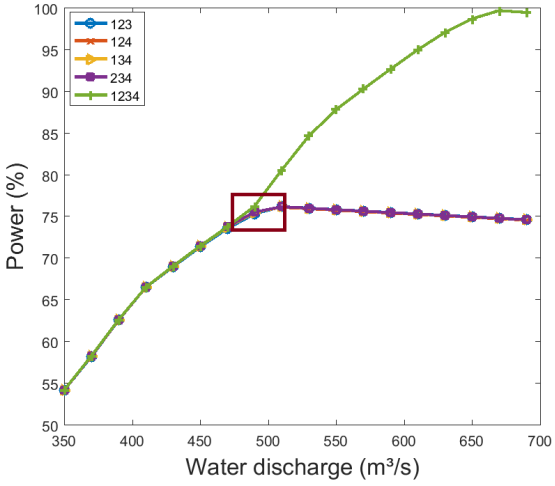


FIGURE 4.1 : Power output for all combinations ©Maissa Daadaa 2023

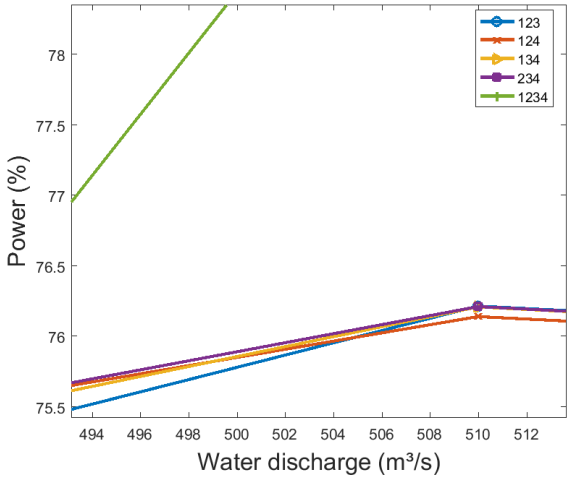


FIGURE 4.2 : Zoom-in of the rectangle in Fig.1 ©Maissa Daadaa 2023

4.2.4 START-UPS

The objective of the problem is to determine for each period 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, there is an important concept that must be taken into account which is the start-ups of the turbines. Frequent start-ups cause maintenance costs and decrease the life cycle of the turbines. In practice, it is recommended to have a limited number of start-ups.

4.2.5 PROBLEM DESCRIPTION

The objective is to maximize energy production and penalize unit start-ups. Hydropower production function can be modelled in many different ways. Polynomial equations, which determines an analytic expression of a polynomial of degree n passing through data points (p,h_n). The wrong choice of the degree influences the results. Splines allow to split the points into subsets and use a polynomial approximation for each subset of points, then connect them. Other methods can be used such as interpolation, linearization, etc. These methods give good results and allow to find the value of water discharge that maximize the energy produced. But in operational reality, instead of approximating the hydropower production functions and optimizing on all the values of the water discharge to find the best values, the best values are already known. These values can be determined using the efficiency curves of water discharge for each possible combination. A set of points corresponding to the maximum water discharge and the maximum efficiency of each available combinations group is determined. However, selecting only these points can limit the optimization and lead to infeasible solutions. In this regard, a set of adjacent points to the maximum should be defined. To do so, the convex hull of the efficiency curves for each combinations group is traced. Using the convex hull allows to define the maximum efficiency points. The adjacent points are chosen as $\pm\ \zeta\ (m^3/s)$ of the

maximum efficiency of water discharge where ζ are integer parameters. The choice of these parameters depends on the management of the powerhouses and the operational precision that can be obtained.

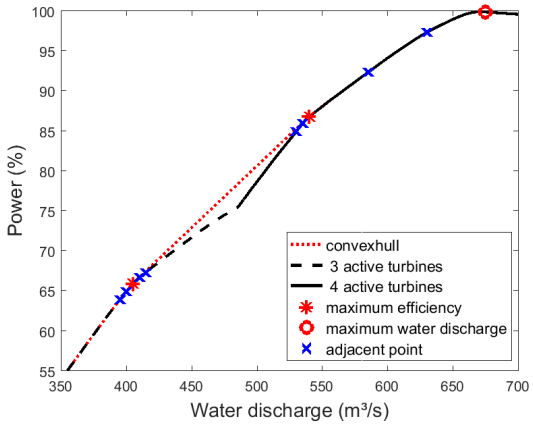


FIGURE 4.3 : Pairs of efficiency points (P,q) ©Maissa Daadaa 2023

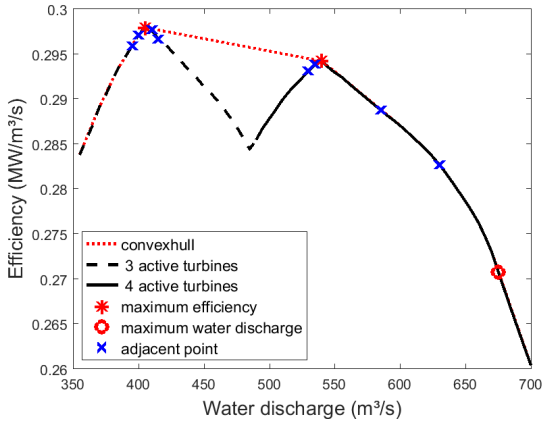


FIGURE 4.4 : The set of efficiency points of water discharge ©Maissa Daadaa 2023

For example, Fig.4.4 shows the efficiency curve of the water discharge (q) for a powerhouse that has 4 available turbines. As explained in section 4.2.3, the number of active turbines can be 3 (curve with dashed line) or 4 (curve with solid line). For each curve, the maximum efficiency (marked as $'*'$) is defined. The maximum water discharge (marked as $'o'$) is defined only for the curve with the maximum number of turbines because it is the only instance where spillage is an option. The adjacent points (marked as $'X'$) are defined as -3 m^3/s and -6 m^3/s and two equidistant points between the maximum water discharge and the efficiency water discharge. For the curve with a dashed line (3 active turbines) the adjacent points are defined with $\pm$ 3 m^3/s and $\pm$ 6 m^3/s. Once these efficiency points of water discharge are defined, the power produced P associated to the points is determined. Finally

the pairs of water discharge and power produced (P, q) associated with each efficiency point can be determined for each possible combination as shown in Fig.4.3.

At period t, the best efficiency point of the set is selected respecting some constraints like the water balance to maximize the energy produced. At period $t+1$ another efficiency point can be selected and the combination can be changed. To decrease the number of changes, a penalty factor must be added for each start-up. This penality factor corresponds to the energy loss caused by the the start-up of the unit in (MW).

The set of pairs of efficiency points presented in the previous curves is obtained with a given forebay elevation. The idea is to fix the value of the forebay elevation at the maximum volume of the reservoir V_{max}. Using V_{max} reduces the number of variables and facilitates the resolution of the problem. However, in reality, the reservoir is not always full. The model presented in this paper takes into account this aspect and makes a correction between the produced power at V_{max} and the power obtained at the optimized volume. To do so, a coefficient θ that approximates the power losses due to the volume difference, in proportion to the maximum volume of the reservoir is given by Eq.(4.3) :

$$p(q,h_n) \approx p(q,h_n^{\max}) - \Psi \times (h_n^{\max} - h_n). \tag{4.3}$$

Where $h_n^{\max}$ is the maximum net water head, hence the volume is equal to the maximum volume of the reservoir ($V_{\max}$). Several curves (p,q) were defined with different volume levels : the minimum, the average, etc. Using Eq.(4.3), for each curve, a coefficient Ψ corresponding to the solution of the least squares between the true production values and their approximations at V_{max} is determined. The average of all the obtained coefficients gives the value of θ. Since each reservoir has its volume limits, the coefficient θ is calculated for each powerhouse.

4.2.6 MATHEMATICAL MODEL

This section presents the mathematical model used to solve the short-term unit commitment problem. The goal of the short-term hydropower problem is to define the optimal unit commitment for the next days up to one week in order to maximize the energy production and penalize unit start-ups. The problem is formulated as a MILP, that trades off the maximum efficiency of generation versus start-up of units. The formulation is given by a three-terms objective function but remains mono objective since they are not conflicting. The first term computes the power output $P^c_{k,t}$ at each point for each combination at each period. This power is determined at the maximum reservoir volume V_{max}. Since the power output is calculated at V_{max}, the second term makes a correction between the power produced at the current volume and V_{max}. In this regard, the second term of the objective function equals the difference between the volumes, multiplied by an estimate of the energy losses θ^c. The last term of the objective function reduces the number of changes by penalizing unit start-ups. Finally the sum of the three terms is multiplied by Δ_t to convert power to energye, which is maximized. The MILP is given by :

$$max_{y,v,z}\Delta_t \times [\sum_{c\in C}\sum_{t\in T}\sum_{b\in B}\sum_{k\in K^c_b} P^c_{k,t} \times\ y^c_{k,t} - \sum_{c\in C}\sum_{t\in T}\theta^c \times (V^c_{max} - v^c_t) - \sum_{c\in C}\sum_{t\in T}\sum_{j\in J_t}\varepsilon^c \times z^c_{j,t}] \tag{4.4}$$

Subject to :

$$v^c_{t+1} = v^c_t + \Delta_t \times [(\delta^c_t \times \beta) - \sum_{b\in B}\sum_{k\in K^c_b}(q^c_{t,k} \times y^c_{k,t} \times \beta) - (d^c_t \times \beta) + \sum_{l\in U^c}\sum_{b\in B}\sum_{k\in k^l_b}(q^l_{t,k} \times\ y^l_{k,t} \times \beta) + (d^l_t \times \beta)] \quad ,\forall c \in C, \forall t \in T \tag{4.5}$$

$$\sum_{b\in B}\sum_{k\in K_b^c} y_{k,t}^c = 1 \quad ,\forall c \in C, \forall t \in T \tag{4.6}$$

$$\sum_{b\in B}\sum_{k\in K_b^c} y_{k,t+1}^c \times A_{t+1,k,j}^c - \sum_{b\in B}\sum_{k\in K_b^c} y_{k,t}^c \times A_{t,k,j}^c \leq z_{j,t}^c \quad ,\forall c \in C\ ,\forall t \in T\ ,\forall j \in J_t \tag{4.7}$$

$$\sum_{t\in T}\sum_{j\in J_t} z_{j,t}^c \leq N_{max}^c \quad ,\forall c \in C \tag{4.8}$$

$$v_{min}^c \leq v_t^c \leq V_{max}^c \quad ,\forall c \in C, \forall t \in T \tag{4.9}$$

$$v_1^c = v_{ini}^c \quad ,\forall c \in C \tag{4.10}$$

$$v_T^c \geq v_{final}^c \quad ,\forall c \in C \tag{4.11}$$

$$y_{k,t}^c, y_{k,t}^l, z_{j,t}^c \in \mathscr{B} \quad ,\forall c \in C, \forall t \in T, \forall b \in B^c, \forall k \in K_b^c, \forall l \in U^c \tag{4.12}$$

$$d_t^c, d_t^l, v_t^c \in \mathscr{R}^+ \quad ,\forall c \in C, \forall t \in T, \forall l \in U^c. \tag{4.13}$$

Constraints (4.5) ensure the water balance of the powerhouses. In the instance where the powerhouses are in series, the water discharge $q_{k,t}^l$ and the water spillage d_t^l of the upstream powerhouses to the volume v_t^c were added. Otherwise, there will not be taken into account. Constraints (4.6) force the model to choose only one operating point at each time period for each powerhouse. Constraints (4.7) are the link between start-up variables and chosen combination considering the set of points. To limit the number of turbine changes during the planning horizon, a maximum number of start-ups N_{max} is imposed with constraints (4.8). Constraints (4.9) limit the water in the reservoir of each powerhouse, and (4.10)-(4.11) specify initial and final volumes. Finally, (4.12) defines the binary variables and (4.13) the real variables.

4.3 CASE STUDY

The MILP model presented in the previous section was tested with data from the Saguenay-Lac-St-Jean hydroelectric system owned by Rio Tinto. Rio Tinto is a company

that produces aluminium in the Saguenay region of the province of Quebec in Canada. The production of aluminium requires a lot of electric energy. Rio Tinto has a hydropower system that fills 90% of its energy needs. Rio Tinto has six powerhouses with 44 turbines. Four reservoirs are available with different capacities.

Two powerhouses in the company's system were considered : Chute-Du-Diable (CD) and Chute-Savane (CS). These two powerhouses are in series and each has five turbines. To schedule hydropower production, Rio Tinto aims to use the maximum efficiency of their turbines in order to maximize the energy produced by defining the best combination of turbines operating at each time period. Currently, they use a dynamic programming algorithm and rely on their staff's experience. The selection of the quantity of water discharged and the combination of active turbines are determined by hand. Moreover, the correlation between the powerhouses is not taken into account. The model proposed in this paper is used to define an automatic procedure which allows to determine the best schedule by defining the best combination of active turbines to maximize the energy production. Turbine start-ups are penalized in the objective function to account for the corresponding power losses. The price of the energy is not considered because in the province of Quebec, where this model is tested, the market is regulated by Hydro-Québec, therefore all producers must buy and sell to them and negociate fixed price contracts every year.

4.3.1 EXPERIMENTAL SETUP

For both powerhouses, operational restrictions require a minimum of three active turbines. The number of possible combinations is thus 16 for each powerhouse as shown in Table 4.3 : 10 combinations of three turbines, 5 combinations of four turbines and 1 combination of five turbines. For each combination, the pair of points of water discharge and power produced (P,q), as explained in Section 4.2.1, are defined. The number of adjacent points was defined

TABLEAU 4.3 : Combination of 5 available turbines ©Maissa Daadaa 2023

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	

by making several tests. First, only the maximum point was chosen, but in some cases, this leads to infeasible solutions. Other points were then added. Firstly, two adjacent points were added and feasible solutions were found. To avoid the risk of infeasibility, two other adjacent points were added to assure that the solution is feasible and to allow the model to be more flexible. Since the discretization of the water discharge is done in steps of 5 m^3/s, the adjacent points are defined with $\pm$ 5 m^3/s and $\pm$ 10 m^3/s of water discharge and two equidistant points between the maximum water discharge and the efficient water discharge are also selected (where the number of active turbines is equal to the number of available turbines).

The model was tested on 65 test instances. The duration of each instance is four days partitioned into periods of 1 hour, for a total of 96 hours. The quantity of natural inflows and the availability of the turbines are known and obtained from the historical database. The model is deterministic, therefore there is no uncertainty in the inflows. Initial and final volumes of each reservoir are also provided.

The Xpress [93] solver accessed via Python was used to solve the proposed formulation. The problem was solved to optimality and the MIP gap was equal to 0.01%. The resulting MILP problem has 10560 binary variables, 386 real variables and 2060 constraints. To solve this unit commitment problem, the computational time ranged between 5 and 15 minutes for each instance, on a laptop computer with an Intel Core i5 Processor and 8 GB of RAM.

4.3.2 COMPUTATIONAL RESULTS

The results from the proposed model are compared with the real operational decisions taken from the database. Periods with turbines unavailable due to maintenance or repair, with a contingency energy request, and the periods with high water spillage were not selected for comparison. The initial volume, the final volume and the states of turbines at $t = 0$ are equal to those in the database.

Table 4.4 reports for each instance and for the 96 hours, the energy production optimized by the proposed model, the energy obtained from the real operational decisions, and the average improvement as the difference between these energies divided by the energy obtained from the real operational decisions :

$$\textit{Average improvement(\%)} = \frac{\textit{(energy production optimized (MW) - energy in database (MW))} \times \textit{100}}{\textit{energy in database (MW)}}$$

Note that all the results are reported as percentages because the actual values are confidential. To do so, all energies are divided by the maximum value of the energy produced for the 65 instances and multiplied by 100 :

$$\textit{Percentage(\%)} = \frac{\textit{energy production (MW)} \times \textit{100}}{\textit{max (energy production optimized (MW), energy in database (MW))}}$$

A positive improvement indicates that the optimized solution is better than the real operational decision in the database.

TABLEAU 4.4 : Total energy production and average improvements ©Maissa daadaa 2023

Instance	Optimized (as a %)	Operational decision (as a %)	Improvement (%)	Instance	Optimized (as a %)	Operational decision (as a %)	Improvement (%)
1	96,00	95,71	0,31	34	81,54	80,20	1,64
2	92,74	91,56	1,27	35	82,36	81,64	0,87
3	74,70	74,19	0,68	36	78,25	77,43	1,05
4	80,92	80,14	0,95	37	84,07	83,60	0,57
5	72,28	71,99	0,40	38	82,81	81,99	0,99
6	79,24	78,76	0,61	39	61,09	60,47	1,01
7	71,95	71,70	0,35	40	69,48	68,77	1,01
8	71,64	71,03	0,85	41	70,25	70,05	0,28
9	87,51	87,00	0,59	42	72,68	72,17	0,71
10	96,23	95,52	0,74	43	71,14	71,09	0,07
11	88,04	87,09	1,08	44	74,63	74,39	0,33
12	80,63	80,03	0,74	45	75,62	75,37	0,34
13	85,44	84,38	1,24	46	77,29	76,95	0,43
14	75,60	75,13	0,62	47	77,53	77,22	0,40
15	75,64	75,29	0,47	48	78,31	77,93	0,49
16	74,43	74,05	0,50	49	77,29	76,57	0,94
17	79,64	79,01	0,79	50	77,19	76,59	0,78
18	72,45	72,01	0,61	51	100,00	99,27	0,73
19	69,25	69,21	0,06	52	81,34	80,59	0,92
20	72,34	72,08	0,36	53	88,74	88,51	0,26
21	72,82	72,00	1,13	54	84,99	84,17	0,97
22	56,39	55,91	0,85	55	97,89	96,73	1,18
23	95,99	95,43	0,59	56	93,25	92,58	0,72
24	94,95	93,94	1,06	57	88,92	87,30	1,82
25	96,49	95,84	0,67	58	81,33	79,92	1,74
26	90,59	89,75	0,92	59	82,88	82,37	0,61
27	90,24	89,66	0,65	60	83,66	83,35	0,37
28	88,05	87,38	0,77	61	79,31	78,56	0,94
29	83,66	82,79	1,04	62	80,21	79,36	1,06
30	62,75	62,34	0,66	63	80,51	79,77	0,92
31	80,30	78,57	2,16	64	82,37	82,02	0,42
32	83,22	80,85	2,85	65	73,33	72,78	0,75
33	88.24	87.30	1.07				

Fig.4.5 is a histogram comparing the average improvements of energy production between the optimized solutions and the real operational decisions. Fig.4.5 shows that 32 instances have an average improvement between 0.5% and 1% and 17 instances have an average improvement greater than 1%. Given the current prices of energy in the province of Quebec, an increase of 1% in the energy production means approximately $14000 of savings in just 4 days (96 hours). Therefore, on an annual basis and considering that this study considers only half of the power plants of the company, the potential savings are substantial.

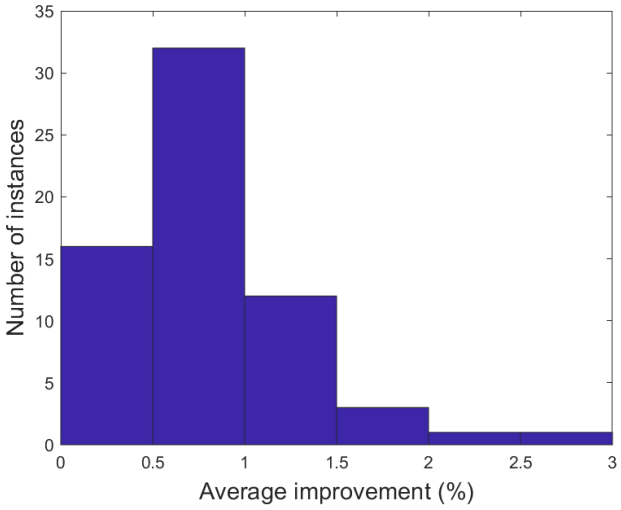


FIGURE 4.5 : Histogram of average improvements of energy ©Maissa daadaa 2023

4.3.3 VALIDATION OF THE MODEL

To validate the values of the objective function, the energy estimated by our model was compared with the real value of the production function at Rio Tinto for the same volume of the reservoir, the same water discharge and the same combination of active turbines. Table 4.5 shows the difference between the energy approximated and the real values at Rio Tinto for some instances. The results show that the values are very close, which confirms the accuracy of our model.

TABLEAU 4.5 : Comparison between the energy approximated and the real values ©Maissa daadaa 2023

instance	difference (%)
31	0.01
32	0.01
36	0.07
65	0.03

4.3.4 INTERPRETATION OF THE RESULTS

Let us analyze instance 32 where the average improvement reaches 2.85% (as shown in Table 4.4). Fig.4.6 and 4.7 show that for the same initial and final volumes, the optimized solutions (solid line) and the the real operational decisions (dashed line) are different.

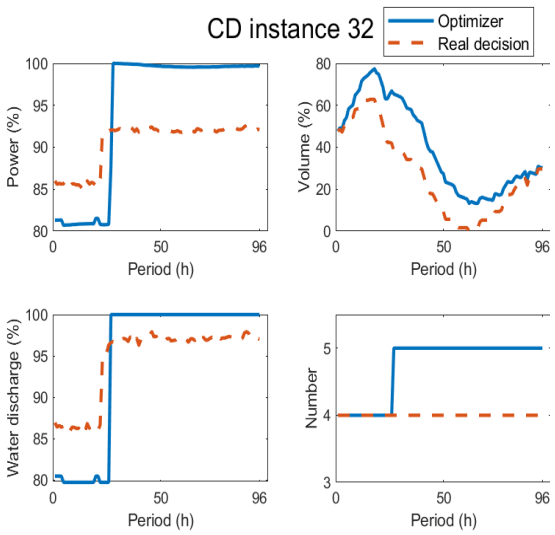


FIGURE 4.6 : Instance 32 for CD
©Maissa Daadaa 2023

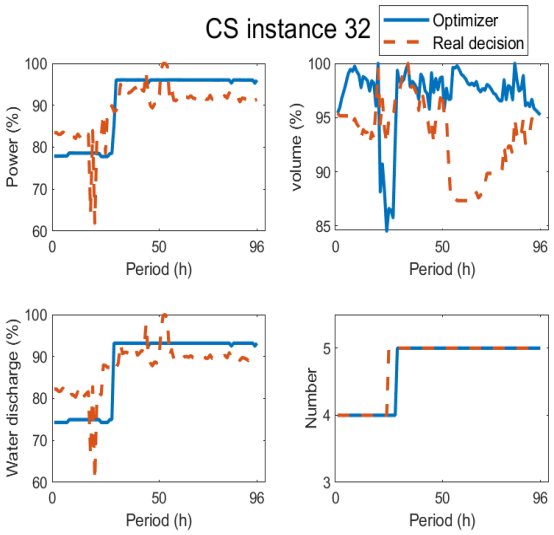


FIGURE 4.7 : Instance 32 for CS
©Maissa Daadaa 2023

For CD, the real operational decision requires the activation of 4 turbines during the whole planning horizon. The optimal solution requires the activation of 4 turbines during the first 27 periods until the volume reaches the maximum value, then the activation of 5 turbines for the next periods. The estimated power loss caused by the start-up of the fifth turbine is negligible compared to the expected power gained when operating with 5 turbines. For CS, the number of active turbines of the optimized and the real operational decisions are equal. However, there is a difference between the produced power of the optimized solution and the real operational decision. This difference is due to the suitable choice of the quantity of water

discharge. In the optimized solution, the value of the water discharge is selected according to the efficiency points. For this reason, the quantity of produced power of the optimized solution is more than that of the real operational decision.

For instances 36 and 22, the average improvement is respectively 1.05% and 0.85%. Fig.4.8 and 4.10 show that for CD in both instances, the number of active turbines and the volume of the reservoir of the optimized solution and the real operational decision are similar. Since the optimized solution uses the efficiency points of water discharge, the produced power is higher than the real operational decision.

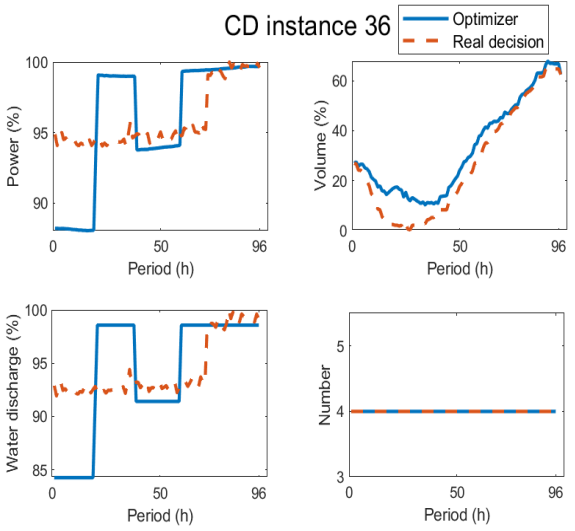


FIGURE 4.8 : Instance 36 for CD
©Maissa Daadaa 2023

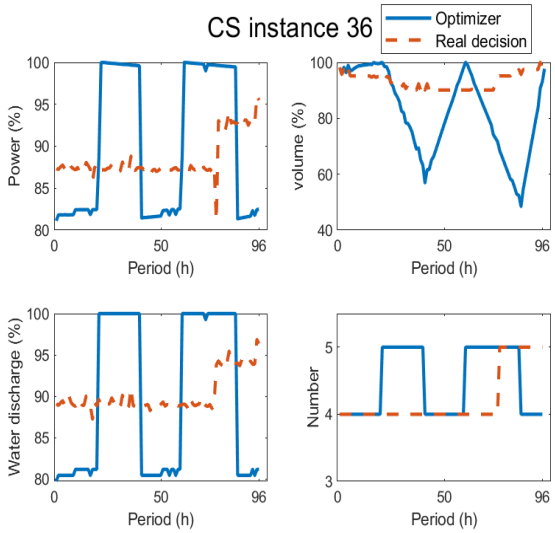


FIGURE 4.9 : Instance 36 for CS
©Maissa Daadaa 2023

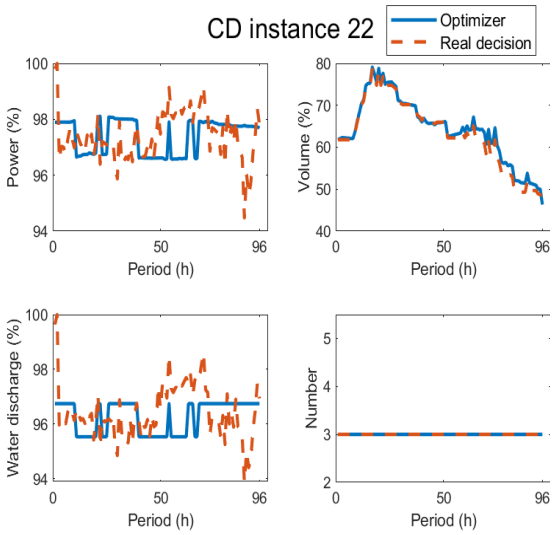


FIGURE 4.10 : Instance 22 for CD
©Maissa Daadaa 2023

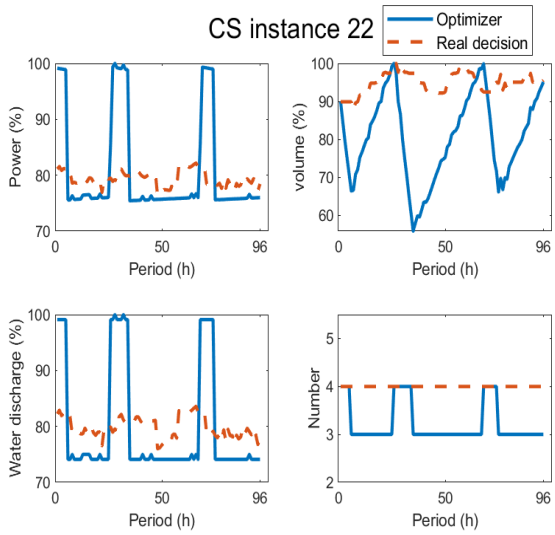


FIGURE 4.11 : Instance 22 for CS
©Maissa Daadaa 2023

An analysis of the results for CS shows that the total power output of the optimized solution is higher than the power output of the real operational decision in both instances. Fig.4.9 and 4.11 show that in the optimized solution, the number of the active turbines switches between two values. For example, for instance 36 (Fig.4.9), the optimized solution requires the activation of 4 turbines during the first 20 periods to reach the maximum volume of the reservoir. To decrease the volume, the number of active turbines is changed from 4 to 5 for the following 19 hours. At period 20, the number of active turbines is changed again to increase the volume. Notice that there is a cycle trend for these two instances. Future research will focus on this phenomenon.

In instance 7, the average improvement is 0.35%. Fig.4.12 and 4.13 show that the number of active turbines, the quantity of water discharge and the produced power of the optimized and the real operational decisions are even similar. Nevertheless, there is a change in the curves at period 49 for CD and at period 85 for CS. At these periods, there is a descent in the volume

of the CS. This is due to the decline of the quantity of the natural inflows in the reservoir. Since the inflows are deterministic and taken from the operational database, there is likely a measurement error in these periods. For this reason, future research will look at implementing a method to generate the inflows and develop a stochastic model.

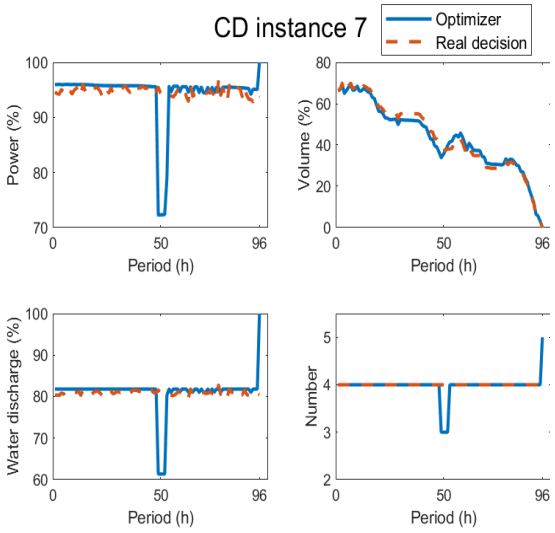


FIGURE 4.12 : Instance 7 for CD
©Maissa Daadaa 2023

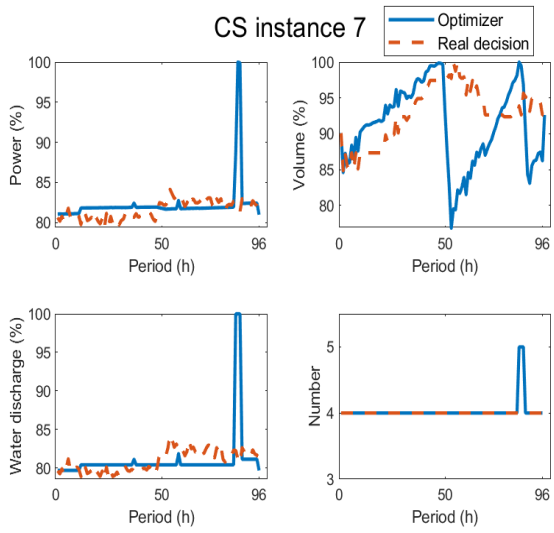


FIGURE 4.13 : Instance 7 for CS
©Maissa Daadaa 2023

4.3.5 DISCUSSIONS

The obtained results for the 65 instances show that the proposed model allows to produce more energy than the energy obtained from the real operational decisions. The accuracy of the model is confirmed since the results are compared to the real value of the production function at Rio Tinto. Therefore, the optimized solutions are always on the efficiency points, as the engineers want to operate the powerhouses. The advantage of this model is that the solution can be directly implemented in practice.

4.4 CONCLUSION

This paper introduced an innovative method to optimize the unit commitment of hydropower generating units. This method allows the producer to implement directly the optimized solution since it reflects the operational reality. The novelty is to determine the pair of points of maximum efficiency for water discharge and the power produced at the maximum storage, then operate the turbines at these points. The points are determined for all possible combinations of turbines. Since the reservoir is not always full, a correction of produced power was done. To avoid start-ups, a maximum number of turbine changes is imposed to find a viable solution in practice. The problem is formulated using a MILP to find the exact combination of turbines and the best operation point that maximizes total energy production while penalizing start-ups. The model was tested using data for two powerhouses from a real world system and the optimal operation schedules were successfully obtained. The solution found by solving the optimization problem can be directly implemented in practice. Future research will consider increasing the number of powerhouses in order to solve a more realistic and challenging system, and developing a stochastic model that takes into account the uncertainty of inflows and study the effects of the cycles.

ACKNOWLEDGEMENT

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

CHAPITRE V

ARTICLE 2 : QUANTIFYING THE IMPACT OF SCENARIO TREE GENERATION AND REDUCTION METHODS ON THE SOLUTION OF THE SHORT-TERM HYDROSCHEDULING PROBLEM

Cet article est publié :

Daadaa, M., Séguin, S., Anjos, M. F., et Demeester, K. Quantifying the impact of scenario tree generation and reduction methods on the solution of the short-term hydroscheduling problem. Energy Systems, 2023. Repéré à : https ://doi.org/10.1007/s12667-023-00623-x.

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.

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

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).

5.1 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 [3, 26, 94]. An overview on mathematical programming approaches for the deterministic short-term hydropower problem is presented in [18]. In [29], 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 [95]. 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 [48, 46]. 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 [96]. 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 [97]. The moment matching methods are used in [50, 98]. 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 [51], 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 [99], 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 [55, 56].

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 [100] where the sum of distances from all objects in that cluster is minimized. In [46], 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 [47]. 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 [48, 58, 49].

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 [95]. 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 [101], 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 [102]. Many other approaches have used backward reduction on the basis that the statistical information is maintained in the best possible way [103, 58].

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 [94]. 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 5.2 presents the methods used to generate and to reduce a full scenario tree. The methodology and the hydropower optimization model are presented in Section 5.3. Numerical results are discussed in Section 5.4 and concluding remarks are presented in Section 5.5.

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.

5.2.1 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 [101]. 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 [58], the Kantorovich distance $D(P,\pi)$ is used and calculated by Eq.(5.1) and Eq.(5.2) :

$$D(P,\pi) = \sum_{s\in I} p_s \min_{i\notin I} C_T(\xi_s,\xi_i) \tag{5.1}$$

$$C_T(\xi_s,\xi_i) = \sum_{t\in T} \parallel \xi_{s_t} - \xi_{i_t} \parallel \tag{5.2}$$

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.5.1. 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.(5.3). 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.(5.4) and Eq.(5.5) :

$$\sum_{s\in I} p_s \min_{i\notin I} C_T(\xi_s,\xi_i) < \varepsilon_\tau \tag{5.3}$$

$$\pi_i = p_i + \sum_{s\in I(i)} p_s \ , \forall i \notin I \tag{5.4}$$

where

$$I(i) = \{s \in I : i = i(s)\},\ i(s) \in \underset{i \notin I}{\operatorname{argmin}} C_T(\xi_s, \xi_i)\ ,\forall s \in I \tag{5.5}$$

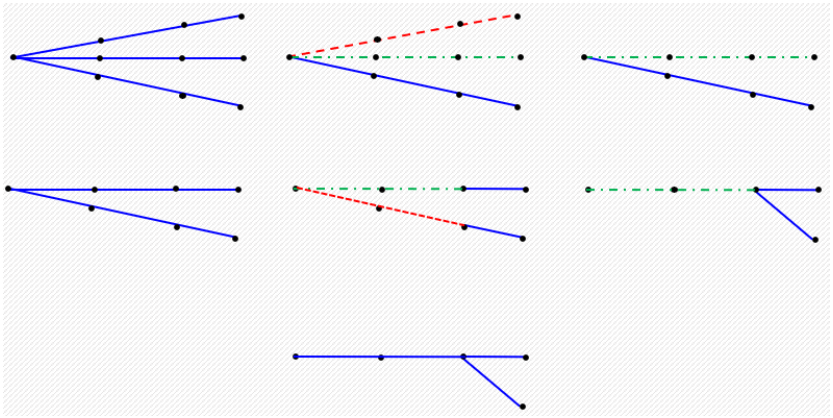


FIGURE 5.1 : Backward reduction method

Fig. 5.1 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 [102].

5.2.2 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 [49]. 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.(5.6) :

$$\xi_{i_t} = H_{rand(),t} \; \forall i \in S, \forall t \in T \tag{5.6}$$

(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.(5.7) :

$$d_{i,w} = \sum_{t \in T} \parallel H_{w_t} - \xi_{i_t} \parallel \tag{5.7}$$

(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.(5.8) :

$$O = order\,(D) \tag{5.8}$$

(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.(5.9),(5.10),(5.11), and (5.12) :

$$\xi_{i_t}^{r+1} = \xi_{i_t}^{r} + \varepsilon(r).h_\lambda(O_i).(H_{w_t} - \xi_{i_t}) \tag{5.9}$$

$$\varepsilon(r) = \varepsilon_0(\varepsilon_f/\varepsilon_0)^{r/r_{max}} \tag{5.10}$$

$$h_\lambda(O_i) = e^{-(O_i/\lambda)} \tag{5.11}$$

$$\lambda = \lambda_0(\lambda_f/\lambda_0)^{r/r_{max}} \tag{5.12}$$

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 [47, 49, 48, 104]. 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.(5.13) :

$$\pi_i = Count\{l' \in [1,L], l' \mid d_{i,l'} = \min_{i' \in [1,S]}(d_{i',l'})\}/L \tag{5.13}$$

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

5.2.3 IMPLEMENTATION

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 5.2.1. 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 5.2.2. 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.

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 :

$$E = \sum_{i \in S} \pi_i \times \sum_{n \in N_i} \xi_n \tag{5.14}$$

$$V = \sum_{i \in S} \pi_i \times (\sum_{n \in N_i} \xi_n - E)^2 \tag{5.15}$$

$$SD = \sqrt{V} \tag{5.16}$$

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.

5.3 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 [94]. 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.5.2. 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 5.2 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.

TABLEAU 5.2 : Combinations of 5 available turbines ©Maissa Daadaa 2023

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.

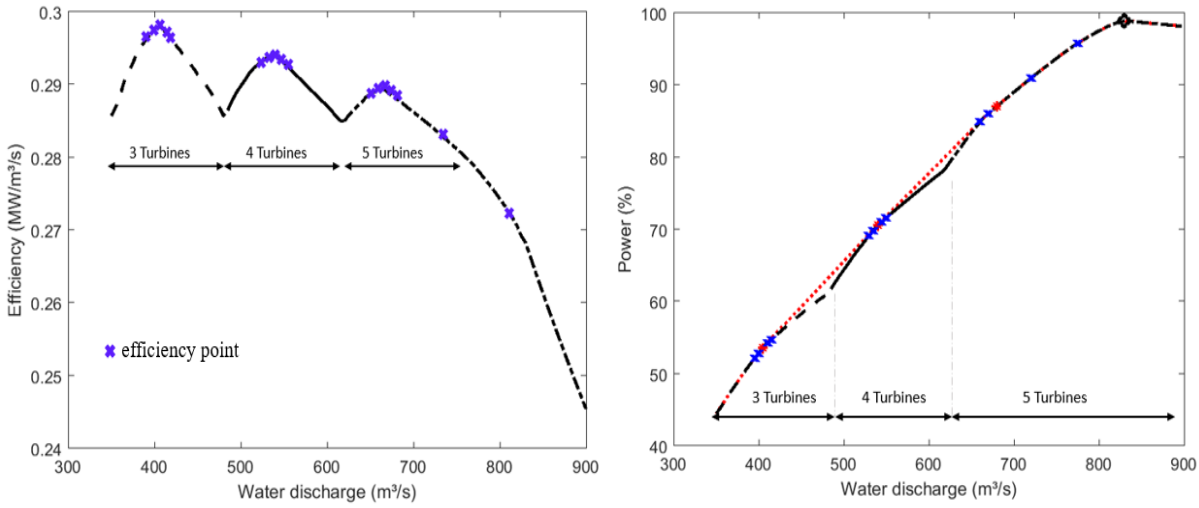


FIGURE 5.2 : The efficiency curves ©Maissa Daadaa 2023

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.

5.3.1 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 :

$$\max_{y,v,z}[\sum_{c\in C}\sum_{b\in B^c}\sum_{k\in K^c_b} P^c_{k,0}\times\ y^c_{k,0} - \sum_{c\in C}\theta^c\times(V^c_{max}-v^c_0) - \sum_{c\in C}\sum_{j\in J}\varepsilon^c\times z^c_{j,0} +$$
$$\sum_{i\in S}\pi^c_i\times\Delta_t\times[\sum_{c\in C}\sum_{n\in N_i}\sum_{b\in B^c}\sum_{k\in K^c_b} P^c_{k,n}\times\ y^c_{k,n} - \sum_{c\in C}\sum_{n\in N_i}\theta^c\times(V^c_{max}-v^c_n)] -$$
$$\sum_{c\in C}\sum_{n\in N_i}\sum_{j\in J}\varepsilon^c\times z^c_{j,n}] \tag{5.17}$$

Subject to :

$$v^c_{n+1} = v^c_n + \Delta_t\times[(\xi^c_n\times\beta) - \sum_{b\in B}\sum_{k\in K^c_b}(q^c_{n,k}\times y^c_{k,n}\times\beta) - (d^c_n\times\beta) +$$
$$\sum_{l\in U^c}\sum_{b\in B^c}\sum_{k\in k^l_b}(q^l_{n,k}\times\ y^l_{k,n}\times\beta) + (d^l_n\times\beta)] \tag{5.18}$$
$$,\forall c\in C, \forall n\in N_i, \forall i\in S$$

$$\sum_{b\in B^c}\sum_{k\in K^c_b} y^c_{k,n+1}\times A^c_{n+1,k,j} - \sum_{b\in B^c}\sum_{k\in K^c_b} y^c_{k,n}\times A^c_{n,k,j} \le z^c_{j,n} \tag{5.19}$$
$$,\forall c\in C\ ,\forall n\in N_i\ ,\forall i\in S$$
$$,\forall j\in J$$

$$\sum_{b\in B^c}\sum_{k\in K_b^c} y_{k,n}^c = 1 \quad ,\forall c\in C,\forall n\in N_i,\forall i\in S \tag{5.20}$$

$$\sum_{i\in S}\sum_{n\in N_i}\sum_{j\in J} z_{j,n}^c \leq N_{max}^c \quad ,\forall c\in C \tag{5.21}$$

$$v_{min}^c \leq v_n^c \leq V_{max}^c \quad ,\forall c\in C,\forall n\in N_i,\forall i\in S \tag{5.22}$$

$$v_0^c = v_{ini}^c \quad ,\forall c\in C \tag{5.23}$$

$$v_{N_i}^c \geq v_{final}^c \quad ,\forall c\in C,\forall i\in S \tag{5.24}$$

$$y_{k,n}^c, y_{k,n}^l, z_{j,n}^c \in \mathscr{B} \quad ,\forall c\in C,\forall n\in N_i,\forall i\in S,\forall b\in B^c \tag{5.25}$$

$$,\forall k\in K_b^c,\forall l\in U^c$$

$$d_n^c, d_n^l, v_n^c \in \mathscr{R}^+ \quad ,\forall c\in C,\forall n\in N_i,\forall i\in S,\forall l\in U^c. \tag{5.26}$$

Constraints (5.18) ensure water balance at the plants. Constraints (5.19) are the link between start-up variables and the chosen combination considering the set of points. Constraints (5.20) 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 (5.21). Constraints (5.22) are the bounds on reservoir volumes, and constraints (5.23)-(5.24) specify initial and final volumes. Finally, constraints (5.25) define the binary variables and (5.26) the real variables.

5.3.2 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.

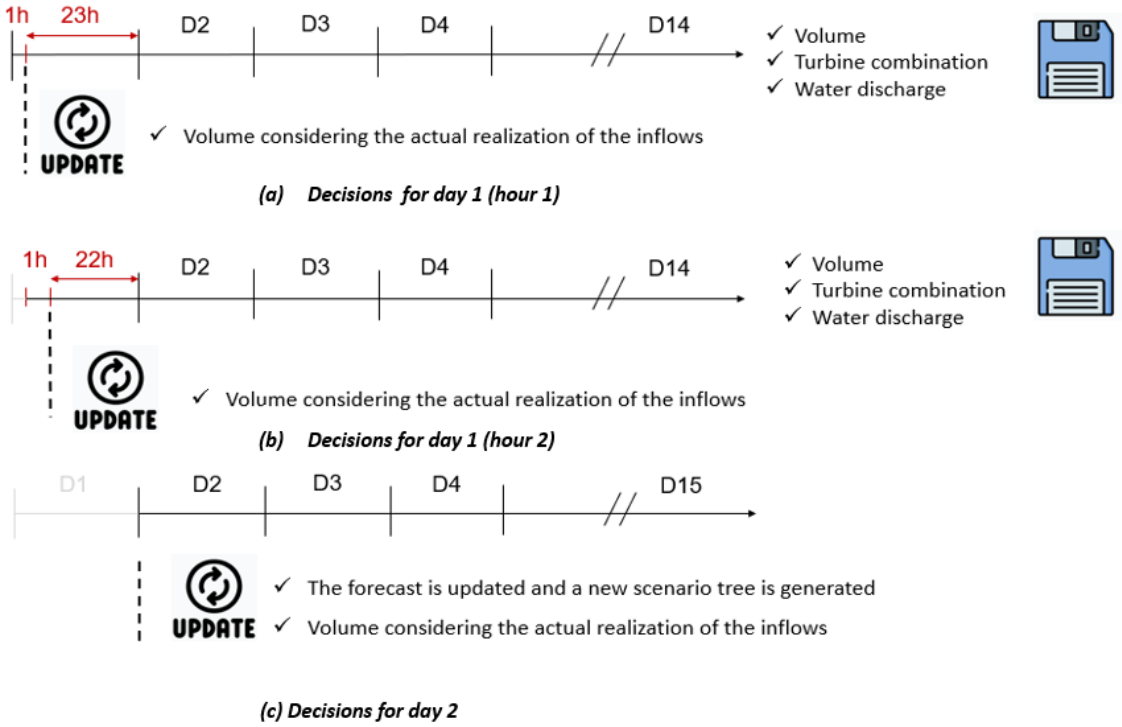


FIGURE 5.3 : The methodology according to the management of the powerhouses ©Maissa Daadaa 2023

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).

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.5.3-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.5.3-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.5.3-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.5.3-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.

5.4 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 [102], and for the neural gas, the scenario trees are obtaining by solving the algorithm using

Python [105]. 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 [106] and Xpress solver accessed via Python [93]. 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.

5.4.1 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.

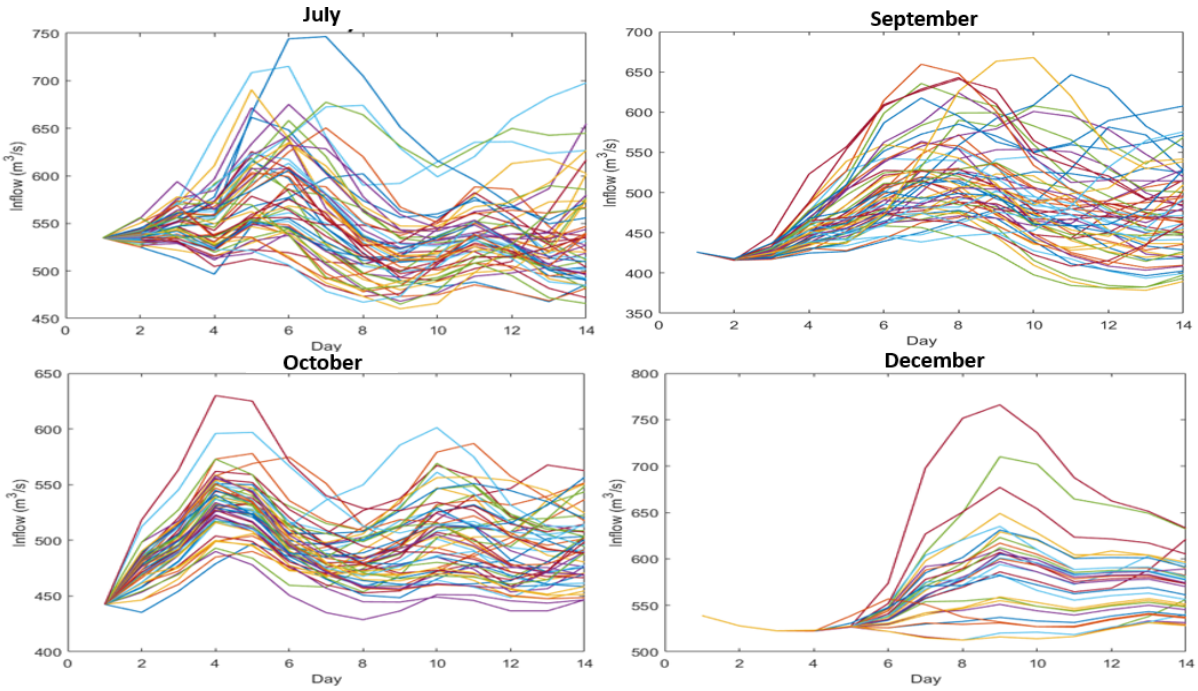


FIGURE 5.4 : The full scenario trees for the first day in the rolling horizon for July, September, October and December ©Maissa Daadaa 2023

Fig.5.4 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.5.5 and Fig.5.6 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 5.2.1, 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.

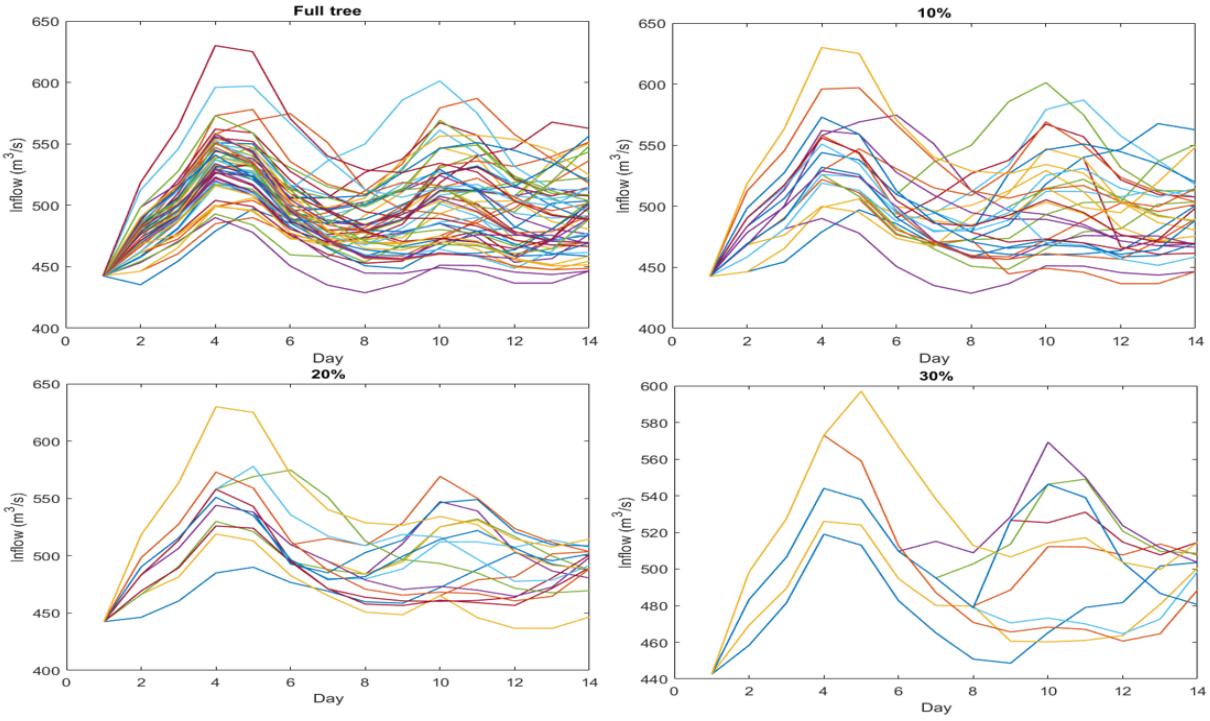


FIGURE 5.5 : An example of the full and the reduced scenario trees with 10%, 20% and 30% of reduction with backward reduction ©Maissa Daadaa 2023

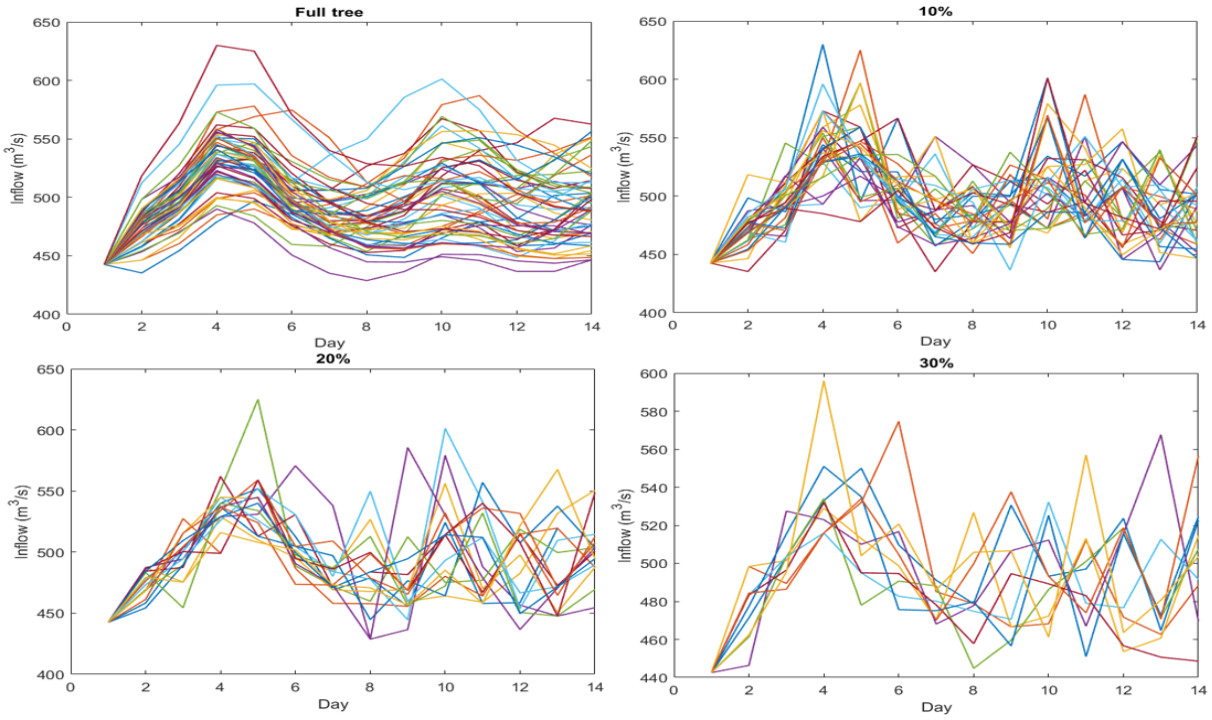


FIGURE 5.6 : An example of the full and the reduced scenario trees with 10%, 20% and 30% of reduction with neural gas ©Maissa Daadaa23

5.4.2 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 5.2.3 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.5.7.

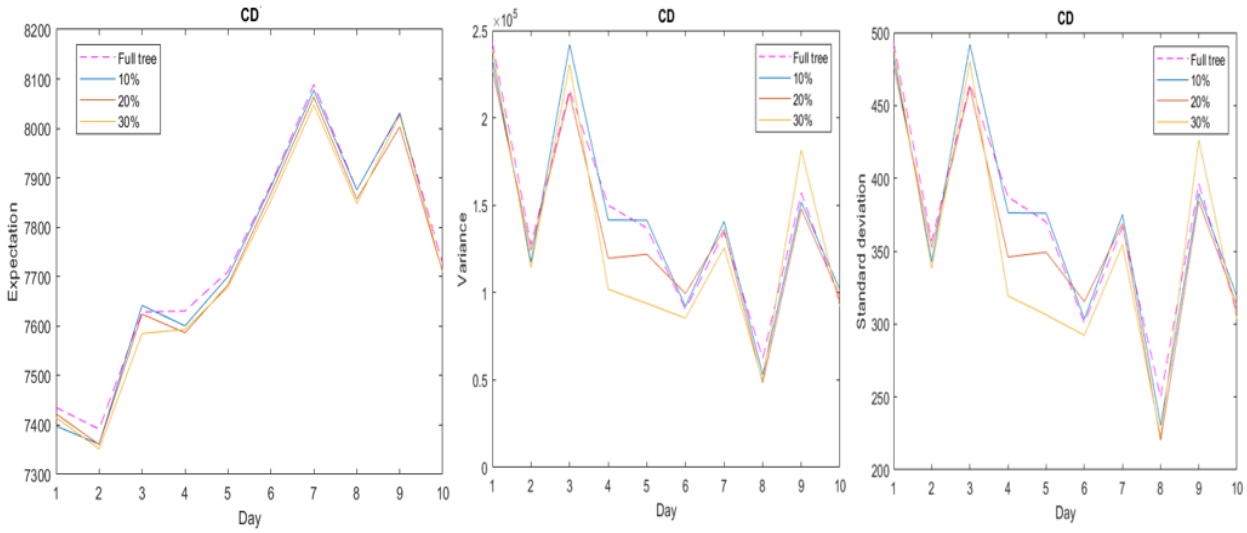


FIGURE 5.7 : An example of the statistical tests for backward reduction for powerhouse CD
©Maissa Daadaa 2023

(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.5.8 and Table A1. This can cause loss in the diversity. All clustering and moment matching methods suffer from the same problem. [48]. 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.

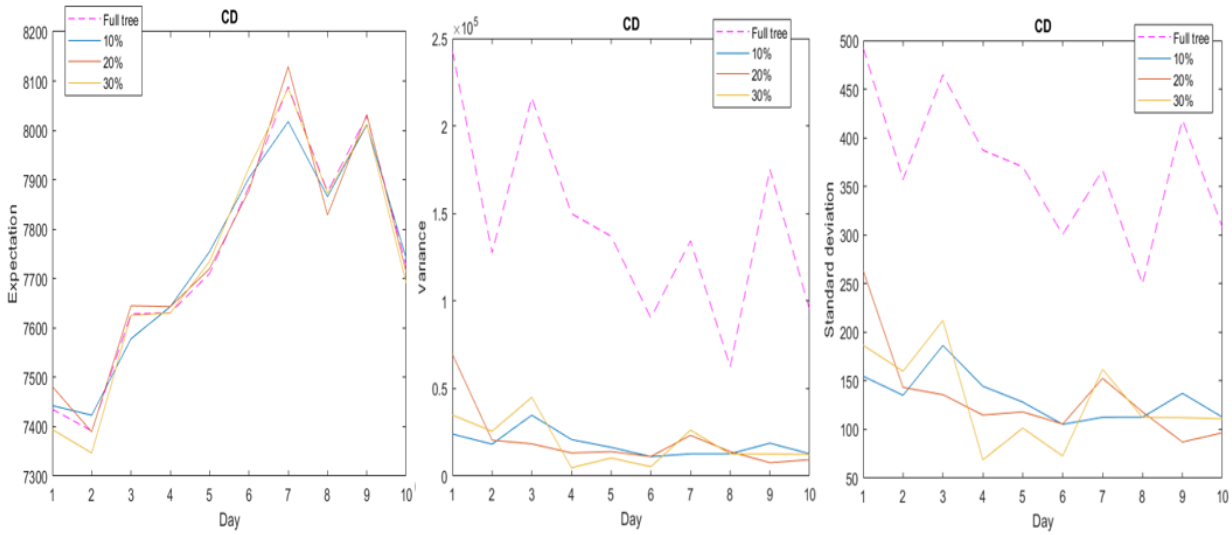


FIGURE 5.8 : An example of the statistical tests for neural gas for powerhouse CD ©Maissa Daadaa 2023

5.4.3 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.

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.

TABLEAU 5.3 : Total energy production for 4 test cases ©Maissa Daadaa 2023

July 2021 (GWh)				
Reduction	Full tree	Backward	Neural gas	Diff-B -NG(%)
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)				
Reduction	Full tree	Backward	Neural gas	Diff-B -NG(%)
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)				
Reduction	Full tree	Backward	Neural gas	Diff-B -NG(%)
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)				
Reduction	Full tree	Backward	Neural gas	Diff-B -NG(%)
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 5.3 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.5.9 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.

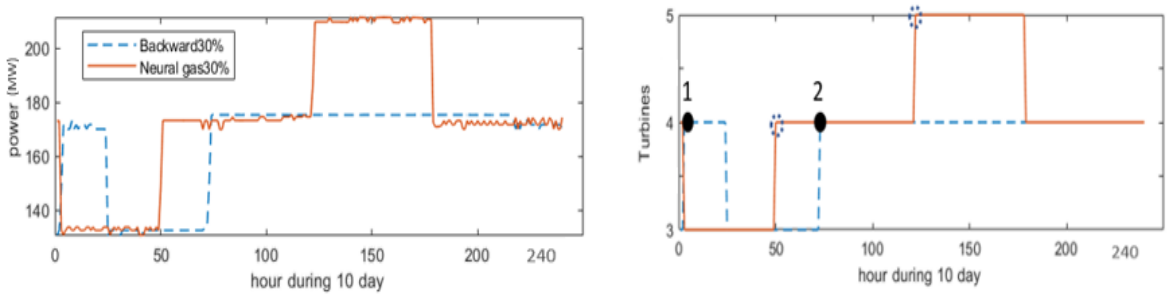


FIGURE 5.9 : Maximum number of start-ups ©Maissa Daadaa 2023

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 5.3 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.

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 5.3 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.

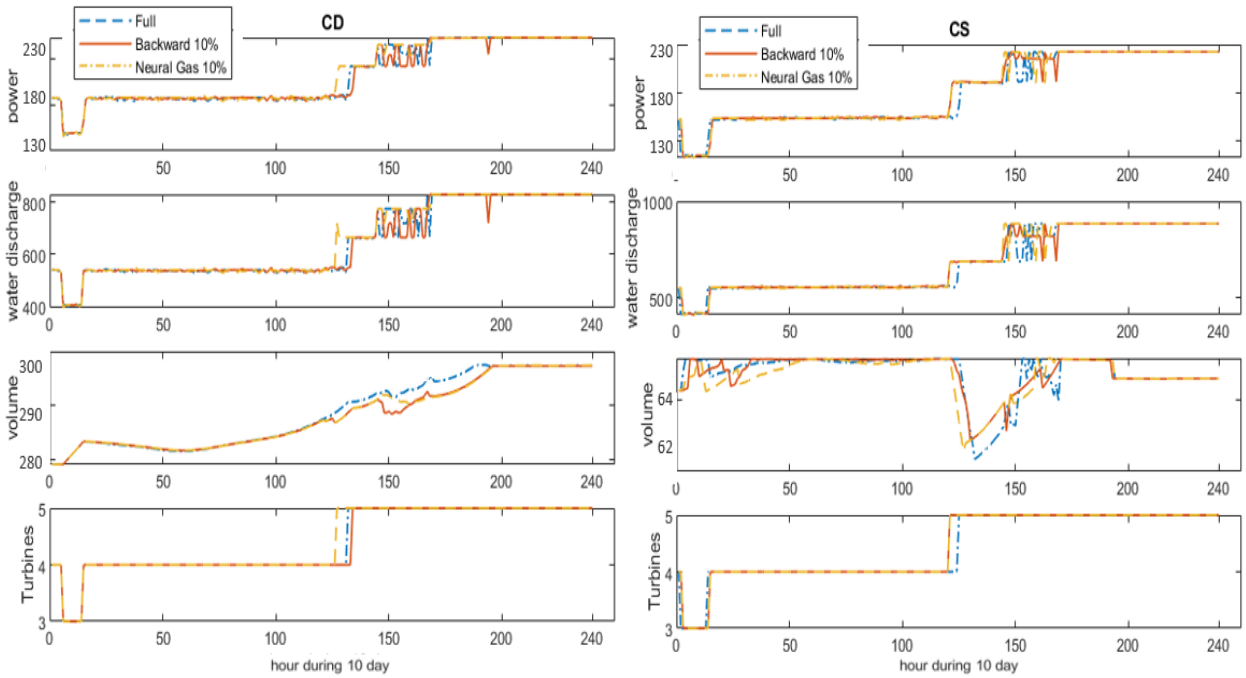


FIGURE 5.10 : Comparison of the solutions for backward reduction, neural gas and the full tree for December with 10% of reduction ©Maissa Daadaa 2023

Let us start with December. Fig.5.10 and 5.11 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.5.10, 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.5.12 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.5.11. This is due to the loss of the information with 30% of reduction.

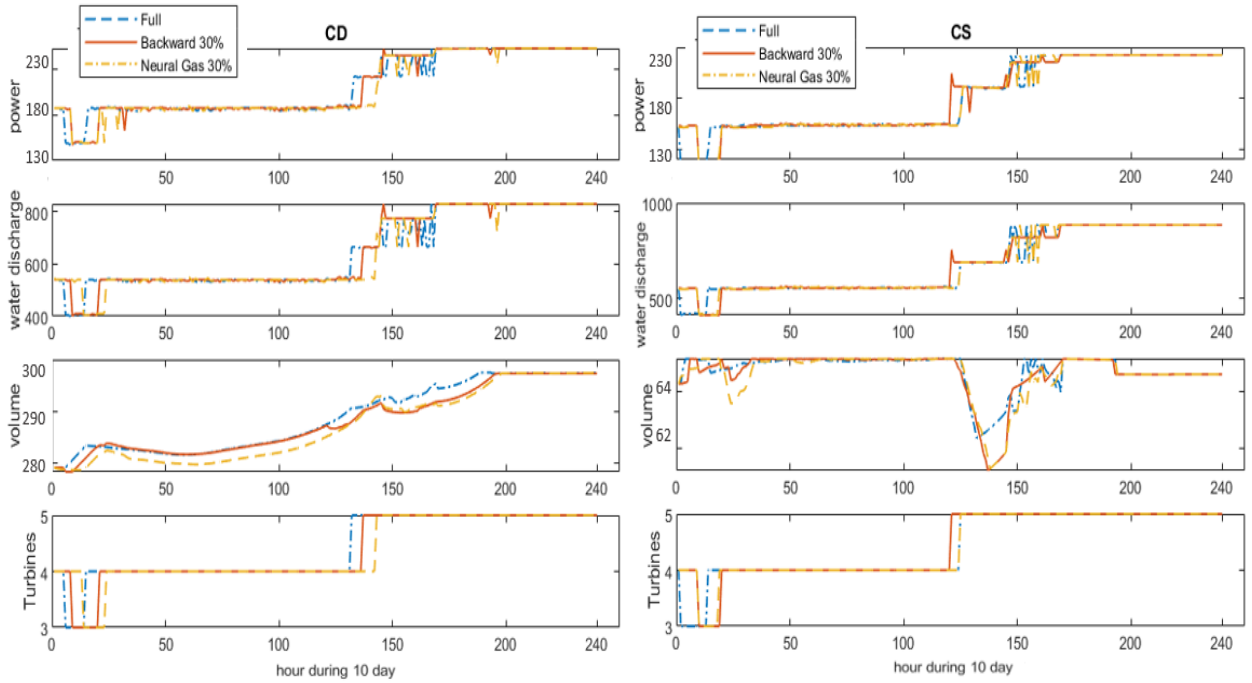


FIGURE 5.11 : Comparison of the solutions for backward reduction, neural gas and the full tree for December with 30% of reduction ©Maissa Daadaa 2023

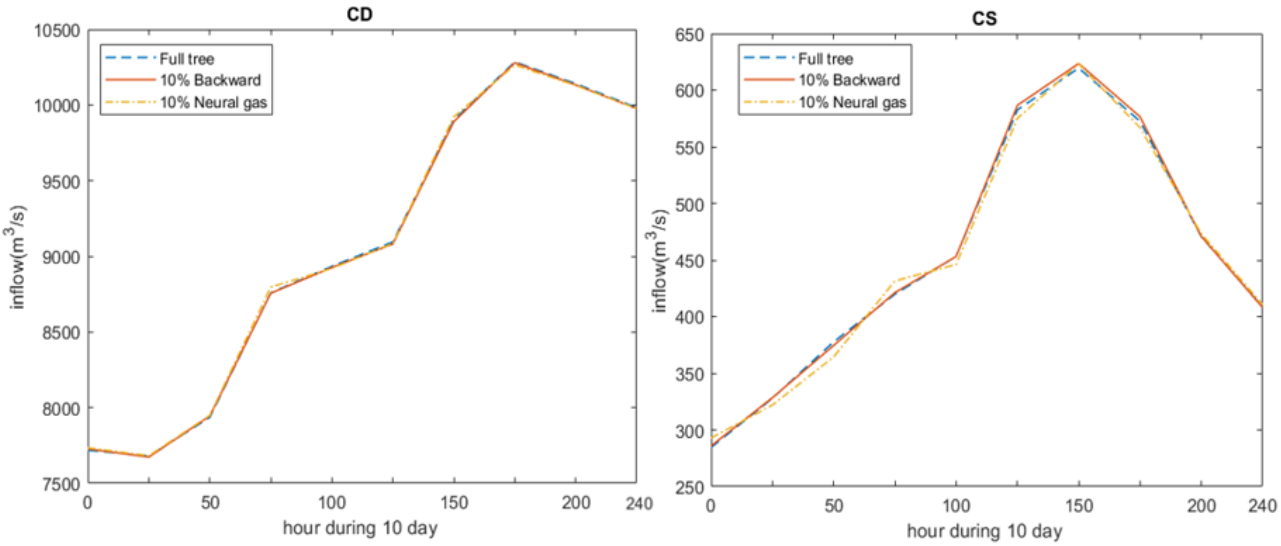


FIGURE 5.12 : The expected volume of the reservoir for the full tree, backward reduction and neural gas with 10% of reduction ©Maissa Daadaa 2023

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.5.13 and 5.14 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.5.13 and Fig.5.14. This is due to the variability of inflow scenarios in this month. For example, Fig.5.4 indicates that the variation in July is greater than in December.

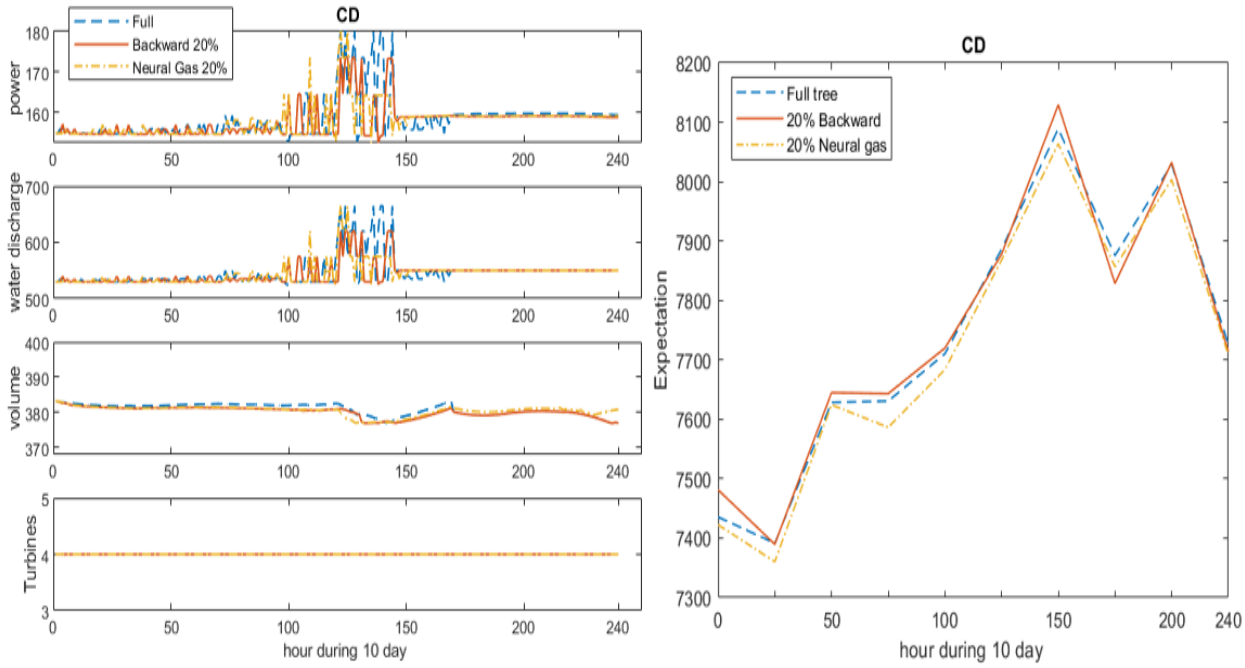


FIGURE 5.13 : Comparison of the solutions for backward reduction, neural gas and the full tree for the powerhouse CD for July with 20% of reduction (left figure) and the expected volume of the reservoir for each method (right figure) ©Maissa Daadaa 2023

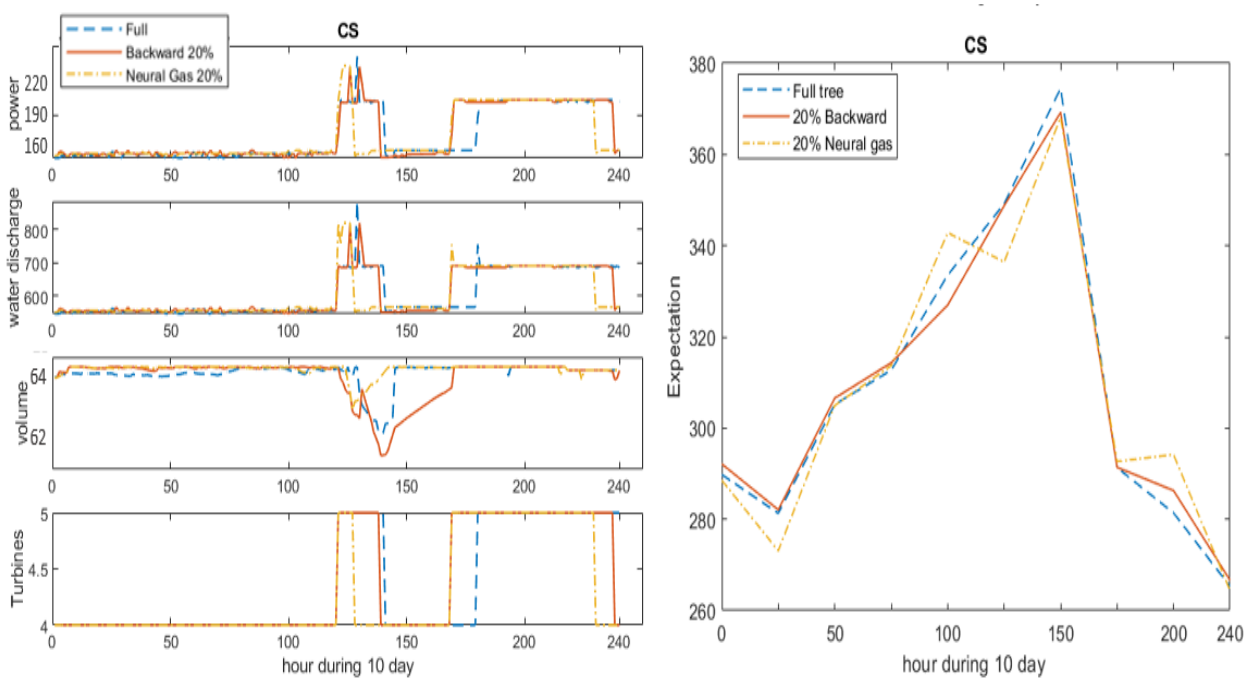


FIGURE 5.14 : Comparison of the solutions for backward reduction, neural gas and the full tree for the powerhouse CS for July with 20% of reduction (left figure) and the expected volume of the reservoir for each method (right figure) ©Maissa Daadaa 2023

5.4.4 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.

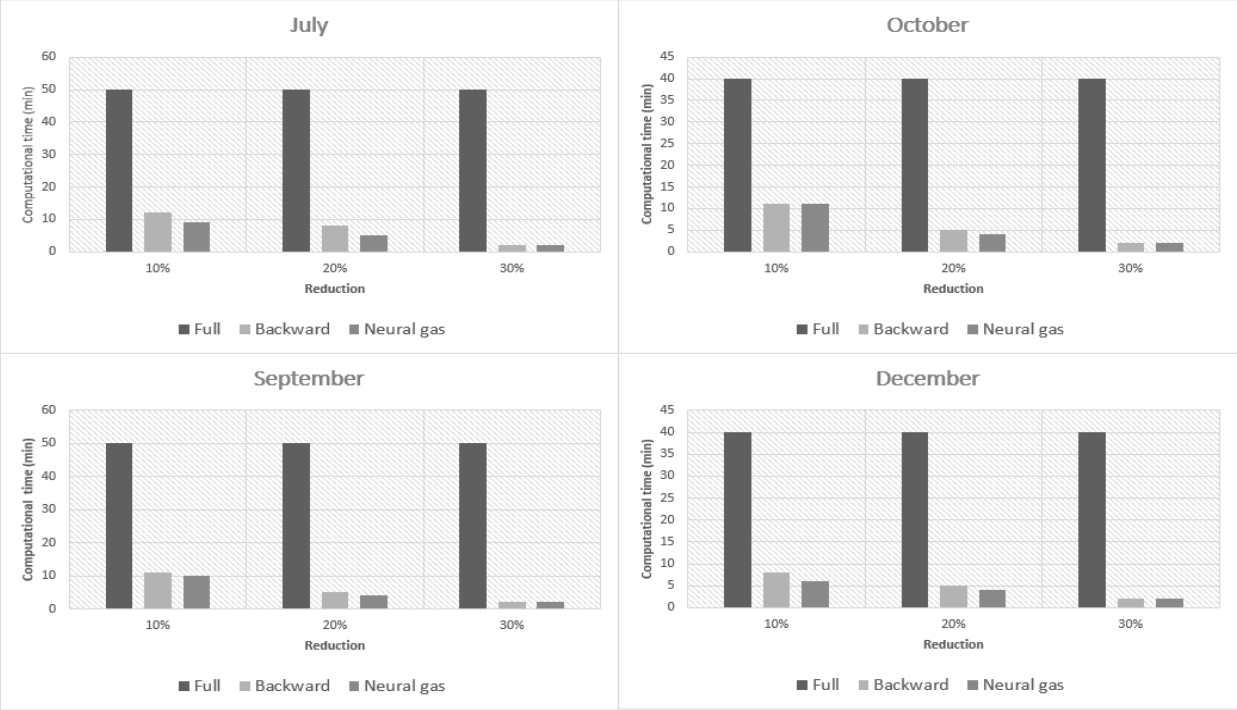


FIGURE 5.15 : Comparison of the average computational time for the full scenario tree and the reduced scenario trees ©Maissa Daadaa 2023

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.5.15. 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.

5.4.5 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.

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.5.16, 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.

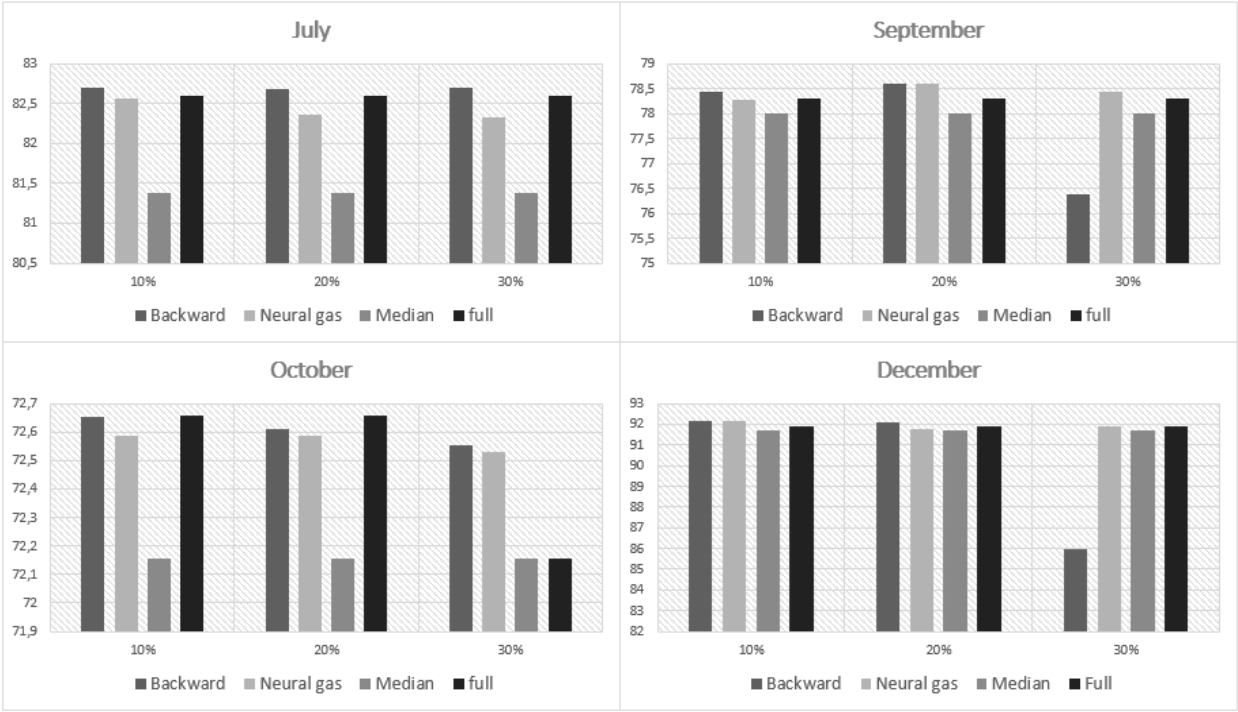


FIGURE 5.16 : Total energy produced for stochastic methods vs median method ©Maissa Daadaa 2023

For example, Fig.5.17 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.

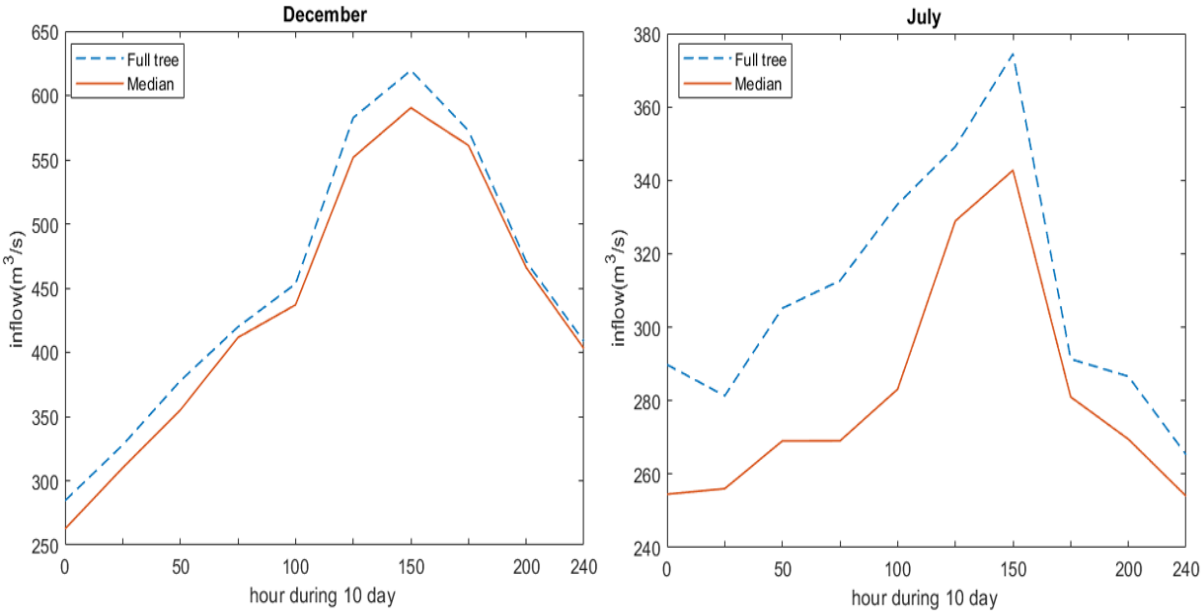


FIGURE 5.17 : A comparison between the expected volume of the reservoir of the median scenario and the stochastic method ©Maissa Daadaa 2023

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.

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 5.4 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%.

TABLEAU 5.4 : The difference of the total energy production between the stochastic methods and the deterministic one ©Maissa Daadaa 2023

	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.5.18 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.5.18, 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.

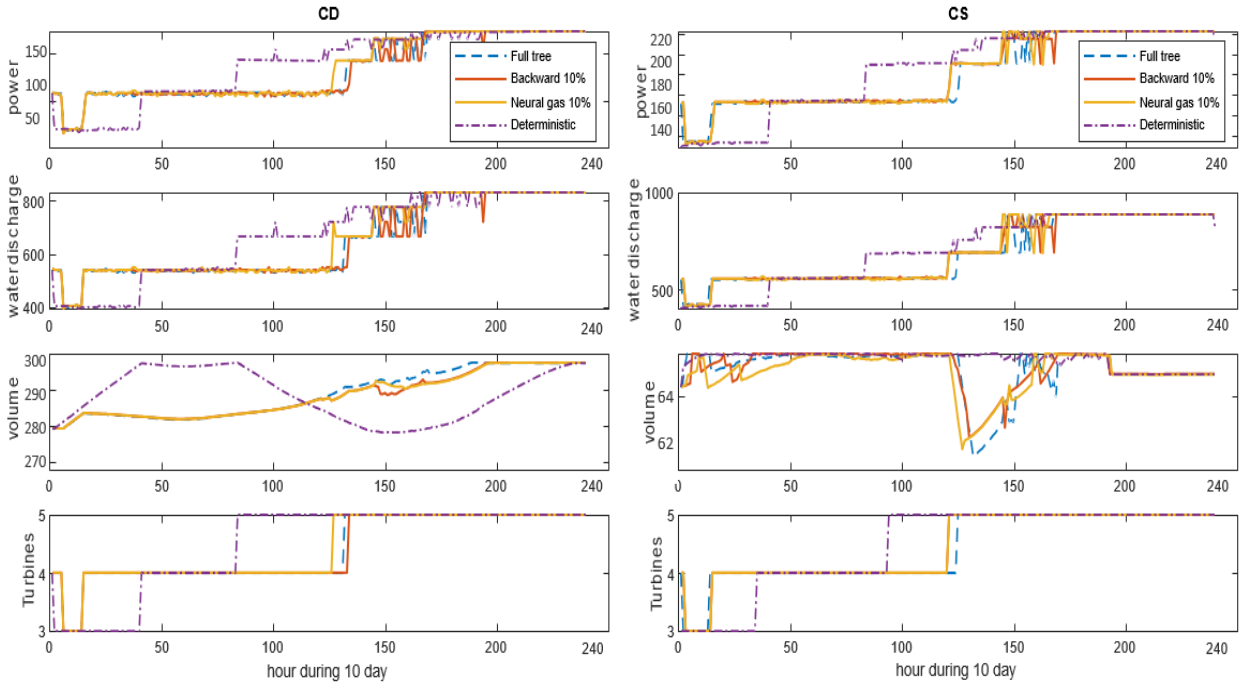


FIGURE 5.18 : A comparison between the deterministic and the stochastic solution ©Maissa Daadaa 2023

5.5 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.

CHAPITRE VI

ARTICLE 3 : OPTIMIZING STRATEGIES FOR SHORT-TERM HYDROPOWER SCHEDULING USING A BLACKBOX OPTIMIZATION FRAMEWORK

Cet article est soumis à la revue IEEE Transactions on Power Systems le 19 septembre 2023.
Le rapport technique suivant est publié :
Daadaa, M., Séguin, S., Anjos, M. F., et Demeester, K. Optimizing strategies for short-term hydropower scheduling using a blackbox optimization framework. Les Cahiers du GERAD, G-2023-38, 2023.

Abstract : This paper presents a study on the utilization of optimization models for short-term hydropower scheduling, which relies on using inflow predictions to select the most suitable model. Different deterministic and stochastic models have been proposed to schedule turbines for an entire planning horizon. In this paper, we take a different perspective on the problem by investigating how to determine the best optimization model during the planning horizon in order to maximize the total energy production based on the inflows, without the need to run over the model. Specifically, we use a mixed-integer linear programming formulation as a deterministic model, where the inflows are represented by a median scenario for each day, and a stochastic multistage mixed-integer linear programming formulation as a stochastic model, where the uncertainty of the inflows is represented using a scenario tree. For the stochastic model, the stages are aggregated to reduce the computation time. We formulate the problem of choosing the optimization model as a blackbox optimization problem with the aim of finding a way to determine the best model to use based on the inflow values. Computational experiments with two powerhouses in series, each with five turbines, for a 10-day rolling-horizon are presented. The solutions provided by the blackbox solver are

analyzed, including an evaluation of the expected inflow volumes. These analyses shed light on the selection of the most appropriate optimization model. The results show that the choice of optimization model is influenced by the observed variability of inflows.

Key words : Short-term hydropower scheduling, mixed-integer linear programming, stochastic multi-stage programming, blackbox optimization.

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

NOMENCLATURE

The following notations are used in the paper :

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.
$d \in \{1,2,..,T\}$:	index of the day in the rolling-horizon.
$m \in [1,2,..,Nc^{max}]$	index of the change time.
$p \in [1,2,..,Nc^{max}+1]$	index of period between times changes
$t_m \in [0,..,T]$	the value of the change time
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).

$$A_{n,k,j}^c = \begin{cases} 1 & \text{if the turbine } j \text{ of powerhouse } c \text{ at the point } k \text{ is activated at node } n. \\ 0 & \text{otherwise.} \end{cases}$$

Nc^{max} maximum number of changes

$$a_p = \begin{cases} 1 & \text{if the deterministic model is used at period p} \\ 0 & \text{otherwise.} \end{cases}$$

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).

6.1 INTRODUCTION

Short-term hydropower scheduling is concerned with defining the optimal strategy for daily energy production by defining the optimal values of the water discharges, the reservoir volumes, and the operational status of the turbines in order to achieve an objective such as maximizing total energy production, minimizing total cost, or maximizing total profit, subject

to operational constraints, including limits on turbine startups. The primary purpose of limiting the number of turbine starts in a powerhouse is to ensure efficient and reliable operation of the turbines and to minimize potential damage or wear to the turbines. A variety of deterministic and stochastic models have been proposed for this problem. A recent overview of the proposed models can be found in [107].

In deterministic models, all the necessary parameters are known in advance. Various formulations and methods have been proposed in the literature [18]. Mixed-integer linear programming (MILP) is widely used to solve hydropower scheduling problem because it involves making decisions on discrete variables, such as when to start and stop the turbines for each powerhouses. MILP can handle these discrete decisions by incorporating integer variables into the optimization model, ensuring that the generated solutions are realistic. In [108], a MILP formulation is used to solve hydroscheduling problem in the case of a deregulated market, taking into account head variations, hydraulic losses, and mechanical and electrical losses in the turbine generators. The formulation was tested on real cases derived from the Brazilian system with constant inflows over the planning horizon. In [94], the authors developed a novel MILP formulation based on the efficiency curves of the turbine generators to solve the hydroscheduling problem in order to maximize the energy produced and penalize the start-ups. The formulation is based on the definition of a pair of points representing the maximum efficiency for water discharge and the power produced for each combination of active turbines since the maximum power generation is reached at these points. The inflow forecasts are considered as known and available from a historical database. The method was tested on a real case from the Saguenay-Lac-St-Jean hydroelectric system in Québec, Canada. In [26], another deterministic MILP approach was developed under the assumption that inflow uncertainties can be ignored in a horizon of days to weeks [109]. Other studies propose MILP

formulations considering irregular forbidden zones of operation [5, 110]. In [111], binary variables are used to handle multiple vibration zones and performance curves of the unit.

Another optimization method commonly used for deterministic hydropower scheduling is dynamic programming. Dynamic programming solves the problem by considering multiple sub-problems so that the optimal solution can be obtained using recursion. In [3], dynamic programming was used to determine the maximum power output generated by a given combination of active turbines. This information is then used as input for a two-phase optimization process that determines the optimal water discharge, reservoir volume and set of turbines in operation at each period of the planning horizon. The time steps are the hours, the states are the number of turbines in operation at each step, and the decision variables are the number of turbine starts and stops at each step. In [112], dynamic programming is used to determine the number of turbines in operation on an hourly basis in order to maximize the power produced. The inconvenience of this method is that the problem quickly becomes difficult to solve as the number of variables increases [113].
Other methods used to solve deterministic hydroscheduling problem include Lagrangian relaxation [114, 18], genetic algorithms [115, 116], and the combination of outer approximation and Benders decomposition [117].

In stochastic models, some parameters of the problem are uncertain. In hydropower systems, uncertain parameters include the natural inflows, the prices (in the case of a deregulated market), the demand, or other aspects of the problem. Considering the uncertainties allow the producer to obtain a robust and precise solution that reflects reality. Different research papers have looked into short-term hydropower scheduling under uncertainty. A multi-stage mixed-integer linear stochastic programming (SMMILP) is one of the more common methods used for this problem. In [43], a SMMILP is developed to solve the hydropower optimization problem subject to the uncertainty of prices and inflows. The objective is to make a balance

between current profits and expected future profits in deregulated market. The model is tested with data from a Norwegian hydropower producer and the Nordic power market. In [118], a two-stage mixed integer stochastic programming is developed and takes the uncertainty in market prices and both production and physical trading aspects. The first stage determines bids for the day-ahead and the second involves trades in the intraday market and production decisions. In [119], a SMMILP is developed in order to maximize the total energy production subject to the uncertainty of the inflows. The decisions are taken at the beginning of the horizon before knowing the realization of uncertainty, then are adjusted once the uncertainties are known. The uncertainty of inflows is represented by a scenario tree. In hydropower management, the usual way of describing uncertain parameters is by scenario trees. It consists of nodes and paths, where each path represents a scenario with an associated probability and the nodes within the scenarios represent the values of uncertainty. The scenario tree serve as an input to the stochastic model. Therefore, several approaches for generating and reducing scenario trees are proposed including clustering methods[120], moment-matching methods[121], backward and forward reduction[58] and Monte Carlo method [122].

Another method proposed in the literature to solve stochastic hydropower optimization models is stochastic dynamic programming. These models are commonly employed to address hydropower problems within long or medium-term planning horizons. In [123], a stochastic hydroscheduling is solved where the uncertainty of the demand is considered in order to minimize the expected cost. A deterministic equivalent linear program is used to define the water flows, and then a stochastic dynamic programming recursion is applied to give an approximately optimal scheduling for a single station. Other methods proposed in the literature to deal with uncertainty include robust optimization [4] and artificial intelligence [124].

The cited works contribute to advancing the field of hydropower optimization by

proposing different techniques for deterministic and stochastic models. The deterministic models offer the advantage of simplicity and computational efficiency compared to the stochastic models. However, they are unable to adequately account for parameter uncertainties, which can lead to suboptimal solutions in the face of unexpected events. To overcome these limitations, an approach that combines deterministic and stochastic methods can be used to find a trade-off between achieving an optimal solution and reducing computational time, since the resolution of the deterministic model is faster than that of the stochastic model. For example, in [119], a comparison of the computational time between the stochastic and deterministic models shows that the resolution with the deterministic model is 12 times faster. To achieve this objective, it is important to determine the appropriate model to use during the planning horizon (either deterministic, stochastic, or both) and the optimal time to transition from one model to the other when a transition is deemed necessary. This decision should be based on some indicators that we aim to determine. To solve this problem, a state-of-the-art blackbox optimization solver is used. Blackbox optimization is a computational tool or algorithm used to find optimal solutions to optimization problems when the underlying objective function or constraints are not explicitly known or easily defined and can only be calculated through a computer code [125]. In [78], blackbox optimization is applied to determine the optimal structure of the scenario tree that maximizes energy production. The scenario tree includes several input parameters, including the number of stages, the number of child nodes for each node, and the aggregation level for each day. Blackbox optimization is used to adjust these parameters in a way that the energy production is maximized.

In this paper, blackbox optimization is used to determine the optimal combination of models and the optimal time to transition from one model to another. For this purpose, a MILP formulation is used as a deterministic model because it provides a flexible framework for modeling and optimizing hydropower scheduling problems. It allows the inclusion of

various constraints that enable the representation of real-world operational requirements such as turbine start-up constraints, reservoir limits, etc. Inflows are represented using the median scenario for each day. A SMMILP formulation is used as a stochastic model because it can handle both discrete and continuous decision variables, making it suitable for the complex nature of hydropower scheduling. This modeling approach enables optimization of decisions such as turbine start-up. Moreover, the use of a SMMILP facilitates the application of aggregation techniques to reduce computational complexity, which is the case in this work. The uncertainty of the inflows is represented by a scenario tree that is generated and reduced using the backward reduction method. This method aims to iteratively identify specific scenarios that are eliminated from a full scenario tree in a way that minimizes the distance of the probability distribution between the reduced scenario tree and the original full scenario tree [101]. Moreover, the implementation of the backward reduction method can be done directly using Scenred2/GAMS [102].

For this purpose, a 10-day rolling horizon benchmark test is developed based on real data from Rio Tinto in the Saguenay region of the province of Québec in Canada. The solution provided by the blackbox solver is analyzed, including an evaluation of the expected inflow volumes, to investigate the choice of optimization models. These analyzes provide valuable insights into the choice of the optimal optimization model. The results show that the choice of optimization model is indeed influenced by the observed variability of inflows.

The paper is organized as follows. Section 6.2 presents the hydropower optimization models. Blackbox optimization is presented in section 6.3. The results are discussed in Section 6.4. Finally, concluding remarks are presented in Section 6.5.

6.2 SHORT TERM HYDROPOWER SCHEDULING

The objective of the hydroscheduling problem is to optimize energy generation by efficiently utilizing the available resources. This entails determining the optimal water discharge, the volume of the reservoirs, and the state of each turbine (on or off) for each powerhouse and at each time stage, subject to some constraints and taking into account different factors such as inflow uncertainty, efficiency, and turbine start-ups. However, the hydropower production function is characterized by nonconvexity and nonlinearity, which poses significant modeling and optimization challenges. To address this complexity, various approximation and linearization techniques have been proposed to effectively represent and analyze the production function in different ways.

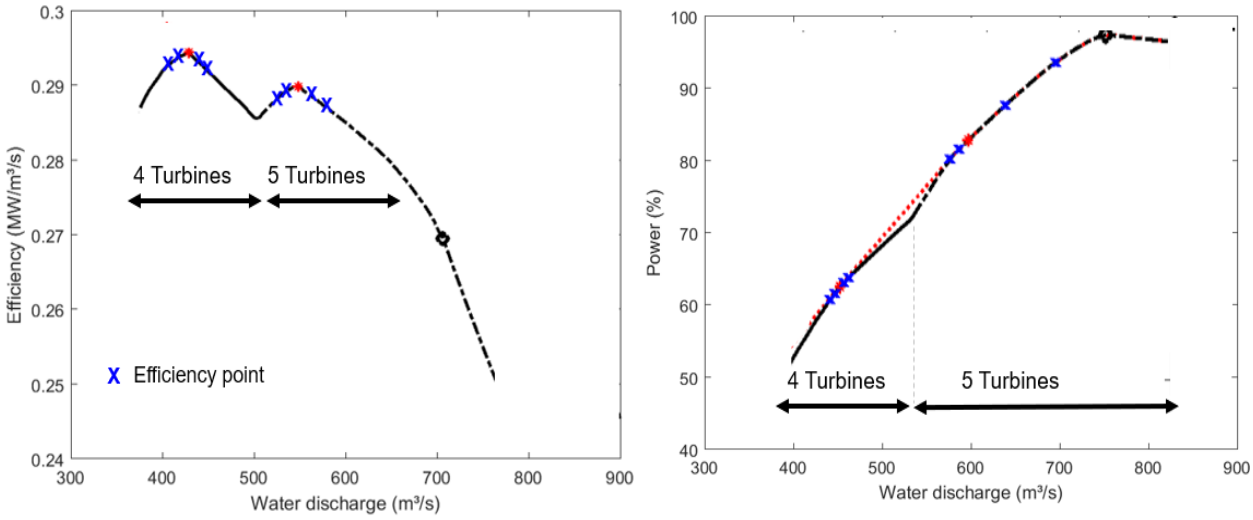


FIGURE 6.1 : The efficiency curves ©Maissa Daadaa 2023

In this paper, a novel technique based on efficiency curves of water discharge and produced power is adopted to develop deterministic and stochastic models. The idea is to use the efficiency curves to determine the efficiency points of water discharge for each combination of active turbines, since the maximum generated power is obtained at these points. Therefore,

pair of points of water discharge and produced power is determined. Fig.6.1 shows an example of the efficiency points for water discharge and power produced for a single combination of 4 and 5 turbines. The model selects one of these points at each decision time to maximise the energy produced. However, to avoid turbine start-ups, a maximum number of turbine changes is imposed to find a viable solution in the practice. This technique was proposed and tested in [94]. The results show that the proposed model outperforms real operational decisions in terms of generated energy. Moreover, by using the efficiency points the number of parameters and variables is reduced and the problem becomes easier to solve. In addition, the producer can directly implement the optimized solution since it is obtained on the efficiency points and thus matches what the engineers want for powerhouses operations. Since our objective is to find a solution that reflects the operational reality, this technique is applied. In this section, the formulations of the proposed deterministic and stochastic models based on this technique are presented.

6.2.1 DETERMINISTIC MODEL

For the deterministic model, a MILP formulation is developed using the efficiency curves to maximize energy produced and penalize turbine start-ups. The objective is to select a pair of points of maximum efficiency of water discharge and power produced and to find the best combinations of active turbines that maximize the total energy produced. Since no uncertainties are considered, only one scenario is used to represent the inflow. The problem is solved and the solutions for the volume, the water discharge, the produced power, and the combination of active turbines are determined. The volume of the reservoir is updated considering the actual realization of the inflow and the optimization process is repeated during the planning horizon. The decision variables are $y^c_{k,n}$, which allow the model to select the efficiency point k for a given water discharge and power produced for each node n and powerhouse c, the volume of

the reservoir v^c_n, and the state of the turbines $z^c_{j,n}$. The proposed formulation is developed in [94] and is employed in this study. The MILP formulation is given by :

$$\max_{y,v,z}[\sum_{c\in C}\sum_{n\in N}\sum_{b\in B^c}\sum_{k\in K^c_b} P^c_{k,n}\times\ y^c_{k,n} - \sum_{c\in C}\sum_{n\in N}\theta^c\times(V^c_{max}-v^c_n) - \sum_{c\in C}\sum_{n\in N}\sum_{j\in J}\varepsilon^c\times z^c_{j,n}]\times\Delta_\tau \tag{6.1}$$

Subject to :

$$v^c_{n+1} = v^c_n + \Delta_\tau\times[(\xi^c_n\times\beta) - \sum_{b\in B}\sum_{k\in K^c_b}(q^c_{n,k}\times y^c_{k,n}\times\beta) - (d^c_n\times\beta) + \sum_{l\in U^c}\sum_{b\in B^c}\sum_{k\in k^l_b}(q^l_{n,k}\times\ y^l_{k,n}\times\beta) + (d^l_n\times\beta)] \tag{6.2}$$
$$,\forall c\in C,\forall n\in N$$

$$\sum_{b\in B^c}\sum_{k\in K^c_b} y^c_{k,n+1}\times A^c_{n+1,k,j} - \sum_{b\in B^c}\sum_{k\in K^c_b} y^c_{k,n}\times A^c_{n,k,j} \le z^c_{j,n}$$
$$,\forall c\in C\ ,\forall n\in N\ ,\forall j\in J \tag{6.3}$$

$$\sum_{b\in B^c}\sum_{k\in K^c_b} y^c_{k,n} = 1 \quad ,\forall c\in C,\forall n\in N \tag{6.4}$$

$$\sum_{i\in S}\sum_{n\in N}\sum_{j\in J} z^c_{j,n} \le N^c_{max} \quad ,\forall c\in C \tag{6.5}$$

$$v^c_{min} \le v^c_n \le V^c_{max} \quad ,\forall c\in C,\forall n\in N \tag{6.6}$$

$$v^c_0 = v^c_{ini} \quad ,\forall c\in C \tag{6.7}$$

$$v^c_N \ge v^c_{final} \quad ,\forall c\in C \tag{6.8}$$

$$y^c_{k,n}, y^l_{k,n}, z^c_{j,n} \in \mathscr{B} \quad ,\forall c\in C,\forall n\in N,\forall b\in B^c$$
$$,\forall k\in K^c_b,\forall l\in U^c \tag{6.9}$$

$$d^c_n, d^l_n, v^c_n \in \mathscr{R}^+ \quad ,\forall c\in C,\forall n\in N,\forall l\in U^c. \tag{6.10}$$

Constraints (6.2) ensure the water balance of the powerhouses. When the powerhouses are connected, the water discharge $q^l_{k,n}$ and the water spillage d^l_n of the upstream powerhouses

are added to the volume v_n^c. Constraints (6.3) establish the relationship between start-up variables and combination choice. Constraints (6.4) ensure that only one operating point is chosen for each powerhouse at each node. The number of start-ups are limited using constraints (6.5). Constraints (6.6)-(6.8) specify the bounds on the reservoir volumes. Finally, (6.9) defines the binary variables and (6.10) the real variables.

6.2.2 STOCHASTIC MODEL

A SMMILP formulation is developed for the stochastic model to account for uncertain inflows in order to maximize energy produced and penalize turbine start-ups. The uncertain inflows are represented using scenario trees. For each day of the planning horizon, a set of scenarios is provided to present the forecast of inflows. Based on the distribution of these inflows, a full scenario tree is constructed. This full scenario tree is generated and reduced using the backward reduction method [48, 58]. This method aims to minimize the probability distribution distance between the full scenario tree and the reduced tree by iteratively selecting the scenarios to be removed. The concept is to calculate the distance between the reduced and the full scenario trees for each time period and delete scenarios when the reduced tree is sufficiently close to the full tree within a certain accuracy. This accuracy is defined as the reduction percentage. It defines the desired reduction in terms of the measured distance between the full and reduced scenario trees. For example, if the reduction percentage is 10%, it means that the distance between the reduced tree and the full tree is less than 10%. This implies that the reduced tree retains 90% of the information present in the full tree.
Fig.6.2 shows an example of a full scenario tree with 55 scenarios which is generated and reduced using the backward reduction method, resulting in a final reduced scenario tree with 21 scenarios. Backward reduction method is used in several papers [103, 101] and the results prove its effectiveness. In [119] a comparison is made between backward reduction and neural

gas methods. A full scenario tree is generated and reduced for each day using the backward reduction and the neural gas methods. Different percentage of reduction (10%, 20% and 30%) are used to build the reduced scenario trees. These reduced scenario trees are used as parameters in the stochastic model. The solutions obtained with both methods are compared, and the results show that the backward method outperforms the neural gas algorithm in terms of preserving statistical information. Moreover, the solutions obtained with 10% of reduction are similar to those obtained with the full tree, but the computation time is higher.

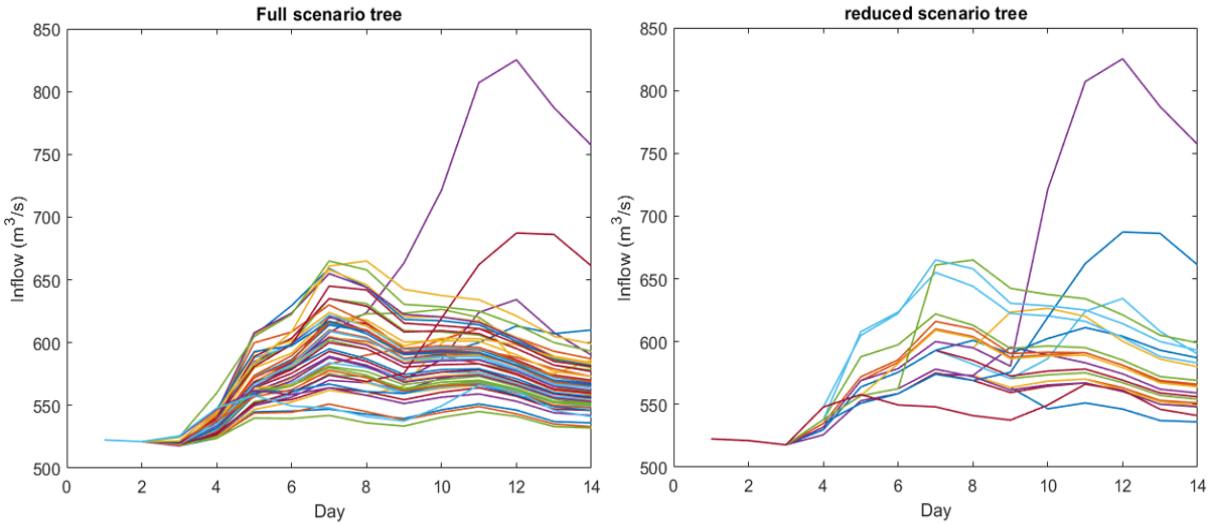


FIGURE 6.2 : An example of the full and the reduced scenario trees ©Maissa Daadaa 2023

In this paper, the backward reduction method with 10% of reduction is used to generate and reduce the scenario tree. Since the inflow forecasts are updated daily, a scenario tree is generated and then the model is solved for each scenario tree node. Once the actual realizations of the inflows are known, the volume of the reservoir is updated and the scenario tree generation and reduction process and optimization are repeated. The objective is to maximize the total energy production in stage 0 and the expected energy production in future stages. The same technique is used as in the deterministic model, where a pair of points of maximum efficiency

of water discharge and power produced, and the best combinations of active turbines are determined. The decision variables are divided into different stages, ensuring that the decisions made in one stage are independent of the information available in subsequent stages. These variables are $y^c_{k,n}$, which allow the model to select the efficiency point k for a given water discharge and power produced for each node n in the scenario tree and powerhouse c, the volume of the reservoir v^c_n, and the state of the turbines $z^c_{j,n}$. The SMMILP formulation is developed in [94] and is given by :

$$\max_{y,v,z}[\sum_{c\in C}\sum_{b\in B^c}\sum_{k\in K^c_b} P^c_{k,0}\times\ y^c_{k,0} - \sum_{c\in C}\theta^c\times(V^c_{max}-v^c_0) - \sum_{c\in C}\sum_{j\in J}\varepsilon^c\times z^c_{j,0} +$$

$$\sum_{i\in S}\pi^c_i\times\Delta_\tau\times[\sum_{c\in C}\sum_{n\in N_i}\sum_{b\in B^c}\sum_{k\in K^c_b} P^c_{k,n}\times\ y^c_{k,n} - \sum_{c\in C}\sum_{n\in N_i}\theta^c\times(V^c_{max}-v^c_n)] -$$

$$\sum_{c\in C}\sum_{n\in N_i}\sum_{j\in J}\varepsilon^c\times z^c_{j,n}] \tag{6.11}$$

Subject to :

$$v^c_{n+1} = v^c_n + \Delta_\tau\times[(\xi^c_n\times\beta) - \sum_{b\in B}\sum_{k\in K^c_b}(q^c_{n,k}\times y^c_{k,n}\times\beta) - (d^c_n\times\beta) +$$

$$\sum_{l\in U^c}\sum_{b\in B^c}\sum_{k\in k^l_b}(q^l_{n,k}\times\ y^l_{k,n}\times\beta) + (d^l_n\times\beta)] \tag{6.12}$$

$$,\forall c\in C,\forall n\in N_i,\forall i\in S$$

$$\sum_{b\in B^c}\sum_{k\in K^c_b} y^c_{k,n+1}\times A^c_{n+1,k,j} - \sum_{b\in B^c}\sum_{k\in K^c_b} y^c_{k,n}\times A^c_{n,k,j} \le z^c_{j,n}$$

$$,\forall c\in C,\forall n\in N_i,\forall i\in S,\forall j\in J \tag{6.13}$$

$$\sum_{b\in B^c}\sum_{k\in K_b^c} y_{k,n}^c = 1 \quad ,\forall c \in C, \forall n \in N_i, \forall i \in S \tag{6.14}$$

$$\sum_{i\in S}\sum_{n\in N_i}\sum_{j\in J} z_{j,n}^c \leq N_{max}^c \quad ,\forall c \in C \tag{6.15}$$

$$v_{min}^c \leq v_n^c \leq V_{max}^c \quad ,\forall c \in C, \forall n \in N_i, \forall i \in S \tag{6.16}$$

$$v_0^c = v_{ini}^c \quad ,\forall c \in C \tag{6.17}$$

$$v_{N_i}^c \geq v_{final}^c \quad ,\forall c \in C, \forall i \in S \tag{6.18}$$

$$y_{k,n}^c, y_{k,n}^l, z_{j,n}^c \in \mathscr{B} \quad ,\forall c \in C, \forall n \in N_i, \forall i \in S, \forall b \in B^c \tag{6.19}$$
$$,\forall k \in K_b^c, \forall l \in U^c$$

$$d_n^c, d_n^l, v_n^c \in \mathscr{R}^+ \quad ,\forall c \in C, \forall n \in N_i, \forall i \in S, \forall l \in U^c. \tag{6.20}$$

Constraints (6.12) ensure water balance at the powerhouses. Constraints (6.13) are the link between start-up variables and the chosen combination of active turbines considering the set of efficiency points. Constraints (6.14) force the model to select only one operating efficiency point at each node for each powerhouse. A maximum number of start-ups N_{max} is imposed with constraints (6.15). Constraints (6.16)-(6.18) are the bounds on the reservoir volumes. Finally, constraints (6.19) define the binary variables and (6.20) the real variables.

6.2.3 OPTIMIZATION PROBLEM

In the realm of optimization modeling, deterministic models are simpler and more computationally efficient than stochastic models, but they cannot adequately account for parameter uncertainty. On the other hand, stochastic models can effectively capture uncertainty, but are generally more computationally expensive. To overcome these limitations, a deterministic and a stochastic methods can be used in alternance during the planning horizon to find a trade-off between achieving an optimal solution and reducing computation time. To achieve this objective, it is important to determine the appropriate model to use during the planning

horizon (either deterministic, stochastic, or both) and the optimal time to transition from one model to the other when a transition is required.

For example, Fig.6.3 illustrates different possibilities for the models used in the optimization process. A deterministic model can be used for the whole planning horizon, a stochastic model can be used for the whole planning horizon, or an alternation between the two models can be used. In this example, a single change time t_1 is proposed. While it is theoretically possible to alternate between models every day, but implementing such a solution may not be practical in practice. Therefore, it is crucial to identify the optimal time for transitioning between the different models. To solve this problem, blackbox optimization can be used to determine the appropriate model choice and the optimal transition times between different models.

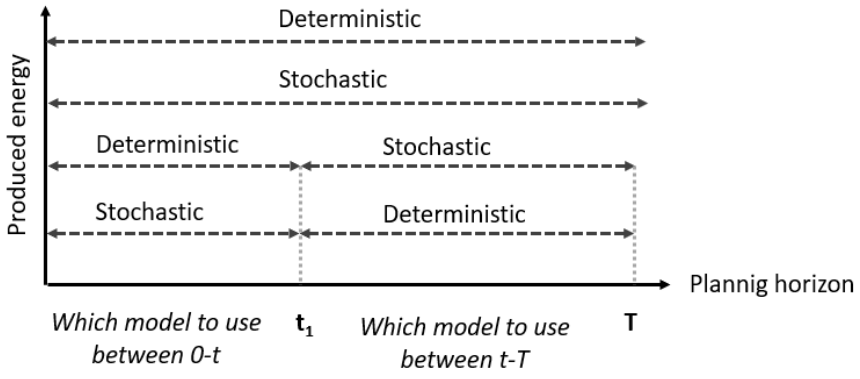


FIGURE 6.3 : Optimization models ©Maissa Daadaa 2023

6.3 BLACKBOX OPTIMIZATION

In this section, the concept of the blackbox optimization is described, then the formulation of the problem is exposed.

6.3.1 BLACKBOX DESCRIPTION

In blackbox optimization (BBO), the internal operations or details of the blackbox functions are not explicitly used or modeled. Rather, the optimization algorithm interacts with the blackbox functions by providing input values and returning corresponding output values. The goal is to iteratively search for the optimal solution by exploring the feasible solution space through a series of function evaluations. The general form of an optimization problem is :

$$min_{x\in\Omega}\ f(x). \tag{6.21}$$

where f is the blackbox function.

The mesh adaptive direct search algorithm (MADS) is an algorithm that solves BBO problems [126]. It is an iterative algorithm that evolved from the generalized pattern search (GPS) [127] which in turn evolved from the coordinate search (CS) [73]. The algorithm evaluates the blackbox at some trial points on a spatial discretization called the mesh. The aims is to generate a finite number of trial points on the mesh in order to find a feasible trial point that improves the current best solution. In MADS algorithm, the selection of points is guided by descent directions. The algorithm aims to find points in the search space that move in a direction of descent, leading to improvement in the objective function value. For those purposes, two steps *search* and *poll* are repeated at each iteration until a predefined stopping criterion is reached. In *search* step, trial points are generated anywhere on the mesh with the aim of finding a better solution than the current best one. In *the poll* step, a list of trial points in the vicinity of the current best solution is generated. If a better solution than the current one is found, it is set as the best solution and the mesh size is increased. If the iteration is fails, the next iteration initiated on a finer mesh to improve the solution. MADS is implemented using Nomad software [128] to solve the BBO problem.

6.3.2 BLACKBOX FORMULATION

In this paper, BBO is used to determine both the choice of model and the timing of transition from one model to another (when a change is deemed necessary) in order to maximize the total energy produced and penalize turbine start-ups. BBO explores various possibilities and evaluates the performance of the different models to determine the most effective combination. To avoid frequent model changes and to obtain a solution that matches the operational reality, a maximum number of changes Nc^{max} is imposed. According to Nc^{max}, the possible number of the change times $t_m \; \forall \; m \in [1,..,Nc^{max}]$ is defined. The choice of the used model between the change times is determined using binary variables $a_p, \forall \; p \in [1,..,Nc^{max}+1]$: 1 means that a deterministic model is chosen during such a period, otherwise, a stochastic model is chosen and therefore the blackbox optimization variables are the times of changes $t_m, \forall \; m \in [1,..,Nc^{max}]$ and the binary variables $a_p, \forall \; p \in [1,..,Nc^{max}+1]$.
For example, in the case where only one change is allowed during the planning horizon T (Nc^{max} =1) :

- t_1 : is the change time from one model to another (if a change is deemed necessary) as shown Fig.6.3.
- $a_p \; \forall \; p \in [1,2]$: a_1 represents the proposed model between [0, t_1] and a_2 represents the proposed model between [t_1+1, T]. For example, if a_1 = 1 and $a_2 = 0$, this means that during [0,t_1] the deterministic model is used and during [t_1+1, T] the stochastic model is used, or, if a_1 = 1 and $a_2 = 1$, this means that the deterministic model is used throughout the planning horizon as shown in Fig.6.3.

Therefore, the blackbox optimization process is described as follows :

1. ***Initialization*** : The optimization process starts by initializing the inputs for the blackbox function. These inputs are the change time $t_m \in [0,T], \forall \; m \in [1,..,Nc^{max}]$ and binary variables $a_p, \forall \; p \in [1,..,Nc^{max}+1]$.

2. ***Evaluation*** : The blackbox function EBB(t_m,a_p), $\forall\ p \in [1,..,Nc^{max}+1]$, $\forall\ m \in [1,..,Nc^{max}]$ that maximizes the total energy produced and penalizes the start-ups whole the planning horizon is evaluated according to the initial input set. If the binary variable a_p,$\forall\ p \in [1,..,Nc^{max}+1]$ is equal to 1, the MILP formulation is used for a specific period based on the given value of the change time $t_m \in [0,T]$. Otherwise, the SMMILP formulation is used. This involves running the function and obtaining the corresponding output value as shown in Fig.6.4.
3. ***Solution Update*** : The obtained output value is compared with the current best solution. If the new output value is better, it becomes the new best solution. The blackbox solver then updates its internal state to reflect this improvement.
4. ***Search for Better Solutions*** : The blackbox solver produces new inputs in order to find a better solution than the best one.
5. ***Check for Convergence*** : The blackbox solver analyzes the newly obtained output values and compares them to the best solution. If a better solution is found, the process continues to search for further improvements. However, if no better solution is found or a convergence criterion is met (10000 evaluations in this work), the optimization process is terminated.
6. ***Output*** : Once the optimization process is completed, the best optimization model to use during the planning horizon and the optimal change times are returned.

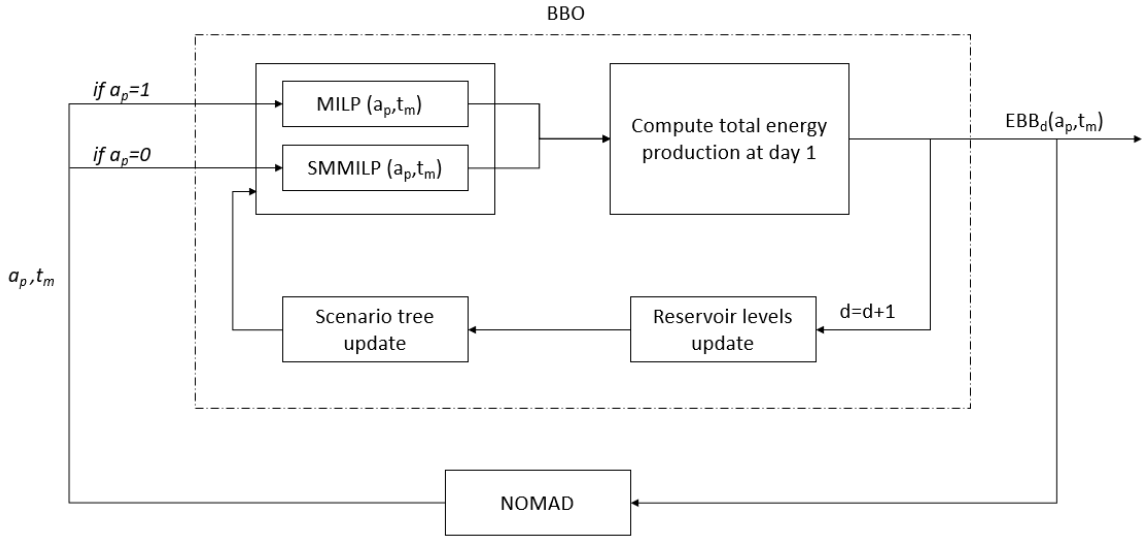


FIGURE 6.4 : Blackbox process ©Maissa Daadaa 2023

The objective function of the BBO problem is denoted by :

$$\max_{a_p,t} \quad \sum_{d\in T} EBB_d(t_m,a_p) \quad ,\forall p \in [1,..,Nc^{max}+1] \\ ,\forall m \in [1,..,Nc^{max}] \tag{6.22}$$

Subject to :

$$\sum_{m\in[1,..,Nc^{max}]} t_m \leq Nc^{max} \tag{6.23}$$

$$t_{m-1} - t_m \leq 1 \quad ,\forall m \in [1,..,Nc^{max}] \tag{6.24}$$

$$t_m < T \quad ,\forall m \in [1,..,Nc^{max}] \tag{6.25}$$

$$a_p \in \mathscr{B} \quad ,\forall p \in [1,..,Nc^{max}+1] \tag{6.26}$$

$$t_m \in \mathbb{N} \quad ,\forall m \in [1,..,Nc^{max}] \tag{6.27}$$

where $d \in T$ is the period (day) in the rolling horizon.

Constraints (6.23) ensure that the maximum number of change times is imposed. Constraints (6.24) and (6.25) are the bounds on change times. Constraints (6.26) define the binary variables and (6.27) the natural variables.

6.4 COMPUTATIONAL RESULTS

This section details the case study and the results on which the BBO problem is solved with the objective of defining the optimization model used during the planning horizon to maximize the energy produced and penalize turbine start-ups.

6.4.1 CASE STUDY

In this paper, the proposed approach is tested with real data from the Saguenay-Lac-St-Jean hydroelectric system owned by Rio Tinto. It is a company that produces aluminium in the Saguenay region. They own a hydroelectric system with five powerhouses that provides 90% of its energy needs. The remaining energy needs are purchased at a known price from Hydro-Québec, a public company responsible for the generation, transmission and distribution of electricity. This means that all producers must buy and sell to them and negotiate fixed price contracts every year. Therefore, the energy price is not considered and only the uncertainties of the inflows are considered. For the purpose of this paper, the analysis focuses on two in-series powerhouses, chute-savane (CS) and chute-du-diable (CD). Each of these powerhouses is equipped with 5 turbines and has two reservoirs with different capacities. A rolling-horizon methodology is used to validate the optimization approach. The planning horizon of the rolling-horizon is of 10 days. The forecast is for 14 days. For day 1 of the rolling-horizon, the forecasts are for days 1 through 14, for day 2 of the rolling-horizon, the forecasts are for days 2 through 15, and so on. For each day of the rolling-horizon, a set of scenarios is provided

to represent the forecast of inflows and the aim is to optimally dispatch the amount of water available in the reservoirs in order to maximize the energy produced and penalize turbine start-ups. Decisions are made hourly for the first day and daily for the next 13 days, and 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. Then, the volume of the reservoir is updated considering the actual realization of the inflow and the optimization process is repeated for the remaining 9 days. The objective of this work is to determine the appropriate optimization model, either deterministic, stochastic, or both, that can be used during the rolling-horizon in order to maximize the total energy produced and ultimately provide a metric to choose the model based on the results. Therefore, the blackbox framework is used to provide insight on the type of the model that should be used, without running the models altogether. The BBO problem is developed to solve this problem, where the blackbox is used everyday during the 10-day rolling horizon. For the deterministic model, the median scenario of all available scenarios is used for every day to represent the inflows, and a scenario tree of 1 node per stage is solved using a MILP formulation. For the stochastic model, the inflow scenarios are generated and reduced using backward reduction method for every day of the rolling-horizon. The problem is solved using a SMMILP formulation. Since stochastic models suffer from long computation times, a small study is conducted to determine the feasibility of using an aggregate stochastic model in order to reduce the computational time. The optimal solutions provided by the BBO solver are analysed to investigate the choice of optimization models during the planning horizon. The formulations are solved using the servers of compute Canada [106] and Xpress solver accessed via Python [93], and the BBO problem is solved using the solver NOMAD [128]. The proposed model is tested using five datasets (July, September, October, and December) for the year 2021 and September for the year 2020. Inflow forecasts are available for each month. The method is evaluated over a rolling period of 10 days, during which decisions are made based on forecasts for the subsequent 14

days. For example, decisions on the first day within the rolling horizon are based on forecasts spanning from day 1 to day 14 of the month. Similarly, on the 10th day, decisions are made based on forecasts ranging from day 10 to day 24 of the month. Therefore, 10 days of decisions are made, and a total of 24 days in the dataset are used for each month.

6.4.2 AGGREGATED STOCHASTIC MODEL

To determine the appropriate optimization model, either deterministic, stochastic, or a combination of both, a BBO problem is used. The main difficulty of the BBO problem resides in the computational time. This is due to the fact that, the SMMILP or/and the MILP formulations are called at each evaluation. The problem is solved at each node of the scenario tree during the 10 days of the rolling-horizon and for 14 forecast days. For the MILP formulation, only one scenario (the median scenario) is used, while for the SMMILP the scenario tree contains several scenarios. Moreover, for every day of the rolling-horizon, the decisions are made hourly for the first day and daily for the next 13 days, which increases the number of optimization variables and the problem become hard to solve. To reduce the computational time, the stages in the SMMILP formulation can be aggregated. Instead of optimizing the decisions for each day, the optimization algorithm focuses on optimizing the decisions for the aggregated intervals. Aggregating days reduces the number of decision variables and constraints in the optimization problem. This leads to faster computations. For that, a small study is conducted in this paper to determine the feasibility of using an aggregated model instead of the stochastic model. However, it is important that the aggregation achieves a balance between computational efficiency and solution accuracy. Since aggregation can be applied in different ways, two different options are tested :

- Hourly for the first stage, every 2 days for the next 5 stages and 3 days for the last stage (Op1)

- Hourly for the first stage, every 3 days for the next 3 stages and 4 days for the last stage (Op2).

A comparison between the energy production and computational time for each available option is made to choose the best option that achieves a trade-off between computational time and solution accuracy. Five data sets (July, September, October and December) for the year 2021 and September for the year 2020 are used. Tab.6.2 illustrates the energy produced for the whole planning horizon obtained from the aggregated model with both options (Op1 and Op2) and the non-aggregated model. Negative values indicate that the first option (Op1) provides slightly higher energy compared to the second option (Op2). The results show that the difference in generated energy between the two options is very small. For example, in September 2021, the difference is only 0.02%.

TABLEAU 6.2 : The difference of the total energy production (GWh) between non-aggregated and aggregated models ©Maissa Daadaa 2023

Data sets	Non-aggregated (10^3)	Aggregated Op1 (10^3)	Aggregated Op2 (10^3)	Difference Op1/Op2 (%)
September 2020	101.989	101.901	101.831	-0.06
July 2021	82.6986	82.2410	82.0337	-0.25
September 2021	78.4234	78.3810	78.3653	-0.02
October 2021	72.6545	72.5838	72.5116	-0,09
December 2021	92.045	91.8205	91.751	-0.07

TABLEAU 6.3 : The difference of the computational time between non-aggregated and aggregated models ©Maissa Daadaa 2023

Data sets	Non-aggregated (min)	Aggregated Op1 (min)	Aggregated Op2 (min)
September 2020	12	10	1
July 2021	12	11	1.2
September 2021	11	9	1.1
October 2021	12	10	1.2
December 2021	12	10	1.2

The average time required to optimize a single day in the rolling-horizon process for the different models is calculated and presented in Tab.6.3. The results indicate that

optimize with option (Op2) is significantly faster when compared to option (Op1). For example, in September 2020, using the aggregated model with the option (Op2) is 12 times faster than the non-aggregated model and 10 times faster than the option (Op1), moreover the difference between the energy produced with both options is only 0.06%. Based on these results, the decision is made to aggregate the stages using option (Op2), where decisions are made hourly for the first day, every 3 days for the next 3 stages, and 4 days for the last stage. However, only the decisions for the volume of the reservoir, the water discharge, the power generated and the turbines in operation of the first stage are kept.

Let us now compare the solutions of the aggregated model using option (Op2) and the non-aggregated model to evaluate the accuracy of the solutions. Fig.6.5 shows an example of the proposed decisions for the month of December : 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 the aggregated model with option (Op2) (dashed line) and for the non-aggregated model (solid line) for two powerhouses CD and CS. As shown in Fig.6.5, the solutions proposed with the aggregated and the non-aggregated models are broadly similar. Moreover, the results are obtained 10 times faster.

The results of this small study show that the aggregation of the stage with the option (Op2) offers advantages in terms of computational efficiency and solution quality. This aggregated model is used when solving the BBO problem.

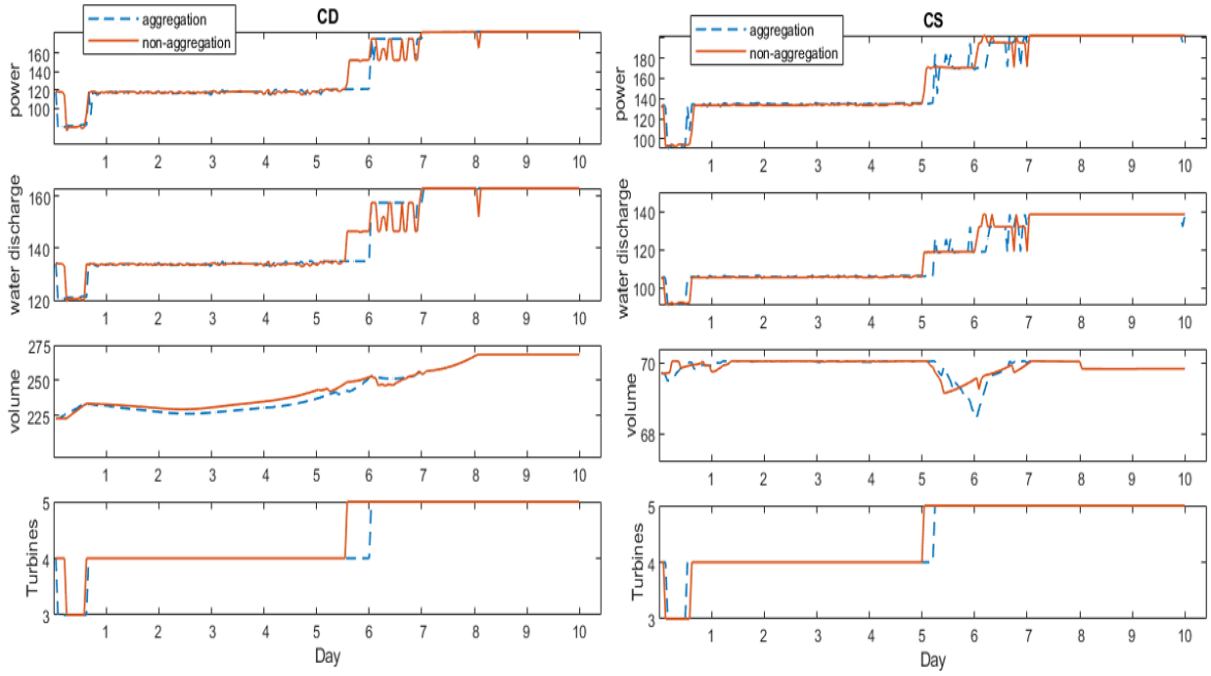


FIGURE 6.5 : Comparison of the solutions obtained from the aggregated and non-aggregated models ©Maissa Daadaa 2023

6.4.3 BLACKBOX SOLUTION

In this section we report the results obtained with the proposed BBO approach. The objective is to define the used optimization model during the planning horizon in a way that maximizes the total energy produced. For that, at each iteration, the solver provides inputs to the blackbox where the MILP and/or SMMILP formulations are called based on the blackbox variables. Then, the solver collects the output (total energy produced) and uses it to generate new inputs to the blackbox in order to find a better solution than the best one. The process is repeated until no improved solution is found or a convergence criterion is met (10000 evaluations in this work). As explained in Section 6.3.2, the maximum number of change time is defining at the beginning. In this work, two tests are made : first, only one change is allowed

during the whole planning horizon. Second, two changes are allowed. The computational experiments are presented in next sections.

COMPUTATIONAL EXPERIMENTS

The BBO model is tested using five data sets (July, September, October and December) for the year 2021 and September for the year 2020. Tab.6.4 illustrates the outputs of the blackbox. BBO1 present the obtained energy produced where only one change time is allowed, and BBO2 when 2 change times are allowed.

TABLEAU 6.4 : Total energy production (GWh) from the optimization models ©Maissa Daadaa 2023

Data sets	Deterministic (10^3)	Stochastic (10^3)	BBO1 (10^3)	BBO2 (10^3)	Proposed model
July 2021	81.3712	82.0337	82.0337	82.0337	Stochastic
October 2021	72.1565	72.5116	72.5116	72.5116	Stochastic
September 2021	78.0841	78.3653	78.3759	78.3759	Deterministic & Stochastic
September 2020	101.451	101.831	101.831	101.999	Deterministic & Stochastic
December 2021	91.727	91.751	91.820	91.827	Deterministic & Stochastic

The results show : for July and October 2021, using the stochastic model during the planning horizon is the best model because maximum energy production is achieved when only the stochastic model is used. For September 2020, December, and September 2021, more energy can be generated by using both models throughout the planning horizon. Tab.6.5 shows the proposed solutions for the cases where both stochastic and deterministic models are used, allowing for a maximum of two changes.

TABLEAU 6.5 : Proposed solutions from blackbox optimization ©Maissa Daadaa 2023

Data sets	Number of changes	Solutions
September 2021	1	6 days stochastic - 4 days deterministic
December 2021	2	6 days deterministic - 2 days stochastic - 2 days deterministic
September 2020	2	1 day stochastic- 1 day deterministic - 8 days stochastic

For example, for September 2021, the blackbox suggests using a stochastic model for the first 6 days and a deterministic model for the next 4 days. Therefore, a change time (at the end of day 6) is proposed, although the maximum number of changes is 2. For December 2021, the blackbox proposes two changes, using the deterministic model for the first 6 days, the stochastic model for 2 days, and finally the deterministic model for the last 2 days.

VALIDATION OF THE METHOD

To evaluate the quality of the solutions provided by the BBO solver, two tests are carried out. First, the solutions obtained by solving the BBO problem with the aggregated model are compared with those obtained by the BBO with the non-aggregated model in order to evaluate the accuracy of the BBO.. Tab.6.6 illustrates an example of the results obtained for the dataset of the month of September 2020.

TABLEAU 6.6 : Proposed solutions from blackbox model using the aggregated and non-aggregated model ©Maissa Daadaa 2023

Data sets	energy produced in GWh (10^3)	Solutions
Aggregated	101.9994	1 day stochastic- 1 day deterministic - 8 days stochastic
non-aggregated	102.0767	1 day stochastic- 1 day deterministic - 8 days stochastic

Tab.6.6 shows that the total energy produced from the BBO with the aggregated model is quite similar to the total energy produced from the BBO with the non-aggregated model. In addition, both models propose the same solutions. The difference in the total energy produced is only 0.07%. However, the solution of the BBO with the aggregated model is 18 times faster than the non-aggregated model and allows 6 times less use of memory capacity.
Secondly, a comparison is made between the results obtained with the BBO method and the results derived from the actual inflow realizations in order to assess the quality of the methods. Table 6.7 illustrates the difference between the energy produced with the BBO method and those resulting from the real inflow realizations. The results show that the difference is small and does not exceed 0.68%.

TABLEAU 6.7 : Energy produced in GWh using BBO method and the real realization
©Maissa Daadaa 2023

Data sets	energy produced with real realization (10^3)	energy produced (BBO) (10^3)	Difference(%)
December 2021	92.054	91.827	0.24
September 2021	78.909	78.3653	0.68
September 2020	102.302	101.999	0.29

ANALYSIS OF THE RESULTS

The solutions obtained from the BBO1, BBO2, deterministic and stochastic models for the 5 test cases are analyzed and compared. For example, Fig.6.6 shows the power produced, the amount of water discharge, the volume of the reservoir and the number of turbines in operation for the two powerhouses CD and CS for the month of December 2021.

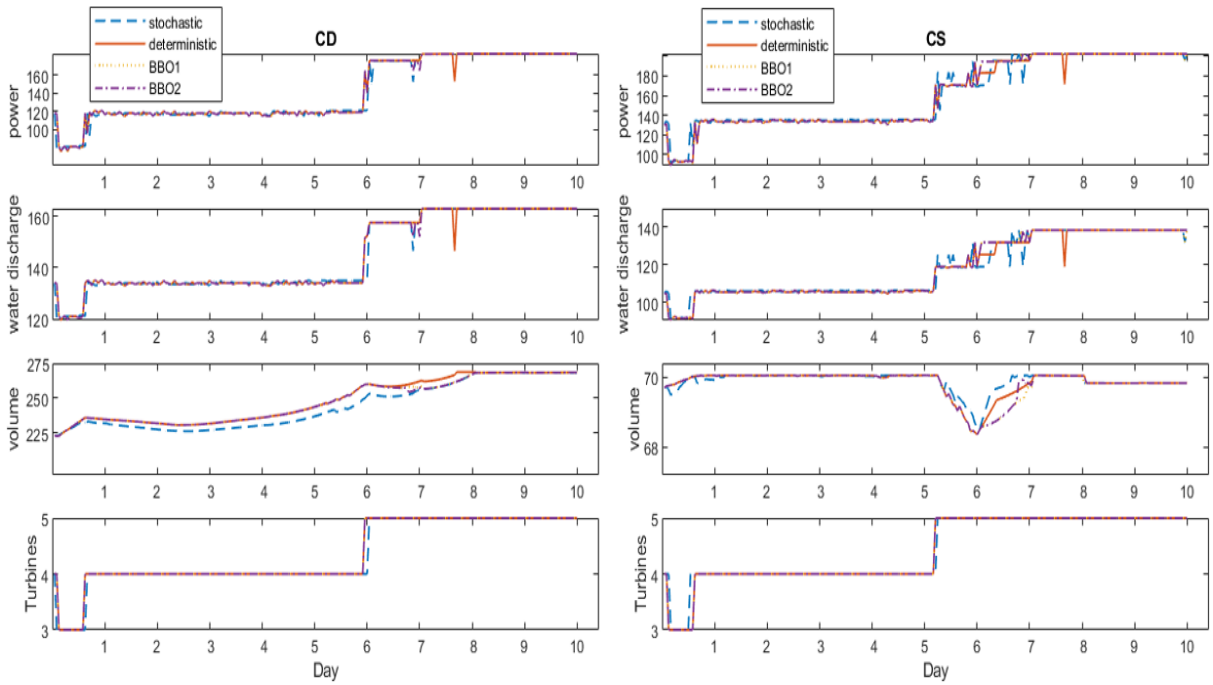


FIGURE 6.6 : Solutions for stochastic, deterministic, BBO1 and BBO2 models for the December 2021 case ©Maissa Daadaa 2023

According to Tab.6.5, for December 2021, a deterministic model is used for the first 6 days, followed by a stochastic model for the next 2 days, and a deterministic model for the last 2 days. This distinction can be seen in Fig.6.6, where the curves of the BBO2 and deterministic models match up to day 6. From day 6 to 8, a noticeable change in the curve's shape occurs and after day 8, the curves of the BBO2 and deterministic models overlap again. The same observations hold for the other cases.

To gain insight into the factors influencing the choice of the optimization model, an analysis is conducted on the expected volume of inflows for both the stochastic and deterministic models. For the deterministic model, the expected volume of the median scenario is calculated at each day of the rolling horizon. For the stochastic model, since different scenarios are available, the expected volume is calculated as follows :

$$EV = \sum_{i \in S} \pi_i \times \sum_{n \in N_i} \xi_n \tag{6.28}$$

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

Fig.6.7 illustrates the expected volume of inflows for both models for the powerhouses CD and CS for December 2021. The results show that the expected volumes are quite similar for the first 6 days before a significant discrepancy occurs between the 6th and 8th day, especially for the powerhouse CS. After the eighth day, the expected volume of inflows for the stochastic and deterministic models converge again and become quite similar which is consistent with the results obtained in Tab.6.5.

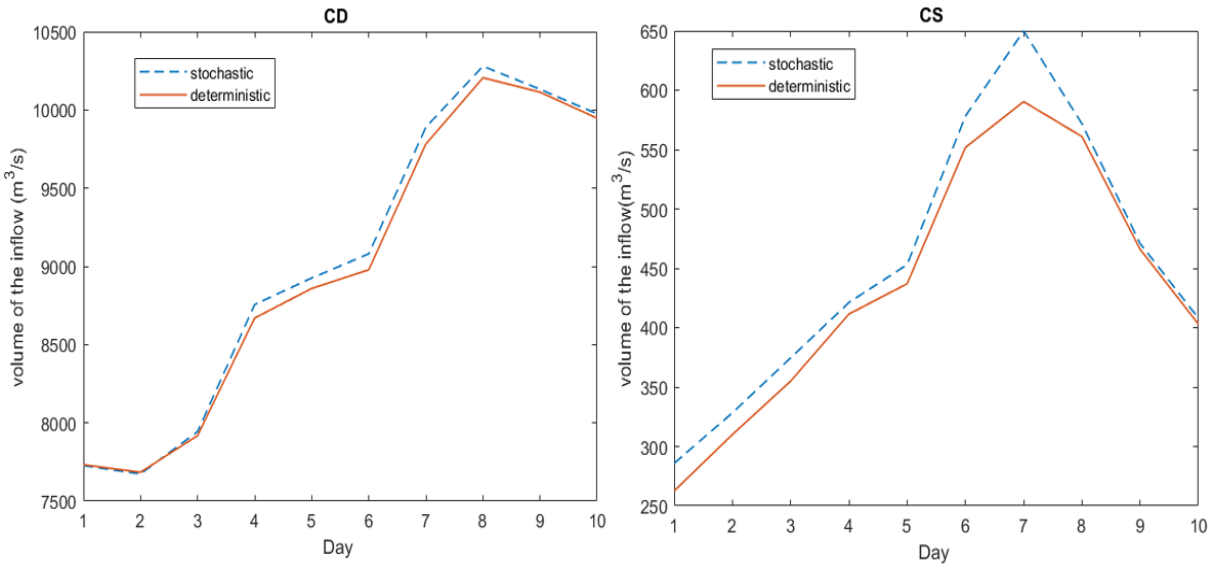


FIGURE 6.7 : The expected volume of inflows for the stochastic and deterministic models for the December 2021 case ©Maissa Daadaa 2023

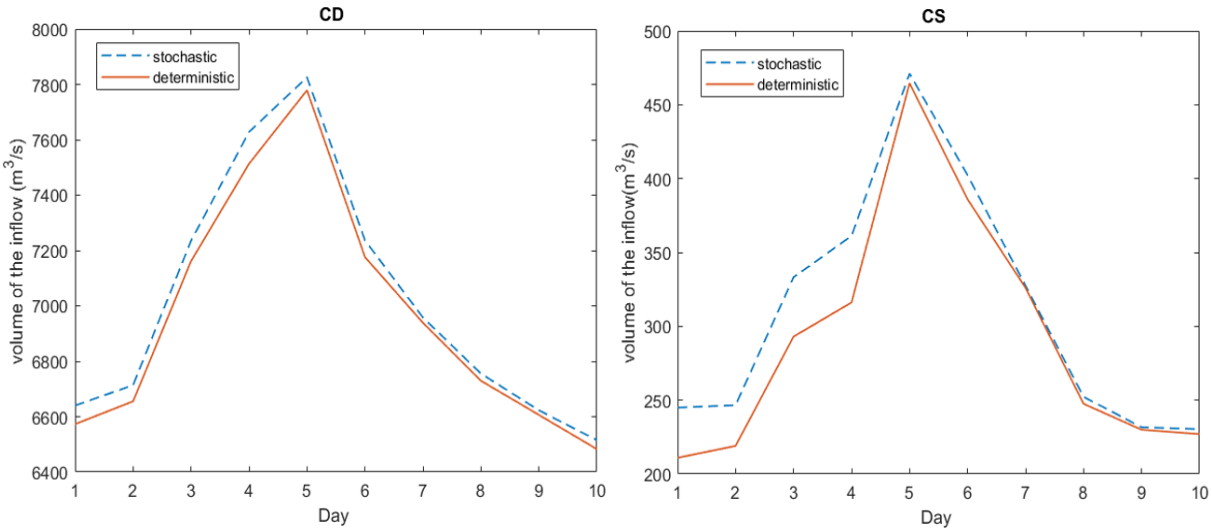


FIGURE 6.8 : The expected volume of inflows for the stochastic and deterministic models for the September 2021 case ©Maissa Daadaa 2023

For September 2021 and September 2020, the same observations hold. For example, Fig.6.8 illustrates the expected volume of inflows for both models for September 2021. It is worth noting that Tab.6.5 confirms the use of a stochastic model for the first six days, and this distinction is clearly evident in Fig.6.8 where a noticeable difference in the expected volume of inflows can be observed during this period for the powerhouse CS.

Let us analyze the cases where the stochastic model is throughout the entire planning horizon. For example, Fig.6.9 illustrates the expected volume of inflows for the month of October. For both powerhouses, it is noticed that a substantial disparity exists between the expected volume of inflows for the stochastic and deterministic models.

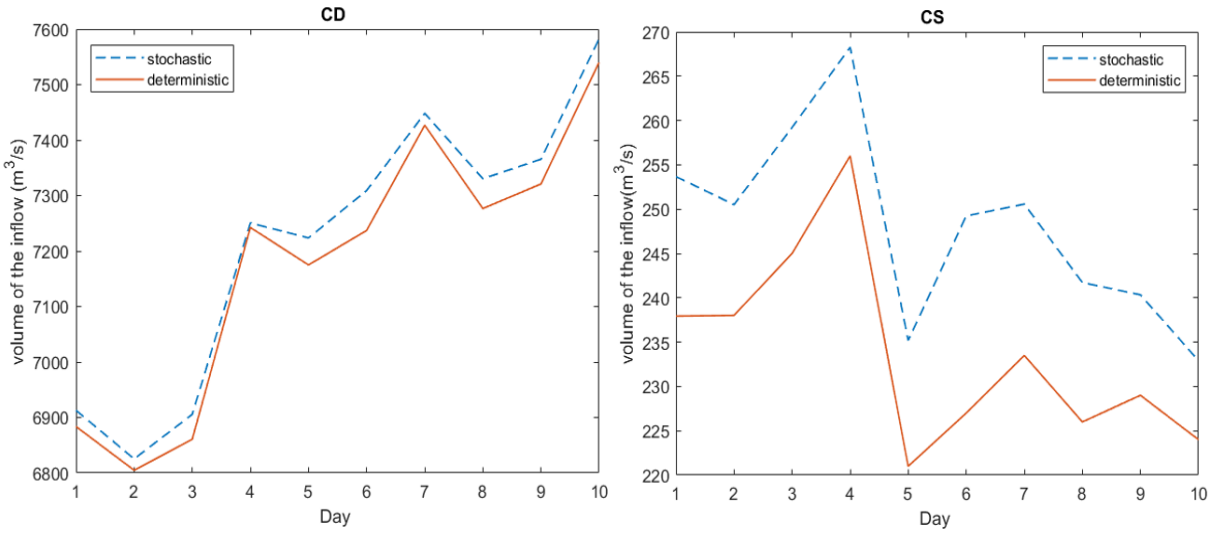


FIGURE 6.9 : The expected volume of inflows for the stochastic and deterministic models for the October 2021 case ©Maissa Daadaa 2023

According to Fig.6.7, 6.8, 6.9 and Tab.6.5, we notice that the deterministic model is used when the expected volume of inflows from deterministic and stochastic models are quite similar. Otherwise, a stochastic model is used. To gain further clarity, a study of inflow variability was conducted to understand its impact on the choice of optimization model. This analysis can provide valuable insight into how the choice of optimization model may be affected by the observed variability of inflows. For this purpose, a boxplot is used as shown in Fig.6.10.

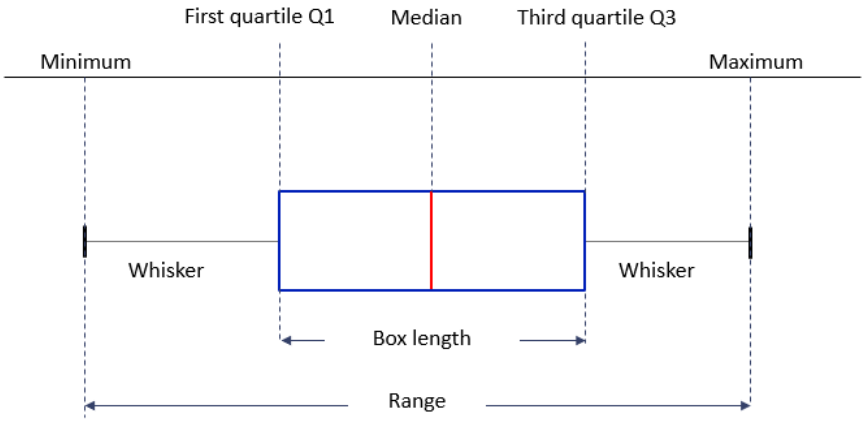
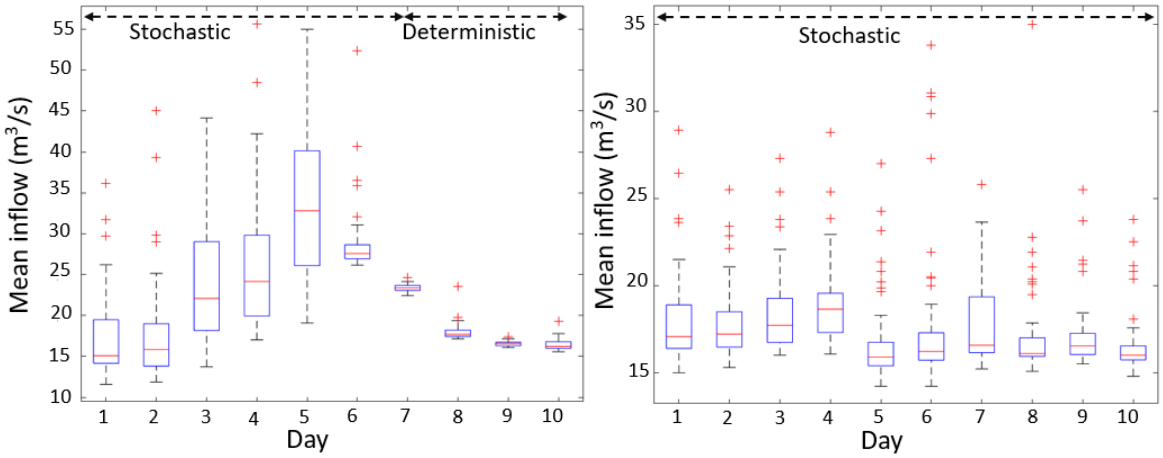


FIGURE 6.10 : A boxplot ©Maissa Daadaa 2023

A boxplot is a graph that indicates how the values in the data are spread out based on five sample statistics : the minimum, the maximum, the first quartile(Q1), the median and the third quartile(Q3). The lower box represents the first quartile and the top of the box represent the third quartile. The middle of the box represents the median. The top of the line linking the box to the whisker is the maximum of the data and the bottom is the minimum of the data. Except for the five sample statistics, the boxplot shows the outliers which ares the values that are much higher or much lower compared to the other values in the set (usually 1.5 times the interquartile range IQR =Q3-Q1). Using the boxplot gives an indication of the sample variability by examining the range (length of the entire box whose the length of the whiskers), IQR, or by the outliers.

In this paper, the boxplot is used to define the variability of the inflows for the 5 test cases. For that, the mean of each scenario in the scenario tree is calculated and presented with a boxplot for every day in the rolling-horizon. For example, Fig.6.11 shows the boxplots for September and October 2021. As shown in Tab.6.5, for September 2021 the stochastic model is used for the first 6 days and the deterministic one for the last 4 days. According to Fig.6.11a, it is observed that the size of the boxplot up to day 6 is very important compared to the sizes of the

boxplots from day 7 onward. For example, for day 5, the range of the boxplot =*37 (55-18)* and for day 7, the range = *2*. This means that the variability of the inflows in the first six days is very large. For this reason, the stochastic model is used for these days. The same is noticed for December 2021 and September 2020. In addition, it is noted that the blackbox of the mean inflow suggests the use of the stochastic model when the range of the boxplot exceeds 10.



(a) September 2021 (b) October 2021

FIGURE 6.11 : The boxplot for the mean inflow for September and October 2021 cases for the powerhouse CS ©Maissa Daadaa 2023

For October 2021, the stochastic model is used for the entire planning horizon, although the size of the boxplot is not important for these days as shown Fig.6.11b. This is because there are many outliers during this period (the red dots in the Fig.6.11b). These outliers represent the extreme scenarios. For this reason, the stochastic model generates more energy as the extreme scenarios are considered unlike the deterministic model. The same is noticed for the month of July 2021.

Based on the analyses conducted throughout this study, it appears that the choice

between a deterministic model, a stochastic model, or both, depends on the expected inflows. When the expected volume of inflows has a high similarity between the deterministic and stochastic models and the variability of inflows is relatively low (i.e., the range of the boxplot of the mean inflows is less than 10 or the number of extreme scenarios is very small), the deterministic model proves to be a viable option, offering the advantage of low computational time. However, when the expected volume of inflows has significant differences and the variability of inflows is high, the stochastic model is strongly recommended. These results demonstrate the importance of carefully considering the degree of variability of uncertainties when selecting appropriate models.

6.5 CONCLUSION

This paper presents a study on the choice of optimization models. The objective is to determine the best model that maximizes the energy produced during the planning horizon. Therefore, deterministic and stochastic models can be used in different periods of the planning horizon. For the deterministic model, the median scenario of all available scenarios is determined for every day and a scenario tree with one node per stage is solved using a MILP formulation. For the stochastic model, the inflow scenarios are generated and reduced using backward reduction method for every day of the planning horizon. The problem is solved using a SMMILP formulation. The stages are aggregated to reduce the computation time. The best optimization model is formulated as a blackbox optimization problem. Therefore, at each iteration, the solver provides inputs to the blackbox where the MILP and/or SMMILP formulations are called. Then, the solver collects the output (total energy produced) and uses it to generate new inputs to the blackbox in order to find a better solution than the best one. The process is repeated until no improved solution is found or a convergence criterion is met (10000 evaluations). The problem is tested for five test cases. First, the aggregated model is

compared to the non-aggregated one in order to evaluate the quality of the solution. The results show that the solutions are quite similar. Second, the proposed solutions from the blackbox problem are presented and analyzed. The results show that, for July and October 2021, using the stochastic model for the whole planning horizon allows to produce more energy. This is due to the fact that in these months the expected volume of inflows from the scenario tree method is significantly different from that of the median scenario. Moreover, the variability of inflows is very high (the range of the boxplot of the mean inflows exceeds 10 or numerous extreme scenarios are observed). For the months of September, December 2021, and September 2020, more energy can be generated with both the deterministic and stochastic models. Moreover, the results show that in some periods, where the expected volume of inflows of the scenario tree method is quite similar to that of the median scenario and the variability of inflows is low, the deterministic model is used for these periods. Therefore, it is important to thoroughly consider the degree of variability of uncertainties when selecting the appropriate models for optimal hydropower management. For future research, we intend to conduct experiments with a more extensive data set. Although the current approach was tested with 5 datasets, expanding to a larger and more diverse data set will provide valuable insights and explore a metric that can effectively reveal the best optimization model.

CONCLUSION

Ce chapitre présente une conclusion sur les travaux effectués durant cette thèse. Des limites et des perspectives sont exposées aussi afin d'améliorer les travaux futurs.

Synthèse des travaux

Le premier objectif de cette thèse consiste à développer un modèle déterministe afin de résoudre le problème du chargement des unités. Le but est de trouver une planification optimale afin de maximiser l'énergie produite. Les modèles d'optimisation à court terme sont complexes et dépendent de plusieurs facteurs : de la fonction de production qui est une fonction non linéaire et non convexe, de l'efficacité des turbines qui se diffère d'une turbine à une autre, et des démarrages qui réduisent la durée de vie des turbines et entraînent des pertes en termes d'énergie et des coûts de maintenance. Ces facteurs sont considérés dans ce premier modèle déterministe. Pour ce faire, des courbes d'efficacité des débits turbinés et de puissances produites pour chaque combinaison des turbines possible sont utilisées. Ces courbes permettant de déterminer les points (les bosses) d'efficacité des débits turbinés et c'est en ces points que le maximum de la puissance produite est atteint. Un modèle linéaire mixte en nombres entiers est développé afin de choisir le meilleur couple débit/puissance à chaque période afin de maximiser l'énergie produite en limitant les démarrages des turbines. Les résultats numériques démontrent que le modèle proposé performe bien comparant aux décisions opérationnelles historiques. De plus, une comparaison entre l'énergie estimée par le modèle proposé et la valeur réelle de la fonction de production chez Rio Tinto a montré l'efficacité de la méthode proposée.

Le deuxième objectif adopte le premier modèle, mais est étendu afin de prendre en compte l'incertitude des apports. La représentation des apports naturels est effectuée en utilisant des arbres de scénarios. La génération de ces arbres de scénarios est faite en utilisant

deux méthodes : le gaz neuronal et backward reduction. Une comparaison entre ces deux méthodes est effectuée afin d'étudier l'impact du choix de la méthode sur la solution. Pour ce faire, les prévisions d'apports sont mises à jour quotidiennement et les arbres de scénarios sont générés et réduits. Ces arbres de scénarios sont utilisés comme des paramètres pour résoudre l'équivalent déterministe d'un modèle stochastique où le modèle est résolu à chaque noeud de l'arbre et seulement la solution du premier noeud est conservée. Une fois les vrais apports sont connus, les volumes des réservoirs sont mis à jour et le processus est répété durant un horizon de planification roulant. Les résultats ont montré qu'il est recommandé d'utiliser une méthode de génération qui conserve la variance lorsque la variabilité des scénarios est élevée, car elle permet d'avoir plus d'énergie. De plus, une comparaison entre les modèles stochastique et déterministe (où le scénario médian des apports est utilisé) est effectuée. Cette comparaison démontre que l'utilisation du modèle stochastique permet d'augmenter la production d'énergie, mais dans certains cas, l'énergie produite du modèle déterministe est assez similaire à celle du modèle stochastique.

Le dernier objectif de cette thèse étudie le choix d'utilisation des modèles d'optimisation pour le problème d'optimisation hydroélectrique à court terme. Le but est de déterminer le meilleur modèle d'optimisation à utiliser dans chaque période de l'horizon de planification afin de maximiser l'énergie totale produite en se basant sur les prévisions. Pour ce faire, les deux modèles proposés dans les deux premiers objectifs sont utilisés avec une agrégation pour le modèle stochastique afin de minimiser le temps de calcul. Le modèle agrégé est comparé au modèle non agrégé afin d'évaluer la qualité de la solution. Les résultats montrent que les solutions sont assez similaires. Le problème est formulé comme un problème d'optimisation boite noire. Les résultats numériques démontrent que, pour les mois où les volumes espérés dans les réservoirs fournis par la méthode de l'arbre des scénarios sont différents à ceux du

scénario médian et la variabilité des apports est très importante, le modèle stochastique permet d'augmenter la production d'énergie. Autrement le modèle déterministe peut être utilisé.

Limitations et perspectives

Dans le cadre de la maximisation de l'énergie produite, les modèles l'optimisation cherchent à turbiner le maximum d'eau afin d'augmenter la production de l'énergie. Dans le monde réel, il est important d'empêcher les réservoirs de se vider à la fin de l'horizon de planification. Pour ce faire, deux façons sont possibles. La première consiste à imposer une contrainte du volume final afin d'assurer que les réservoirs ne seront pas vidés. Cette contrainte permet de maintenir un niveau minimum d'eau dans les réservoirs, assurant ainsi la continuité de la production hydroélectrique. La deuxième approche consiste à utiliser des fonctions de valorisation finale de l'eau. Ces fonctions attribuent une valeur à l'eau restante dans les réservoirs à la fin de l'horizon de planification, en cherchant à maximiser cette valeur. Dans le cadre de cette thèse, une contrainte sur le volume final est imposée dans les modèles d'optimisation pour les trois objectifs considérés. Les valeurs des volumes finaux des réservoirs pour les centrales Chute-de-Diable et Chute-de-Savane sont déterminées à partir de données historiques. Cependant, pour les modèles d'optimisation stochastique, il est recommandé d'utiliser des fonctions de valorisation finale de l'eau. Étant donné que ces modèles prennent en compte plusieurs scénarios d'apports, l'imposition d'une cible de volume pour tous les scénarios pourrait conduire à des solutions potentiellement sous-optimales ou non réalisables. Ainsi, l'utilisation de fonctions de valorisation finale de l'eau permet de prendre en compte de manière plus réaliste et précise l'état final des réservoirs dans le processus d'optimisation stochastique.

Dans les modèles d'optimisation proposés dans cette thèse, les contraintes de bilan hydrique permettant de déterminer le volume réel des réservoirs à la fin de chaque période

sont imposées. Le volume final est calculé à partir du volume actuel, des quantités des apports entrant dans les réservoirs, des débits turbinés et des déversements. Les déversements correspondent à des quantités d'eau déversées et non utilisées pour la production d'électricité. Une mauvaise gestion lors de la prise de décision quant à la quantité d'eau à déverser peut entraîner une perte d'énergie. Il serait intéressant de prendre en considération cette perte en pénalisant les déversements dans la fonction objectif.

Concernant l'étude du choix du modèle d'optimisation abordée dans le troisième objectif, les analyses ont révélé que lorsque l'étendue de la boîte à moustaches dépasse 10, l'utilisation du modèle stochastique est préférable. Ces résultats ont été obtenus et validés à partir de 5 études de cas spécifiques. Cependant, il serait intéressant de tester les modèles sur un ensemble plus large de données. En intégrant des données supplémentaires, la sélection du modèle d'optimisation peut être encore affinée et une métrique plus robuste peut être déterminée. De plus, augmenter le nombre de centrales hydroélectriques dans l'analyse augmenterait la complexité du problème et permettrait ainsi une évaluation plus approfondie des performances et de l'adaptabilité des modèles d'optimisation dans divers contextes. Dans le même ordre d'idées, il serait avantageux de développer un outil d'analyse qui examine la distribution et la variabilité des apports en se basant sur une métrique, afin d'aider à la prise de décision.

BIBLIOGRAPHIE

[1] I. E. Agency, “Renewables 2022,” https://www.iea.org/fuels-and-technologies/renewables, 2023, consulté le : 2023-04-07.

[2] H.-Q. C. energy provider, “Canada, deuxième producteur mondial d’hydroélectricité,” https://www.hydroquebec.com/clean-energy-provider/clean-energy.html, 2023, consulté le : 2023-05-03.

[3] S. Séguin, P. Côté, et C. Audet, “Self-scheduling short-term unit commitment and loading problem,” *IEEE Transactions on Power Systems*, vol. 31, n° 1, pp. 133–142, 2015.

[4] M. F. Anjos, A. J. Conejo *et al.*, “Unit commitment in electric energy systems,” *Foundations and Trends® in Electric Energy Systems*, vol. 1, n° 4, pp. 220–310, 2017.

[5] Z. Zhao, C. Cheng, X. Jin, L. Liu, et L. Yan, “A milp model for hydro unit commitment with irregular vibration zones based on the constrained delaunay triangulation method,” *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106241, 2020.

[6] J. Kumar, I. U. Khalil, A. U. Haq, A. Perwaiz, et K. Mehmood, “Solver-based mixed integer linear programming (milp) based novel approach for hydroelectric power generation optimization,” *IEEE Access*, vol. 8, pp. 174 880–174 892, 2020.

[7] S. Maheswari et C. Vijayalakshmi Seshathri, “An optimal design to schedule the hydro power generation using lagrangian relaxation method,” dans *Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012*. Springer, 2012, pp. 723–730.

[8] F. Rouhi et R. Effatnejad, “Unit commitment in power system t by combination of dynamic programming (dp), genetic algorithm (ga) and particle swarm optimization (pso),” *Indian Journal of Science and Technology*, vol. 8, n° 2, p. 134, 2015.

[9] Y. Shang, S. Lu, J. Gong, R. Liu, X. Li, et Q. Fan, “Improved genetic algorithm for economic load dispatch in hydropower plants and comprehensive performance comparison with dynamic programming method,” *Journal of hydrology*, vol. 554, pp. 306–316, 2017.

[10] S. Dreyfus, “Richard bellman on the birth of dynamic programming,” *Operations Research*, vol. 50, n° 1, pp. 48–51, 2002.

[11] J.-F. Hêche, T. M. Liebling, et D. De Werra, *Recherche opérationnelle pour ingénieurs*. PPUR presses polytechniques, 2003, vol. 2.

[12] A. Arce, T. Ohishi, et S. Soares, “Optimal dispatch of generating units of the itaipú hydroelectric plant,” *IEEE Transactions on power systems*, vol. 17, n° 1, pp. 154–158, 2002.

[13] J. I. Pérez-Díaz, J. R. Wilhelmi, et L. A. Arévalo, “Optimal short-term operation schedule of a hydropower plant in a competitive electricity market,” *Energy Conversion and Management*, vol. 51, n° 12, pp. 2955–2966, 2010.

[14] P. K. Singhal et R. N. Sharma, “Dynamic programming approach for solving power generating unit commitment problem,” dans *2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011)*. IEEE, 2011, pp. 298–303.

[15] J. Yi, J. W. Labadie, et S. Stitt, “Dynamic optimal unit commitment and loading in hydropower systems,” *Journal of Water Resources Planning and Management*, vol. 129, n° 5, pp. 388–398, 2003.

[16] C. Cheng, S. Liao, Z. Tang, et M. Zhao, “Comparison of particle swarm optimization and dynamic programming for large scale hydro unit load dispatch,” *Energy Conversion and Management*, vol. 50, n° 12, pp. 3007–3014, 2009.

[17] P. Van den Bosch et G. Honderd, “A solution of the unit commitment problem via decomposition and dynamic programming,” *IEEE Transactions on Power Apparatus and Systems*, n° 7, pp. 1684–1690, 1985.

[18] R. Taktak et C. D’Ambrosio, “An overview on mathematical programming approaches for the deterministic unit commitment problem in hydro valleys,” *Energy Systems*, vol. 8, n° 1, pp. 57–79, 2017.

[19] J. I. Pérez et J. R. Wilhelmi, “Nonlinear self-scheduling of a single unit small hydro plant in the day-ahead electricity market,” *Proceedings of ICREPQ*, vol. 7, 2007.

[20] J. Catalão, H. M. I. Pousinho, et V. M. F. Mendes, “Scheduling of head-dependent cascaded hydro systems : Mixed-integer quadratic programming approach,” *Energy Conversion and Management*, vol. 51, n° 3, pp. 524–530, 2010.

[21] F. J. Diaz, J. Contreras, J. I. Muñoz, et D. Pozo, “Optimal scheduling of a price-taker cascaded reservoir system in a pool-based electricity market,” *IEEE Transactions on Power Systems*, vol. 26, n° 2, pp. 604–615, 2010.

[22] J. Catalão, H. M. I. Pousinho, et V. M. F. Mendes, “Hydro energy systems management in portugal : Profit-based evaluation of a mixed-integer nonlinear approach,” *Energy*, vol. 36, n° 1, pp. 500–507, 2011.

[23] J. A. Lopez, J. L. Ceciliano-Meza, I. G. Moya, et R. N. Gomez, “A miqcp formulation to solve the unit commitment problem for large-scale power systems,” *International Journal of Electrical Power & Energy Systems*, vol. 36, n° 1, pp. 68–75, 2012.

[24] M. Kadowaki, T. Ohishi, L. S. Martins, et S. Soares, “Short-term hydropower scheduling via an optimization-simulation decomposition approach,” dans *2009 IEEE Bucharest PowerTech*. IEEE, 2009, pp. 1–7.

[25] X. Ge, L. Zhang, J. Shu, et N. Xu, “Short-term hydropower optimal scheduling considering the optimization of water time delay,” *Electric Power Systems Research*, vol. 110, pp. 188–197, 2014.

[26] L. S. Guedes, P. de Mendonça Maia, A. C. Lisboa, D. A. G. Vieira, et R. R. Saldanha, “A unit commitment algorithm and a compact milp model for short-term hydro-power generation scheduling,” *IEEE Transactions on Power Systems*, vol. 32, n° 5, pp. 3381–3390, 2016.

[27] A. J. Conejo, J. M. Arroyo, J. Contreras, et F. A. Villamor, “Self-scheduling of a hydro producer in a pool-based electricity market,” *IEEE Transactions on power systems*, vol. 17, n° 4, pp. 1265–1272, 2002.

[28] X. Li, T. Li, J. Wei, G. Wang, et W. W. Yeh, “Hydro unit commitment via mixed integer linear programming : A case study of the three gorges project, china,” *IEEE Transactions on Power Systems*, vol. 29, n° 3, pp. 1232–1241, 2013.

[29] A. Borghetti, C. D'Ambrosio, A. Lodi, et S. Martello, "An milp approach for short-term hydro scheduling and unit commitment with head-dependent reservoir," *IEEE Transactions on power systems*, vol. 23, n° 3, pp. 1115–1124, 2008.

[30] P. Xia, C. Deng, Y. Chen, et W. Yao, "Milp based robust short-term scheduling for wind–thermal–hydro power system with pumped hydro energy storage," *IEEE Access*, vol. 7, pp. 30 261–30 275, 2019.

[31] C. Su, C. Cheng, et P. Wang, "An milp model for short-term peak shaving operation of cascaded hydropower plants considering unit commitment," dans *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. IEEE, 2018, pp. 1–5.

[32] J. A. Rodriguez, M. F. Anjos, P. Côté, et G. Desaulniers, "Milp formulations for generator maintenance scheduling in hydropower systems," *IEEE Transactions on Power Systems*, vol. 33, n° 6, pp. 6171–6180, 2018.

[33] M. Bavafa, H. Monsef, et N. Navidi, "A new hybrid approach for unit commitment using lagrangian relaxation combined with evolutionary and quadratic programming," dans *2009 Asia-Pacific Power and Energy Engineering Conference*. IEEE, 2009, pp. 1–6.

[34] V. N. Dieu et W. Ongsakul, "Augmented lagrange hopfield network based lagrangian relaxation for unit commitment," *International Journal of Electrical Power & Energy Systems*, vol. 33, n° 3, pp. 522–530, 2011.

[35] X. Yu et X. Zhang, "Unit commitment using lagrangian relaxation and particle swarm optimization," *International Journal of Electrical Power & Energy Systems*, vol. 61, pp. 510–522, 2014.

[36] A. Bhardwaj, V. K. Kamboj, V. K. Shukla, B. Singh, et P. Khurana, "Unit commitment in electrical power system-a literature review," dans *2012 IEEE International Power Engineering and Optimization Conference Melaka, Malaysia*. IEEE, 2012, pp. 275–280.

[37] H. Zeynal, L. X. Hui, Y. Jiazhen, M. Eidiani, et B. Azzopardi, "Improving lagrangian relaxation unit commitment with cuckoo search algorithm," dans *2014 IEEE International Conference on Power and Energy (PECon)*. IEEE, 2014, pp. 77–82.

[38] F. Y. Takigawa, E. Finardi, et E. L. da Silva, “A decomposition strategy to solve the short-term hydrothermal scheduling based on lagrangian relaxation,” dans *2010 IEEE/PES Transmission and Distribution Conference and Exposition : Latin America (T&D-LA)*. IEEE, 2010, pp. 681–688.

[39] J. Wang et Y. Zhang, “Short-term optimal operation of hydropower reservoirs with unit commitment and navigation,” *Journal of Water Resources Planning and Management*, vol. 138, n° 1, pp. 3–12, 2011.

[40] M. L. Fisher, “The lagrangian relaxation method for solving integer programming problems,” *Management science*, vol. 50, n° 12_supplement, pp. 1861–1871, 2004.

[41] A. Frangioni, C. Gentile, et F. Lacalandra, “Sequential lagrangian-milp approaches for unit commitment problems,” *International Journal of Electrical Power & Energy Systems*, vol. 33, n° 3, pp. 585–593, 2011.

[42] P. Thipwiwatpotjana, *Linear programming problems for generalized uncertainty*. University of Colorado Denver, 2010.

[43] S.-E. Fleten et T. K. Kristoffersen, “Short-term hydropower production planning by stochastic programming,” *Computers & Operations Research*, vol. 35, n° 8, pp. 2656–2671, 2008.

[44] Y. Vardanyan et M. Amelin, “A sensitivity analysis of short-term hydropower planning using stochastic programming,” dans *2012 IEEE Power and Energy Society General Meeting*. IEEE, 2012, pp. 1–7.

[45] M. Kazemi Zanjani, M. Nourelfath, et D. Ait-Kadi, “A multi-stage stochastic programming approach for production planning with uncertainty in the quality of raw materials and demand,” *International Journal of Production Research*, vol. 48, n° 16, pp. 4701–4723, 2010.

[46] S. Séguin, S.-E. Fleten, P. Côté, A. Pichler, et C. Audet, “Stochastic short-term hydropower planning with inflow scenario trees,” *European Journal of Operational Research*, vol. 259, n° 3, pp. 1156–1168, 2017.

[47] J. M. Latorre, S. Cerisola, et A. Ramos, "Clustering algorithms for scenario tree generation : Application to natural hydro inflows," *European Journal of Operational Research*, vol. 181, n° 3, pp. 1339–1353, 2007.

[48] B. Xu, P.-A. Zhong, R. C. Zambon, Y. Zhao, et W. W. Yeh, "Scenario tree reduction in stochastic programming with recourse for hydropower operations," *Water Resources Research*, vol. 51, n° 8, pp. 6359–6380, 2015.

[49] J. Li, F. Zhu, B. Xu, et W. W. Yeh, "Streamflow scenario tree reduction based on conditional monte carlo sampling and regularized optimization," *Journal of Hydrology*, vol. 577, p. 123943, 2019.

[50] I. Aravena et E. Gil, "Hydrological scenario reduction for stochastic optimization in hydrothermal power systems," *Applied Stochastic Models in Business and Industry*, vol. 31, n° 2, pp. 231–240, 2015.

[51] B. Vitoriano, S. Cerisol, et A. Ramos, "Generating scenario trees for hydro inflows," dans *Proceedings of the 6th International Conference Probabilistic Methods Applied to Power Systems PMAPS*, vol. 2, 2000.

[52] K. Høyland et S. W. Wallace, "Generating scenario trees for multistage decision problems," *Management science*, vol. 47, n° 2, pp. 295–307, 2001.

[53] M. Olsson et L. Söder, "Optimal regulating market bidding strategies in hydropower systems," dans *Proc. PSCC*. Citeseer, 2005.

[54] T. K. Kristoffersen, "Stochastic programming with applications to power systems," *Working Papers, Department of Mathematical Sciences, University of Aarhus*, vol. 2007, n° 1, 2007.

[55] H. Leövey et W. Römisch, "Quasi-monte carlo methods for linear two-stage stochastic programming problems," *Mathematical Programming*, vol. 151, n° 1, pp. 315–345, 2015.

[56] I. Rios, R. J. Wets, et D. L. Woodruff, "Multi-period forecasting and scenario generation with limited data," *Computational Management Science*, vol. 12, n° 2, pp. 267–295, 2015.

[57] F. Zhu, P.-a. Zhong, B. Xu, W. Liu, W. Wang, Y. Sun, J. Chen, et J. Li, "Short-term stochastic optimization of a hydro-wind-photovoltaic hybrid system under multiple uncertainties," *Energy Conversion and Management*, vol. 214, p. 112902, 2020.

[58] N. Growe-Kuska, H. Heitsch, et W. Romisch, "Scenario reduction and scenario tree construction for power management problems," dans *2003 IEEE Bologna Power Tech Conference Proceedings,*, vol. 3. IEEE, 2003, pp. 7–pp.

[59] Y. C. Albers, "Stochastic programming tools with application to hydropower production scheduling," *Diss. Bachelor's Thesis, ETH Zurich*, vol. 1, n° 1, p. 102, 2011.

[60] P. Côté et R. Leconte, "Comparison of stochastic optimization algorithms for hydropower reservoir operation with ensemble streamflow prediction," *Journal of Water Resources Planning and Management*, vol. 142, n° 2, p. 04015046, 2016.

[61] C. Gauvin, *Modèles d'optimisation stochastique pour le problème de gestion de réservoirs*. Ecole Polytechnique, Montreal (Canada), 2017.

[62] C. Zhao, J. Wang, J.-P. Watson, et Y. Guan, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Transactions on Power Systems*, vol. 28, n° 3, pp. 2708–2717, 2013.

[63] A. Soroudi, "Robust optimization based self scheduling of hydro-thermal genco in smart grids," *Energy*, vol. 61, pp. 262–271, 2013.

[64] B. Mohammadi-Ivatloo, H. Zareipour, N. Amjady, et M. Ehsan, "Application of information-gap decision theory to risk-constrained self-scheduling of gencos," *IEEE Transactions on Power Systems*, vol. 28, n° 2, pp. 1093–1102, 2012.

[65] A. Street, F. Oliveira, et J. M. Arroyo, "Contingency-constrained unit commitment with $n-k$ security criterion : A robust optimization approach," *IEEE Transactions on Power Systems*, vol. 26, n° 3, pp. 1581–1590, 2010.

[66] G. Liu et K. Tomsovic, "Robust unit commitment considering uncertain demand response," *Electric Power Systems Research*, vol. 119, pp. 126–137, 2015.

[67] L. Wu, M. Shahidehpour, et Z. Li, "Comparison of scenario-based and interval optimization approaches to stochastic scuc," *IEEE Transactions on Power Systems*, vol. 27, n° 2, pp. 913–921, 2011.

[68] W. van Ackooij, R. Henrion, A. Möller, et R. Zorgati, "Joint chance constrained programming for hydro reservoir management," *Optimization and Engineering*, vol. 15, n° 2, pp. 509–531, 2014.

[69] U. A. Ozturk, M. Mazumdar, et B. A. Norman, "A solution to the stochastic unit commitment problem using chance constrained programming," *IEEE Transactions on Power Systems*, vol. 19, n° 3, pp. 1589–1598, 2004.

[70] X. Wu, C. Cheng, J. R. Lund, W. Niu, et S. Miao, "Stochastic dynamic programming for hydropower reservoir operations with multiple local optima," *Journal of Hydrology*, vol. 564, pp. 712–722, 2018.

[71] P. Côté, D. Haguma, R. Leconte, et S. Krau, "Stochastic optimisation of hydro-quebec hydropower installations : a statistical comparison between sdp and ssdp methods," *Canadian Journal of Civil Engineering*, vol. 38, n° 12, pp. 1427–1434, 2011.

[72] E. Fermi, "Numerical solution of a minimum problem," Los Alamos Scientific Lab., Los Alamos, NM, Rapport Technique, 1952.

[73] N. Amaioua, *Modèles quadratiques et décomposition parallèle pour l'optimisation sans dérivées*. Ecole Polytechnique, Montreal (Canada), 2018.

[74] V. Torczon, "On the convergence of pattern search algorithms," *SIAM Journal on optimization*, vol. 7, n° 1, pp. 1–25, 1997.

[75] C. Audet et J. E. Dennis Jr, "Mesh adaptive direct search algorithms for constrained optimization," *SIAM Journal on optimization*, vol. 17, n° 1, pp. 188–217, 2006.

[76] S. Le Digabel, *NOMAD : Nonlinear optimization with the MADS algorithm*. Groupe d'études et de recherche en analyse des décisions, 2010.

[77] S. Alarie, C. Audet, V. Garnier, S. Le Digabel, et L.-A. Leclaire, “Snow water equivalent estimation using blackbox optimization,” *Pac J Optim*, vol. 9, n° 1, pp. 1–21, 2013.

[78] S. Séguin, C. Audet, et P. Côté, “Scenario-tree modeling for stochastic short-term hydropower operations planning,” *Journal of Water Resources Planning and Management*, vol. 143, n° 12, p. 04017073, 2017.

[79] M. Minville, D. Cartier, C. Guay, L.-A. Leclaire, C. Audet, S. Le Digabel, et J. Merleau, “Improving process representation in conceptual hydrological model calibration using climate simulations,” *Water Resources Research*, vol. 50, n° 6, pp. 5044–5073, 2014.

[80] I. E. Agency, “Renewables 2018,” www.iea.org/renewables2018/power, 2018, consulté le : 2018-10-26.

[81] P. Liu, J. Zhao, L. Li, et Y. Shen, “Optimal reservoir operation using stochastic dynamic programming,” *Journal of Water Resource and Protection*, vol. 4, n° 06, p. 342, 2012.

[82] R. Egging, S.-E. Fleten, I. Grønvik, A. Hadziomerovic, et N. Ingvoldstad, “Linear decision rules for hydropower scheduling under uncertainty,” *IEEE Transactions on Power Systems*, vol. 32, n° 1, pp. 103–113, 2016.

[83] P. Côté et R. Leconte, “Comparison of stochastic optimization algorithms for hydropower reservoir operation with ensemble streamflow prediction,” *Journal of Water Resources Planning and Management*, vol. 142, n° 2, p. 04015046, 2015.

[84] M. N. Hjelmeland, J. Zou, A. Helseth, et S. Ahmed, “Nonconvex medium-term hydropower scheduling by stochastic dual dynamic integer programming,” *IEEE Transactions on Sustainable Energy*, vol. 10, n° 1, pp. 481–490, 2018.

[85] J. Catalão, S. Mariano, V. M. F. Mendes, et L. Ferreira, “Nonlinear optimization method for short-term hydro scheduling considering head-dependency,” *European Transactions on Electrical Power*, vol. 20, n° 2, pp. 172–183, 2010.

[86] S. Liao, J. Liu, B. Liu, C. Cheng, L. Zhou, et H. Wu, “Multicore parallel dynamic programming algorithm for short-term hydro-unit load dispatching of huge hydropower stations serving multiple power grids,” *Water Resources Management*, pp. 1–18, 2020.

[87] N. Amjady et M. R. Ansari, “Hydrothermal unit commitment with ac constraints by a new solution method based on benders decomposition,” *Energy Conversion and Management*, vol. 65, pp. 57–65, 2013.

[88] R. Fang et Z. Popole, “Multi-objective optimized scheduling model for hydropower reservoir based on improved particle swarm optimization algorithm,” *Environmental Science and Pollution Research*, pp. 1–9, 2019.

[89] M. Basu, “Hybridization of bee colony optimization and sequential quadratic programming for dynamic economic dispatch,” *International Journal of Electrical Power & Energy Systems*, vol. 44, n° 1, pp. 591–596, 2013.

[90] G. E. Alvarez, M. G. Marcovecchio, et P. A. Aguirre, “Unit commitment scheduling including transmission constraints : a milp formulation,” dans *Computer Aided Chemical Engineering*. Elsevier, 2016, vol. 38, pp. 2157–2162.

[91] B. Fu, C. Ouyang, C. Li, J. Wang, et E. Gul, “An improved mixed integer linear programming approach based on symmetry diminishing for unit commitment of hybrid power system. energies. 2019 ; 12 (5). article no. 833,” 2019.

[92] B. Tong, Q. Zhai, et X. Guan, “An milp based formulation for short-term hydro generation scheduling with analysis of the linearization effects on solution feasibility,” *IEEE Transactions on Power Systems*, vol. 28, n° 4, pp. 3588–3599, 2013.

[93] F. I. Corporation, “FICO Xpress Optimization Suite,” Available at :https://www.msi-jp.com/xpress/learning/square/optimizer-2015.pdf, 2014.

[94] M. Daadaa, S. Séguin, K. Demeester, et M. 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*, vol. 125, p. 106419, 2021.

[95] V. Yelena, “On stochastic optimization for short-term hydropower planning,” Mémoire de maîtrise, KTH Royal Institute of Technology, 2012.

[96] C. Li et I. E. Grossmann, “A review of stochastic programming methods for optimization of process systems under uncertainty,” *Frontiers in Chemical Engineering*, vol. 2, p. 622241,

2021.

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

[98] E. Moiseeva et M. R. Hesamzadeh, "Strategic bidding of a hydropower producer under uncertainty : Modified benders approach," *IEEE Transactions on Power Systems*, vol. 33, n° 1, pp. 861–873, 2017.

[99] R. Hochreiter et G. C. Pflug, "Financial scenario generation for stochastic multi-stage decision processes as facility location problems," *Annals of Operations Research*, vol. 152, n° 1, pp. 257–272, 2007.

[100] S. Adhau, R. Moharil, et P. Adhau, "K-means clustering technique applied to availability of micro hydro power," *Sustainable Energy Technologies and Assessments*, vol. 8, pp. 191–201, 2014.

[101] T. Follestad, O. Wolfgang, et M. M. Belsnes, "An approach for assessing the effect of scenario tree approximations in stochastic hydropower scheduling models," dans *Proc. of the 17th Power System Computation Conference*, 2011, pp. 271–277.

[102] G. A. M. System, "SCENRED2," Available at : https://www.gams.com/latest/docs/T_SCENRED2.html, 2023.

[103] Y. Vardanyan et M. R. Hesamzadeh, "The coordinated bidding of a hydropower producer in three-settlement markets with time-dependent risk measure," *Electric Power Systems Research*, vol. 151, pp. 40–58, 2017.

[104] F. Questier, Q. Guo, B. Walczak, D. Massart, C. Boucon, et S. De Jong, "The neural-gas network for classifying analytical data," *Chemometrics and Intelligent Laboratory Systems*, vol. 61, n° 1-2, pp. 105–121, 2002.

[105] G. Van Rossum et F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA : CreateSpace, 2009.

[106] D. R. A. of Canada, "Technical documentation," Available at : https://docs.alliancecan.ca/wiki/Technical_documentation, 2022.

[107] I. Benkalai et S. Séguin, "Hydropower optimization," *Les Cahiers du GERAD ISSN*, vol. 711, p. 2440, 2020.

[108] E. Finardi et M. R. Scuzziato, "Hydro unit commitment and loading problem for day-ahead operation planning problem," *International Journal of Electrical Power & Energy Systems*, vol. 44, n° 1, pp. 7–16, 2013.

[109] A. L. Diniz et T. M. Souza, "Short-term hydrothermal dispatch with river-level and routing constraints," *IEEE Transactions on Power Systems*, vol. 29, n° 5, pp. 2427–2435, 2014.

[110] E. Finardi et E. Da Silva, "Unit commitment of single hydroelectric plant," *Electric Power Systems Research*, vol. 75, n° 2-3, pp. 116–123, 2005.

[111] C. Cheng, J. Wang, et X. Wu, "Hydro unit commitment with a head-sensitive reservoir and multiple vibration zones using milp," *IEEE Transactions on Power Systems*, vol. 31, n° 6, pp. 4842–4852, 2016.

[112] D. E. M. Wapet, S. N. Essiane, R. Wamkeue, P. J. Gnetchejo *et al.*, "Hydropower production optimization from inflow : case study of songloulou hydroplant," *Journal of Power and Energy Engineering*, vol. 8, n° 08, p. 37, 2020.

[113] H. I. Arvidsen, "Hydropower production scheduling using stochastisc dual dynamic programming subject to environmental constraints," Mémoire de maîtrise, The University of Bergen, 2019.

[114] E. Finardi, E. L. d. Silva, et C. Sagastizábal, "Solving the unit commitment problem of hydropower plants via lagrangian relaxation and sequential quadratic programming," *Computational & applied mathematics*, vol. 24, pp. 317–342, 2005.

[115] E. Gil, J. Bustos, et H. Rudnick, "Short-term hydrothermal generation scheduling model using a genetic algorithm," *IEEE Transactions on power systems*, vol. 18, n° 4, pp. 1256–1264, 2003.

[116] B. Postolov et A. Iliev, “Adaptive genetic algorithm for hydro-thermal unit commitment considering the security constraints,” *International Journal of Electrical Engineering and Computing*, vol. 4, n° 2, pp. 61–69, 2020.

[117] C. J. López-Salgado, O. Añó, et D. M. Ojeda-Esteybar, “Stochastic unit commitment and optimal allocation of reserves : A hybrid decomposition approach,” *IEEE Transactions on Power Systems*, vol. 33, n° 5, pp. 5542–5552, 2018.

[118] E. Faria et S.-E. Fleten, “Day-ahead market bidding for a nordic hydropower producer : taking the elbas market into account,” *Computational Management Science*, vol. 8, pp. 75–101, 2011.

[119] M. Daadaa, S. Séguin, M. Anjos, et K. Demeester, “Quantifying the impact of scenario tree generation meth-ods on the solution of the short-term hydroscheduling problem,” *Les Cahiers du GERAD ISSN*, vol. 711, p. 2440, 2022.

[120] B. Li, K. Sedzro, X. Fang, B. Hodge, et J. Zhang, “A clustering-based scenario generation framework for power market simulation with wind integration,” *Journal of Renewable and Sustainable Energy*, vol. 12, n° 3, p. 036301, 2020.

[121] S. Backe, M. Ahang, et A. Tomasgard, “Stable stochastic capacity expansion with variable renewables : Comparing moment matching and stratified scenario generation sampling,” *Applied Energy*, vol. 302, p. 117538, 2021.

[122] F. G. Dires, M. Amelin, et G. Bekele, “Inflow scenario generation for the ethiopian hydropower system,” *Water*, vol. 15, n° 3, p. 500, 2023.

[123] A. B. Philpott, M. Craddock, et H. Waterer, “Hydro-electric unit commitment subject to uncertain demand,” *European Journal of Operational Research*, vol. 125, n° 2, pp. 410–424, 2000.

[124] A. Akbari Asanjan, T. Yang, K. Hsu, S. Sorooshian, J. Lin, et Q. Peng, “Short-term precipitation forecast based on the persiann system and lstm recurrent neural networks,” *Journal of Geophysical Research : Atmospheres*, vol. 123, n° 22, pp. 12–543, 2018.

[125] C. Audet et W. Hare, “Derivative-free and blackbox optimization,” *Springer*, 2017.

[126] S. Le Digabel, “Algorithm 909 : Nomad : Nonlinear optimization with the mads algorithm,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 37, n° 4, pp. 1–15, 2011.

[127] C. Audet et J. E. Dennis Jr, “Analysis of generalized pattern searches,” *SIAM Journal on optimization*, vol. 13, n° 3, pp. 889–903, 2002.

[128] S. Le Digabel, C. Tribes, V. R. Montplaisir, et C. Audet, “Nomad user guide version 3.9. 1,” 2019.

APPENDICE A

COMPARAISON STATISTIQUE ENTRE LES MÉTHODES BACKWARD REDUCTION ET GAZ NEURONAL

TABLEAU 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