UNIVERSITÉ DE MONTRÉAL

MODÈLES D'OPTIMISATION STOCHASTIQUE POUR LE PROBLÈME DE GESTION DE RÉSERVOIRS

CHARLES GAUVIN
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(MATHÉMATIQUES DE L'INGÉNIEUR)
MAI 2017

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

MODÈLES D'OPTIMISATION STOCHASTIQUE POUR LE PROBLÈME DE GESTION DE RÉSERVOIRS

présentée par : GAUVIN Charles
en vue de l'obtention du diplôme de : Philosophiæ Doctor
a été dûment acceptée par le jury d'examen constitué de :

M. AUDET Charles, Ph. D., président
M. GENDREAU Michel, Ph. D., membre et directeur de recherche
M. DELAGE Erick, Ph. D., membre et codirecteur de recherche
M. BASTIN Fabian, Ph. D., membre
M. GEORGHIOU Angelos, Ph. D., membre externe

REMERCIEMENTS

Bien que la réalisation de cette thèse n'ait pas toujours été une expérience facile ni agréable, je suis reconnaissant de la richesse des expériences personnelles et professionnelles que j'ai vécues au cours de ces quatre dernières années. J'ai eu l'opportunité d'en apprendre plus sur un très large spectre de sujets en plus de développer une certaine expertise dans des domaines passionnants.

Avant toute chose, j'aimerais mettre l'accent sur la chance que j'ai eu au cours de mon parcours. Sans chance, je n'aurais pu intégrer la maîtrise à Polytechnique, car je n'avais pas les compétences de base nécessaires. Je n'aurais pas non plus pu poursuivre au doctorat ni rencontrer Erick, qui est devenu mon codirecteur.

Je tiens à remercier Michel de m'avoir supervisé au cours de ces quatre années. Bien que j'aie grandement bénéficié de sa très vaste expérience, je retiens particulièrement son côté humain et ses encouragements à certains moments clés.

Je suis également reconnaissant à Erick de m'avoir encadré. J'apprécie maintenant sa grande rigueur et minutie. Ses suggestions ont considérablement amélioré la qualité de mes travaux et m'ont très bien guidé. Je suis privilégié d'avoir pu le côtoyer, car il représente assurément l'un des meilleurs chercheurs en recherche opérationnelle que j'ai eu la chance de rencontrer. C'est aussi grâce à lui que j'ai pu visiter le groupe du professeur Kuhn à Lausanne, ce qui a été un voyage extraordinaire.

I want to thank Daniel Kuhn as well as everyone in the RAO group at EPFL, namely Dirk, Grani, Napat, Peyman, Soroosh and Viet (ordered arbitrarily). My stay in Switzerland was memorable thanks to your kindness and the dynamic atmosphere of the lab.

Mes pensées se tournent ensuite vers tous les gens du GERAD que j'ai pu rencontrer au cours de ma maîtrise et de mon doctorat. Comme j'ai passé beaucoup de temps au laboratoire, je ne pourrais nommer tous les gens que j'ai eu la chance de côtoyer sans oublier de noms. Leur présence a cependant fortement contribué à enrichir et agrémenter mon expérience.

Merci aux gens d'Hydro-Québec et de l'IREQ de m'avoir fourni les données et la problématique mais surtout de judicieux conseils. Je suis particulièrement reconnaissant à Grégory Émiel, Pierre-Marc Rondeau, Louis Delorme, Pierre-Luc Carpentier, Laura Fagherazzi ainsi qu'Alexia Marchand. Je suis également reconnaissant du support financier d'Hydro-Québec et du CRSNG.

Je tiens à souligner l'importance des discussions que j'ai eu avec Pascal Côté et Sara Séguin

de Rio Tinto. Leur expertise et intérêt pour le problème de gestion hydroélectrique m'ont permis de clarifier et de remettre en question plusieurs points.

Merci à Charles Audet, Fabian Bastin et Richard Gourdeau d'avoir lu ma thèse avec autant de minutie et d'intérêt en plus de m'avoir fait des commentaires d'une grande pertinence.

I would like to thank Angelos for accepting my invitation to be on my Ph.D. defense committee and taking the time to give me such insightful comments. I am privileged to have such a prolific researcher review my work.

J'ai également été touché par l'accueil que j'ai reçu au CIRRELT à Québec. Merci à Yan Cimon et Pierre Marchand de m'avoir trouvé un endroit pour travailler ainsi que de leur bonne humeur. Merci également à Jean-François Côté, Géraldine Heilporn, Leandro Coelho et Bernard Lamond pour les nombreuses opportunités qu'ils m'ont présentées.

Je remercie également tous les gens qui m'ont encouragé à poursuivre des études supérieures et qui m'ont donné des opportunités d'effectuer de la recherche, notamment Paul Intrevado, Geneviève Bassellier et Liette Lapointe. Merci également à Guy Desaulniers pour son excellent encadrement durant ma maîtrise.

I would also like to take the time to thank Jeroen Struben for the support and enthusiasm he displayed while I was still an undergraduate student at McGill. His supervision and teaching had a profound impact on the way I think in general as well as my desire to pursue research.

Merci à Didier, Éric, Charles, Philippe, Cédric, Étienne, Carl-André et Marc (dans un ordre arbitraire) pour leur amitié au cours de ces nombreuses années.

Merci à toute ma famille pour leur support constant et leurs nombreuses attentions. Je suis choyé de toujours pouvoir compter sur eux.

Finalement, je remercie Caroline, l'amour de ma vie, pour ses nombreux encouragements au cours de ces dernières années. Elle est sans contredit la personne qui a le plus contribué à cette thèse et je lui en serai toujours reconnaissant.

RÉSUMÉ

La gestion d'un système hydroélectrique représente un problème d'une grande complexité pour des compagnies comme Hydro-Québec ou Rio Tinto. Il faut effectivement faire un compromis entre plusieurs objectifs comme la sécurité des riverains, la production hydroélectrique, l'irrigation et les besoins de navigation et de villégiature. Les opérateurs doivent également prendre en compte la topologie du terrain, les délais d'écoulement, les interdépendances entre les réservoirs ainsi que plusieurs phénomènes non linéaires physiques. Même dans un cadre déterministe, ces nombreuses contraintes opérationnelles peuvent mener à des problèmes irréalisables sous certaines conditions hydrologiques.

Par ailleurs, la considération de la production hydroélectrique complique considérablement la gestion du bassin versant. Une modélisation réaliste nécessite notamment de prendre en compte la hauteur de chute variable aux centrales, ce qui mène à un problème non convexe.

En outre, de nombreuses sources d'incertitude entourent la réalisation d'un plan de production. Les prix de l'électricité sur les marchés internationaux, la disponibilité des turbines, la charge/demande du réseau ainsi que les apports en eau sont tous incertains au moment d'établir les soutirages et les déversés pour un horizon temporel donné. Négliger cette incertitude et supposer une connaissance parfaite du futur peut mener à des politiques de gestion beaucoup trop ambitieuses. Ces dernières ont tendance à engendrer des conséquences désastreuses comme le vidage ou le remplissage très rapide des réservoirs, ce qui conduit ensuite à des inondations ou des sécheresses importantes.

Cette thèse considère le problème de gestion de réservoirs avec incertitude sur les apports. Elle tente spécifiquement de développer des modèles et des algorithmes permettant d'améliorer la gestion mensuelle de la rivière Gatineau, notamment en période de crue. Dans cette situation, il est primordial de considérer l'incertitude autour des apports, car ces derniers ont une influence marquée sur l'état hydrologique du système en plus d'être la cause d'évènements désastreux comme les inondations.

La gestion des inondations est particulièrement importante pour la Gatineau, car la rivière coule près de la ville de Maniwaki qui a déjà vécu des inondations dans le passé et continue de présenter des risques importants. Cette rivière représente également une excellente étude de cas, car elle possède plusieurs barrages et réservoirs. La grande dimension du système rend difficile l'application de certains algorithmes populaires comme la programmation dynamique stochastique.

Afin de minimiser le risque d'inondations, on propose initialement un modèle de programmation stochastique multi-étapes (*multi-stage stochastic program*) basé sur les règles de décision affine et les règles de décision affines liftées. On considère l'aversion au risque en évaluant la valeur à risque conditionnelle (*conditional value-at-risk*) aussi connue comme "CVaR". Ce travail considère une représentation polyhédrale de l'incertitude très simple basée sur la moyenne et la variance d'échantillon.

Le deuxième article propose d'améliorer cette représentation de l'incertitude en considérant explicitement la corrélation temporelle entre les apports. À cet effet, il introduit les modèles de séries chronologiques de type ARIMA et présente une manière de les incorporer efficacement dans un modèle multi-étapes avec règles de décision. On étend ensuite l'approche pour évaluer les processus GARCH, ce qui permet d'incorporer l'hétéroscédasticité.

Le troisième travail raffine la représentation de l'incertitude utilisée dans le deuxième travail en s'appuyant sur un modèle ARMA calibré sur le logarithme des apports. Cette représentation non linéaire mène à un ensemble d'incertitude non convexe qu'on choisit d'approximer de façon conservatrice par un polyhèdre. Ce modèle offre néanmoins plusieurs avantages comme la possibilité de dériver une expression analytique pour l'espérance conditionnelle. Afin de considérer la hauteur de chute variable, on propose un algorithme de région de confiance très simple, mais efficace.

Ces travaux montrent qu'il est possible d'obtenir de bons résultats pour le problème de gestion de réservoir en considérant les règles de décision linéaires en combinaison avec une représentation basée sur les processus ARIMA.

ABSTRACT

The problem of designing an optimal release schedule for a hydroelectric system is extremely challenging for companies like Rio Tinto and Hydro-Québec. It is essential to strike an adequate compromise between various conflicting objectives such a riparian security, hydro-electric production as well as navigation and irrigation needs. Operators must also consider the topology of the terrain, water delays, dependence between reservoirs as well as non-linear physical phenomena. Even in a deterministic framework, it may be impossible to find a feasible solution under given hydrological conditions.

Considering hydro-electricity generation further complicates the problem. Indeed, a realistic model must take into account variable water head, which leads to an intractable bilinear non-convex problem.

In addition, there exists various sources of uncertainty surrounding the elaboration of the production plan. The price of electricity on foreign markets, availability of turbines, load of the network and water inflows all remain uncertain at the time of fixing water releases and spills over the given planning horizon. Neglecting this uncertainty and assuming perfect foresight will lead to overly ambitious policies. These decisions will in turn generate disastrous consequences such as very rapid emptying or filling of reservoirs, which in turn generate droughts or floods.

This thesis considers the reservoir management problem with uncertain inflows. It aims at developing models and algorithms to improve the management of the Gatineau river, namely during the freshet. In this situation, it is essential to consider the randomness of inflows since these drive the dynamics of the systems and can lead to disastrous consequences like floods.

Flood management is particularly important for the Gatineau, since the river runs near the town of Maniwaki, which has witnessed several floods in the past. This river also represents a good case study because it comprises various reservoirs and dams. This multi-dimensionality makes it difficult to apply popular algorithms such as stochastic dynamic programming.

In order to minimize the risk of floods, we initially propose a multi-stage stochastic program based on affine and lifted decision rules. We capture risk aversion by optimizing the conditional value-at-risk also known as "CVaR". This work considers a simple polyhedral uncertainty representation based on the sample mean and variance.

The second paper builds on this work by explicitly considering the serial correlation between inflows. In order to do so, it introduces ARIMA time series models and details their in-

corporation into multi-stage stochastic programs with decision rules. The approach is then extended to take into account heteroscedasticity with GARCH models

The third work further refines the uncertainty representation by calibrating an ARMA model on the log of inflows. This leads to a non-convex uncertainty set, which is approximated with a simple polyhedron. This model offers various advantages such as increased forecasting skill and ability to derive an analytical expression for the conditional expectation. In order to consider the variable water head, we propose a successive linear programming (SLP) algorithm which quickly yields good solutions.

These works illustrate the value of using affine decision rules in conjunction with ARIMA models to obtain good quality solutions to complex multi-stage stochastic problems.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

(S)AR(I)MA	(Seasonal) Autoregressive (Integrated) Moving Average
CVaR	Conditional Value At Risk
GARCH	Generalized Autoregressive Conditional Heteroscedastic
MSE	Mean Square Error
SDDP	Stochastic Dual Dynamic Programming
SDP	Stochastic Dynamic Programming
SLP	Successive Linear Programming
SOCP	Second Order Cone Programming
SSDP	Sampling Stochastic Dynamic Programming
VaR	Value At Risk

CHAPITRE 1 INTRODUCTION

Hydro-Québec représente l'un des plus grands producteurs d'électricité au monde. Tirant la quasi-totalité de son énergie de l'hydro-électricité, la société d'état est responsable de la gestion de plus de 80 centrales hydroélectriques et quelques unités thermales avec une capacité installée de plus de 36 000 MW (Hydro-Québec, 2015). Elle dispose également d'une vingtaine de réservoirs de forte contenance et plus de 600 barrages répartis en quatre zones géographiques distinctes.

L'optimisation d'un tel parc de production représente un problème d'une complexité énorme. Il faut effectivement considérer un problème de très grande dimension avec des objectifs multiples parfois contradictoires, des contraintes non convexes pouvant être difficiles à modéliser ainsi que de nombreuses sources d'incertitude s'étalant sur un horizon temporel parfois très long.

Afin de trouver des solutions opérationnelles à un tel problème, on procède à une décomposition du problème en plusieurs sous-étapes, de façon similaire à ce qui est fait dans d'autres domaines comme le transport aérien (Saddoune, 2010) et la foresterie (Rönnqvist, 2003). Dans le cas des problèmes d'optimisation hydroélectriques, on considère spécifiquement la modélisation court (opérationnel), moyen (tactique) et long terme (stratégique) (voir le tableau 1.1)

Tableau 1.1 Classification typique des problèmes de gestion hydroélectrique.

Type	**Horizon temporel typique**	**Pas de temps**	**Type de décisions**	**Représentation du parc**	**Importance de l'aléa**
Court terme	1 semaine - 1 mois	horaire	opérationnel ex : -vente et achat sur les marchés spot et jour d'avant -production journalière	détaillée	faible
Moyen terme	1-2 ans	hebdomadaire	tactique ex : -contrats à moyen terme -évaluation des stocks saisonniers	moyenne	moyenne
Long terme	5-10 ans	mensuel	stratégique ex : -investissements -localisation de centrales	agrégée	élevée

Ces problèmes sont liés entre eux et les résultats des modèles plus agrégés sont utilisés afin de guider la résolution de modèles plus fins. Par exemple, la résolution de modèles moyen terme peut permettre d'obtenir des hyperplans afin de valoriser les stocks finaux d'eau dans

les réservoirs. Ces hyperplans peuvent ensuite être utilisés par des modèles court terme dans le but d'éviter de vider les réservoirs à la fin de l'horizon relativement court.

Cette thèse se penche spécifiquement sur un problème d'horizon court terme, à savoir la gestion mensuelle de la rivière Gatineau en présence d'incertitude sur les apports. Le modèle a notamment été développé dans l'optique de faciliter la gestion de la rivière en période de crue. Les sections qui suivent donnent plus de détails sur la gestion de la rivière ainsi que sur la production hydroélectrique en général.

1.1 Production hydroélectrique [1]

La production hydroélectrique exploite le cycle naturel d'évapotranspiration, condensation, précipitation et ruissellement/infiltration afin de convertir l'énergie potentielle de l'eau en énergie mécanique puis en énergie électrique (en plus de pertes d'énergie thermique). Ces transformations sont décrites à l'aide de la figure 1.1. [2]

Le processus consiste initialement à stocker de l'énergie potentielle sous forme d'eau dans des réservoirs (A). [3] En ouvrant les vannes (*intake*) (E), les opérateurs peuvent contrôler la quantité d'eau soutirée (*released*) qui sera ensuite amenée vers la conduite forcée (*penstock*), ce qui engendrera certaines pertes thermiques. Cette eau fera ensuite tourner les pales de la turbine (C), ce qui permettra de produire de l'électricité grâce au générateur (D) dans la centrale (*powerhouse/power plant*) (B). L'électricité est finalement transportée par les lignes haute tension (G), alors que l'eau s'écoule vers la rivière en aval (H).

1. Cette section est inspirée des travaux de Séguin (2016); Côté (2010).

2. Pour assurer la cohésion et faciliter la lecture des chapitres 4 5 et 6, certains termes techniques anglais correspondants sont indiqués en *italique*.

3. Selon la capacité du réservoir, on dira que la centrale est une centrale à réservoir (*conventional/dam plant*) ou une centrale au fil de l'eau (*run-of-the-river plant*).

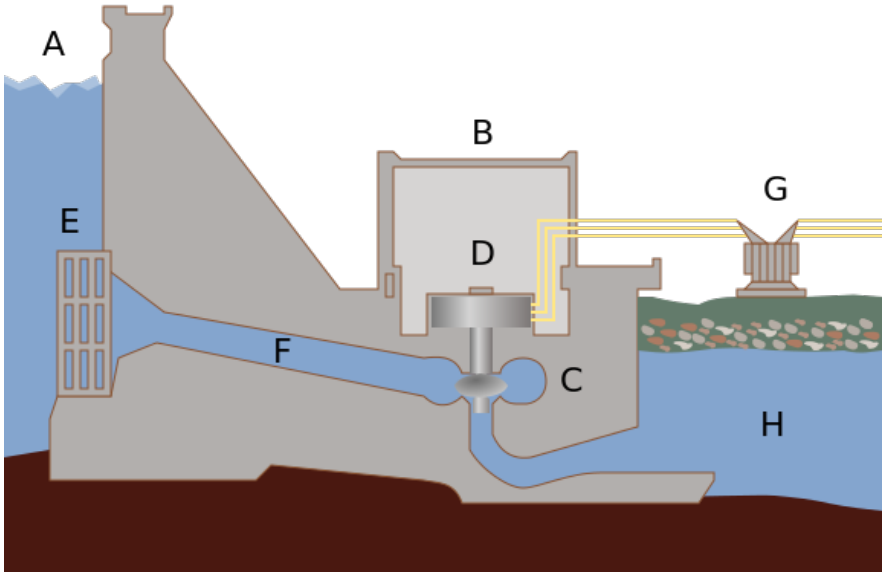

Figure 1.1 Vue en coupe d'une centrale hydroélectrique. Adaptée avec permission de "Hydroelectric dam" par Tomia, 2008. Image sous license GFDL et CC-BY-2.5.

La puissance générée (en MW) à un instant donné par une centrale est une fonction de la hauteur de chute (*water head*), c'est-à-dire de la différence (en m) entre le niveau d'eau ou bief amont et le niveau d'eau ou bief aval. La puissance dépend également des caractéristiques des turbines individuelles et de l'eau soutirée et déversée (en m^3/s). Dans cette thèse, on considère une formulation plus agrégée que celle de Séguin (2016) et les groupes de turbines sont traités comme un seul bloc homogène. Par ailleurs, on suppose que toutes les turbines sont disponibles et fonctionnent à leur capacité normale. Cette hypothèse n'est pas toujours vérifiée, car la pression de l'eau sur les pales ainsi que certains phénomènes de cavitation peuvent engendrer des bris qui nécessitent des arrêts de maintenance. On ignore également les zones interdites, car ces dernières causent des difficultés numériques considérables et notre solution ne les considère jamais. La formule utilisée pour représenter $\mathcal{P}_{it}^{totale}$, la puissance produite (en MW) à tout instant durant le jour t à la centrale i prend la forme[4] :

$$\mathcal{P}_{it}^{total}(\mathcal{R}_{it}, \mathcal{L}_{it}, \mathcal{H}_{it}) = \mathcal{P}_{it}(\mathcal{R}_{it}, \mathcal{L}_{it})\mathcal{H}_{it}. \tag{1.1}$$

où $\mathcal{P}_{it}(\mathcal{R}, \mathcal{L})$ représente la puissance de référence, qui est fonction des débits soutirés $\mathcal{R}_{it}$ et déversés $\mathcal{L}_{it}$ pour une hauteur de chute fixée. L'expression $\mathcal{H}_{it}$ indique quant à elle la hauteur

4. Cette représentation est notamment utilisée par Gjelsvik et al. (2010).

de chute relative, donnée par :

$$\mathcal{H}_{i\tau} = \frac{\mathcal{N}_{j^-\tau} - \mathcal{N}_{j^+\tau}}{\mathcal{H}_i^{ref}} \tag{1.2}$$

où $\mathcal{H}_i^{ref}$ représente la hauteur de chute de référence et $\mathcal{N}_{j^-t}$ et $\mathcal{N}_{j^+\tau}$ représentent le niveau d'eau amont (*forebay*) et aval (*tailrace*) (en m) à la centrale. Le niveau d'eau amont $\mathcal{N}_{j^-t}$ est une fonction concave croissante du volume du réservoir $\mathcal{S}_{jt}$. Pour la plupart des réservoirs, le niveau aval est considéré comme constant. Cependant, lorsque deux centrales sont très près l'une de l'autre, le niveau aval à une centrale est une fonction du volume à la centrale aval.

Pour des déversés $\mathcal{L}_{it}$ constants, la fonction $\mathcal{P}_{it}(\cdot, \mathcal{L}_{it})$ peut être bien approximée par une fonction concave par morceaux (voir la figure 1.2). La pente négative du dernier segment s'explique par le fait que les débits soutirés dépassant un certain seuil réduisent plus la hauteur de chute qu'ils ne contribuent à produire de la puissance. La fonction $\mathcal{P}_{it}(\cdot, \cdot)$ est détaillée dans le chapitre 6. On suppose également que la puissance produite est constante durant la journée, et $\mathcal{P}_{it}^{total}$ représente donc également l'énergie produite (en MWh) durant le jour j.

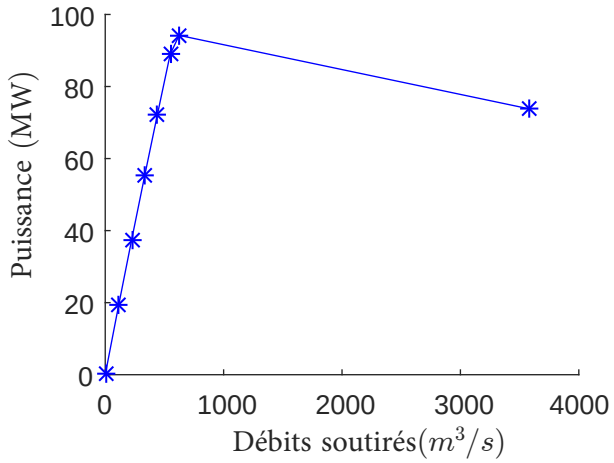

Figure 1.2 Puissance de référence pour une hauteur de chute et débits déversés constants

Bien que ce ne soit pas indiqué sur la figure 1.1, les centrales disposent également d'un évacuateur de crues ou déversoir (*spillway*) permettant de déverser une partie de l'eau vers l'aval sans production électrique. De tels déversements (*spills*) impliquent non seulement une perte d'énergie potentielle, mais causent également une diminution de la hauteur de chute et donc une réduction de la productivité de l'eau effectivement amenée à travers la conduite

pour produire de l'électricité.

1.2 Gestion de la rivière Gatineau

1.2.1 La rivière

La rivière Gatineau fait partie du Bassin versant (*catchment/watershed*) de la rivière des Outaouais. Elle s'étend sur environ 400 km et draine une superficie de près de 23 000 km^2 (Commission de toponymie : Gouvernement du Québec, 2017; Pina et al., 2016). Tirant sa source en Haute-Mauricie, elle traverse une partie des Laurentides et de l'Outaouais à travers du territoire majoritairement boisé. Elle se jette ensuite dans la rivière des Outaouais près de la ville de Gatineau (voir la figure 1.3).

Son bassin versant rivière est composée de 2 réservoirs de forte contenance : Cabonga et Baskatong. Ces réservoirs ont une réserve utile d'environ 1500 et 3000 hm^3 et sont opérés par Hydro-Québec. Le réservoir Baskatong représente le plus gros réservoir du bassin de l'Outaouais et joue un rôle crucial dans la gestion de la rivière. Il permet notamment de contrôler les débits au niveau de la ville riveraine de Maniwaki, qui a déjà subi des inondations en 1929, 1936, 1947 et en 1974. Ce dernier est aussi utilisé occasionnellement pour contrôler les niveaux d'eau jusqu'à Montréal, plusieurs centaines de kilomètres en aval. Par ailleurs, Cabonga et Baskatong assurent l'apport en eau potable aux municipalités avoisinantes, abritent de nombreuses espèces animales et végétales en plus de servir pour la villégiature.

Ces réservoirs servent aussi à réguler la production hydroélectrique. À cet effet, quatre centrales sont présentement en opération sur la rivière. On retrouve d'abord la centrale à réservoir Mercier, située juste en aval de Baskatong et d'une capacité d'environ 55 MW. Viennent ensuite les trois centrales au fil de l'eau Paugan d'une capacité installée de 226 MW, Chelsea d'une capacité de 152 MW et Rapides-Farmers d'une capacité de 104 MW (Hydro-Québec (2016)). [5] Ces centrales ont une capacité relativement petite comparativement à certaines centrales du parc. À titre de comparaison, la centrale Robert-Bourassa sur la rivière La Grande a une capacité d'environ 5600 MW (Hydro-Québec, 2016).

Tel qu'indiqué par la figure 1.3, la distance physique est négligeable entre certaines paires de réservoirs, mais de l'ordre de quelques centaines de kilomètres pour d'autres. Afin de représenter les délais d'écoulement, on introduit le paramètre λ_{jl} indiquant la proportion d'eau relâchée par le réservoir immédiatement an amont au temps t qui se rend au réservoir j immédiatement en aval au temps $t+l$, où $l \in \{\delta_j^{min}, \cdots, \delta_j^{max}\}$, et δ_j^{min} et δ_j^{max} représentent

5. Due au manque de données, les trois articles de la thèse ignorent tous la plus petite centrale Mercier, mise en opération en 2005.

le délai minimal et maximal. Les valeurs de ces paramètres sont calculées par les hydrologues chez Hydro-Québec et sont utilisées comme intrants par notre modèle (voir le tableau 1.2).

Tableau 1.2 Proportion de l'eau soutirée du réservoir immédiatement en amont au jour t se rendant au réservoir au jour $t + l$. Le réservoir Cabonga n'est pas considéré, car c'est le premier réservoir du système. Un délai de $l = 0$ indique que toute l'eau relâchée se rend au réservoir en aval en une seule journée.

	l											
	0	1	2	3	4	5	6	7	8	9	10	11
Baskatong	-	0.02	0.08	0.14	0.15	0.14	0.12	0.09	0.07	0.05	0.04	0.11
Maniwaki	0.91	0.09	-	-	-	-	-	-	-	-	-	-
Paugan	0.46	0.46	0.08	-	-	-	-	-	-	-	-	-
Chelsea	1.00	-	-	-	-	-	-	-	-	-	-	-
Rapides-Farmers	1.00	-	-	-	-	-	-	-	-	-	-	-

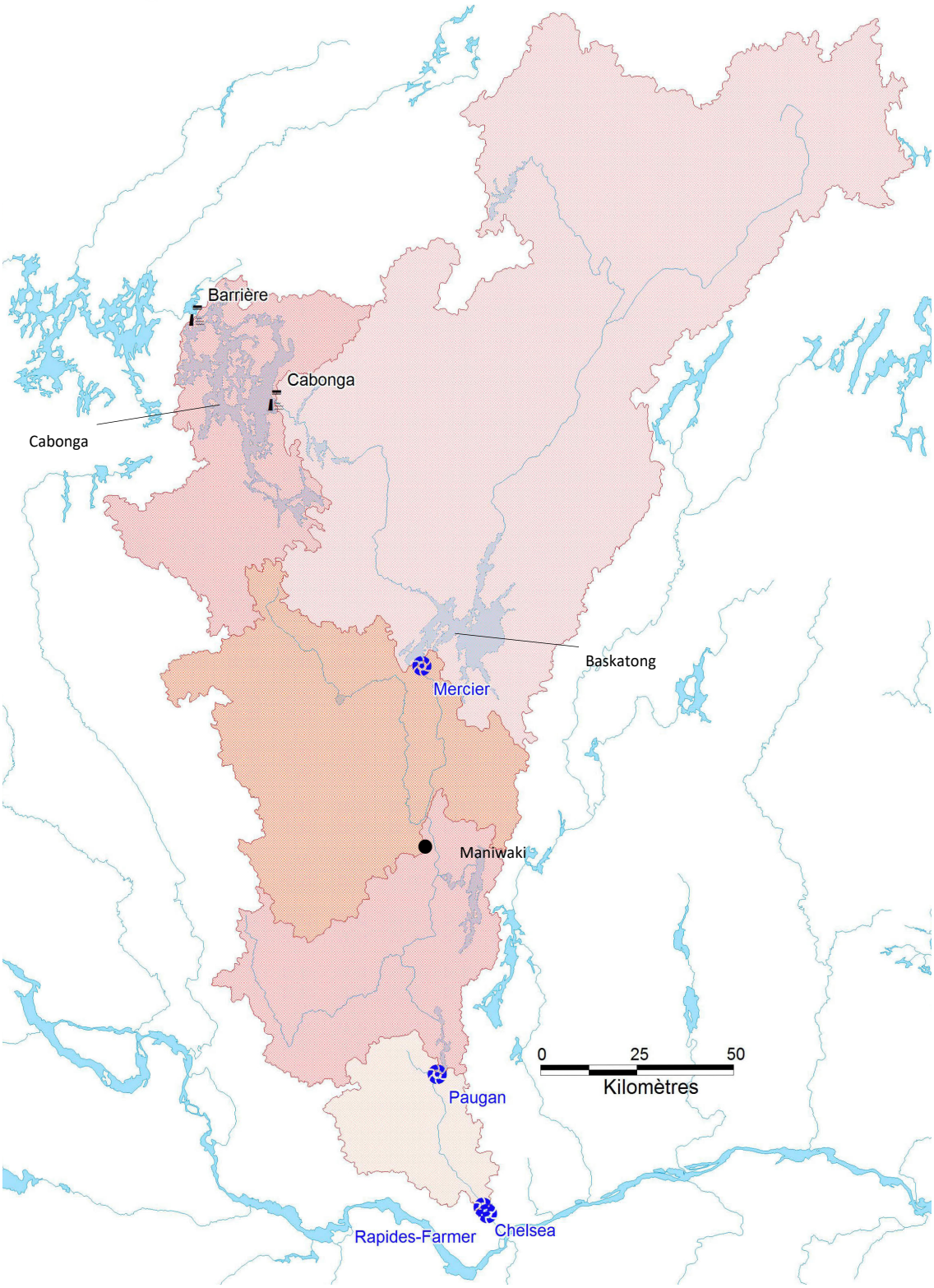

Figure 1.3 Bassin versant de la rivière Gatineau. Image fournie par Hydro-Québec. Adaptée et utilisée avec permission. Les symboles bleus indiquent les centrales, les rectangles noirs les barrages et les points noirs les villes.

1.2.2 Les apports

Dans un contexte de minimisation des inondations, les facteurs les plus importants à considérer sont les stocks initiaux d'eau dans les réservoirs ainsi que les apports en eau, qui demeurent incertains au moment de développer un plan d'opération. Comme l'indique la figure 1.4, ces derniers sont sujets à des variations saisonnières et inter-annuelles très importantes.

Durant l'hiver et le printemps, le régime hydraulique est principalement influencé par la présence puis la fonte de neige. La crue printanière engendre notamment des apports très importants dans un court laps de temps. Cette période est névralgique, car l'hydraulicité moyenne et la variabilité y sont maximisées.

Pendant l'été, les apports sont faibles. Comme ce moment coïncide avec une période de navigation et d'achalandage accru sur la rivière, il est important de s'assurer que les niveaux d'eau minimaux soient respectés.

Suit ensuite une seconde crue en automne due à des précipitations plus importantes que durant le reste de l'année. Pour certains réservoirs et certaines années, la crue automnale est encore plus critique que celle du printemps. La section 6 présente d'ailleurs une étude de cas considérant la crue automnale de 2003.

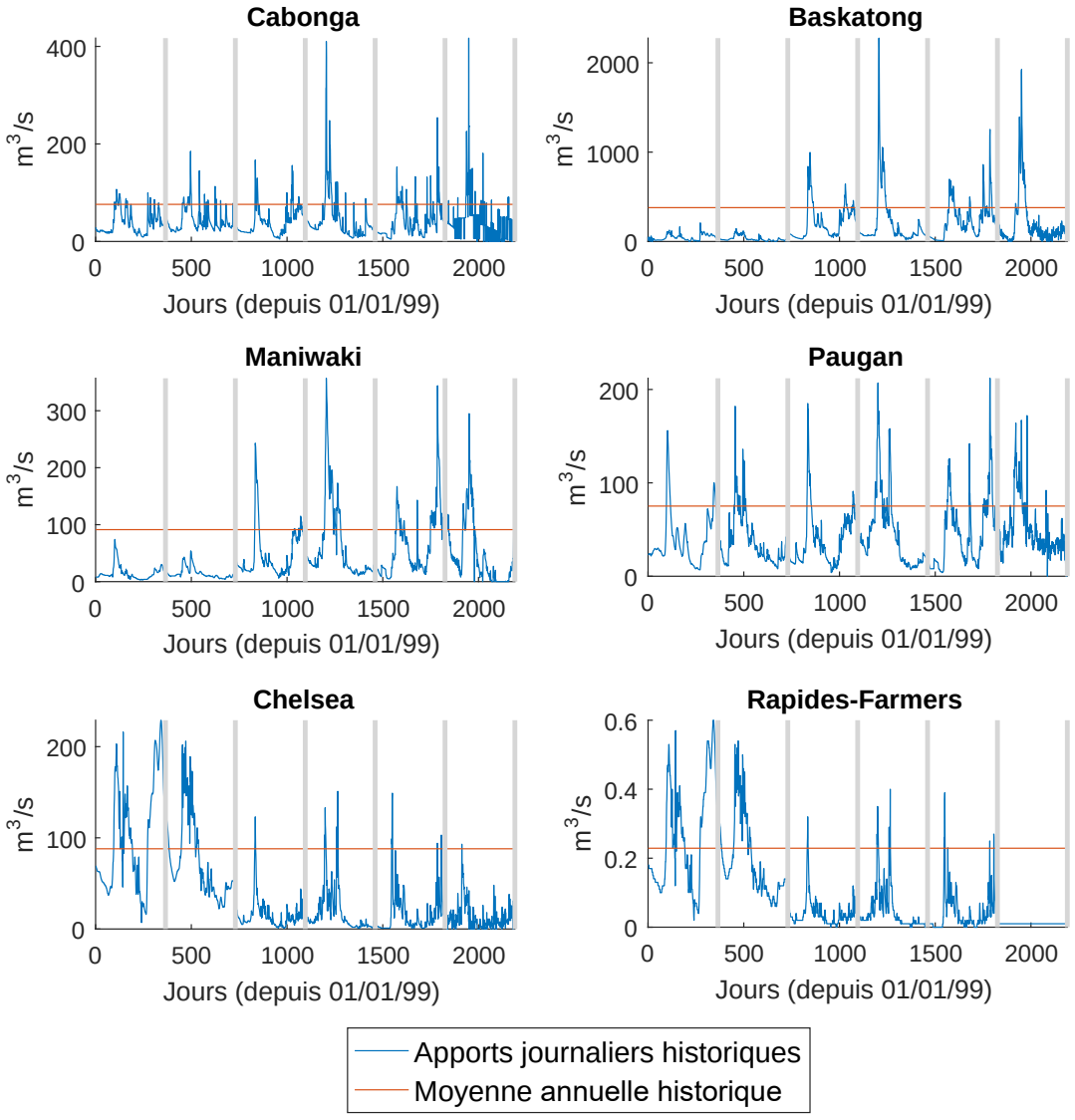

Figure 1.4 Apports historiques journaliers aux différents points névralgiques de la rivière (en m^3/s) de 1999-2004. Les barres grises verticales délimitent les 6 années.

1.2.3 Gestion historique des réservoirs

Hydro-Québec est tenu de gérer ses installations de façon intégrée en s'assurant de la sécurité des opérations et en se concertant avec plusieurs groupes d'intérêts. Le risque d'inondation à Maniwaki est particulièrement important pour les opérateurs de la rivière, notamment en période de crue (*freshet*). Pour ces raisons, la rivière Gatineau est gérée de façon indépendante du reste du parc de production.

Les apports ont une influence capitale sur la gestion des différents réservoirs. En effet, le

volume moyen d'apport en eau durant la crue printanière à Cabonga et Baskatong est de 590 et 3630 hm^3 et a même atteint 710 et 4700 hm^3 en 1997. Ayant des capacités utiles d'environ 1500 et 3000 hm^3, ceci implique que ces réservoirs, particulièrement Baskatong, peuvent se remplir complètement en une seule crue. Les opérateurs utilisent donc un cycle de vidange et de remplissage annuel pour Baskatong (voir la figure 1.5). Le 1er avril, aux alentours du commencement de la crue printanière, on vise un niveau d'eau suffisamment faible permettant de remplir le réservoir tout en évitant les débits trop élevés en aval à Maniwaki.

Comme Cabonga reçoit moins d'eau par rapport à sa capacité et qu'il est possible d'évacuer une partie de son stock hors du bassin versant de la Gatineau vers le lac Barrière, le réservoir offre plus de flexibilité et peut être utilisé afin de stocker ou de soutirer de l'eau selon la quantité d'apports.

Les stocks aux centrales au fil de l'eau sont quant à eux beaucoup plus variables d'une journée et d'une année à l'autre. En effet, ces réservoirs ont une très petite contenance et ils peuvent être vidés ou remplis complètement en quelques jours seulement. Ces niveaux d'eau volatiles permettent aux opérateurs de tirer des avantages opérationnels dus à des variations de hauteur de chute.

On note notamment que le réservoir Paugan est historiquement maintenu à son niveau maximal, ce qui maximise la hauteur de chute. Les niveaux d'eau à Chelsea sont également maintenus à des niveaux assez élevés, bien qu'il semble y exister des cycles interannuels causés notamment par les apports ainsi que le désir des opérateurs de conserver une certaine marge de manœuvre. Finalement, on observe que les niveaux d'eau à Rapides-Farmers ne dépassent jamais un seuil critique. Ceci est dû à la présence d'une crête déversante à la centrale. Cet ouvrage d'évacuation ne permet pas aux opérateurs de contrôler directement les débits déversés. En effet, lorsque les stocks d'eau dépassent un certain niveau, une certaine proportion de l'eau est naturellement évacuée en aval. Comme ce déversement ne conduit pas à de la production hydroélectrique, les opérateurs choisissent de le minimiser.

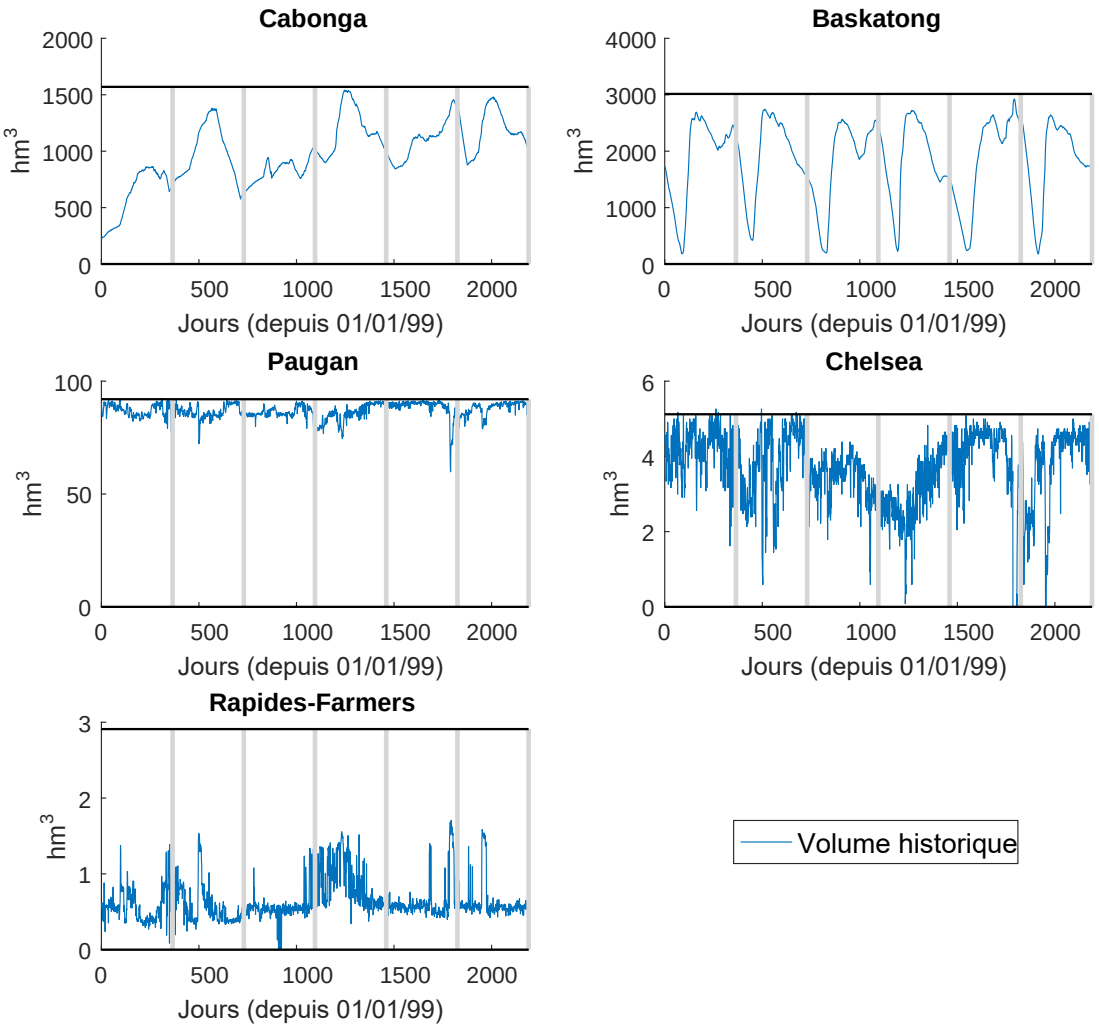

Figure 1.5 Volumes historiques journaliers aux différents réservoirs de 1999-2004 (en hm^3)

1.3 Objectifs de recherche

L'objectif de cette thèse est de concevoir des modèles mathématiques et des algorithmes dans le but de développer un outil d'aide à la décision facilitant la gestion de la rivière pour les opérateurs et permettant de réaliser certaines études exploratoires. À terme, cet outil pourrait même permettre d'améliorer ces opérations, notamment en aidant à concevoir des plans de production générant plus d'électricité tout en maintenant, ou même en réduisant, le risque d'inondations.

1.3.1 Modélisation

La première étape consiste à bien *modéliser* le problème. Il faut spécifiquement identifier une fonction objectif adéquate, la représentation de l'incertitude, les contraintes, ainsi que les décisions à prendre. Il est impératif de proposer une modélisation fidèle à la réalité permettant de s'assurer du respect de toutes les contraintes opérationnelles et satisfaisant les intérêts des différents groupes d'intérêt.

En présence d'aléa, la prise de décision prend la forme d'une politique de gestion, c'est-à-dire un ensemble de fonctions de l'incertitude. On optimise donc sur un espace fonctionnel, plutôt que sur un espace de variables comme c'est le cas en programmation mathématique déterministe.

Il faut ensuite déterminer la bonne fonctionnelle permettant d'évaluer l'objectif pour différents scénarios d'apports. Cette thèse utilise l'espérance mathématique, mais explore également quelques mesures de risque afin d'évaluer l'impact de l'aversion au risque sur les solutions finales.

Tel que mentionné précédemment, la modélisation des apports est d'une importance cruciale pour ce problème. Il faut donc savoir bien utiliser les techniques de modélisation statistique. Cette thèse exploite notamment les processus stochastiques à temps discrets de type ARIMA (*autoregressive integrated moving average*) pour bien représenter la corrélation temporelle. Les arbres de décision, populaires en programmation stochastique multi-étapes, sont également explorés dans le chapitre 4.

1.3.2 Résolution

La deuxième étape consiste à concevoir des algorithmes permettant de *résoudre* efficacement le problème. Il s'agit d'écrire du code informatique permettant de traduire le modèle théorique en format numérique pour enfin utiliser des solveurs commerciaux extrêmement efficaces comme CPLEX et Mosek afin de le résoudre.

L'outil devra être utilisable de façon opérationnelle. Ceci implique entres autres qu'il devra fournir des solutions de bonne qualité à l'intérieur de temps de calcul raisonnables, idéalement quelques secondes. Il devra permettre d'évaluer différents scénarios efficacement et de fournir des réponses rapides afin de guider le choix des opérateurs.

1.3.3 Simulation et ajustement

Finalement, la dernière étape consiste à *ajuster* le modèle après avoir observé les résultats préliminaires et les avoir confrontés aux scénarios réels. Il s'agit ici de concevoir un banc de test de simulation pour bien tester notre modèle. Encore une fois, plusieurs notions de statistiques et de processus hydrologiques deviennent utiles pour concevoir des scénarios d'apport synthétiques réalistes et intéressants. L'étude de cas du chapitre 4 illustre notamment comment ces simulations permettent de mieux comprendre les hypothèses et les limitations de notre modèle. L'utilisation de scénarios historiques se révèle également essentielle pour augmenter la crédibilité de nos travaux. Le chapitre 6 explique notamment que notre modèle pourrait améliorer les décisions historiques.

1.4 Plan du mémoire

Le chapitre 2 présente une revue de littérature des principales méthodes d'optimisation stochastique utilisées pour la gestion de réservoirs. Cette section survole également deux sujets importants pour la thèse : les mesures de risque et les séries chronologiques de type ARIMA. Le chapitre 3 dresse un portrait d'ensemble des 3 articles de la thèse, présentés aux chapitres 4, 5 et 6. Finalement, le chapitre 8 tire des conclusions sur le travail réalisé et propose des pistes de solutions futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Modèles de séries chronologiques de type ARIMA

Une partie importante de cette thèse se base sur la représentation des apports à l'aide des séries chronologiques à temps discret de type ARIMA. Ces processus linéaires paramétriques offrent plusieurs avantages. Ils possèdent une représentation linéaire compacte et facile à comprendre, ont des propriétés bien étudiées, bénéficient de logiciels permettant de réaliser leur calibrage et ont été utilisés avec succès pour modéliser les apports. Cette section s'appuie fortement sur les trois références Box et al. (2008); Brockwell and Davis (1987) et Tsay (2005)

2.1.1 Modèles univariés stationnaires

Par souci de simplicité, on commence par traiter le cas univarié, c'est-à-dire qu'on considère d'abord un processus stochastique $\{\zeta_t\}_{t=-\infty}^{\infty}$ où chaque ζ_t est une variable aléatoire réelle.

Avant de procéder à la description de ces processus, il est important de détailler certains concepts. On dit qu'un processus stochastique $\{\zeta_t\}_{t=-\infty}^{\infty}$ est stationnaire [1] s'il respecte les propriétés suivantes :

$$\mathrm{E}\left[\zeta_t\right] = \mu, \qquad \forall t \in \mathbb{Z}, \tag{2.1}$$

$$\gamma(h) = \mathrm{E}\left[\zeta_t \zeta_{t+h}\right], \qquad \forall t, h \in \mathbb{Z}, \tag{2.2}$$

$$\mathrm{E}\left[\zeta_t^2\right] < \infty, \qquad \forall t \in \mathbb{Z}. \tag{2.3}$$

L'équation (2.1) implique que le processus a une moyenne constante. Dans ce cas, on peut toujours remplacer ζ_t par $\zeta_t' = \zeta_t - \mu$ et on peut donc supposer que $\mu = 0$. L'équation (2.2) indique que la covariance dépend uniquement du déplacement $h \in \mathbb{Z}$ dans le temps et donc en particulier on obtient la variance $\gamma(0) = \sigma^2$, qui est constante par rapport au temps, en évaluant $\gamma(\cdot)$ en 0. Finalement, (2.3) indique que le second moment est borné, ce qui implique que la variance est également bornée.

On s'intéresse notamment au processus stationnaire de bruit blanc $\{\varrho_t\}$ qui possède les propriétés suivantes :

1. Il s'agit en fait d'hypothèse de stationnarité au sens large aussi connue sous le nom de stationnarité faible ou de second ordre.

$$\gamma(h) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{sinon ,} \end{cases} \qquad \forall h \in \mathbb{Z} \tag{2.4}$$

$$\mathrm{E}\left[\varrho_t\right] = 0, \qquad \forall t \in \mathbb{Z}. \tag{2.5}$$

Bien qu'on ait uniquement besoin de ces conditions sur les deux premiers moments pour développer les modèles de séries chronologiques, les chapitres 5 et 6 imposent des contraintes additionnelles sur le support des ϱ_t. Le chapitre 5 supposent que les ϱ_t sont non corrélés et suivent un processus GARCH (*generalized autoregressive conditional heteroscedastic model*), décrit dans les sections qui suivent. Le chapitre 6 fait des hypothèses plus fortes. Il suppose notamment que les ϱ_t suivent une distribution particulière et sont indépendants.

Étant donné le processus de bruit blanc $\{\varrho_t\}$ décrit précédemment, on dit que le processus stochastique $\{\zeta_t\}$ suit un processus ARMA(p, q) s'il est stationnaire et s'il respecte les équations de différence suivantes pour tout $t \in \mathbb{Z}$:

$$\phi(B)\zeta_t = \theta(B)\varrho_t, \tag{2.6}$$

où $\phi(B) = \sum_{i=0}^{p} \phi_i B^i$ et $\theta(B) = \sum_{i=0}^{q} \theta_i B^i$ représentent des polynômes. L'opérateur B permet d'effectuer un décalage dans le temps : $B^i\zeta_t = \zeta_{t-i}$. Les paramètres à calibrer sont les $\phi_i \in \mathbb{R}$ où $i = 1, \cdots, p$ et $\theta_j \in \mathbb{R}$ où $j = 1, \cdots, q$, pour des $p, q \in \mathbb{N}$ fixés. On impose toujours $\phi_0 = \theta_0 = 1$.

La stationnarité de $\{\zeta_t\}$ est assurée si $\phi(z) \neq 0, \forall z \in \mathbb{C} : |z| \leq 1$. Dans ce cas, il existe $\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$ tel que $\sum_{i=0}^{\infty} |\psi_i| < \infty$ et $\psi(B) = \phi^{-1}(B)\theta(B)$ de telle sorte que pour tout t :

$$\psi(B)\varrho_t = \phi^{-1}(B)\theta(B)\varrho_t = \phi^{-1}(B)\phi(B)\zeta_t = \zeta_t \,^2 \tag{2.7}$$

La représentation (2.7) est d'une grande importance, car elle indique que les ζ_t peuvent être exprimés comme une combinaison linéaire infinie des résidus ou chocs passés : $\{\varrho_s : s \leq t\}$. La

2. Pour résoudre cette équation, on considère $\phi(B)\psi(B) = \theta(B) \Leftrightarrow (1 + \psi_1 B + \psi_2 B^2 + \cdots)(1 + \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) = (1 + \theta_1 B + \cdots \theta_q B^q)$ et on s'assure récursivement que la somme des coefficients des B^i soit égale à 0 pour tout $i \in \mathbb{N}$. Par exemple : $\phi_1 + \psi_1 = \theta_1$, $\phi_2 + \phi_1\psi_1 + \psi_2 = \theta_2$, etc.

condition que la somme des coefficients ψ_i converge absolument assure que la limite existe et est unique dans un certain sens (Brockwell and Davis, 1987). Dans la littérature, on dit que $\psi(B)$ est un filtre linéaire ou une fonction de transfert. Tel que l'indique la figure 2.1, $\psi(B)$ prend le bruit blanc $\{\varrho_s\}_{s\leq r}$ non-corrélé en intrant et produit des ζ_t ayant une "corrélation plus structurée" comme extrant.

Cette représentation permet également de dériver une expression décomposable pratique pour effectuer des prévisions. En effet pour tout $t \in \mathbb{Z}$ et $l \in \mathbb{N}$:

$$\zeta_{t+l} = \underbrace{\sum_{i=l}^{\infty} \psi_i \varrho_{t+l-i}}_{\hat{\zeta}_t(l)} + \underbrace{\sum_{i=0}^{l-1} \psi_i \varrho_{t+l-i}}_{\rho_t(l)} \tag{2.8}$$

Le premier terme $\hat{\zeta}_t(l)$ représente la prévision (*forecast*) effectuée au temps t pour un temps futur (*lead time*) de l périodes. Le deuxième terme $\rho_t(l)$ représente donc l'erreur de prévision : $\rho_t(l) = \zeta_{t+l} - \hat{\zeta}_t(l)$. La prévision est donc déterministe sachant $\{\varrho_s\}_{s=-\infty}^{t}$, car elle dépend uniquement des résidus passés, qu'on suppose parfaitement observables. L'erreur de prévision peut quant à elle être exprimée comme un combinaison linéaire des résidus futurs : $\{\varrho_s\}_{s=t+1}^{t+l}$.

Cette stationnarité dépend du polynôme $\phi(B)$. La condition correspondante : $\theta(z) \neq 0, \forall z \in \mathbb{C} : |z| \leq 1$ se nomme inversibilité. Dans ce cas, il existe $\pi(B) = \sum_{i=0}^{\infty} \pi_i B^i$ tel que $\sum_{i=0}^{\infty} |\pi_i| < \infty$ et $\pi(B) = \theta^{-1}(B)\phi(B)$. La propriété suivante tient donc pour tout t :

$$\varrho_t = \pi(B)\zeta_t. \tag{2.9}$$

La propriété d'inversibilité assure entre autres que pour tout $l \in \mathbb{N}$ et $f : \mathbb{R}^{t+l} \to \mathbb{R}$, les espérances conditionnelles $\mathrm{E}\left[f(\zeta_1, \cdots, \zeta_{t+l})|\zeta_{[t]}\right]$ et $\mathrm{E}\left[f(\zeta_1, \cdots, \zeta_{t+l})|\varrho_{[t]}\right]$ sont identiques, car l'information fournie par les $\{\varrho_s\}_{s=-\infty}^{t}$ est identique à celle fournie par $\{\zeta_s\}_{s=-\infty}^{t}$.

Comme c'est presque toujours le cas dans la littérature, cette thèse considère uniquement des processus stationnaires et inversibles. Ces deux conditions sont respectées si $\theta(z)\phi(z) \neq 0, \forall z \in \mathbb{C} : |z| \leq 1$. Il faut donc que les racines des polynômes $\theta(z)$ et $\phi(z)$ se trouvent à l'extérieur du cercle unitaire. Par exemple, on sait que le processus ARMA(1,1) utilisé au chapitre 5 respecte cette conditions, car on a : $\phi(z) = 1 - 0.96z$ et $\theta(z) = 1 - 0.13z$. Les racines sont donc $0.96^{-1} > 1$ et $0.13^{-1} > 1$. Le processus AR(4) utilisé au chapitre 6 a également 4 racines (complexes) à l'extérieur du cercle unitaire.

Sous les hypothèse de stationnarité et d'inversibilité, on a : $\psi(B) = \phi^{-1}(B)\theta(B)$ et $\pi(B) = \theta^{-1}(B)\phi(B)$. On a donc $\psi(B)\pi(B) = 1$ ou encore $\psi^{-1}(B) = \pi(B)$ et $\pi^{-1}(B) = \theta(B)$. Tel que mentionné dans Brockwell and Davis (1987), sous certaines autres hypothèses standards sur les zéros des polynômes $\phi(z)$ et $\theta(z)$, on peut même garantir qu'il existe une seule solution satisfaisant (2.6).

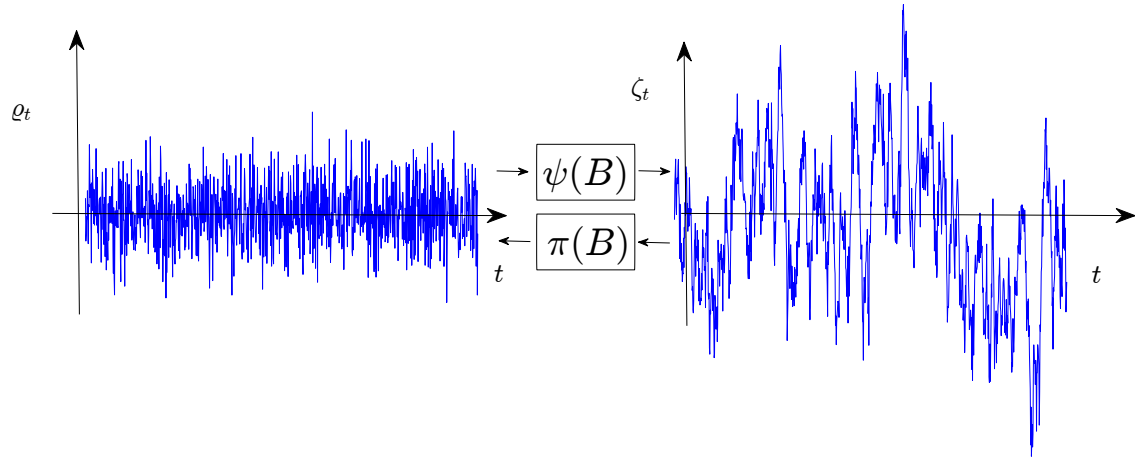

Figure 2.1 Filtres linéaires

2.1.2 Modèles non stationnaires

On peut exprimer le polynôme $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ comme $\phi(z) = (1 - G_1 z)(1 - G_2 z)\cdots(1 - G_p z)$ où G_i^{-1} sont les racines du polynôme $\phi(z)$.

Si $\phi(z)$ a $k \in \mathbb{N}$ racines unitaires et le reste de ses racines à l'extérieur du cercle unitaire, alors il existe $K \subset \{1, \cdots, p\}; |K| = k$ tel que $G_i = 1, \forall i \in K$ ce qui implique $\phi(B) = \phi'(B)(1-B)^k$ où $\phi'(z) \neq 0, \forall z \in \mathbb{C} : |z| \leq 1$. On peut donc appliquer l'analyse précédente sur la série stationnaire : $z'_t = (1-B)^k z_t$. Par exemple si on a une tendance d'accroissement linéaire, alors il serait judicieux d'utiliser $k = 1$ et d'évaluer les différences $(1-B)z_t = z_t - z_{t-1}$ qui sera stationnaire, même si les z_t ne le sont pas Box et al. (2008). L'utilisation de l'opérateur $(1-B)^k$ conduit à un modèle (non-stationnaire) "intégré" d'ordre k, dénoté ARIMA$(\cdot, k, \cdot)$.

Un autre cas commun est celui où le polynôme $\phi(z)$ peut être exprimé comme $\phi(z) = \phi'(z)(1-z^s)$ pour un $s \in \mathbb{N}$ donné où $\phi'(z) \neq 0, \forall z \in \mathbb{C} : |z| \leq 1$. Dans ce cas, $\phi(z)$ a s racines sur le cercle unitaire de la forme : $e^{2\pi i \frac{k}{s}}$ où $\sqrt{i} = -1$ et $k = 0, 1, \cdots, s-1$. La solution à $(1-B^s)z_t = \varrho_t$ où ϱ_t représente du bruit blanc sera donc, en moyenne, une combinaison de cosinus et sinus d'une périodicité de s périodes. L'utilisation de l'opérateur $(1-z^s)$ conduit ici à l'ajout de la saisonnalité, d'où le "S" dans "SARIMA".

2.1.3 Modèles GARCH

Dans le cas le plus simple, les résidus $\{\varrho_t\}$ sont supposés indépendants et on a $\mathrm{E}\left[\zeta_{t+l}^2|\varrho_{[t]}\right] = \mathrm{E}\left[\zeta_{t+l}^2\right] = \sigma^2$, peu importe $l \in \mathbb{N}$. Si on relaxe cette hypothèse et qu'on suppose plutôt que les résidus sont non corrélés et que les variances conditionnelles respectent la relation suivante :

$$\hat{\sigma}_{t-1}^2(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho_{t-i}^2 + \sum_{j=1}^{s} \beta_j \hat{\sigma}_{t-1-j}^2(1) \ , \tag{2.10}$$

où $\hat{\sigma}_t^2(l) = \mathrm{E}\left[\varrho_{t+l}^2|\varrho_{[t]}\right]$ pour tout $l \in \mathbb{N}$, alors on dit que les résidus suivent un modèle GARCH(m, s) (Bollerslev, 1986). Le terme hétéroscédasticité conditionnelle indique simplement que la variance conditionnelle n'est plus constante comme dans le cas où les ϱ_t sont indépendants. Ceci permet de modéliser plusieurs phénomènes tirés du monde réel, notamment en finance, où des périodes de volatilité accrues tendent à être suivies de périodes de variation élevée (Garcia et al., 2005; Tsay, 2005). Bien que les modèles GARCH n'aient pas été utilisés abondamment en gestion de réservoirs, Pianosi and Soncini-Sessa (2009) ont considéré cette forme d'incertitude dans un modèle réduit résolu "en ligne" pour gérer le lac Verbano sur la frontière italo-suisse.

Afin de s'assurer de la non-négativité de la variance conditionnelle, on impose des restrictions sur les coefficients : $\alpha_0, \alpha_i, \beta_j \geq 0, \forall i, j$. En prenant l'espérance des deux côtés de (2.10) et en se rappelant que l'espérance est constante selon les hypothèses de stationnarité (2.4), on dérive également d'autres relations sur les coefficients.

Le chapitre 5 indique qu'en appliquant une transformation affine aux apports, le test statistique Ljung-Box (Ljung and Box, 1978) permet de rejeter l'hypothèse que les ϱ_t^2 sont indépendants, suggérant donc l'utilité d'un modèle GARCH. Par contre, l'utilisation du logarithme au chapitre 6 permet de contrôler ce phénomène et ne nécessite donc pas l'utilisation du modèle GARCH.

2.1.4 Modèles de séries chronologiques multivariées

Cette thèse considère uniquement des processus stochastiques univariés en faisant des hypothèse simplificatrices sur la corrélation spatiale des apports à chaque pas de temps. Le chapitre 8 discute néanmoins de quelques améliorations possibles afin de considérer des modèles plus réalistes. Ces représentations s'appuient sur des processus ARMA multidimensionnels (vectoriels) aussi appelés VARMA. Si ζ_t et ϱ_t représentent des vecteurs aléatoires de dimension $|J|$ pour tout t et $\Phi_i \in \mathbb{R}^{|J|\times|J|}, i = 1, \cdots, p$, alors on pourrait notamment exploiter la

représentation :

$$\zeta_t = \sum_{i=1}^{p} \Phi_i \zeta_{t-i} + \varrho_t \tag{2.11}$$

On note cependant que l'identification, le calibrage ainsi que la vérification des conditions de stationnarité sont relativement plus complexes pour les modèles multivariés (Tsay, 2005). Ces derniers introduisent également certains phénomènes comme la colinéarité en plus de nécessiter la sauvegarde de matrices de dimension $|J| \times |J|$ plutôt que des scalaires et sont susceptibles d'augmenter considérablement la complexité du modèle. Bien que ces modèles soient un peu plus rares dans la littérature, ils ont été considérés par certains chercheurs comme Gjelsvik et al. (2010) qui discute notamment de l'utilisation d'un modèle VAR(1).

2.2 Programmation dynamique stochastique (*Stochastic dynamic programming - SDP*)

La programmation dynamique stochastique représente l'une des méthodes les plus utilisées en optimisation de petits systèmes sur un horizon moyen terme. Cette méthode bénéficie de plusieurs décennies de recherche et a été implantée avec succès pour déterminer les soutirages mensuels et évaluer les stocks d'eau pour certains réservoirs opérés par Hydro-Québec.

Afin de décrire cette approche, on considère le problème de gestion de réservoirs en présence d'incertitude sur les apports. On le formule comme le problème de contrôle optimal suivant adapté de Turgeon (2005) et Séguin (2016) :

$$\max_{\pi_t(\cdot)} \left\{ \mathrm{E}\left[\sum_{t=1}^{T} B_t(x_t, u_t, \xi_t) \right] : x_{t+1} = f_t(x_t, u_t, \xi_t);\ x_t \in \mathbb{X}_t;\ u_t \in \mathbb{U}_t(x_t, \xi_t);\ \pi_t(x_t) = u_t \right\}, \tag{2.12}$$

où on définit les concepts suivants :

- Les états : $x_t \in \mathbb{X}_t \subset \mathbb{R}^s$ comprennent habituellement les volumes de chaque réservoir au début du temps t ainsi qu'un sous-ensemble de l'historique des apports pour tenir compte de la corrélation temporelle. Le vecteur x_0 est connu au temps 0.
- Les actions/décisions : $u_t \in \mathbb{U}_t(x_t, \xi_t) \subset \mathbb{R}^a$ représentent habituellement les débits soutirés (turbinés) et déversés. Le domaine $\mathbb{U}_t(x_t, \xi_t)$ assure le respect des contraintes

opérationnelles.

— L'incertitude : $\xi_t \in \Xi_t \subset \mathbb{R}^r$ représentent habituellement les apports aux différents réservoirs.

— Les fonctions de récompense : $B_t : \mathbb{X}_t \times \mathbb{U}_t \times \Xi_t \to \mathbb{R}$ sont typiquement les fonctions de production auxquelles on soustrait une pénalité pour des inondations et/ou les déversés. $B_t(\cdot)$ peut être une fonction arbitraire, possiblement discontinue et non convexe.

— Les fonctions de transition : $f_t : \mathbb{X}_t \times \mathbb{U}_t \times \Xi_t \to \mathbb{R}^s$ représentent habituellement les équations de conservation de masse (conservation de l'eau).

On cherche la politique optimale, c'est-à-dire les fonctions $\pi_t : \mathbb{R}^s \to \mathbb{R}^a$ pour chaque $t \in \mathbb{T} = \{1, \cdots, T\}$ qui maximisent (2.12). À cet effet, on exprime notre problème de manière récursive en exploitant le principe d'optimalité de Bellman. Ceci nécessite de définir $J_t(x_t)$ comme la fonction de valeur au temps $t \in \{1, \cdots, T\}$ évaluée en x_t, c'est-à-dire le coût espéré des opérations de t à T étant donné l'état actuel x_t. Plus précisément :

$$J_T(x_T) = J^{val.\,eau}(x_T) \tag{2.13}$$

$$J_t(x_t) = \mathrm{E}_{\xi_t}\left[\max_{u_t \in \mathbb{U}_t(x_t, \xi_t)} \left\{B_t(x_t, u_t, \xi_t) + J_{t+1}(f_t(x_t, u_t, \xi_t))\right\} | \xi_{t-1}, \cdots, \xi_1\right], t = 1, \cdots, T-1, \tag{2.14}$$

où $J^{val.\,eau}(\cdot)$ est la fonction valorisant le stock final d'eau. Cette fonction peut notamment être obtenue en résolvant un problème de gestion de réservoirs moyen terme. Les chapitres 4, 5 et 6 considèrent le cas où $J^{val.\,eau}(\cdot)$ est affine, mais en général cette fonction est continue et non convexe.

Comme $J^{val.\,eau}(\cdot)$ et $B_t(\cdot)$ peuvent être des fonctions non convexes, le problème d'optimisation (2.14) est difficile en général et il faut procéder à une approximation. On suppose souvent dans la littérature qu'on discrétise les états, les actions et les réalisations de l'incertitude selon la même granularité à chaque pas de temps pour chaque dimension. La dimension de l'incertitude et des actions reste constante à travers le temps, mais la dimension des états peut croître. On fixe : $|\mathbb{X}_t| = K_x^{\hat{s}} \times |\Xi_t|^{g(t)}, |\mathbb{U}_t| = K_u^a$ et $|\Xi_t| = K_\xi^r$ où $K_x, K_u, K_\xi \in \mathbb{N}$ représente le nombre de discrétisations pour chacune des dimensions des états, des décisions et de l'incertitude pour tout t et $g : \mathbb{T} \to \mathbb{T} \cup \{0\}$. Pour les valeurs manquantes, une interpolation est utilisée. La constante K_x ne tient pas compte de la dimension de l'historique des apports. Elle pourrait notamment représenter le nombre de discrétisations pour chacun des

$\hat{s}$ réservoirs du système.

La tractabilité de cette méthode de résolution dépend principalement de la définition des états, qui elle dépend notamment de la représentation de l'incertitude considérée. On considère quelques choix communs :

2.2.1 Cas général, dépendance arbitraire entre les apports

Dans le cas général où ξ_t et ξ_{t+l} sont des vecteurs aléatoires corrélés de dimension r, on définit $x_t = (\hat{x}_t, \xi_{[t-1]})$ où $\hat{x}_t$ représente le vecteur de volumes initiaux de dimension $\hat{s}$ et $\xi_{[t-1]} = (\xi_1, \cdots, \xi_{t-1})$ représente l'historique d'apports.

Ce problème général mène a une formulation intractable dans la quasi-totalité des cas, car la dimension de la variable d'état est extrêmement grande. Il faut effectivement procéder à $O(\sum_{t=1}^{T} K_x^{\hat{s}} K_u^a K_\xi^{r(t-1)}) = O(K_x^{\hat{s}} K_u^a K_\xi^{(T-1)r})$ évaluations des fonctions de valeurs.

2.2.2 Cas où les apports sont indépendants

Dans ce cadre très particulier, on peut exprimer les fonctions de valeurs uniquement comme des fonctions des stocks d'eau initiaux et l'espérance conditionnelle se réduit à l'espérance. On a donc uniquement $O(\sum_{t=1}^{T} K_x^s K_u^a) = O(T K_x^s K_u^a)$ évaluations à faire, où $s = \hat{s}$. Ceci représente un gain extrêmement important, car on passe d'une complexité exponentielle selon l'horizon T à une complexité linéaire.

2.2.3 Cas (V)AR(p), corrélation d'ordre $p \in \mathbb{N}$

Dans le cas où ξ_t suit un processus (vectoriel) autorégressif d'ordre p, c'est-à-dire que $\Phi(B)\xi_t = \varrho_t$ où $\Phi(B) = 1 - \sum_{i=1}^{p} \Phi_i B^i$ et $\Phi_i \in \mathbb{R}^{r\times r}, \forall i$ et $\{\xi_t\}, \{\varrho_t\}$ sont des processus stochastiques discrets de dimension r, les apports au temps t dépendent uniquement de $(\xi_{t-1}, \cdots, \xi_{t-p})$ de façon linéaire. Ceci peut réduire la dimension du problème considérablement par rapport au cas général.

En supposant que ξ_{t+l} est connu pour tout $l \in \mathbb{Z}_-$, on a $O(\sum_{t=1}^{T} K_x^{\hat{s}} K_u^a K_\xi^{r\min\{t-1,p\}}) = K_x^{\hat{s}} K_u^a (\sum_{t=1}^{p+1} K_\xi^{r(t-1)} + \sum_{t=p+2}^{T} K_\xi^{rp}) = O(K_x^{\hat{s}} K_u^a (K_\xi^{rp} + K_\xi^{rp}(T-p-1))) = O(K_x^{\hat{s}} K_u^a (T-p) K_\xi^{rp})$. Il s'agit donc d'un cas mitoyen entre l'indépendance et la corrélation parfaite.

La cas Markovien avec $p = 1$ a notamment attiré beaucoup d'attention, car il permet de considérer une certaine corrélation temporelle sans toutefois engendrer des difficultés computationelles trop importantes. Bien que ce modèle de l'incertitude se révèle particulièrement utile pour représenter les apports mensuels, les apports journaliers tendent à avoir une cor-

rélation temporelle plus importante (Turgeon, 2005; Gauvin et al., 2017).

Afin d'obtenir une représentation plus fidèle pour les apports journaliers, Turgeon (2005) considère le cas particulier d'un seul réservoir avec des apports unidimensionnels où $(\xi_{t+l}, \cdots, \xi_t) \sim MVN(\mu_{t,l}, \sigma^2_{t,l})$ pour tout $t \in \mathbb{T}$ et $l \in 0 \cup \mathbb{N}$, c'est-à-dire que les apports suivent une loi multinormale peut importe le temps t et l'horizon l. Sous cette hypothèse, il parvient à déduire une formulation où il est possible de remplacer l'historique $(\xi_{t-1}, \cdots, \xi_{t-p})$ par une seule variable hydrologique H_t représentant une combinaison linéaire des apports passés. Bien que l'hypothèse de normalité soit difficile à vérifier en pratique, il est possible d'appliquer des transformations de puissance similaires à celles utilisés au chapitre 6 pour mieux respecter cette hypothèse.

2.2.4 Autres modélisations des variables hydrologiques

Il est également possible de considérer différentes variables hydrologiques afin de mieux prévoir les apports. Quentin et al. (2014) et Côté et al. (2011) considèrent l'ajout de l'équivalent de l'eau de neige (*snow water equivalent*), l'humidité du sol ainsi que différentes combinaisons de ces indicateurs. Kelman et al. (1990) considèrent quant à eux des prévisions saisonnières.

2.2.5 Améliorations permettant de réduire les malédictions de la dimensionalité

L'analyse de complexité précédente illustre que la méthode souffre de la malédiction de la dimensionalité et ne peut permettre de résoudre de gros problèmes. En effet, la discrétisation implique une croissance exponentielle selon le nombre de décisions ou d'états. Pour un système avec 5 réservoirs comme c'est le cas dans aux chapitres 4, 5 et 6, on a besoin de $\Omega(10^5 \times 10^5 \times 29 \times 10^5) = \Omega(10^{15})$ évaluations de la fonction de valeur si on considère uniquement les débits soutirés, un processus autorégressif d'ordre 1 ainsi que 10 discrétisations pour chaque dimension des états, décisions et incertitude.

Comme on le verra aux chapitres 4, 5 et 6, il faudrait ajouter les débits totaux (somme des déversés et soutirés) au temps passé pour modéliser nos contraintes sur les variations des flots. Afin de prendre en compte les délais d'écoulement décrits à la section 1.2.1, il faudrait également considérer ces débits passés sur un horizon allant jusqu'à 11 jours. Ceci mènerait à des variables d'état de taille supérieure à 20. Une telle dimension conduirait assurément à des problèmes de trop grande taille pour pouvoir être résolus directement par la méthode SDP.

Il est néanmoins possible de réduire les effets de cette malédiction de la dimensionalité en effectuant une meilleure discrétisation et interpolation des fonctions de valeurs. Cervellera

et al. (2006); Castelletti et al. (2008) proposent notamment d'interpoler la fonction de valeurs grâce aux réseaux de neurone (*artifical neural network*). Powell (2007) propose plusieurs types d'approximations s'appuyant notamment sur l'apprentissage machine afin de bien approximer les fonctions de valeurs à faible coût.

Dans le cas où $\mathbb{U}_t(x_t, \xi_t)$ est convexe et $B_t(x_t, u_t, \xi_t)$ est concave conjointement en (x_t, u_t) et $J_{t+1}(\cdot)$ est concave, Zéphyr et al. (2016) propose de décomposer l'espace des états à l'aide de simplexes plutôt que par hyperrectangles comme c'est habituellement. Ces derniers suggèrent également une approche d'approximation récursive permettant d'obtenir une meilleure approximation de la fonction de valeur.

2.3 Programmation stochastique dynamique par échantillonnage (*Sampling stochastic dynamic programming - SSDP*)

Tel qu'illustré à la section précédente, la programmation dynamique stochastique a bénéficié de nombreuses améliorations au cours des dernières décennies. Cependant, l'algorithme SDP souffre toujours de plusieurs limitations, notamment la représentation de l'incertitude. En effet, il est difficile de bien représenter la corrélation spatiale pour les systèmes à plusieurs réservoirs ainsi que les corrélations temporelles plus longues que quelques périodes avec l'algorithme SDP de base sans engendrer une explosion de la dimension des états.

Par ailleurs, tel que mentionné par Kelman et al. (1990), la discrétisation des apports créé une distorsion dans la distribution, ce qui a comme effet d'éliminer ou de réduire la probabilité de scénarios extrêmes. Ceci peut avoir des conséquences désastreuses pour la gestion des inondations et des sécheresses.

Afin de pallier à ces limitations, certains chercheurs on introduit la programmation stochastique dynamique par échantillonnage (Côté and Leconte, 2015; Stedinger and Faber, 2001; Kelman et al., 1990). Plutôt que de supposer que les apports ξ_t suivent une distribution discrète particulière, ces derniers proposent d'utiliser $M \in \mathbb{N}$ scénarios d'apports historiques ou prévisions d'ensemble (*ensemble streamflow predictions*) : $\{\bar{\xi}^1, \cdots, \bar{\xi}^M\}$ où $\xi^i \in \mathbb{R}^T$. Ces séries chronologiques tiennent implicitement compte des dépendances spatiales et temporelles complexes et sont faciles à représenter et utiliser.

La difficulté consiste à évaluer les probabilités de transiter d'un scénario i à un autre j. Ces probabilités sont calculées à l'aide du théorème de Bayes et en faisant des hypothèse particulières sur les distributions des variables hydrologiques. Si une variable hydrologique est ajoutée, on peut également évaluer ces probabilités de transition en conditionnant sur l'état hydrologique actuel.

Des variations de cette approche algorithmique offrent d'excellents résultats avec des temps de calcul raisonnables. Ils ont également été testés dans des cadres opérationnels réels, notamment par Rio Tinto (Côté and Leconte, 2015).

2.4 Programmation stochastique dynamique duale (*Stochastic dual dynamic programming - SDDP*)

Dans le cas spécial où $\mathbb{U}_t(x_t, \xi_t)$ est un polyhèdre pour tout (x_t, ξ_t) et $f_t(\cdot), B_t(\cdot)$ sont des fonctions linéaires, il est possible d'utiliser un algorithme spécialisé : la programmation stochastique dynamique duale. Cet algorithme a initialement été conçu afin de résoudre des problèmes de coordination entre la production hydroélectrique et thermale pour la gestion long terme du parc de production brésilien (Pereira and Pinto, 1991). L'algorithme SDDP permet d'obtenir des solutions à des problèmes linéaires stochastiques multi-étapes de très grandes dimensions que des algorithmes de type SDP et SSDP seraient incapables de résoudre.

Dans ce cas, il est possible de montrer que la fonction de valeur $J_{t+1}(\cdot)$ est concave (pour un problème de maximisation comme (2.14)) et linéaire par morceaux (Shapiro, 2011). Grâce à la dualité forte en programmation linéaire, il est possible d'obtenir des hyperplans de support pour ces fonctions de valeurs. Il est possible de raffiner progressivement ces approximations à l'aide d'une phase récursive arrière. Puis, une phase de simulation avant nous permet d'obtenir une solution réalisable étant donné un ensemble de trajectoires possibles du processus stochastique généré.

Ces phases sont expliquées sommairement dans les sections qui suivent. Pour les besoins d'illustration, les variables aléatoires sont supposées indépendantes à chaque temps. Bien que ce ne soit pas le cas en général, il est notamment possible d'ajouter des réalisations passées à la variable d'état pour modéliser des processus ARIMA. On suppose également que les variables aléatoires sont discrètes et peuvent uniquement prendre M_t valeurs distinctes à chaque temps t et qu'il y a $\Pi_{t=1}^{T} M_t = M$ scénarios possibles. On se limite également au cas de recours relativement complet, c'est-à-dire qu'avec probabilité 1, il est toujours possible de trouver une solution réalisable au problème (2.14), peu importe l'état actuel.

2.4.1 Phase arrière - borne supérieure (pour un problème de maximisation)

On suppose qu'on dispose de $\tilde{J}_t(\cdot)$, une borne supérieure sur la fonction de valeur à chaque étape t et où $\tilde{J}_t(x) = J_t(x) = 0, \forall x$. On connaît également $\{\bar{x}_t\}_{t=1}^{T}$, un ensemble d'états à évaluer ainsi que $\{\bar{\xi}_t^i\}_{i=1}^{M_t}$, les réalisations possibles au temps t. Pour chaque $t = T, T-1, \cdots, 1$,

pour tout $\bar{x}_t$ et $\bar{\xi}_t^i$, on résout le *dual* du problème linéaire suivant :

$$\max_{u_t \in \mathbb{U}_t(\bar{x}_t, \bar{\xi}_t^i)} \left\{ B_t(\bar{x}_t, u_t, \bar{\xi}_t^i) + \tilde{J}_{t+1}(f_t(\bar{x}_t, u_t, \bar{\xi}_t^i)) \right\} \tag{2.15}$$

En prenant la moyenne sur les M_t scénarios, la valeur des variables duales optimales ainsi que l'objectif optimal nous permettent ensuite d'obtenir un hyperplan de la forme $\alpha_t x_t + \beta_t$ qui majore $J_{t+1}(x_t)$. On raffine ensuite la représentation en utilisant : $\tilde{J}_t(x_t) = \min\{\tilde{J}_t(x_t), \alpha_t x_t + \beta_t\}$ (pour un problème de maximisation).

2.4.2 Phase avant - borne inférieure

En utilisant les approximations (fonctions majorantes) $\tilde{J}_t(\cdot)$ trouvées à l'étape précédente, on considère ensuite $N \leq M$ trajectoires possibles du processus stochastique : $\{\hat{\xi}^1, \cdots, \hat{\xi}^N\}$ où $\hat{\xi} \in \mathbb{R}^T, \forall i$. Pour chaque $t = 1, \cdots, T$ on résout le problème suivant (pouvant être modélisé comme un programme linéaire en ajoutant des variables de décision additionnelles) :

$$\max_{u_t \in \mathbb{U}_t(\hat{x}_t, \hat{\xi}_t^i)} \left\{ B_t(\hat{x}_t, u_t, \hat{\xi}_t^i) + \tilde{J}_{t+1}(f_t(\hat{x}_t, u_t, \hat{\xi}_t^i)) \right\} \tag{2.16}$$

où $\hat{x}_{t+1} = f_t(\hat{x}_t, u_t, \hat{\xi}_t^i)$ et $\hat{x}_0$ est connu.

Dans le cas où $N < M$, on doit procéder par échantillonnage et la solution optimale trouvée est donc aléatoire. Dans ce cas, l'algorithme peut être stoppé si la différence entre la borne supérieure et la borne inférieure se situe dans un certain intervalle de confiance.

L'algorithme SDDP jouit d'une popularité considérable dans le milieu académique et industriel. Phillpot and Guan (2008); Shapiro (2011) ont étudié ses propriétés théoriques. Les travaux récents de Shapiro et al. (2013) ont permis d'incorporer l'aversion au risque. Tilmant and Kelman (2007); Rougé and Tilmant (2016) ont également utilisé l'algorithme dans divers contextes de gestion des ressources hydriques alors que Maceira and Damázio (2004) l'ont utilisé pour optimiser le parc de production brésilien. L'algorithme a également été incorporé dans des logiciels d'aide à la décision dans de nombreux pays d'Amérique du Sud afin de faciliter la planification énergétique.

2.5 Programmation stochastique sur arbre

Une approche alternative aux algorithmes SDP, SSDP et SDDP est la programmation stochastique multi-étapes sur arbre. Dans sa forme la plus simple, cette approche consiste simplement à indexer chacune des décisions du problème déterministe original par une réalisation du processus stochastique à un temps donné, c'est-à-dire à un nœud de l'arbre de scénario.

Les figures 2.2 et 2.3 présentent deux exemples d'arbre de scénarios sur un horizon de $T = 30$ jours. Alors que l'arbre 2.2 considère deux réalisations différentes à chaque 5 jours, l'autre utilise une structure de branchement asymétrique où l'on considère différentes réalisations aux premiers jours, mais où il n'existe plus d'incertitude à partir du jour 6.

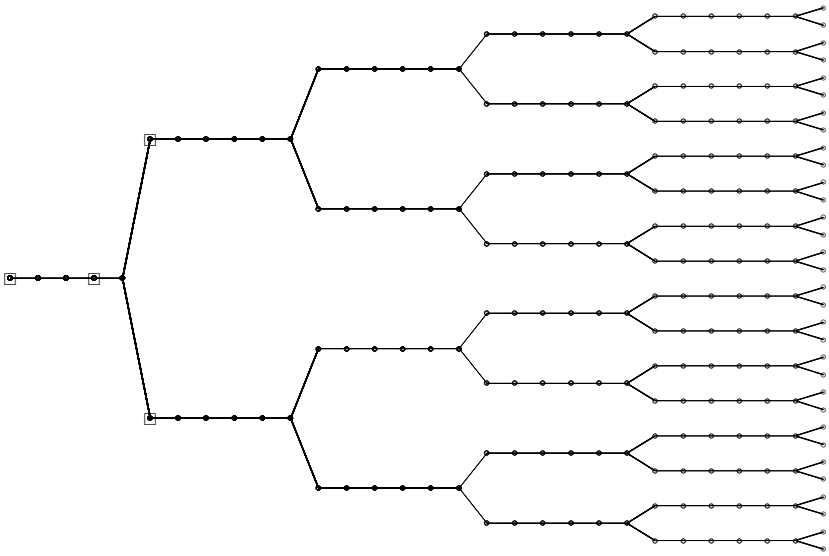

Figure 2.2 Arbre de scénario symétrique

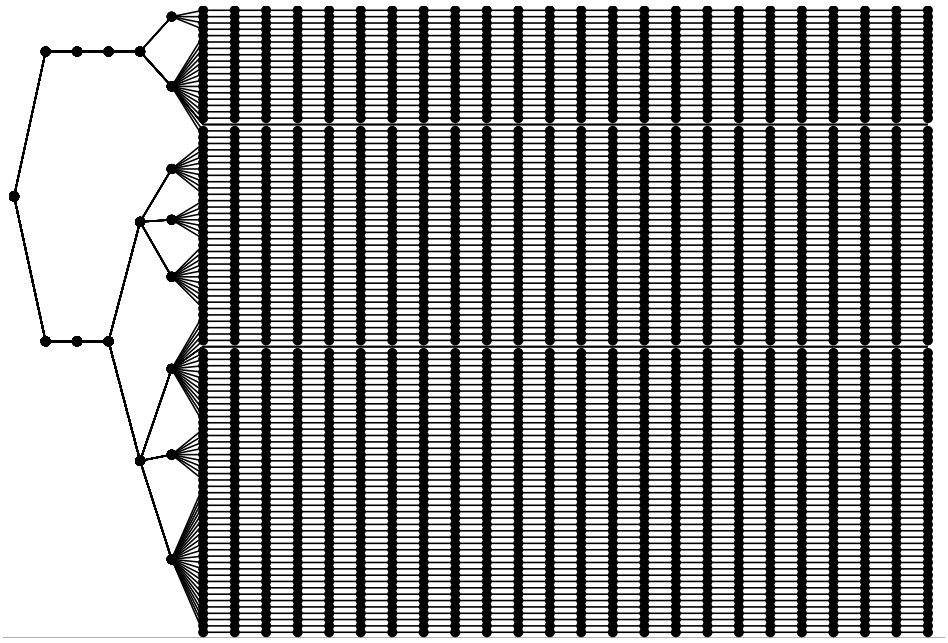

Figure 2.3 Arbre de scénario asymétrique

Les arbres de scénario ont notamment été utilisés pour résoudre des problèmes de production hydroélectrique (Carpentier et al., 2015, 2013a; Gonçalves et al., 2012; Fleten and Kristoffersen, 2008; Growe-Kuska et al., 2003; Aasgård et al., 2014; De Ladurantaye et al., 2009; Séguin et al., 2016b). Ces modèles permettent de représenter les corrélations temporelles et spatiales plus facilement et intuitivement que les modèles ARIMA par exemple. Il est également facile de convertir un modèle d'optimisation déterministe existant en modèle stochastique sur arbre.

La principale limitation des arbres de scénarios est la croissance exponentielle de la taille du modèle selon l'horizon temporel T. À titre d'exemple, en considérant une marche aléatoire avec deux réalisations possibles à chaque étape, on obtient 2^t nœuds à l'étape t et donc $\sum_{t=1}^{T} 2^t = 2^{T+1} - 1$ nœuds au total dans l'arbre. Afin de limiter cet accroissement, il est possible d'utiliser certaines approximations du processus. Il est notamment possible de faire comme à la figure 2.2 et ne brancher qu'à certaines périodes définies. Cependant, ceci engendre une perte d'information, même dans le cas où le processus est réellement discret et représentable par un nombre fini de scénarios.

Dans le but de pallier à cette problématique, plusieurs chercheurs ont développé des méthodologies et des outils informatiques permettant de réduire la taille des arbres de scénarios. Les travaux de Growe-Kuska et al. (2003); Heitsch and Römisch (2009) sur les distances entre les mesures de probabilité ont entre autres mené au développement de l'outil (SCENRED2),

intégré à la librairie GAMS. L'outil permet de construire des arbres de scénarios à partir de peignes ou encore de réduire la taille de l'arbre initial et a notamment été utilisées par Feng and Ryan (2014). Les travaux récents de Pflug and Pichler (2016) ont également proposé une approche théorique basée sur les distances de probabilités imbriquées. Ce concept a été exploité par Séguin et al. (2016b) pour un problème de gestion hydroélectrique court terme réel.

Une approche alternative consiste à utiliser l'algorithme *progressive hedging* introduit initialement par Rockafellar and Wets (1991). Cette technique consiste essentiellement à décomposer l'arbre en un peigne composé de scénarios indépendants et d'imposer des contraintes de non-anticipativité pour lier certains de ces scénarios. On utilise ensuite une méthode de langrangien augmenté pour pénaliser les violations des contraintes de non-anticipativité.

Cette méthode a été étudiée par Carpentier et al. (2013a) dans le cadre d'une étude sur l'évaluation des politiques de gestion des réservoirs au Québec ainsi que par Gonçalves et al. (2012, 2013) pour résoudre le problème de coordination hydro-thermal au Brésil. Toutefois, l'algorithme est sensible aux variations des paramètres et la convergence peut être longue à atteindre. Tel que mentionné brièvement par Côté and Leconte (2015), il est assez lourd d'implémenter le *progressive hedging* pour des problèmes réels, surtout si on veut évaluer les règles de gestion à l'aide de simulation et d'un horizon fuyant.

2.6 Règles de décision linéaires/affines (*Linear/affine decision rules*) [3]

Bien que les méthodes de résolution présentées aux sections précédentes possèdent des avantages indéniables et aient été utilisées avec succès pour résoudre des problèmes industriels réels, elles supposent (presque) toutes que les variables aléatoires suivent une distribution discrète.

Cette hypothèse assez forte se révèle parfois problématique, car elle entraîne des difficultés numériques importantes. Pour les arbres de scénarios, la dimension de l'espace des décisions croît exponentiellement selon l'horizon du problème. Pour les méthodes liées à la programmation dynamique, la complexité croît exponentiellement selon la dimension de l'incertitude et la considération de la corrélation temporelle augmente la taille des variables d'état.

Bien que la méthode SDDP ne nécessite pas de faire l'hypothèse de distributions discrètes, elle nécessite quand même de pouvoir échantillonner à partir de la distribution. De plus, le nombre

3. Dans le reste du texte, on utilise les termes "affine" et "linéaire" de manière interchangeable. En effet, pour une fonction affine $l : \mathbb{R}^p \to \mathbb{R}^q$ de la forme $l(\xi) = L^0 + L^\Delta \xi$ où $p, q \in \mathbb{N}$, $L^0 \in \mathbb{R}^q, L^\Delta \in \mathbb{R}^{q \times p}$ et $\xi \in \Xi \subset \mathbb{R}^p$ on peut trouver une fonctions linéaire correspondante $\hat{l}(\hat{\xi}) = \hat{L}^\Delta \hat{\xi}$ où $\hat{\xi} \in \{(1, \xi^\top)^\top : \xi \in \Xi\} \subset \mathbb{R}^{p+1}$ et $\hat{L}^\Delta = (L^0, L^\Delta) \in \mathbb{R}^{q \times (p+1)}$ (voir Georghiou et al. (2015)).

de problèmes d'optimisation résolus dépend du nombre de réalisations échantillonnées, autant pour la phase avant que la phase arrière.

La discrétisation de l'incertitude complique également l'implémentation des règles de gestion dans le cas usuel où l'incertitude suit une loi continue. Par exemple, il faut procéder à une interpolation des fonctions de valeurs lorsqu'on utilise des méthodes dérivées de la programmation dynamique, car leur valeur est uniquement calculée explicitement à quelques points. De manière semblable, les arbres de scénarios conduisent à des politiques implémentables [4] uniquement lorsqu'un scénario considéré par le modèle se matérialise réellement.

Bien qu'il soit possible de contourner cette limitation en résolvant ces problèmes en horizon fuyant, cette approche peut se révéler assez lourde en temps de calcul lorsqu'on effectue des simulations avec de nombreux scénarios d'apport.

Plutôt que de considérer des règles de gestion arbitraires et de les approximer en les discrétisant comme c'est le cas pour SDP, SSDP et la programmation stochastique sur arbre, il est possible d'étudier un ensemble limité de fonctions ayant une forme particulièrement simple. Cette thèse utilise notamment les règles de décision linéaires (*linear decision rules*) et les règles de décision linéaires par morceaux aussi connues sous le nom de règles de décision "liftées" (*lifted decision rules*).

La méthode basée sur ces règles de décision suppose qu'au temps t, on optimise sur un espace de fonctions de la forme $\mathcal{X}_t : \mathbb{R}^{rT} \to \mathbb{R}^{n_t}$ où $n_t \in \mathbb{N}$, $r \in \mathbb{N}$ et T représentent respectivement la dimension de l'espace des décisions, la dimension de l'incertitude (les apports ξ_t dans cette thèse) et l'horizon temporel (fini). Si $K_t = \{n_{t-1} + 1, \cdots, n_{t-1} + n_t\}$ représente l'ensemble des indices associés aux décisions au temps t et que $\{\xi_t\}_{t=1}^{T}$ représente les apports où ξ_t est un vecteur aléatoire de dimension r pour tout t, alors il est possible d'exprimer les règles de décision linéaires sous la forme :

$$\mathcal{X}_{k,t}(\xi) = \mathcal{X}^0_{k,t} + \sum_{j=1}^{r} \sum_{t'=1}^{T} \mathcal{X}^{t',i}_{k,t} \xi_{t',j}, \qquad \forall k \in K_t \tag{2.17}$$

où $\mathcal{X}^0_{k,t}, \mathcal{X}^{t',i}_{k,t} \in \mathbb{R}$ sont les paramètres de la fonction affine des apports à optimiser. Dans les cas où $\mathrm{E}\left[\xi_{tj}\right] = 0, \forall j, t$, on peut interpréter $\mathcal{X}^0_{k,t}$ comme la décision moyenne ou nominale alors que le terme $\sum_{j=1}^{r} \sum_{t'=1}^{T} \mathcal{X}^{t',i}_{k,t} \xi_{j,t'}$ permet d'ajuster cette solution aux différentes réalisations des

4. On définit l'implémentabilité comme le fait de pouvoir mettre en application une décision réelle et pertinente (par exemple la quantité de débits à soutirer d'un réservoir) uniquement sur la base de l'information disponible à ce moment.

ξ_{jt}.

Afin d'obtenir des règles de décision implémentables [5], on impose $\mathcal{X}_{k,t}^{t',i} = 0, \forall t' \geq t, \forall i = 1, \cdots, r$. Ceci permet naturellement de s'assurer que les décisions à prendre au temps t dépendent uniquement de $\xi_{[t-1]}$, l'historique des apports observés.

Dans le reste de cette section ainsi que le reste de la thèse, on considère le cas particulier $r = 1$. Il est notamment raisonnable de supposer que le processus d'apports est unidimensionel si on considère une seul réservoir ou si on étudie les apports sur toute la rivière.

Il est surprenant de constater que les règles de décision linéaires ont été introduites il y a plus de 40 ans pour optimiser la capacité de réservoirs multi-usage sous certaines contraintes sur la navigation et le risque d'inondation (ReVelle and Kirby, 1970; Loucks and Dorfman, 1975). Ces auteurs considèrent par exemple le cas où les débits soutirés prennent la forme :

$$\mathcal{R}_t = \mathcal{S}_t - \mathcal{B}_t \tag{2.18}$$

où $\mathcal{R}_t, \mathcal{S}_t$ représentent respectivement les débits soutirés, les volumes d'eau dans le réservoir au temps t et $\mathcal{B}_t$ est une variable de décision réelle à optimiser. En négligeant les débits déversés, les contraintes de conservation de l'eau s'expriment comme : $\mathcal{S}_{t+1} = \mathcal{S}_t + \xi_t - \mathcal{R}_t$ où ξ_t sont les apports. En exploitant (2.18) et en explicitant la dépendance sur ξ_t, on a donc :

$$\mathcal{S}_{t+1}(\xi) = \xi_t + \mathcal{B}_t, \qquad t = 1, \cdots, T \tag{2.19}$$

$$\mathcal{R}_{t+1}(\xi) = \xi_t + \mathcal{B}_t - \mathcal{B}_{t+1}, \quad t = 1, \cdots, T \tag{2.20}$$

$$\mathcal{S}_1(\xi) = \hat{\mathcal{S}}_1, \tag{2.21}$$

$$\mathcal{R}_1(\xi) = \hat{\mathcal{S}}_1 - \mathcal{B}_1, \tag{2.22}$$

où l'on suppose que $\hat{\mathcal{S}}_1$ est connu. $\mathcal{S}_1(\cdot)$ et $\mathcal{R}_1(\cdot)$ sont donc des fonctions constantes de l'incertitude.

On peut donc conclure que sous cette forme, les volumes et les débits soutirés sont bien des fonctions affines de l'incertitude ayant la forme (2.17). Cependant, ils représentent un cas relativement limitant et rigide, car les $\mathcal{B}_t$ sont communs à $\mathcal{S}_{t+1}$ et $\mathcal{R}_{t+1}$ et on fixe le coefficients de ξ_t à 1 et celui de ξ_{t-l} à 0 pour $l = 1, \cdots, t-1$. Les débits et les volumes

5. On dit également que les règles de décision sont causales, non anticipatives ou encore adaptées à la filtration $\{\mathscr{F}_t\}_{t=1}^T$ où $\mathscr{F}_t = \sigma(\{\xi_{[t-1]}\})$ représente l'information contenue par l'historique $\xi_{[t-1]}$.

au temps t dépendent donc uniquement des apports à ce temps donné d'une manière très précise.

Une autre limitation importante de ces travaux est le fait de supposer que les apports suivent une distribution particulière connue parfaitement. Sous cette hypothèse, ces auteurs parviennent à modéliser le problème comme un problème avec contraintes probabilistes (*chance-constrained problem*). Ils garantissent ainsi que les contraintes du type $\mathcal{R}_t(\xi) \leq \bar{r}$ sont respectées avec une probabilité $1 - \epsilon$ pour un petit $\epsilon > 0$ fixé.

Certains auteurs comme Stedinger (1984) ont néanmoins critiqué l'utilisation pratique de tels modèles, notamment car ils conduisent à des décisions trop rigides se révélant souvent trop conservatrices lorsqu'évaluées sur des scénarios réels. Les règles de décision linéaires ont donc perdu beaucoup de popularité au cours des décennies suivantes.

Il a fallu attendre les années 2000 avant que les travaux de Ben-Tal et al. (2004) et Kuhn et al. (2011) sur les problèmes stochastiques et robustes [6]multi-étapes redonnent une impulsion aux règles de décision linéaires. [7] Depuis cette période, les règles de décision linéaires ont connu un regain d'intérêt considérable. Bertsimas et al. (2010) ont étudié leurs propriétés et ont établi certaines conditions garantissant leur optimalité. Kuhn et al. (2011) proposent d'utiliser ces règles de décision sur le problème dual et primal afin de pouvoir évaluer leur degré de sous-optimalité. Afin d'améliorer leur performance, plusieurs auteurs ont également développé des extensions afin de considérer des règles de décision linéaires sur un espace lifté (Georghiou et al., 2015; Goh and Sim, 2010) ou encore des règles de décision entières (Bertsimas and Georghiou, 2015).

Les règles de décision linéaires ont également été utilisées avec succès dans de nombreux domaines comme la gestion d'inventaires, la localisation d'entrepôts, la gestion de projets, etc. (Delage and Iancu). Ces techniques ont également été exploitées pour résoudre des problèmes en énergie comme les problèmes de chargement (*unit commitment problems*) (Lorca and Sun, 2015; Lorca et al., 2016; Bertsimas et al., 2013) et des problèmes de gestion de portefeuille

7. On définit l'optimisation robuste comme une méthodologie de prise de décision en présence d'incertitude faisant un nombre minimal d'hypothèses sur l'aléa et axée sur la tractabilité des problèmes résolus (Ben-Tal et al., 2009). Un problème d'optimisation robuste est habituellement formulé comme un problème de la forme : $\min_{x \in X} \max_{\xi \in \Xi} \{f(x, \xi) : g_i(x, \xi) \leq 0,\ \forall i = 1, \cdots, m, \forall \xi \in \Xi\}$ où X et Ξ représentent des ensembles déterministes convexes et $f(\cdot, \cdot)$ et $g_i(\cdot, \cdot)$ sont des fonctions convexes en x et concave en ξ. On suppose également couramment que le problème de maximisation interne puisse se formuler comme un programme linéaire, conique de second ordre ou semi-défini afin de pouvoir utiliser la dualité pour le reformuler comme un problème de minimisation. L'approche peut s'inscrire dans un cadre non-probabiliste ou probabiliste, comme c'est le cas dans ce mémoire. Dans ce second cas, on suppose que l'aléa (ξ) peut être représenté par des variables aléatoires dont on connait certaines informations comme le support et l'espérance. L'optimisation robuste probabiliste peut donc être considérée comme une branche de l'optimisation stochastique.

dans un marché de l'électricité dérégulé (Rocha and Kuhn, 2012). De nombreuses versions des règles de décision ont plus particulièrement été utilisées pour résoudre des problèmes de gestion de réservoirs (Apparigliato, 2008; Pan et al., 2015; Tandberg and Vefring, 2012; Braaten et al., 2015; Egging et al., 2017).

Ces travaux supposent que les règles de décisions prennent le forme (2.17) et que les variables aléatoires prennent des valeurs dans un certain ensemble convexe et compact Ξ avec probabilité 1. Bien qu'elle ne soit pas faite en toute généralité, cette hypothèse sur le support est motivée par des notions de probabilités formelles (Ben-Tal et al., 2009). Elle permet également de dériver l'équivalent déterministe à travers la dualité conique et ainsi de formuler un problème d'optimisation pouvant être résolu efficacement par des algorithmes de points intérieurs ou même par la méthode du simplexe.

Dans cette thèse, on considère le cas où $\Xi = \{\xi \in \mathbb{R}^T : A\xi \leq b\}$ pour $A \in \mathbb{R}^{m\times T}$ et $b \in \mathbb{R}^m$ pour un certain $m \in \mathbb{N}$, c'est-à-dire que le support est polyhédral. Dans ce cas, pour assurer le respect de la contrainte $\mathcal{R}_t(\xi) \leq \bar{r}$ avec probabilité 1 à un t donné, il suffit de s'assurer que la valeur du problème d'optimisation (linéaire) suivant :

$$\max_{\xi\in\Xi} \mathcal{R}_{t+1}(\xi) = \max_{\xi: A\xi\leq b} \{\mathcal{R}_t^0 + \sum_{t'=1}^{T} \mathcal{R}_t^{t'}\xi_{t'}\} = \mathcal{R}_t^0 + \max_{\xi: A\xi\leq b} \{\sum_{t'=1}^{T} \mathcal{R}_t^{t'}\xi_{t'}\} \tag{2.23}$$

est bien plus petite ou égal à $\bar{r}$. Par dualité, c'est le cas s'il existe π_t dans l'ensemble

$$\{\pi_t \in \mathbb{R}^m \ : \ \mathcal{R}_t^0 + \pi_t^\top b \leq \bar{r} \ , \ A^\top \pi \leq \mathcal{R}_t^\Delta\} \tag{2.24}$$

où $\mathcal{R}_t^\Delta = (\mathcal{R}_t^1, \cdots, \mathcal{R}_t^T)^\top$. Pour dériver l'équivalent déterministe du problème original, il faut donc ajouter les variables π_t ainsi que les contraintes $\mathcal{R}_t^0 + \pi_t^\top b \leq \bar{r}$ et $A^\top \pi_t \leq \mathcal{R}_t^\Delta$ pour *chaque* contrainte $\mathcal{R}_t(\xi) \leq \bar{r}$. L'optimisation se fait ensuite sur $\pi_t, \mathcal{R}_t^\Delta, \mathcal{R}_t^0$.

Il est donc clair que la taille de l'équivalent déterministe augmente par rapport à celle du problème déterministe original. Par contre, cette augmentation est polynomiale selon le nombre de variables et de contraintes utilisées pour représenter le support Ξ. On note également que dans le cas où le problème déterministe est linéaire et que Ξ est polyhédral l'équivalent déterministe demeure un problème linéaire. Il est aussi possible de représenter Ξ comme une ellipse ou encore à l'aide d'un ensemble d'inégalités matricielles linéaires. Ceci mènerait à un

problème conique de second ordre ou un problème semi-défini.

2.6.1 Mesures de risque

Dans la grande majorité des travaux cités aux sections précédentes, l'objectif à optimiser est la *valeur espérée* d'une fonction de récompense, typiquement une combinaison linéaire de la production hydroélectrique totale, de pénalisations pour des violations de contraintes et/ou de déversés ainsi qu'une valorisation du stock final d'eau.

Bien que l'espérance mathématique et l'espérance conditionnelle possèdent des avantages considérables comme la linéarité et la propriété de l'espérance totale en plus d'être bien étudiées, ces fonctionnelles souffrent également de limitations importantes. On considère souvent l'espérance dans une optique intuitive où l'on cherche à s'assurer de la performance d'un système en moyenne sur un grand nombre de réalisations. Cependant, tel que soulevé par Ben-Tal et al. (2009), cette interprétation par la loi des grands nombres est uniquement satisfaisante pour un preneur de décision si ce dernier est prêt à échantillonner de nombreuses observations [8].

Une telle hypothèse est peut-être vérifiée dans le cas de la production hydroélectrique court terme où la planification doit être effectuée de très nombreuses fois. Dans ce cas, il est probablement acceptable de produire moins durant certains jours si cette réduction est compensée par une production accrue pendant une autre période. Toutefois, ce n'est pas nécessairement le cas pour la minimisation d'inondations, car les conséquences humaines et environnementales d'une seule occurrence d'un tel désastre sont extrêmement importantes.

Afin de modéliser cette aversion au risque, il est possible d'utiliser les travaux importants sur la théorie de l'utilité espérée (Neumann and Morgenstern, 1953; Föllmer and Schied, 2011). Toutefois, il est souvent difficile de bien codifier la fonction d'utilité pour un preneur de décision particulier (Armbruster and Delage, 2015). Par ailleurs, plusieurs travaux suggèrent que les fonctions d'utilité seraient non convexes et non concaves en réalité, ce qui mènerait à des difficultés numériques considérables. Les lacunes de la théorie de l'utilité espérée ont par

7. Il existe des similitudes très marquées entre l'histoire des règles de décision linéaires et celle de l'optimisation robuste. En effet, l'optimisation robuste a initialement été introduite dans les années 70 par Soyster (1973) pour résoudre des problèmes linéaires incertains. Toutefois, cette méthodologie est uniquement devenue populaire dans les années 2000 suite notamment aux développements en optimisation conique (Ben-Tal and Nemirovski, 1998; El Ghaoui and Lebret, 1997; El Ghaoui et al., 1998). Ces recherches ont a leur tour généré de l'attention sur les problèmes avec incertitude linéaire (Sim and Bertsimas, 2004) et ont même mené au développement d'outils de formulation automatiques (Goh and Sim, 2011). Ces avancées ont grandement facilité l'utilisation et la popularité de la méthodologie.

8. Si ces observations sont faites à travers le temps, alors ce système doit également se comporter de façon stationnaire.

ailleurs mené à la formulation de paradoxes comme celui d'Ellsberg (Ellsberg, 1961). [9]

Le paradoxe d'Ellsberg est intéressant, car il démontre que la plupart des individus ne respectent pas les axiomes classiques de la théorie de l'utilité espérée. Ces derniers font plutôt preuve d'aversion à l'ambiguïté. En effet, ils font face non seulement à de l'incertitude au niveau des réalisations des variables aléatoires, mais également à une ambiguïté sur la distribution à considérer. En d'autres mots, il existe un risque sur le risque.

Les travaux de Artzner et al. (1999) sur la théorie des mesures de risque cohérentes permettent de répondre à certaines de ces critiques. Cette approche consiste à définir un ensemble de fonctionnelles possédant certaines propriétés computationelles désirables fondées sur des axiomes intuitifs et rigoureux.

Si on considère $U, W \in \mathcal{L}^\infty$ où $\mathcal{L}^\infty$ est un espace de variables aléatoires bornées avec probabilité 1, alors on définit une mesure de risque comme une fonctionnelle $\rho : \mathcal{L}^\infty \to \mathbb{R}$ telle qu'une position aléatoire U est acceptable si $\rho(U) \leq 0$, c'est-à-dire si le risque $\rho(U)$ de la variable aléatoire U est non positif. On cherche à minimiser $\rho(U)$ et on peut donc interpréter U comme une perte ou une violation. Les mesures de risque cohérentes respectent les quatre axiomes suivant :

$$U \geq W \Rightarrow \rho(U) \geq \rho(W), \tag{2.25}$$

$$\rho(U + c) = \rho(U) + c, \forall c \in \mathbb{R}, \tag{2.26}$$

$$\rho(\lambda U + (1 - \lambda)W) \leq \lambda\rho(U) + (1 - \lambda)\rho(W), \forall \lambda \in [0, 1], \tag{2.27}$$

$$\rho(\lambda U) = \lambda\rho(U), \forall \lambda \geq 0, \tag{2.28}$$

où l'inégalité $U \geq W$ tient avec probabilité 1 et c et λ sont des constantes réelles.

9. Le paradoxe d'Ellsberg est illustré à l'aide d'une expérience assez simple. On considère une urne remplie d'un tiers de balles rouges et de deux tiers de balles jaunes et noires dont on ignore les proportions exactes. On considère ensuite 2 rondes où l'on propose chaque fois 2 paris à une personne. Au cours de la première ronde, on offre le choix entre A : 100 $ si une balle rouge est tirée et B : 100 $ si une balle noire est tirée. Pendant la deuxième ronde, on propose ensuite C : 100 $ si on tire une balle jaune ou rouge et D : 100 $ si on tire une balle jaune ou noire. selon la théorie de l'utilité espérée, une personne rationnelle qui préfère strictement (n'est pas indifférente) le choix A estime que la proportion de balles rouges est supérieure à celle de balles noires. Ainsi, pour être cohérente avec son premier choix, cette personne devrait également préférer strictement le choix C au choix D. Hors, en pratique, les participants préfèrent majoritairement le choix A par rapport au B ainsi que le choix D par rapport au C. Cette contradiction apparente découle possiblement du fait que l'on sait exactement que la probabilité de gagner les paris A et D est de $\frac{1}{3}$ et $\frac{2}{3}$, respectivement, alors que la probabilité de gagner les paris B et C se situe dans les intervalles $[0, \frac{2}{3}]$ et $[\frac{1}{3}, 1]$, respectivement. En d'autres termes, il existe de l'ambiguïté sur les paris B et C et la pire probabilité de B ou C est inférieure à la probabilité certaine du pari correspondant.

La propriété (2.25) de monotonicité signifie qu'une plus grande perte implique plus de risque. L'axiome (2.26) de transitivité implique que l'ajout d'une quantité déterministe à une position incertaine augmente exactement le risque de cette valeur. L'inégalité (2.27) assure la convexité de la mesure de risque, ce qui est justifié intuitivement par l'idée que la diversification ne devrait pas augmenter le risque. Finalement, la propriété (2.28) d'homogénéité assure que la multiplication d'une variable aléatoire par un facteur positif multiplie le risque exactement par ce facteur.

Les mesures de risque cohérentes considèrent implicitement l'aversion à l'ambiguïté mentionné précédemment et démontrée par plusieurs individus en pratique. En effet, il est possible de les exprimer sous la forme (Leitner, 2005; Föllmer and Knispel, 2013; Föllmer and Schied, 2011) :

$$\rho(X) = \sup_{P \in \mathcal{P}} E_P[X] \tag{2.29}$$

où $\mathcal{P}$ représente un certain ensemble de mesures de probabilité. Cette représentation permet de répondre au paradoxe d'Ellsberg, car elle assure que le risque d'une position incertaine correspond à la pire espérance sur un ensemble de mesures de probabilité.

Le chapitre 4 considère particulièrement le cas de la CVaR_α où $\alpha \in [0, 1]$ est un paramètre contrôlant l'aversion au risque. Cette mesure de risque possède de nombreuses propriétés intéressantes. Étant donné une mesure du probabilité de base P, elle peut notamment être formulée comme le problème d'optimisation suivant : $\text{CVaR}_\alpha(X) = \min_{t \in \mathbb{R}} t + \frac{1}{1-\alpha}\text{E}_P\left[\max\{X - t, 0\}\right]$ (Rockafellar and Uryasev, 2002). Dans le cas où il existe N scénarios possibles : $(\omega_1, \cdots, \omega_N)$ et que $P(\omega_i) = p_i$ et $X(\omega_i) = x_i$, alors on peut calculer $\text{CVaR}_\alpha(X)$ à l'aide du problème linéaire suivant et de son dual :

$$\begin{array}{lll|lll} \min\limits_{t,y} & t + \frac{1}{1-\alpha}\sum\limits_{i=1}^{N} y_i p_i & & \max\limits_{q} & \sum\limits_{i=1}^{N} \pi_i x_i & \\ \text{s. à :} & t + y_i \geq x_i, \quad \forall i = 1, ..., N & & \text{s. à :} & \sum\limits_{i=1}^{N} \pi_i = 1 & \\ & y_i \geq 0 \quad \forall i = 1, ..., N & & & 0 \leq \pi_i \leq p_i/(1-\alpha), & \forall i = 1, ..., N \end{array}$$

Ainsi, on obtient $\mathcal{P} \equiv \mathcal{P}_\alpha = \{\pi \in \mathbb{R}_+^N : \sum_{i=1}^N \pi_i x_i = 1 \ ; \ \pi_i \leq p_i/(1-\alpha), \quad \forall i = 1, \cdots, N\}$ dans la représentation (2.29).

On peut également écrire $\text{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 q_X(t)dt$ où $q_X(\alpha) = \inf\{t : F_X(t) \geq \alpha\}$ est la fonction de quantile évaluée en $\alpha \in (0,1)$. La fonction $q_X(\cdot)$ est la pseudo-inverse de la fonction de répartition $F_X(\cdot)$ où $F_X(x) = P(X \leq x)$ pour $x \in \mathbb{R}$ (Acerbi and Tasche, 2002). Dans le cas où F_X est continue, on obtient donc $\text{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\text{E}\left[X\mathbb{I}_{X>q_X(\alpha)}\right] = \text{E}\left[X|X > q_X(\alpha)\right]$

Cette interprétation à l'aide de l'espérance conditionnelle est intéressante, car elle indique intuitivement que l'on considère uniquement les pires $100(1-\alpha)$ % des scénarios. Par ailleurs, on peut démontrer que $\text{E}[X] \leq \text{CVaR}_\alpha(X) \leq \text{ess sup } X$ pour tout $\alpha \in (0,1)$ ainsi que $\lim_{\alpha\to 0}\text{CVaR}_\alpha(X) = \text{E}[X]$ et $\lim_{\alpha\to 1}\text{CVaR}_\alpha(X) = \text{ess sup } X$ où ess sup X indique la plus grande valeur (le supremum) que X peut prendre avec probabilité 1. Le paramètre α représente donc une manière simple et efficace de contrôler l'aversion au risque.

Les mesures de risque ont gagné en popularité au cours des dernières années. Bien qu'elles aient initialement été appliquées au domaine financier, notamment pour évaluer le risque de crédit et gérer des portefeuille (Acerbi and Tasche, 2002; Bertsimas et al., 2004), elles ont trouvé de nombreuses applications pratiques et théoriques en recherche opérationnelle (Chen et al., 2010; Natarajan et al., 2009; Brown and Bertsimas, 2009; Brown, 2007), plus spécifiquement en optimisation de la production énergétique (Gonçalves et al., 2013; Shapiro et al., 2013).

CHAPITRE 3 ORGANISATION DE LA THÈSE

Les trois prochains chapitres sont associés à trois article publiés ou soumis dans des revues scientifiques. Ces travaux considèrent tous le problème de gestion de réservoir court terme en présence d'incertitude. On s'intéresse plus particulièrement à la gestion mensuelle de la rivière Gatineau avec pas de temps journaliers. L'accent est mis sur les nombreuses contraintes opérationnelles spécifiques au bassin versant comme le risque d'inondations à différents segments de la rivière.

Les modèles et les algorithmes ont été développés dans le but de produire un outil d'aide à la décision permettant de faciliter le travail des gestionnaires de la rivière. À cet effet, tous les modèles sont relativement faciles à implémenter, permettent d'obtenir de bonnes solutions très rapidement et peuvent être utilisés dans un contexte de simulation pour procéder à des études de cas.

Le chapitre 4 s'intéresse à la minimisation du risque d'inondations. Après avoir introduit le problème de gestion de réservoir, on propose une modélisation du problème déterministe. On introduit ensuite les règles de décision linéaires et les règles liftées. Ceci nous permet de formuler l'équivalent déterministe comme un problème linéaire qu'on résout avec CPLEX.

Le chapitre présente également une étude numérique assez exhaustive basée sur diverses simulations et règles de gestion. On s'intéresse notamment à la résolution d'un modèle déterministe en horizon roulant, un modèle de programmation stochastique sur arbre construit avec SCENRED2, quelques variantes des règles de décision affines ainsi que les règles de décision liftées. Le chapitre conclut qu'en général, les règles de décision affines qui considèrent uniquement l'incertitude sur un horizon limité offrent les meilleurs compromis entre temps de calcul et performance.

On y étudie aussi la performance de quelques représentations polyhédrales utilisées couramment dans la littérature sous l'influence de différents processus hydrologiques. On explique notamment qu'en présence de forte corrélation temporelle, il peut être néfaste d'utiliser un ensemble basé sur un budget d'incertitude (Sim and Bertsimas, 2004) trop petit.

Finalement, cet article analyse l'impact de l'utilisation de la mesure de risque CVaR_α avec différentes tolérances $\alpha \in [0, 1]$. Les résultats de simulation n'apportent pas de conclusion probante applicable à tous les modèles. Ceci est peut-être dû au fait que les simulations considèrent des réalisations de l'incertitude plus réalistes que le modèle d'optimisation. Cette simulation hors-échantillon mène à des conditions pour lesquelles le modèle d'optimisation

n'a pas été calibré.

Le chapitre 5 jette les bases de l'utilisation de séries chronologiques de type ARIMA pour des problèmes stochastiques multi-étapes avec règles de décision. On considère initialement un processus ARMA(1,1) stationnaire où les chocs sont indépendants, mais on étend les travaux au cas où ils sont uniquement non corrélés et suivent un processus GARCH(1,1). On suppose ici que les apports sont une fonction linéaire des chocs et que ces derniers appartiennent à un polyhèdre motivé par l'inégalité de Chebychev.

Contrairement au chapitre 4, on minimise une fonction quadratique convexe des inondations afin de refléter les conséquences marginales croissantes des inondations. Ceci mène à un problème quadratique où l'objectif varie dans le temps selon les réalisations passées lorsqu'on ajoute l'hétéroscédasticité.

Bien que les résultats de simulation soient prometteurs, cette section ne semble pas indiquer un avantage important d'une modélisation basée sur les processus ARMA par rapport à une représentation ignorant la corrélation temporelle similaire à celle utilisée dans le chapitre 4 dans le cas des apports historiques. Effectivement, les modèles GARCH(1,1) + ARMA(1,1), ARMA(1,1) ainsi que le modèle naïf donnent des performances semblables en termes de minimisation des inondations. Cette faiblesse est attribuable au choix de modèle ARMA linéaire qui viole certaines propriétés de base comme la non-négativité des apports en plus d'offrir un potentiel prédictif faible.

Contrairement aux deux premiers chapitres 4 et 5, le chapitre 6 considère également la valorisation du stock d'eau, la production hydroélectrique ainsi que le risque d'inondations. On considère une combinaison convexe des différents objectifs et on explore l'espace des solutions selon différents scénarios hydrologiques.

Afin de modéliser réalistement la production hydroélectrique, on considère une hauteur de chute variable, ce qui mène à un problème non convexe qu'on résout approximativement à l'aide d'un algorithme de région de confiance très simple, mais efficace.

Ce chapitre présente par ailleurs un processus ARMA qui lie les apports aux chocs de façon non linéaire. Cette représentation plus réaliste mène à un ensemble d'incertitude non convexe, qu'on choisit d'approximer de façon conservatrice par un polyhèdre. Cette forme permet également de dériver une expression analytique pour l'espérance conditionnelle.

La section 8 effectue un retour sommaire sur l'ensemble du travail réalisé et propose quelques pistes d'améliorations et de réflexions. On évalue notamment les différentes façons d'améliorer la modélisation de la rivière et des apports.

CHAPITRE 4 ARTICLE 1: DECISION RULE APPROXIMATIONS FOR THE RISK AVERSE RESERVOIR MANGEMENT PROBLEM

Cet article a été publié : Gauvin, C., Delage, E., and Gendreau, M. (2017). Decision rule approximations for the risk averse reservoir management problem. *European Journal of Operational Research*, 261 :317–336. (Gauvin et al., 2017). [1]Auteurs : Charles Gauvin, Erick Delage et Michel Gendreau.

Abstract : This paper presents a new formulation for the risk averse stochastic reservoir management problem. Using recent advances in robust optimization and stochastic programming, we propose a multi-stage model based on minimization of a risk measure associated with floods and droughts for a hydro-electrical complex. We present our model and then identify approximate solutions using standard affine decision rules commonly found in the literature as well as lifted decision rules. Finally, we conduct thorough numerical experiments based on a real river system in Western Québec and conclude on the relative performance of families of decision rules.

Keywords : Stochastic programming, Robust optimization, Risk analysis, OR in energy

4.1 Introduction

The problem of designing an optimal release schedule for a set of interconnected reservoirs is extremely challenging. Operators must often make decisions for various dams or other water control structures in each period of a given time horizon while taking into account the dynamical structure and topology of the system. It may also be important to consider complex non-linear physical phenomenon such as the effect of water volumes on outflow or water delays. This generally leads to high dimensional dynamic and non-convex problems.

Furthermore, operators must deal with various conflicting objectives of varying importance. Hydro-electrical complexes must namely balance criteria such as recreational and environmental needs with electricity generation, irrigation and flood control (Labadie, 2004). The

1. L'écriture de ce chapitre a été légèrement modifiée par rapport au rapport technique afin de standardiser la notation pour l'ensemble de la thèse. Après avoir reçu les rapports d'arbitres pour le deuxième article, nous avons effectué plusieurs changements de notation pour se rapprocher de l'écriture standard en gestion de ressources hydriques. Ces changements ne nous avaient pas été demandés pour le premier manuscrit soumis précédemment à *European Journal of Operational Research.*

importance of considering multiple criteria in stochastic reservoir management problems is namely illustrated in Hajkowicz and Higgins (2008); Tilmant and Kelman (2007); Castelletti et al. (2010).

Considerable uncertainty surrounds various factors such as price of electricity, turbine breakage, inflows and demand. It is virtually impossible to perfectly identify and represent the true multi-dimensional stochastic process. However, failure to take into account this stochasticity can lead not only to severely suboptimal solutions, but even disastrous consequences such as floods or droughts. Incorporating uncertainty also poses serious numerical limitations as this leads to multi-stage stochastic programs, which are generally intractable (see Dyer and Stougie (2006)).

There is a need to develop scalable and flexible operational tools that can take this uncertainty into account. In Québec for instance, current operating practices for short-term planning include use of simplified linear programming models or more complex heuristics. Although both approaches evaluate policies on extensive test-beds in order to assess their performance under various inflow scenarios, they do not consider stochasticity explicitly.

Existing methods capable of handling decision under uncertainty suffer from various limitations. Stochastic dynamic programming (SDP), which has historically been one of the most popular algorithms in academia and in practice for reservoir management (Saad and Turgeon, 1988; Saad et al., 1996; Turgeon and Charbonneau, 1998; Stedinger and Faber, 2001), can typically only be applied for problems that ignore water flow delays and consider a limited number of reservoirs, typically 1 or 2. In contrast, our case study considers 5 reservoirs and would require using a state of dimension 18 when considering the sum of past releases over all possible water delays. The discretization required by SDP may also lead to approximations and numerical difficulties. Finally, the algorithms often assume normally or log-normally distributed random variables (Turgeon, 2005), which is not verified empirically for daily inflows.

In the last decades, the stochastic dual dynamic programming algorithm (SDDP) (Pereira and Pinto, 1991; Tilmant and Kelman, 2007; Rougé and Tilmant, 2016) has emerged as viable alternative to traditional SDP to tackle multidimensional problems. This algorithm consists of a backward recursive phase where the future value function is approximated with Benders cuts though the solution of a dual problem and a forward sampling and simulation phase. This alleviates the curses of dimensionality associated to the space of controls and is therefore useful for large reservoir systems. However, the algorithm does not provide implementable policies as would an SDP algorithm, can display relatively slow convergence and relies on a fixed sampling distribution such as the log-normal Shapiro et al. (2013).

Stochastic programming models based on scenario trees have also attracted attention (Gon-

çalves et al., 2012; Carpentier et al., 2013a; Aasgård et al., 2014). This tree representation is useful to model a wide range of distributions while capturing complex correlations of the multi-dimensional processes. However, it also fails to provide an implementable policy and generally leads to very large deterministic equivalent programs. In order to mitigate this computational issue, various tree-reduction approaches based on Heitsch and Römisch (2009) have been used. Although useful, the loss of information caused by the aggregation may fail to capture more extreme scenarios, which may be essential for flood management purposes. Other scenario aggregation and partitioning schemes (Carpentier et al., 2013b) as well as the progressive hedging algorithm (Rockafellar and Wets, 1991) have also been investigated. Nonetheless, convergence may be long to reach for large problems and these algorithm display sensitivity to parameters and scenario representation (Carpentier et al., 2013a; Gonçalves et al., 2012).

This paper addresses these issues by proposing a multi-stage multi-reservoir model based on robust optimization. Robust optimization is a rapidly expanding paradigm that has gained in popularity over the years (Bertsimas et al., 2011; Ben-Tal et al., 2009). By restricting uncertainty to reside within a given parametrized and deterministic convex set, large stochastic models can be converted into their robust counterparts while maintaining tractability through conic programming duality.

The robust optimization paradigm has been extended to dynamic problems with affine decision rules (Ben-Tal et al., 2004) as well as more sophisticated decision rules by various authors to provide more flexibility and precision (Chen et al., 2008; Goh and Sim, 2010; Georghiou et al., 2015). Some of these robust optimization methods have recently been used for reservoir management problems. The authors of Apparigliato (2008) and Pan et al. (2015) namely use this framework to derive operating policies to maximize the expected electric production for a multi-period and multi-reservoir hydro-electric complex. Both works consider low-dimensional systems of up to 3 reservoirs. They focus on systems where the risk of floods are relatively modest and therefore only consider bounds on total releases and volumes.

This paper differs significantly from the previous work in that it considers more realistic operating conditions. Our model explicitly considers water delays, non-linear functions related to the physical structure of the evacuation structures and flow variation constraints. Moreover, we focus on a difficult problem instance without complete recourse with tight constraints that may be significantly violated under certain inflow scenarios.

For this reason, we formulate a novel multi-dimensional problem that minimizes the risk of floods at various critical locations on the river. Unlike Pan et al. (2015), we do not use over-the-shelf software like ROME (Robust Optimization Made Easy) (Goh and Sim, 2011),

because of the computational overhead and because our approach can be tailored to our specific needs. This is namely important because the built-in lifted decision rules cannot be adapted to the ones used in this paper to the best of our knowledge.

Unlike competing methods like SDP, SDDP and stochastic programs based on scenario trees, robust optimization reduces the need for distributional assumptions. In particular, it does not require fixing a pre-specified distribution from which it is possible to sample. As observed by Pan et al. (2015), this increases the resilience of the solutions when tested on out-of-sample data while providing similar performance for in-sample data.

Although these decision rules provide a conservative approximation of the objective value, they remain tractable and provide rules that can easily be used in simulations, which is not necessarily the case of SDDP or tree-based stochastic programs, except when used in a rolling horizon fashion. The ability to simulate river management policies is of prime importance for practical reservoir management. Difficulty to evaluate solutions based on multi-stage decision trees has namely impeded their implementation in practice.

Contrary to an overwhelming majority of models mentioned previously, including Apparigliato (2008) and Pan et al. (2015), we do not assume risk neutrality. We explicitly consider risk aversion and perform sensitivity analysis on the α parameter of the coherent risk measure CVaR_α. This allows us to easily interpret the risk of these natural disasters and perform sensitivity analysis.

Like Pan et al. (2015), we evaluate the use of linear and piecewise affine decision rules. Our work draws heavily from both state-of-the art techniques in robust optimization and the work of Georghiou et al. (2015) on stochastic programming. We conduct extensive numerical experiments based on a real river system to evaluate and compare these different policies and also consider a rolling horizon approach embedded in a realistic simulation environment.

The rest of the paper is structured as follows. We introduce our deterministic model and the objective in Section 4.2. Section 4.3 focuses on the uncertainty surrounding inflows while section 4.4 discusses risk analysis. Section 4.5 discusses affine and piecewise-affine decision rules. We then present a case study based on a real hydroelectric reservoir management problem in Québec in Section 4.6 and offer concluding remarks in Section 4.7.

4.2 Description of the deterministic reservoir management problem (RMP)

4.2.1 Basic model

We consider a discrete time horizon of T periods with decisions at each time $t \in \mathbb{T} = \{1, \cdots, T\}$. If we completely omit all sources of uncertainty, the reservoir management problem we study consists in finding a feasible solution to the following program :

$$\text{(Vol. Bounds)} \quad \underline{s}_j \leq \mathcal{S}_{jt} \leq \bar{s}_j \qquad \forall j \in J, t \in \mathbb{T} \tag{4.1}$$

$$\text{(Flow cons.)} \quad \mathcal{S}_{jt} = s_{j0} + \sum_{\tau \leq t} \left(\sum_{i^- \in I^-(j)} \sum_{l=\delta_{i^-}^{min}}^{\min\{\delta_{i^-}^{max}, \tau-1\}} \lambda_{i^-l} \mathcal{F}_{i^-\tau-l} - \sum_{i^+ \in I^+(j)} \mathcal{F}_{i^+\tau} + \xi_{j\tau} \right) \beta \quad \forall j \in J, t \in \mathbb{T} \tag{4.2}$$

$$\text{(Flow Bounds)} \quad \underline{f}_i \leq \mathcal{F}_{it} \leq \bar{f}_i, \qquad \forall i \in I, t \in \mathbb{T} \tag{4.3}$$

$$\text{(Evac. Curve)} \quad \mathcal{L}_{it} \leq \mathscr{C}_i(\mathcal{S}_{j^-(i)t}), \qquad \forall i \in I^{evac}, t \in \mathbb{T} \tag{4.4}$$

$$\text{(Var. Bound)} \quad |\mathcal{F}_{it} - \mathcal{F}_{it-1}| \leq \bar{\Delta}_i \qquad \forall i \in I, t \in \{2, \cdots, T\} \tag{4.5}$$

$$\text{(Spill. Bounds)} \quad \underline{l}_i \leq \mathcal{L}_{it} \leq \bar{l}_i \qquad \forall i \in I, t \in \mathbb{T} \tag{4.6}$$

$$\text{(Turb. Bounds)} \quad \underline{r}_i \leq \mathcal{R}_{it} \leq \bar{r}_i \qquad \forall i \in I, t \in \mathbb{T} \tag{4.7}$$

$$\text{(Flow def.)} \quad \mathcal{F}_{it} = \mathcal{R}_{it} + \mathcal{S}_{it} \qquad \forall i \in I, t \in \mathbb{T} \tag{4.8}$$

In this formulation, the decision variables $\mathcal{S}_t, \mathcal{L}_t, \mathcal{R}_t, \mathcal{F}_t$, respectively represent average volumes (hm^3) during time t, average spillage (non-productive water discharge) (m^3/s) over time t, average turbined outflow (productive water discharge) (m^3/s) over time t, and average total flow (m^3/s) over time t. The sum of the spillage and the turbined outflow is equal to the total flow.

We distinguish between the spillage and turbined controls that represent real implementable decisions and the volume and total flow that are analysis variables [2] used to enhance model readability (refer to Table 4.1) .

We let J represent the set of reservoirs ; some of which are simply portions of the river with little or no capacity. The set I represents the water controlling sites. Not all sites have plants that may produce electricity by letting water flow though turbines. The sets $I^-(j), I^+(j) \subset I$

2. We define implementability as the property that decisions at a given time can be realistically controlled by human intervention and depend only on the available information. On the other hand, analysis variables are completely determined uniquely once the implementable decisions are fixed. They only serve to track the evolution of the system.

Tableau 4.1 Decision variables.

Notation	Name	Implementable	Unit
$\mathcal{S}_t$	Volume	No	(hm^3)
$\mathcal{L}_t$	Spillage	Yes	(m^3/s)
$\mathcal{R}_t$	Turbined outflow	Yes	(m^3/s)
$\mathcal{F}_t$	Total flow	No	(m^3/s)

represent the set of upstream (incoming) and downstream (outgoing) plants with respect to reservoir j.

For each reservoir j, lower volume bounds ($\underline{s}_j$) represents a minimum amount of water required namely for health, environmental and recreational purposes. These bounds are particularly important during the hot summer months when there are less precipitations and more recreational activities on the water. They are often enforced through contracts and agreements with local communities. The bounds may be fixed or variable over the entire year, but it is reasonable to keep them fixed since the time horizon is sufficiently small.

Upper bounds ($\bar{s}_j$) represent hard maximum levels and are based on the physical capacity of each river segment or reservoir. Those of reservoirs close to human habitations that display a high likelihood of flooding are particularly important. These bounds are also critical for large "head" reservoirs, since operators will lose the capacity to manage the river adequately if the volumes exceed these thresholds.

Flow conservation constraints ensure conservation of water across the system. Since plants are located at different distances from one another on heterogeneous terrain, a drop of water takes a different amount of time to flow down to the next reservoir depending on the plant from which it was released. This phenomenon is modelled with the parameter $\lambda_{i^-(j)l}$, which represents the fraction of the total water discharged from plant $i^-(j)$ reaching the unique downstream reservoir j after l time steps. $\delta^{min}_{i^-(j)}$ and $\delta^{max}_{i^-(j)}$ represent the minimum and maximum number of days required to transport one hm^3 of water from plant $i^-(j)$ to the downstream reservoir. The parameter ξ_{jt} represents the inflows in reservoir j at time t due to precipitations, snow melt and natural water flow.

Operators realize the importance of considering these delays and planning several weeks or months ahead, even in the case of certain inflows. This is an important advantage over stochastic dynamic programming (SDP), which is one of the most popular frameworks used to solve stochastic reservoir management problem (see Labadie (2004)).

The constant β allows conversion of m^3/seconds to hm^3/days while s_{j0} represents the fixed known amount of water (in hm^3) at the beginning of the time horizon in reservoir j. We

reasonably neglect evaporation and other water losses by assuming a relatively cold and humid climate, small water surface area, relatively high pressure and non-porous soil.

Minimum flow bounds $\underline{f}_i$ represent the smallest quantity of water required to flow down site i to produce electricity, for environmental purposes, as well as to ensure navigation and recreation. The upper bound $\bar{f}_i$ represents a critical threshold based on historical observations and used for safety and operational purposes. For instance, operators may prefer to maintain relatively small water flows during the winter season to prevent ice coming into contact with the installations at high speed and to ensure a smooth ice formation.

Evacuation curves $\mathscr{C}_i(\cdot)$ represent the maximal amount of water $\mathscr{C}_i(s)$ that can be physically discharged (spilled) from the evacuator at site $i \in I^{evac}$ for a given amount of (average) volume s during any given time t at the unique upstream reservoir where $I^{evac} \subset I$ represents the set of plants with such constraints. These functions are typically smooth, increasing, non-linear and non-convex. We choose to approximate them with affine functions as this provides a reasonable approximation while ensuring the linear structure of theses constraints.

Variation bounds Δ_i are added to ensure that the total flow of water at each plant i stays relatively constant from one time period to the next. Maintaining a relatively constant water flow is important to prevent turbine breakage as well as to ensure navigation.

4.2.2 Considering floods

Under very wet scenarios and given initial conditions, the tight operational constraints on flows, volumes, spillage and turbined outflow cannot be respected. Even with perfect foresight, numerical experimentation on a real river basin reveals that the wait-and-see solution may be infeasible if the initial volumes are sufficiently high (see section 4.6.1). Hence, we modify the constraints of the original problem to allow controlled violations of volume constraints :

$$\text{(Vol. Bounds}') \qquad \underline{s}_j \leq \mathcal{S}_{jt} - \mathcal{E}_{jt} \leq \bar{s}_j \quad \forall j \in J, t \in \mathbb{T} \tag{4.9}$$

$$\text{(Floods)} \qquad 0 \leq \mathcal{E}_{jt} \quad \forall j \in J, t \in \mathbb{T} \tag{4.10}$$

The overflows/floods $\mathcal{E}_{jt}$ represent the quantity of water (hm^3) exceeding the maximum volume thresholds. It is reasonably assumed that excess water will remain in the reservoirs and may eventually be released downstream. It is straightforward to extend this to droughts as well. However, for our particular application, violating lower volume bounds is highly undesirable and only very limited violations are physically possible. We therefore omit them in

our formulation.

4.2.3 Evaluating floods over the entire horizon for the whole hydro electrical complex

We seek a solution that minimizes these floods over all reservoirs and time periods. These objectives may be conflicting as variations in volumes upstream may have important cascading effects on downstream reservoirs. This approach is tantamount to numerous hydro electrical complexes in the presence of nearby riparian human populations and infrastructure as well as fauna and flora.

It is also crucial to consider the relative importance of floods as they may have widely varying consequences depending on the segments of the river where they occur, the time of the year and their magnitude. We tackle this problem by considering the positive linear combinations of floods $\sum_{j\in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}_{jt})$ with parameters $\kappa_{jt} > 0$, $\forall j, t$ and $\mathcal{E} = (\mathcal{E}_{11}, \cdots, \mathcal{E}_{|J|T})^\top \in \mathbb{R}^{|J|T}$.

To reflect the importance of reservoirs located near riparian populations and high risks of floods as well as those with critical importance, we define the set J^{crit} and fix $\kappa_{jt} = W\kappa$, $\forall j \in J^{crit}, t$ for some large $W \in \mathbb{N}$ and $\kappa \in \mathbb{R}$. Since the relative importance of upper volume bounds at each of the reservoirs in $J \setminus J^{crit}$ is similar, we fix $\kappa_{jt} = \kappa$, $\forall j \in J \setminus J^{crit}, t$. For our specific application, this categorization is clearly defined, but identifying the relative importance of violations may require using preference assessment tools such as MACBETH (Bana e Costa and Vansnick, 1997). We choose κ such that the sum of weights is equal to 1 and keep it constant with respect to time since we consider a small time horizon.

Although we use the simple linear penalization term $\kappa_{jt}\mathcal{E}_{jt}$ for fixed j and t, it would be straightforward to consider convex piecewise linear functions. It would also be possible to use other convex functions such as convex quadratic functions and then obtain conservative polyhedral approximations as we do in Section 4.6.1 for the evacuation curves. Considering more general non-convex functions such as the ones used in Pianosi and Soncini-Sessa (2009) could perhaps be done through non-convex piecewise linear functions, but would require using a mixed-integer program, which would significantly increase the computational requirements compared to our proposed approach and reduce its practical applicability for river management. After discussions with Hydro-Québec, the simple linear penalization function was deemed appropriate for our needs.

4.3 Incorporating inflow uncertainty

In most river systems, there is a considerable amount of uncertainty surrounding inflows. This randomness is one of the main driving forces in the variability of the state of the system. For these reasons, we let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbb{P})$ be a filtered probability space where $\xi_t : \Omega \to \mathbb{R}$, $\forall t \in \{1, \cdots, T\}$ represents the *total inflows* (in m^3/s) for the river basin at time t, $\mathscr{F}$ is a σ-algebra and $\{\mathscr{F}_t\}$ is the natural filtration induced by the stochastic process $\{\xi_t\}_{t=1,\cdots,T}$ where $\mathscr{F}_t = \sigma(\xi_s, s < t) \subset \mathscr{F} = \mathscr{F}_T$. The random vector $\xi_{[t]} = (\xi_1, \cdots, \xi_t)^\top$ represents the past inflows up time t and $\xi \equiv \xi_{[T]}$.

We denote E the expectation operator and assume that $E[\xi_t] = \mu_t < \infty$ and $E[(\xi_t - \mu_t)^2]^{\frac{1}{2}} = \sigma_t < \infty$ for all t. These moments are estimated with their corresponding non-parametric sample statistics. For the random variable $\xi_t : \Omega \to \mathbb{R}$, we denote $\operatorname{ess\,sup} \xi_t = \inf\{a \in \mathbb{R} : \mathbb{P}(\xi_t > a) = 0\}$. We assume $\zeta_t = \xi_t - \mu_t$ is essentially bounded and that the compact and convex set $\mathscr{Z}_\nu \subset \mathbb{R}^T$ represents the true support of the probability measure $P_\zeta = \mathbb{P} \circ \zeta^{-1}$ with fixed parameter $0 \leq \nu < \infty$. More precisely, we assume $\mathscr{Z}_\nu = \{\zeta \in \mathbb{R}^T : \zeta_t \in [-\min\{\nu\sigma_t, \mu_t\}, \nu\sigma_t], \forall t\}$ since $\xi_t = \zeta_t + \mu_t \geq 0$ with P_ζ a.s. $\forall t$.

We choose this uncertainty set because of its simplicity and numerical efficiency. It also allows us to obtain simple probabilistic bounds on individual random variables. Indeed, for any fixed t, Tchebyshev's inequality yields $P(\zeta_t \in [-\min\{\nu\sigma_t, \mu_t\}, \nu\sigma_t]) = P(|\zeta_t| \leq \nu\sigma_t) \geq 1 - \nu^{-2}$. In any case, this can be considered a basic assumption of our model. Nonetheless, we tested the performance of our method with the budgeted uncertainty and present some results in Sections 4.6.3 and 4.6.6.

We assume a relatively small river basin where inflows at each reservoir are perfectly correlated. Although our method can easily consider the case of independent or partially spatially correlated inflows, this would increase the computational burden and complicate the exposition of the problem. In the case of perfectly spatially correlated inflows, we can represent ξ_{jt}, the total water inflow in m^3/s from natural precipitations and spring thaw going in reservoir j at time t as $\xi_{jt}(\zeta) = \mu_{jt} + \zeta_t \alpha_{jt}$ where $\alpha_{jt} = \mu_{jt}(\sum_j \mu_{jt})^{-1}$ is the expected proportion of the total inflows entering the reservoir at that time.

4.4 Risk analysis

4.4.1 Conditional value-at-risk

To evaluate the riskiness of floods over all reservoirs and time periods, we evaluate the conditional value at risk of the total weighted floods over the entire horizon for the hydro-

electrical complex that was presented earlier in Section 4.2.3 for a fixed $\alpha \in [0,1]$:

$$\mathrm{CVaR}_\alpha(\sum_{j\in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}_{jt})) \tag{4.11}$$

The conditional value at risk is a coherent risk measure (Artzner et al., 1999; Rockafellar and Uryasev, 2002; Acerbi and Tasche, 2002). If $\mathcal{U}$ is a continuous random variable bounded with probability 1, then $\mathrm{CVaR}_\alpha(\mathcal{U}) = E[\mathcal{U}|\mathcal{U} > \mathrm{VaR}_\alpha(\mathcal{U})]$ holds for $\alpha \in (0,1)$. In other words, $\mathrm{CVaR}_\alpha(\mathcal{U})$ represents the conditional expectation of $\mathcal{U}$ given that it is greater than the $(1-\alpha)$ left quantile $\mathrm{VaR}_\alpha(\mathcal{U}) = \inf\{u \in \mathbb{R} : \mathbb{P}(\mathcal{U} \leq u) \geq 1-\alpha\}$ for a given $\alpha \in (0,1)$. We can thus interpret the risk of $\mathcal{U}$, as the average of the worst $\alpha 100\%$ of values for some $\alpha \in (0,1)$(see Acerbi and Tasche (2002)). As observed in financial applications, CVaR_α offers the important property of considering all the realizations of a random variable in the "worst" α-tail of the distribution. As $\alpha \to 1$, we consider more realizations, but reduce the weight of extremely bad realizations. For a continuous random variable $\mathcal{U}$, we also have :

$$\lim_{\alpha\to 1}\mathrm{CVaR}_\alpha(\mathcal{U}) = E[\mathcal{U}] \qquad \text{and} \qquad \lim_{\alpha\to 0}\mathrm{CVaR}_\alpha(\mathcal{U}) = \text{ess sup}\,\mathcal{U} \tag{4.12}$$

that correspond to the natural risk neutral approach and to the totally risk averse approach used in robust optimization. Both E and ess sup also correspond to coherent risk measures. Furthermore, we have the relationship $E[\mathcal{U}] \leq \mathrm{CVaR}_\alpha(\mathcal{U}) \leq \text{ess sup}\,\mathcal{U}$, $\forall \alpha \in (0,1)$.

The conditional value-at-risk can also be written as the following optimization problem : $\mathrm{CVaR}_\alpha(\mathcal{U}) = \inf_{t\in\mathbb{R}}\{t + \alpha^{-1}E[\max\{\mathcal{U}-t, 0\}]\}$. CVaR_α therefore represents a very natural and flexible risk measure and offers the possibility to explore the impact of various risk tolerance and model conservativeness through sensitivity analysis on the α parameter.

Recent studies based on coherent risk measures and other statistical functionals have shown the value of incorporating risk aversion into the decision process (Shapiro et al., 2013; Babaei et al., 2015; Artus et al., 2006). Like some of these works, we highlight the fact that risk aversion can be introduced at little or no computational cost compared with the traditional risk neutrality.

4.4.2 Dynamic risk measures and time consistency

Our optimization model evaluates the risk measure CVaR_α at the beginning of the entire horizon once in a static fashion. For $\alpha \in (0,1)$, this approach may lead to solutions that violate the concept of time-consistency. This notion is related to Bellman's principle and can loosely be interpreted as the requirement that optimal decisions taken at a given time for a fixed horizon should remain optimal when the problem is solved again at any other later date (see Acciacio and Penner (2011) for more details on dynamic risk measures). Although this issue is important in a dynamic decision making environment, we omit to address it in our model because there has been little work on tractable representations of time-consistent stochastic problems using dynamic risk measures except in simple cases such as serially independent random variables when considering the risk neutral expectation (see Shapiro (2009)).

4.5 Stochastic model and choice of decision rules

We consider a dynamic setting were the true realization of the random process is progressively revealed as time unfolds over the horizon of T days. A sequence of controls $\mathcal{X}_1, \cdots, \mathcal{X}_t$ must be fixed at each stage $t = 1, \cdots, T$ after observing the realized history $\zeta_{[t-1]}(\omega) = (\zeta_1, \cdots, \zeta_{t-1})^\top(\omega)$, but before knowing the future values $(\zeta_t, \cdots, \zeta_T)^\top(\omega)$. At each time $t = 1, \cdots, T$, we optimize over the class of bounded functions of the form $\mathcal{X}_t : \mathbb{R}^T \to \mathbb{R}^{n_t}$, $\sum_{t=0}^{T-1} n_t = n$. Hence, we convert the decision variables of our deterministic model [3]into more flexible functions.

Real implementable decisions must be non-anticipative such that $\mathcal{X}_t \circ \zeta$ must be $\mathscr{F}_{t-1}$ measurable. Once $\zeta_{[t-1]}(\omega)$ is known, they can be implemented to yield the actual decisions $\mathcal{X}_t(\zeta_{[t-1]}(\omega)) \in \mathbb{R}^{n_t}$. We may interchangeably refer to $\mathcal{X}$ as controls or decision rules and use the notation $\mathcal{X} = (\mathcal{X}_1^\top, \cdots, \mathcal{X}_T^\top)^\top$. The aggregate decision rule $\mathcal{X}_t = (\mathcal{S}_t^\top, \mathcal{L}_t^\top, \mathcal{R}_t^\top, \mathcal{F}_t^\top, \mathcal{E}_t^\top)^\top$ is simply the stacking of each decision at time t. Refer to Table 4.1 for identification of implementable decisions.

4.5.1 Affine decision rules

In order to obtain an upper bound on the optimal value of our stochastic programming problem, we first consider decision rules that are affine in ζ. If $\mathscr{K}_t = \{n_{t-1}+1, \cdots, n_{t-1}+n_t\}$ represents the indices associated with decisions at time $t \geq 1$ and $n_0 = 0$, we can write the controls in the form :

3. We can see the decision variables of our deterministic model as constant functions of the uncertainty.

$$\mathcal{X}_{kt}(\zeta) = \mathcal{X}^0_{kt} + \sum_{t'=1}^{T} \mathcal{X}^{t'}_{kt}\zeta_{t'}, \quad k \in \mathscr{K}_t \tag{4.13}$$

where $\mathcal{X}^0_{kt}, \mathcal{X}^{t'}_{kt} \in \mathbb{R}$, $\forall k \in \mathscr{K}_t, t' \in \{1, \cdots, T\}$. In order to ensure that controls that represent real implementable decisions are non-anticipative, it is sufficient to force $\mathcal{X}^{t'}_{kt} = 0, \forall k \in \mathscr{K}^I_t, t = 1, \cdots, T, t' \geq t$ where $\mathscr{K}^I_t \subset \mathscr{K}_t$ represents the set of indices associated with real implementable decisions.

The volumes, droughts and floods can also be anticipative, since they represent analysis/auxiliary variables that are only required for computation (refer to Table 4.1). Because of the full-dimensionality of $\mathcal{Z}_\nu$ and the additive uncertainty, flow conservation constraints (4.2) are satisfied by solving a non-homogeneous linear system of equations in $\mathcal{X}^{t'}_{kt}, \mathcal{X}^0_{kt}$ (see Lorca et al. (2016) for example and Section 4.9.3 of the Appendix for a detailed formulation).

We reformulate the variation of flow constraints across time by writing :

$$(\mathcal{F}_{it} - \mathcal{F}_{it-1})(\zeta) \leq \bar{\Delta}_i, \qquad \forall i, t = 2, \cdots, T \tag{4.14}$$

$$-(\mathcal{F}_{it} - \mathcal{F}_{it-1})(\zeta) \leq \bar{\Delta}_i, \qquad \forall i, t = 2, \cdots, T \tag{4.15}$$

where all constraints must hold $\forall \zeta \in \mathcal{Z}_\nu$. As mentioned previously, we consider affine approximations of the real maximum spillage constraints : $\mathscr{C}_i : v \to \mathscr{C}^0_i + \mathscr{C}^\Delta_i v$ with $\mathscr{C}^0_i, \mathscr{C}^\Delta_i \in \mathbb{R}$ for all plants $i \in I^{evac}$. Hence if $j^-(i)$ indicates the unique reservoir upstream of plant i, the following evacuation curve constraints must hold $\forall \zeta \in \mathcal{Z}_\nu$:

$$\mathcal{L}_{it}(\zeta) \leq \mathscr{C}^0_i + \mathscr{C}^\Delta_i \mathcal{S}_{j^-(i)t}(\zeta), \qquad \forall i \in I^{evac}, t \tag{4.16}$$

4.5.2 Objective function formulation for robust problem

To express our objective function, we exploit the formulation of CVaR_α as an optimization problem. Indeed, for $\alpha \in (0, 1)$ and random variable $X \in \mathcal{L}^\infty$, we have $\text{CVaR}_\alpha(X) = \inf_{t \in \mathbb{R}}\{t + \frac{1}{\alpha}E[\max\{X - t, 0\}]\}$. This corresponds to the definition of Pflug (2000), where we have replaced $1 - \alpha$ by α so that the CVaR_α we consider is non-increasing in α.

To obtain a conservative (upper bound) on $\text{CVaR}_\alpha(\sum_{j \in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}^+_{jt}(\zeta)))$, we consider the anticipative decision rule $\mathcal{D}(\zeta) = \mathcal{D}^0 + \sum_{t'=1}^{T}\mathcal{D}^{t'}\zeta_{t'}$ where $E[\mathcal{D}(\zeta)] = \mathcal{D}^0$ since we consider

affine decision rules and $E[\zeta_t] = 0, \forall t$ without loss of generality. We then add the decision variables $\varphi, t \in \mathbb{R}$ as well as the following constraints :

$$t + \frac{1}{\alpha}\mathcal{D}^0 \leq \varphi \tag{4.17}$$

$$\mathcal{D}(\zeta) \geq 0 \tag{4.18}$$

$$\mathcal{D}(\zeta) \geq \sum_{j\in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}_{jt}(\zeta)) - t \tag{4.19}$$

that must hold $\forall \zeta \in \mathcal{Z}_\nu$. We finally set φ as the objective function to minimize. Constraints (4.18)-(4.19) model $\mathcal{D}(\zeta) \geq \max\{\sum_{j\in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}_{jt}(\zeta)) - t, 0\}, \forall \zeta \in \mathcal{Z}_\nu$.

The case $\alpha = 0$ is handled separately by adding the variable φ and the constraint $\varphi \geq \sum_{j\in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}_{jt}(\zeta)), \forall \zeta \in \mathcal{Z}_\nu$. Hence minimizing φ yields ess $\sup(\sum_j \sum_{t=1}^{T} \mathcal{E}_{jt}(\zeta)\kappa_{jt}) = \max_{\zeta\in\mathcal{Z}_\nu} \sum_{j\in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}_{jt}(\zeta))$, which is the worst case if we only consider support information and corresponds to the classical robust objective.

The case $\alpha = 1$ (expected value) is treated by adding φ and $\varphi \geq \sum_{j\in J}\sum_{t=1}^{T}(\kappa_{jt}\mathcal{E}_{jt}^{0+}) = E[\sum_j \sum_{t=1}^{T} \mathcal{E}_{jt}(\zeta)\kappa_{jt}]$. In all cases, we use this epigraph form and set φ as the objective value to minimize. Since we restrict ourselves to a limited class of decision rules, it follows that the optimal solution of our problem (given by φ^*) bounds the true risk from above.

The rest of the constraints are readily obtained from the deterministic problem where we replace decision variables with their associated affine decision rules. Deriving the deterministic equivalent program is straightforward by using well known methods from robust optimization. For all the inequality constraints that are meant to hold $\forall \zeta \in \mathcal{Z}_\nu$, we namely use strong linear programming duality, which is possible since $\mathcal{Z}_\nu$ is a non-empty compact polyhedron. The robust equivalent is explicitly computed for a representative inequality in Section 4.9.1 of the Appendix.

4.5.3 Affine decision rules on lifted probability space

In order to obtain a better upper bound on the optimal solution of our risk averse stochastic program, we consider a special case of the lifted decision rules of Georghiou et al. (2015). Since we follow their work closely, we give a brief treatment of these decisions rules. However, we provide a detailed discussion with an example in the Appendix to help build the intuition of the reader.

The general idea is to apply a non-linear transformation $L' : \mathbb{R}^T \to \mathbb{R}^{rT}$ for some $r \in \mathbb{N}$ to $\mathcal{Z}_\nu$

in order to lift the random vector $\zeta \in \mathbb{R}^T$ onto a higher dimensional space : $L'(\zeta) = \zeta'' \in \mathbb{R}^{rT}$. Under our choice of L', which we describe shortly, the set $L'(\mathcal{Z}_\nu)$ is non-convex. It therefore becomes convenient to consider an exterior convex approximation $\mathcal{Z}''_\nu \supset L'(\mathcal{Z}_\nu)$. We then define the retraction operator $R' : \mathbb{R}^{rT} \to \mathbb{R}^T$ such that $(R' \circ L') = I$ and impose that $R'\zeta'' \in \mathcal{Z}_\nu$ for all points $\zeta'' \in \mathcal{Z}''_\nu$.

By considering controls that are affine functions of the *new* lifted random variables $\zeta'' \in \mathbb{R}^{rT}$, the decisions rules (4.13) become :

$$\mathcal{X}_{kt}(\zeta'') = \mathcal{X}^0_{kt} + \sum_{t'=1}^{T} \sum_{s=1}^{r} \mathcal{X}^{t's}_{kt} \zeta''_{t',s} \qquad k \in \mathscr{K}_t \tag{4.20}$$

and we can express the robust counterpart directly in terms of ζ''. By restricting the coefficients of our new lifted affine decision rules (4.20), we can always obtain the initial (unlifted) affine decision rules. Hence, with this choice of $\mathcal{Z}''_\nu$, L' and R', the lifted problem provides an upper bound that is at least as good as the one with the original affine decisions.

The heart of the problem is then to judiciously define L' and the corresponding R' in order to maintain the tractability of the new lifted program while ensuring an efficient representation of the feasible domain of the lifted random variables and improving the upper bound on the optimal solution of our program. We fix $L^{'} = L \circ F$ where L is a non-linear bijective mapping and $F : \mathbb{R}^T \to \mathbb{R}^T$ is a simple bijective linear mapping (F^{-1} is a $T \times T$ matrix that exists) preserving the non-anticipativity of decision rules ($F_{ij} = 0, \forall i < j$) and that can be interpreted as a change of basis. It is possible to consider non-invertible mappings, but we ignore this case for sake of simplicity (see Georghiou et al. (2015)).

The non-linear operator $L : \mathbb{R}^T \to \mathbb{R}^{rT}$ is defined through the partial lifting $L_t : \mathbb{R} \to \mathbb{R}^r$ as follows :

$$L_{tk}(\zeta^{'}) = \begin{cases} \min\{\zeta^{'}_t, z^t_1\} & \text{if } k = 1 \\ \max\{0, \min\{\zeta^{'}_t, z^t_k\} - z^t_{k-1}\} & \text{if } k = 2, \cdots, r-1 \\ \max\{\zeta^{'}_t - z^t_{k-1}, 0\} & \text{if } k = r \end{cases}$$

for all t where $z^t_1 < z^t_2 < \cdots < z^t_{r-1}$ are breakpoints dividing the range of the original $\zeta^{'}_t$ in r dimensions. Similarly, we define the partial (linear) retraction operator $R_t : \mathbb{R}^r \to \mathbb{R}$, $\forall t$ as follows :

$$R_t(\zeta'') = \sum_{k=1}^{r} \zeta''_{tk} \equiv \zeta'_t = [F\zeta]_t = F_t^\top \zeta$$

We therefore see that and R' and R are *linear*. Since we also consider a *polyhedral* $\mathcal{Z}''_\nu$, the lifted problem to remain a linear program. We present the specific choices for F and L together with numerical results in Section 4.6.3 as well as a detailed example in Section 4.9.4.

4.6 Case study

4.6.1 The river system

We apply the risk averse reservoir management problem to the case of the Gatineau river in Québec. This hydro electrical complex is part of the larger Outaouais river basin and is managed by Hydro-Québec, the largest hydroelectricity producer in Canada (Hydro-Québec, 2012). It is composed of 3 run-of-the river plants and 5 reservoirs, only 2 of which have significant capacity (see Figure 4.1).

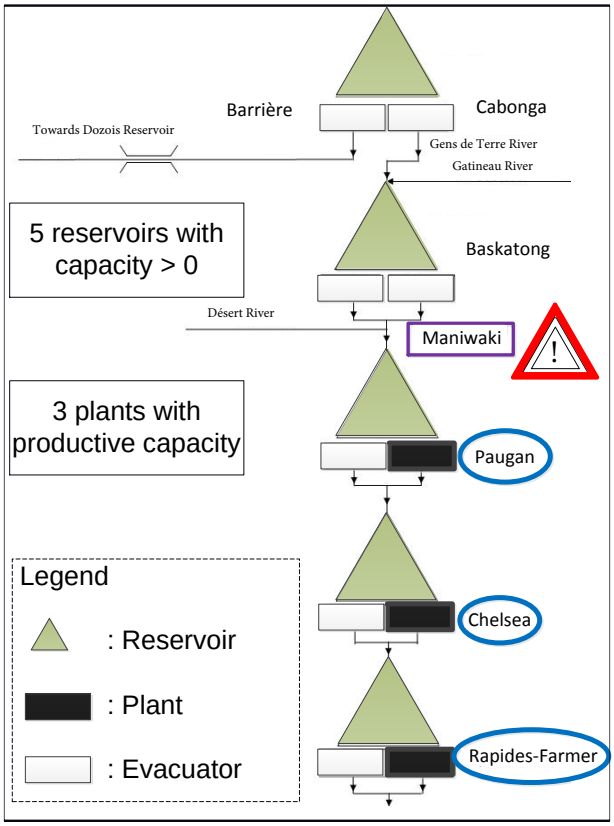

Figure 4.1 Simplified representation of the Gatineau river system.

The Baskatong reservoir is the largest of the larger Outaouais-Gatineau catchment and plays a critical role in the management of the river. It is used to manage risk of floods during the freshet period as well as droughts during the summer months. It has been used to control baseflow at the greater Montréal region several hundreds of kilometres downstream. As such, respect of minimum and maximum water volume thresholds is essential for river operators. We consider a time horizon of 30 days with daily time steps to reflect the real decision process for decisions such as setting water target releases for this reservoir during the freshet.

The maximum spillage at facilities Cabonga, Baskatong, Barrière and Rapides-Farmer is bounded by evacuation curves that reflect the particular structure of the dams and the associated reservoirs. As mentioned previously, we use affine functions to approximate the "real" curves used by Hydro-Québec (see Figure 4.2). Our approximations are conservative on most of the domain of interest, since they bound the maximum spillage from below for high volumes, which is the case in our test study. Being completely conservative for all feasible volumes would be impossible as this would lead to a very conservative or even infeasible solution. The final results are not significantly affected by the approximations errors since the simulation uses the true spillage curves. Moreover, the poor affine approximation for Rapides-Farmer dam has limited effect in general since the reservoir is extremely small, fills up almost instantly, and therefore offers little manoeuvrability and leverage for reservoir decisions.

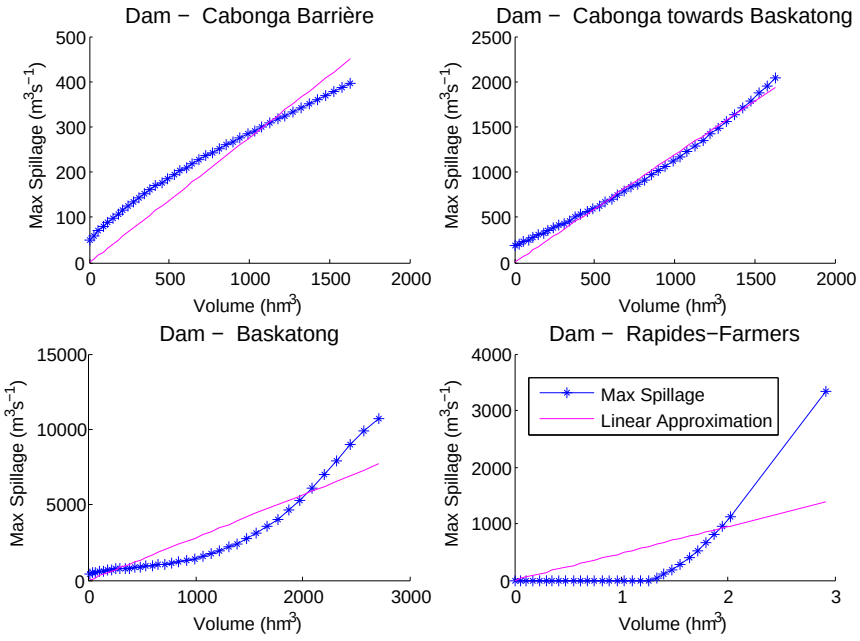

Figure 4.2 Evacuation curves.

Although there are no structures to control water outflow near Maniwaki, there are extremely tight constraints on maximum and minimum flow at that river segment. These bounds represent critical outflow level used to control the risk of drought and flood for the neighbouring town and are therefore essential.

The Gatineau represents an excellent case study as it has relatively small productive capacity compared to the rest of the park. More importantly, the river runs near the small town of Maniwaki that is subject to high risks of flooding, particularly during the spring freshet. Indeed, the city has suffered 4 significant floods in 1929, 1936, 1947 and 1974.

Even today, there exists a need to mitigate the risk of these natural disasters. As Figure 4.3 illustrates, it is possible that we fail to find a solution eliminating all floods even when we have *perfect foresight*. This is namely the case when $\zeta_t = \mu_t + 2\sigma_t, \forall t$. Although this synthetic scenario is wet over the entire month, it remains plausible because of the strong serial correlation of inflows as well as evidence from past observations. In this case, even if we spill as much as possible at Baskatong so that the total flow at Maniwaki downstream is constantly at its upper bound, the deterministic program cannot prevent the reservoir from filling up and eventually flooding.

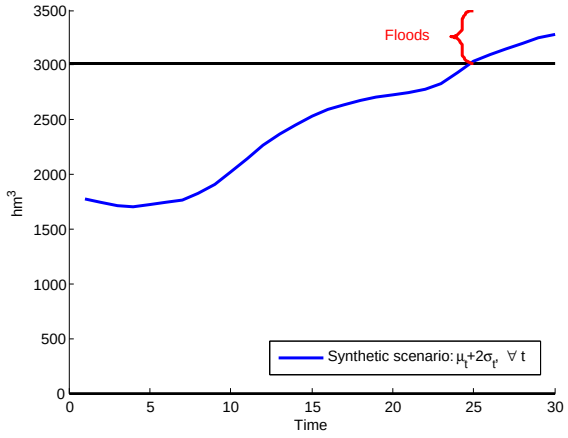

Figure 4.3 Violations with perfect foresight.

4.6.2 The inflows

Based on our sample observations, the assumption of perfectly spatially correlated inflows is not unreasonable for the Gatineau river. As is clear from the Figure 4.4, water inflows are particularly important during the months of March through April (freshet) as snow melts. There is a second surge during fall caused by greater precipitations and finally there are very

little inflows during the winter months. For this reason, we tested our model on the first 30 days of winter (lowest inflows, lowest variability) and the first 30 days of the spring freshet (highest inflows and highest variability). We estimate the true population mean and standard deviation by computing the sample mean and sample variance with these 6 years of data.

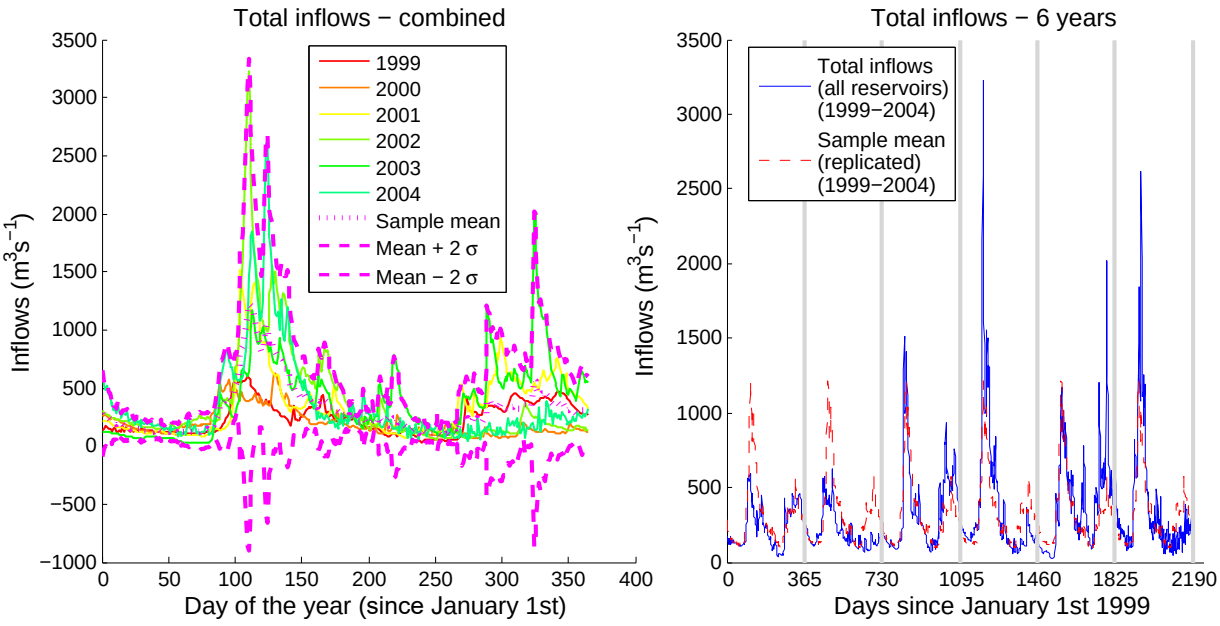

Figure 4.4 Sample observations (6 years) for inflows.

4.6.3 Upper bounds on the risk of floods

We begin by presenting the optimal solutions obtained by solving the robust equivalents with specific classes of decision rules. Since we restrict ourselves to specific family of functions, these values represent upper bounds on the optimal value of the "true" problem with general decision rules when our hypothesis on the support and the mean of the random variables are verified.

We specifically consider 4 classes of decision rules : 1) the standard affine (dimension 1 i.e. $r = 1$), 2) the lifted decision rules with 1 breakpoint at 0 and the identity mapping ($r = 2$), 3) the lifted decision rules with 2 breakpoints at the 2/6 and 5/6 quantiles of the empirical distribution of $\zeta_t, \forall t$ and the identity mapping ($r = 3$), 4) the lifted decision rules with 1 breakpoint at 0 and the lower triangular mapping $F_{ij} = 1$ for $i \geq j$ and 0 when $i < j, \forall j = 1, \cdots, T$ ($r = 2$). In the case of the last mapping F, we can interpret each $F_t^\top \zeta = \zeta_t'$ as the cumulated history from time 1 until t and we observe that this respects

non-anticipativity. We attempted to solve the model with the same mapping F and $r = 3$, but ran into computational difficulties since the problem could not be solved in less than 3 days. As we discuss later in this Section, this is likely a consequence of the high degeneracy of the linear program.

All solutions where obtained by setting $\nu = 3$, which ensured the support contained all historical inflows from 1999-2004 and provided good quality solutions across all models. We tested our models with varying values of ν and found that increasing its value above a certain threshold leads to infeasibility issues as all bounds cannot be simultaneously respected for very large uncertainty sets. As expected, small ν's lead the model to predict small violations. However, the solutions displayed poor performance when tested on realistic inflows that did not belong to the uncertainty set. Our sensitivity analysis did not seem to reveal any clear link between the size of ν and the performance of our models when tested with $\zeta \notin \mathcal{Z}_\nu$.

The upper bounds are displayed in the line "Optimal solution" of Table 4.2. These values are good indicators of the CVaR of volume violations at Baskatong predicted by the model. It is therefore interesting to compare them with the empirical CVaR of violations presented in Figure 4.7 of Section 4.6.6.

Tableau 4.2 Comparison of optimal values and solution times

	Affine - 1 dim			**Lifted - I - 2 dim**			**Lifted - I - 3 dim**			**Lifted - F - 2 dim**		
	Expected Value ($\alpha = 1$)	$\text{CVaR}_{0.5}$	Ess Sup ($\alpha = 0$)	Expected Value ($\alpha = 1$)	$\text{CVaR}_{0.5}$	Ess Sup ($\alpha = 0$)	Expected Value ($\alpha = 1$)	$\text{CVaR}_{0.5}$	Ess Sup ($\alpha = 0$)	Expected Value ($\alpha = 1$)	$\text{CVaR}_{0.5}$	Ess Sup ($\alpha = 0$)
Optimal solution	2.78E+03	5.57E+03	9.77E+03	1.11E+03	2.23E+03	9.77E+03	3.26E+02	6.52E+02	9.77E+03	1.09E+03	2.18E+03	3.74E+03
Optimal solution improvement (%)	-	-	-	60%	60%	0%	88%	88%	0%	61%	61%	62%
Time (seconds)	1.17E+01	6.86E+01	5.37E+02	3.09E+03	8.79E+03	9.95E+03	2.93E+03	4.07E+04	1.03E+05	1.66E+05	2.43E+05	2.30E+05
Time increase (%)	0%	0%	0%	26,367%	12,722%	1,751%	24,966%	59,256%	19,057%	1,422,516%	35,4334%	42,611%

4. All results were obtained by using AMPL with solver CPLEX 12.4 with default choice of solver and basic presolve from CPLEX and AMPL. Tests were conducted on computers with 16.0 GB RAM and i7 CPU's @ 3.4GHz.

The line "Optimal solution improvement (%)" of Table 4.2 shows the improvement with respect to the base affine policies for any fixed risk measure. We observe that the optimal solution of the lifted decision rules strictly improves when considering CVaR_α for $\alpha \in \{0.5, 1\}$ as the number of dimensions rises from 1 to 3 with the identity mapping. However, the optimal solution is constant with respect to the number of breakpoints with the identity mapping for the ess sup. The F mapping with dimension 2 offers the best theoretical bounds for the ess sup (62% improvement compared to the affine policies) while the identity mapping with 2 breakpoints offers the largest improvement (88%) for $\alpha \in \{0, 0.5\}$.

In all cases, the computational burden imposed by the lifted decision rules is much greater than the standard affine decision rules. Problems took up to 67 hours to solve with the F mapping and $r = 2$ while the same problem took only 68.6 *seconds* with the standard affine rules. Although it is surely beneficial to reduce the violations because of the potentially dramatic consequences of floods, computing times of several hours are unacceptable for practical operational purposes.

This dramatic increase in computational times might appear strange at first, since the increase in problem size attributed to the liftings is relatively modest compared to the problem with the standard affine decision rules. Indeed, Table 4.3 shows the number of equalities/inequalities in the robust equivalent corresponding to *each* equality/inequality in the original deterministic formulation. With standard affine decision rules, the number of variables and constraints in the robust equivalent corresponding to each inequality and equality is $O(T)$ compared with $O(Tr)$ for the lifted decision rules where T represents the dimension of the uncertainty and r is the number of lifted components. These values can be obtained by looking at the duals presented explicitly in Sections 4.9.1,4.9.3 and 4.9.4 of the Appendix.

Tableau 4.3 Size of robust equivalents (standard affine vs. lifted).

	Standard affine (box support)			**Lifted affine (box support)**		
	# constraints (rob. model)		# (dual) variables (rob. model)	# constraints (rob. model)		# (dual) variables (rob. model)
	Equalities	Inequalities		Equalities	Inequalities	
Inequality (Det. model)	T	1	2T	Tr	1	2T+T(r+1)
Equality (Det. model)	T+1	-	-	Tr+1	-	-

However, Figure 4.5 and Table 4.4 suggests that the increase in computing times is mainly due to the geometry of the problems and the path followed by the simplex and barrier algorithms called by default by CPLEX. Indeed, there seems to be a significant amount of degeneracy as the optimal solution is composed of a large number of variables fixed at 0 or at very small values. Indeed, most of the dual variables as well as the variables associated with the

linear/piecewise linear part of the decision rules, *i.e.* the $\mathcal{X}_{kt}^{t'}$ in equation (4.13) or the $\mathcal{X}_{kt}^{t's}$ in equation (4.20), end up being zero in the optimal solution. Therefore, there seems to be a large number of basis with the same objective value and it is likely that various simplex pivots lead to zero or negligible improvements.

Finally, by fixing a type of solver, say barrier, primal simplex or dual simplex, way obtain different optimal solutions, which also suggests presence of degeneracy. Surprisingly, trying to limit the consequences of degeneracy by using only barrier rarely sped up computations. We therefore kept the default CPLEX options.

Tableau 4.4 Number of Simplex/Barrier iterations and solution times.

	Affine - 1 dim	**Lifted - I - 2 dim**	**Lifted - I - 3 dim**	**Lifted - F - 2 dim**
# of Simplex iter.	5.10E+04	5.97E+05	1.37E+06	8.68E+06
# of Barrier iter.	43	309	124	170
Total time (secs)	6.86E+01	8.79E+03	4.07E+04	2.43E+05

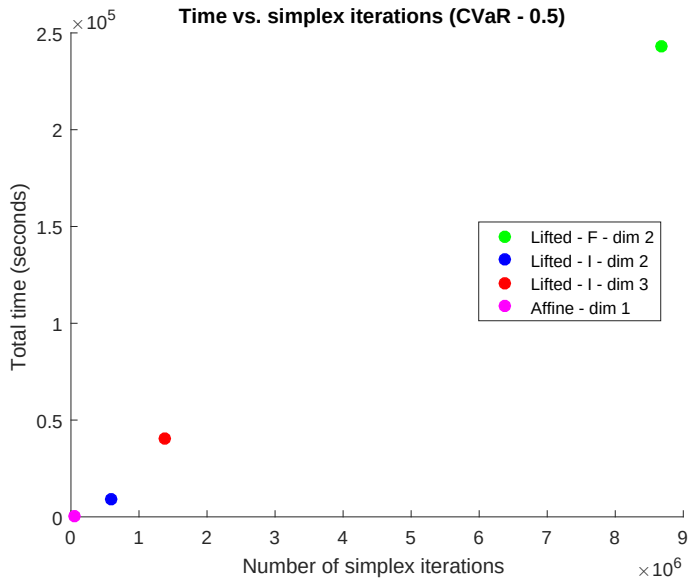

Figure 4.5 Impact of number of simplex iterations on total computing times.

We explored acceleration techniques based on cutting plane methods as suggested by Lorca et al. (2016). However, our higher dimension uncertainty set made it difficult to identify and enumerate extreme scenarios explicitly a priori, even with the standard affine decision rules. Dynamic identification of violated inequalities therefore proved longer than full reformulation through duality.

4.6.4 Evaluating the policies through simulation

Following Chen et al. (2010), the optimal solution of the problem can be useful to derive probabilistic bounds on the risk of floods and droughts. However, a small optimal value does not in itself guarantee the quality of the solutions provided by the problem since there are no guarantees when the uncertainty falls outside of the parametrized support considered in the model. It is imperative to simulate the behaviour of the system with the optimal decision rules under various scenarios to see how well it would perform in different situations.

For this reason, we perform various tests using different scenario generators. We consider independent non-negative truncated normal random variable as well as independent log normal random variable. These random processes are serially independent. As illustrated by Figure 4.4, this is not an acceptable hypothesis since there exists strong cyclical effects combined with multi-lag autocorrelation. To remedy some of these problems and use a simulator closer to real observed inflows, we calibrate a SARIMA model on the raw inflow data (Box et al., 2008). Once again, we fix a $\nu = 3$ for all our models.

4.6.5 Comparison with stochastic model based on a scenario tree

In order to compare the value of affine decision rules, we consider a multi-stage stochastic programming model based on scenario trees. Inspired by similar work by Fleten and Kristoffersen (2008); Feng and Ryan (2014); Growe-Kuska et al. (2003), we begin by constructing a simple fan (2 stage scenario tree) comprised of the equally likely 10,000 scenarios generated by the SARIMA process used in our simulation. We then use SCENRED2 from the GAMS library (https://www.gams.com) to aggregate scenarios and construct a multi-stage trees from the fan (see Heitsch and Römisch (2009)). We specifically force the resulting tree to branch only at each of the first 6 periods (days) and fix the number of scenarios to 100 (see Figure 4.6).

Since it is reasonable to suppose that inflows are jointly continuous, any of the scenarios considered in a stochastic tree model have probability 0 of materializing exactly. Using the solutions found by a single solution of the stochastic tree model in simulation is therefore highly likely to lead to poor quality solutions. To mitigate this effect, we simulate the use of this model by solving it in a rolling horizon fashion. Therefore, we construct 30 different trees, each with different structure, probabilities and inflows. Considering trees with branching occurring only at the first stages has limited effect on the overall quality of the solutions, since the rolling horizon allows the model to adapt to the current state of the system determined by the realized inflows. Preliminary results based on different tree structures in a rolling horizon

framework suggests that branching at the early stages is a more important performance driver than the total number of scenarios considered.

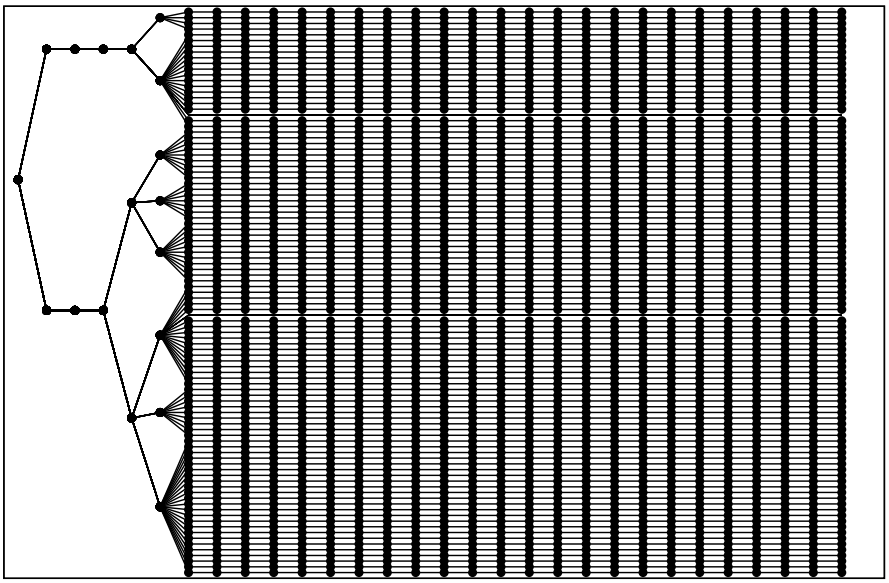

Figure 4.6 Sample multistage tree generated by SCENRED2 from scenario fan.

We also tested affine decision rules in a rolling horizon, but with a shorter lookahead period, namely 1 week. These problems took on average 5 seconds to solve. In comparison, each problem based on scenario trees took almost 4 times more time to solve. Because of these time limitations, we were only able to simulate the behaviour of the scenario tree with rolling horizon based on 1,000 random scenarios, rather than the full 10,000. This also explains why we limited the size and structure of the tree.

4.6.6 Simulation results

We report the empirical conditional value at risk of violations for the three most important constraints in Figure 4.7. We namely consider volume violations at Baskatong and Cabonga, the two large reservoirs, as well as the flow violations at the town of Maniwaki, which is a critical flood indicator. We compute the CVaR at different levels of risk $\alpha' = 100(1-\alpha)$ with $\alpha \in \{0.01, \cdots, 0.99\}$. We also show the expected value at $\alpha' = 0$ and the ess sup at $\alpha' = 100$. All results are obtained by considering 10,000 scenarios, except for the scenario tree, which considers only 1,000, since using a rolling horizon heuristic over the 30-day horizon for 10,000 scenarios would take on average more than 40 days.

In all cases, we find that the theoretical upper bound found by solving the stochastic model is loose. Even if we do not restrict the random variables to lie within the support considered by our model, we find that the empirical CVaR_α is usually smaller than the theoretical upper bound.

For sake of space, we present only results for the independent log-normally distributed random variables with mean equal to μ_t and variance set to $\nu'\sigma_t$ at every $t = 1, \cdots, T$ for a given $\nu' \in \mathbb{R}_+$ [5]. The μ_t and σ_t are the same values as the ones considered in the uncertainty sets. Although we use $1 \leq \nu' < \nu$, we consider the full support of the random variables $(0, \infty)$ and do not restrict the realisations to the box considered by the model. Results for the normal independent variables and the SARIMA time series are similar in nature, even if the scenario tree is constructed with some knowledge of the inflows' serial correlation.

Increasing the value of $\nu^{'}$ from its standard value changes the shape of the empirical distribution, but the relative performance of each model stays relatively constant. Moreover, excessively increasing $\nu^{'}$ may lead to scenarios that are more extreme than natural and that may have limited practical interpretation. The treatment of rare extreme events is an open question we do not wish to tackle here.

Figure 4.8 presents the empirical CVaR_α of violations for the three reservoirs for a fixed policy. These can be seen as parametric curves where the α' is progressively increased from 0 to 100. Thus the curves can be read by starting at the $(0, 0, 0)$ point and moving to the upper $(10^4, 6 \times 10^3, 60)$ point.

As Figure 4.8 illustrates, when considering a fixed policy, different risk measures may lead to solutions that are empirically dominated by others. For instance with the lifted affine decision rules, mapping I and 2 dimensions, the solution provided by minimizing CVaR_α seems to be dominated by the two others since it lies above the curves. However, the situation is reversed for the lifted affine decision rules, mapping F and 2 dimensions. In this case, the solution provided by minimizing CVaR_α seems better than the two others.

In most cases, Figure 4.8 suggests that the policy results are similar for all three risk measures. The policies often have the most noticeable differences in the most extreme scenarios, which rarely occur. However, for the scenario tree with rolling horizon, the policy performances are considerably different, particularly for the violations at Baskatong. This does not seem to be a consequence of the reduced number of simulations, since most results stabilized after a relatively limited number of simulations. It is possible that by exploiting the probabilities

5. Here ν' is a parameter controlling the variability of the simulated random variables. We always performed simulations with $1 \leq \nu'$ to test the robustness of our solutions.

and correlations, the tree based model derived very different outcomes for the different risk measures. On the other hand, the conservative approximations used by the robust model only used the first moment of the random variables.

It is difficult to infer reliable and general conclusions from these observations. Nonetheless, we notice that the affine decision rules with limited 7 day lookahead horizon used in a rolling horizon context compare favourably with other policies. Indeed, violations are 0 for Baskatong and Cabonga with the affine policies with 7 day lookahead horizon used in a rolling horizon fashion for all risk measures.

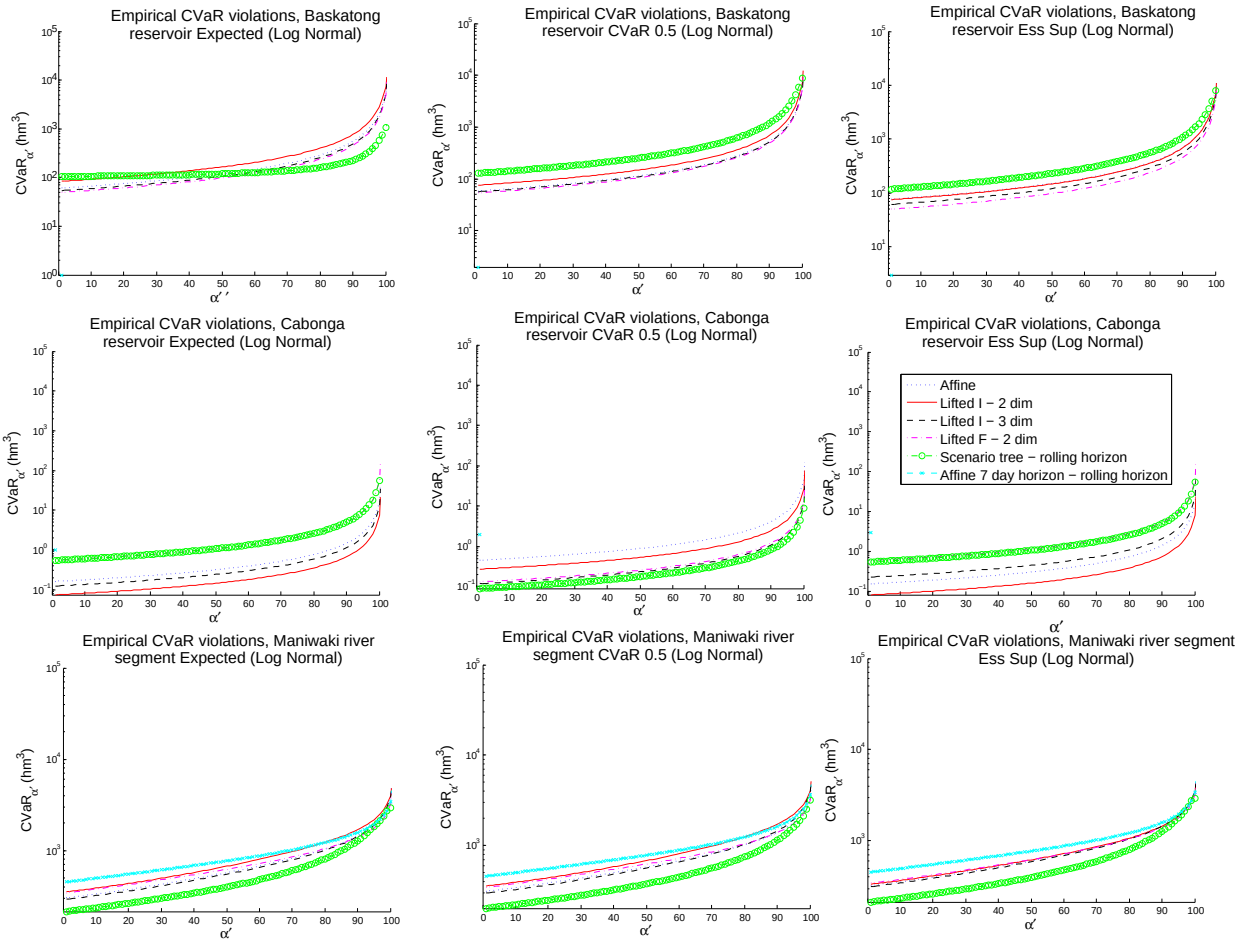

Figure 4.7 Empirical CVaR_α of violations across policies with log-normally simulated random variables and $\alpha \in \{0, 0.5, 1\}$ for a fixed reservoir.

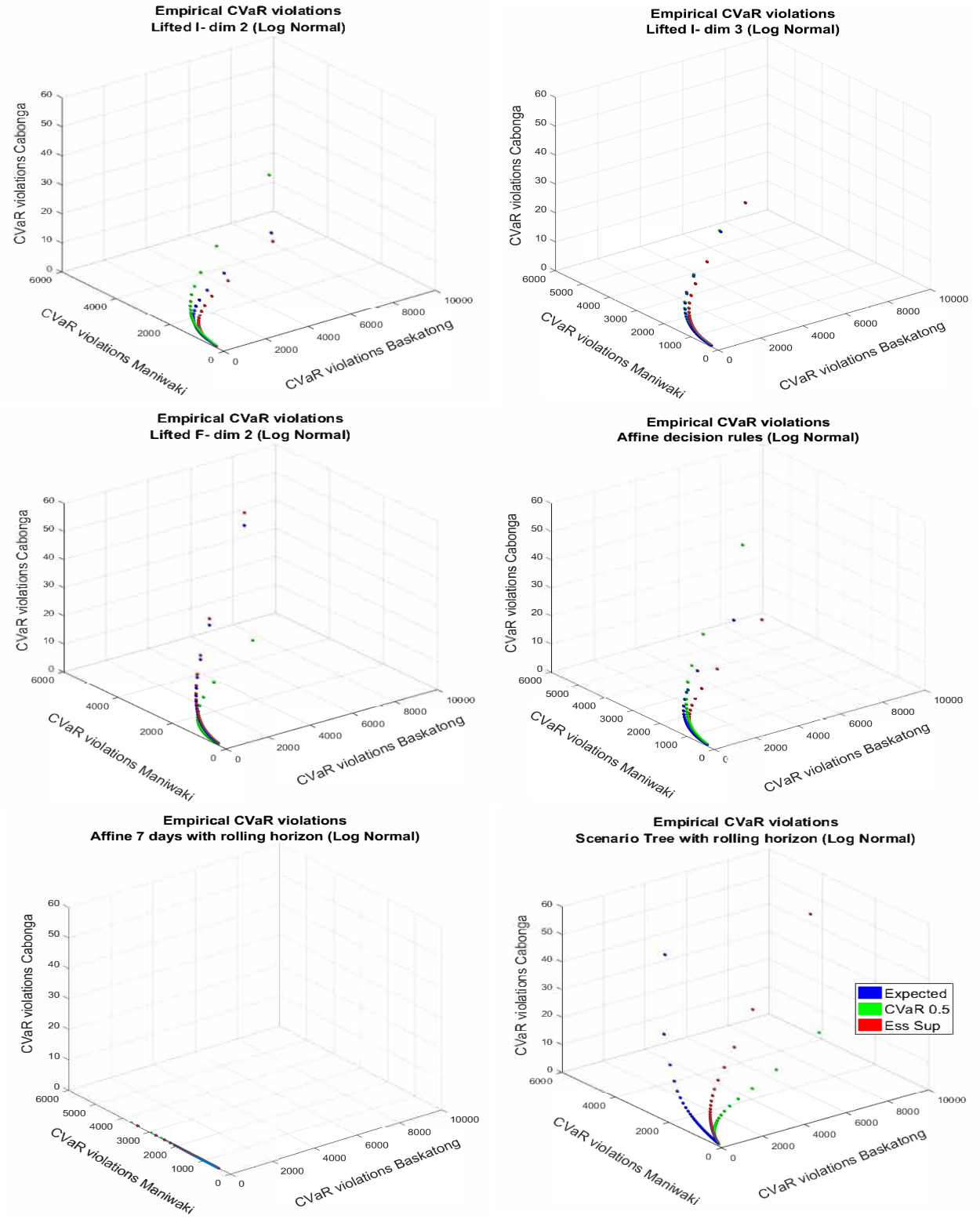

Figure 4.8 Empirical CVaR_α of violations across reservoirs with log-normally simulated random variables for a fixed policy.

Plotting the results for a given reservoir for all risk measures and decision rules does not clearly indicate superiority of any policy. Indeed, Figure 4.7 makes it hard to discriminate between decision rules. In particular, these results do not indicate that lifted decision rules are more performant than the basic affine decision rules. This is surprising as the the theoretical upper bounds computed by lifted decision rules are sizeably smaller than those returned by the affine ones.

This may be due to the fact that the lifted decision rules use the same information and representation of uncertainty as the standard affine controls. Moreover, the empirical CVaR_α results are chiefly influenced by a few (<4) extreme scenarios that fall outside of the support considered by the model. Lifted decision rules do not seem to provide better coverage than the basic affine controls against these extreme events.

However, our results clearly indicate the superiority of the affine and lifted decision rules with regards to the scenario tree model. Indeed, Figure 4.7 highlights the systematically higher violations at Baskatong, Maniwaki and Cabonga for all risk measures compared to any decision rules. Even when solved in a rolling horizon fashion, the scenario tree provide unsatisfactory results.

Although they do not dominate all decision rules for all risk measures and risk levels, the 7 day lookahead affine policies used in rolling horizon provide excellent results. Limited computing resources and time make it difficult to simulate policies considering a longer lookahead horizon, but it is likely that they would yield even better results.

4.6.7 Real scenarios results

We evaluate each policy on real historical observed scenarios. We consider the first 6 years 1999-2004 used to calibrate the model as well as 6 additional out-of-sample years 2008-2013 for validation. Once again, we only present the important reservoirs and river segments, since the other constraints at other reservoirs are always respected. We focus on the first 30 days of the freshet period since other periods of the year are usually free of any violations and we always present results for models using an objective of $\text{CVaR}_{0.5}$ since results are relatively similar for different choice of α.

As Figure 4.9 illustrates, all types of decision rules violate some of the constraints for one of the 12 scenarios. Affine and lifted decision rules only violate constraints on the last 6 out-of-sample years (in red) since they have higher inflows while the scenario tree violates constraints on the 2 sets of data.

The policy provided by the scenario tree model produces a flow that is smaller than the

others on average. Moreover, together with the deterministic model, it is the only policy that violates lower flow bounds. Consequently, the upstream reservoir Baskatong fills up more on average, which increases the risk of floods upstream. This leads to particularly important violations for the year 2013.

On the other hand, affine and lifted decision rules provide a generally adequate solution. Although there are maximum volume violations at the end of the time horizon for very wet years as well as maximum flow violations at periods of high inflow, these violations are usually more acceptable than those of the tree model. It is also rather surprising to see that lifted decision rules do not consistently outperform the simple affine decision rules.

We also include the trajectories for the deterministic model used in a rolling horizon fashion as well as the standard affine decision rules with a limited lookahead horizon of 1 week. The deterministic model does very poorly with regards to flow bound violations at Maniwaki. This result is consistent with the empirical observations of Apparigliato (2008) who observed superiority of the robust approach relative to deterministic optimization. The limited affine model in rolling horizon takes a similar amount of computing resources as the decision tree, but seems to achieve a better trade-off between violations of volume bounds at Baskatong and flow bounds at Maniwaki.

Although we do not push the volumes towards a desired goals at the end of the time horizon, our solution never completely empties the reservoirs nor fills them up when physically possible. This is an advantage of the model since it yields more realistic water release schedules and does not necessarily require taking into account a future monetary value of water.

The implemented flow values at Maniwaki presented in Figure 4.9 also highlight on the presence of a larger number of null values in the linear/piecewise linear part of the optimal controls, which was previously mentioned in Section 4.6.3. Indeed, the realized trajectories of flows at Maniwaki reveal that the controls provided by the various models move together across time. Therefore, the adjustable linear or piecewise linear part of the decision rules must be very small and the bulk of the decisions resides in the nominal decision.

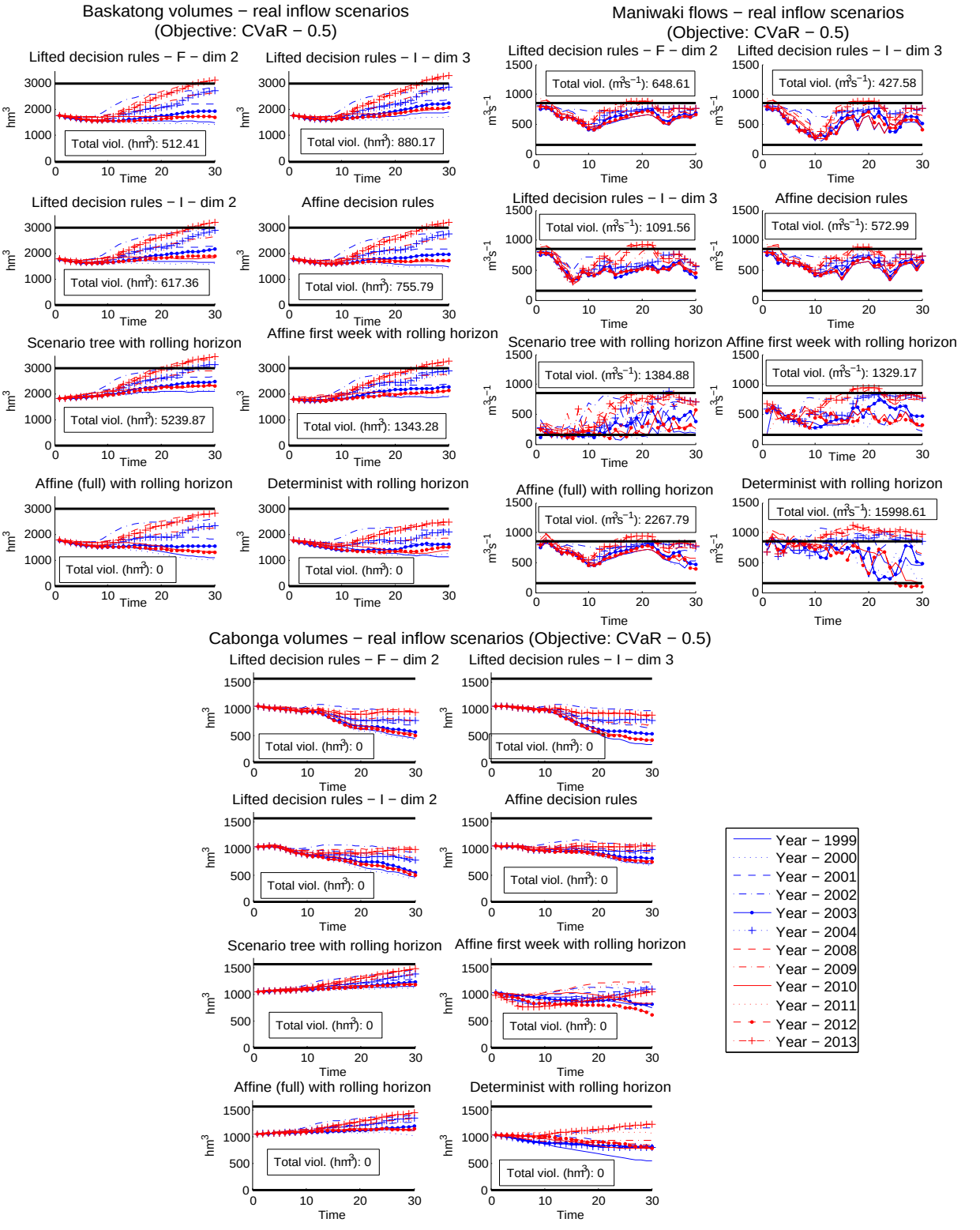

Figure 4.9 Actual values - 12 historical scenarios.

4.6.8 Brief comparison of budgeted and box uncertainty

We also tested our model with the following commonly use budgeted uncertainty set (see namely Ardestani-Jaafari and Delage (2014)) [6] :

$$\left\{ \zeta \in \mathbb{R}^T \;\middle|\; \begin{array}{ll} \exists (\zeta^{+\top}, \zeta^{-\top})^\top \in \check{\mathcal{Z}}_{\nu,\Gamma} & \text{(4.21a)} \\ \zeta_{t'} = \zeta^+_{t'} - \zeta^-_{t'}, \forall t' & \text{(4.21b)} \end{array} \right\}$$

where :

$$\check{\mathcal{Z}}_{\nu,\Gamma} = \left\{ (\zeta^{+\top}, \zeta^{-\top})^\top \in \mathbb{R}^{2T}_+ \;\middle|\; \begin{array}{ll} \zeta^+_{t'} \leq \nu\sigma_{t'}, \quad \forall t' & \text{(4.22a)} \\ \zeta^-_{t'} \leq \min\{\mu_{t'}, \nu\sigma_{t'}\}, \; \forall t' & \text{(4.22b)} \\ \displaystyle\sum_{t'=1}^{T} \frac{\zeta^+_{t'}}{\nu\sigma_{t'}} + \frac{\zeta^-_{t'}}{\min\{\mu_{t'}, \nu\sigma_{t'}\}} \leq \Gamma & \text{(4.22c)} \end{array} \right\}$$

Constraints (4.22a)-(4.22b) correspond to the basic box constraints, equality (4.21b) ensures we get the original ζ_t and can be seen as the retraction operator while (4.22c) is the budget constraint determining the number of random variables that can take their extremal values over the entire horizon. To focus the comparison on the geometry of the uncertainty set rather than the power of the liftings, we fixed the coefficients of $\zeta^+_{t'}$ and $\zeta^-_{t'}$ at the same value in the affine decision rules (see 4.9.2 of the Appendix for more details).

Figure 4.10 suggests that there is not a significant difference between the performance of the affine decision rules with and without the budget constraint with $\Gamma > 0$ when considering independent log-normally distributed random variables. In this case, the graphs of the empirical CVaR of the total violation cross and we cannot identify a dominant policy. This is true for all budgets except for the case $\Gamma = 0$ which performs very poorly and is dominated by all policies. As intuition might suggest, the policy considering only box support hedges against very extreme scenarios. As such, it does better in the worst case, but at the price of degraded $\text{CVaR}_{\alpha'}$ performance compared to the policy with the budget constraint with $\Gamma > 0$ for $\alpha' \in (0, 100]$.

However, when we consider the seasonal time series model with strong correlation across time, we find more significant variations in performance across policies. We namely see that considering $\Gamma \leq \sqrt{T} \approx 5.48$ leads to poor results with empirical risk higher than the one of the box support for all $\alpha' \in [0, 100]$. In this case, the budgeted uncertainty set fails to take

6. For all experiments with the budgeted uncertainty set, we fixed $\nu = 3$ as in the rest of our numerical experiments.

into account the possibility of persistently high or low inflows. For $\Gamma = \sqrt{T} + i\frac{(T-\sqrt{T})}{5}$ and $i \in \{1, 2, 3, 4, 5\}$ however, neither the box nor the policy with the budget dominate the other. Only $i = 5$ is shown in Figure 4.10, but the results are very similar for $i \in \{1, 2, 3, 4\}$. As in the independent case, the box leads to better worst case performance, but degraded $\text{CVaR}_{\alpha'}$ performance for $\alpha' \in (0, 100]$.

It is also interesting to note that although the theoretical models with the box and the non-binding budget constraint with $\Gamma = 30$ are equivalent, the solutions found and the simulation results differ. This is another consequence of the multiplicity of the optimal solutions and the degeneracy of the problem. Solutions with the same objective value can yield different results when tested against specific scenarios. In future work, it may be interesting to consider multiple objectives to resolve this multiplicity. For instance, we could compare different optimal solutions obtained by minimizing a fixed CVaR_{α} based on their performance for different $\text{CVaR}_{\alpha'}$.

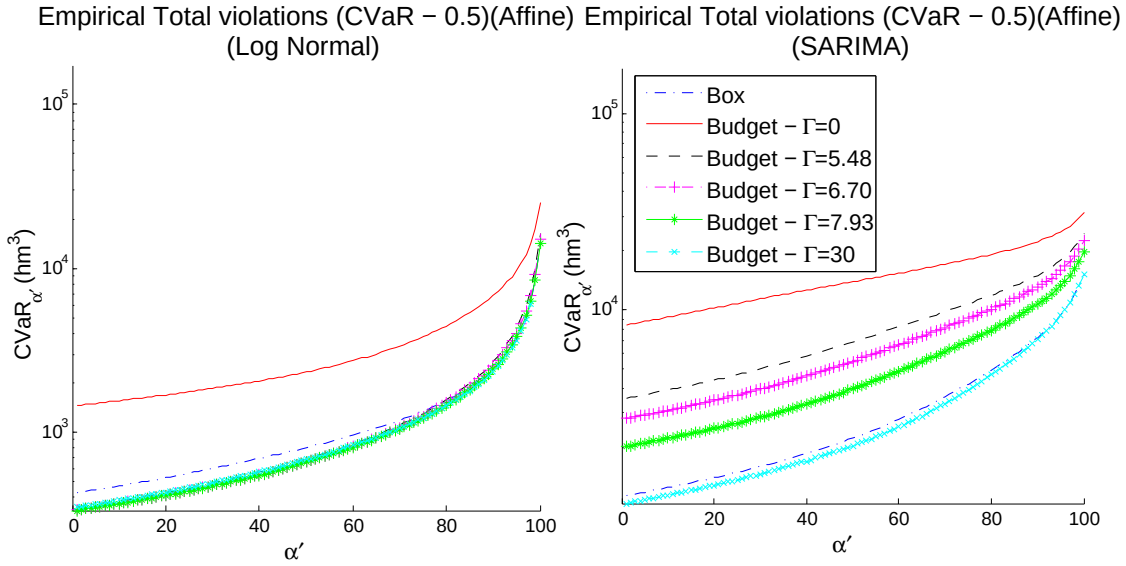

Figure 4.10 Impact of different uncertainty sets on simulation results.

Computational experiments also revealed that by considering the budgeted uncertainty set with $\Gamma = T$ so that the constraint is not binding, the computational time grew from 3.9 minutes with our standard box formulation to 153.9 minutes ($\approx$ 2.5 hours) with the budgeted uncertainty sets when considering the ess sup as the risk measure. This is an increase of more than 39 times.

Table 4.5 sheds light on this initially surprising result. We first observe that the budgeted formulation generates T additional inequalities and equalities for each inequality in the ori-

ginal deterministic formulation. We then note that the robust equivalents have very different structure. The robust equivalent of the box model generates significantly more equalities than inequalities. This is a consequence of the dualization procedure and is described explicitly in Sections 4.9.1 and 4.9.2 of the Appendix. The consequences of this large number of equalities is that the presolve engine of both AMPL and CPLEX can make dramatic reductions in model size by appropriate substitutions. Turning this option off namely resulted in much larger models and increased computing times with the box support.

Tableau 4.5 Size of robust equivalents (box vs. budget).

	Standard affine (box support)			**Standard affine (budget support)**		
	# constraints (rob. model)		# (dual) variables	# constraints (rob. model)		# (dual) variables
	Equalities	Inequalities	(rob. model)	Equalities	Inequalities	(rob. model)
Inequality (Det. model)	T	1	2T	-	2T+1	2T+1
Equality (Det. model)	T+1	-	-	T+1	-	-

To validate these results, we also tested the following equivalent uncertainty set :

$$\left\{ \zeta \in \mathbb{R}^T \;\middle|\; \begin{array}{ll} \exists \varrho \in \mathbb{R}^T_+ & \\ \zeta_{t'} \leq \varrho_{t'} \nu \sigma_{t'}, & \forall t' \\ -\zeta_{t'} \leq \varrho_{t'} \min\{\mu_{t'}, \nu\sigma_{t'}\}, & \forall t' \\ \varrho_{t'} \leq 1, & \forall t' \\ \displaystyle\sum_{t'=1}^{T} \varrho_{t'} \leq \Gamma & \end{array} \right\} \tag{4.23a–e}$$

Contrary to formulation (4.21), the ζ_t are unrestricted in sign in (4.23). This uncertainty set therefore leads to the same number of equalities as the box, but with an additional T inequalities for each original inequality. Considering this alternative formulation reduced the computing times by a factor close to 4 compared with formulation (4.21) for tests performed with CVaR_α and $\alpha \in \{0, 0.5, 1\}$. However, when setting $\Gamma \geq T$, the standard box still performed significantly better than both budgeted formulations in terms of computing times. These findings corroborate the importance of the number of equalities in the resulting deterministic equivalent.

Nonetheless, Figure 4.11 shows that the model with budgeted uncertainty can provide better quality policies than those of the standard box for fixed scenarios. Indeed, if we consider the 12 historical inflow scenarios and use $\Gamma = \sqrt{T} + 2\frac{(T-\sqrt{T})}{5} \approx 15.29$, which is very close to 15.90, the minimal value that ensures that all 6 years of real data are contained within the

budgeted uncertainty set. we obtain lower flow and volume violations over all reservoirs.

On the other hand, the majority of policies obtained with budget $\Gamma = \sqrt{T} + i\frac{(T-\sqrt{T})}{5}$ and $i \in \{1, 2, 3, 4, 5\}$ did not dominate the policy with box support. Furthermore, considering a small budget will deteriorate the quality of the solutions. Indeed, using $\Gamma \leq \sqrt{T}$ yields higher volume and flow violations than using the standard box. As illustrated in Figure 4.11, the model with $\Gamma = \sqrt{T}$ neglects persistently high inflows and therefore generates significant flooding at Baskatong. The volume violations at that reservoir are more than 13 times larger the ones with the affine decision rule considering box uncertainty. [7]

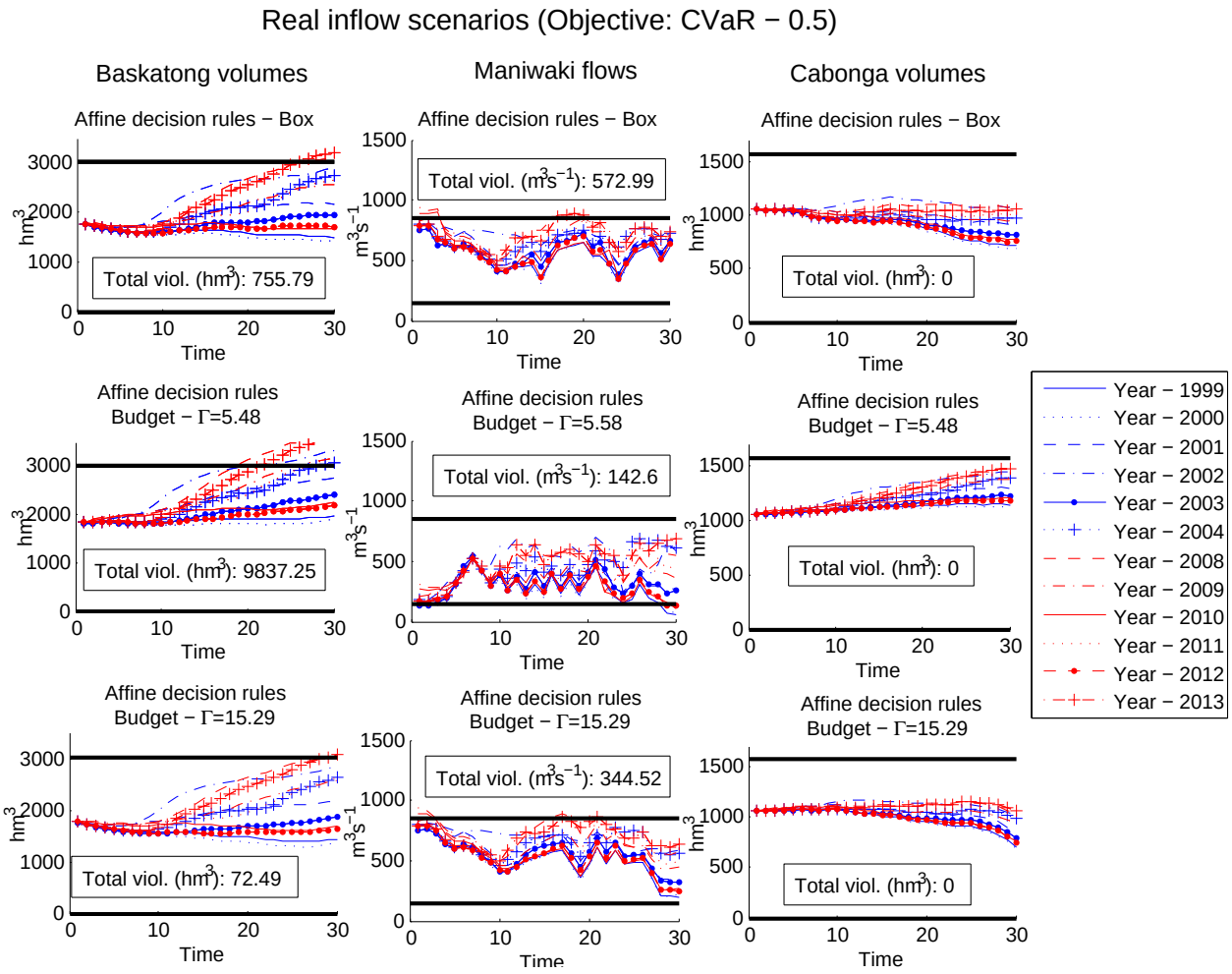

Figure 4.11 Impact of different uncertainty sets on real historical inflows.

7. In all cases, the models were solved once in a static fashion and the decisions were then implemented.

4.7 Conclusion

In conclusion, we propose a novel robust optimization formulation for the risk averse stochastic reservoir management problem. Our model can easily be obtained from an existing deterministic model and is relatively easy to understand and implement.

We also formalize risk-aversion through the coherent risk measure CVaR_α. By performing sensitivity analysis on the α parameter, we study the theoretical and simulated impact of different risk profiles.

Contrary to other popular methods such as stochastic dynamic programming, we can easily incorporate constraints involving different time steps, such as flow variation constraints. Our method also allows the consideration of objective functions such as CVaR that do not permit time decomposition.

Our various numerical experiments confirm the value of using affine decision rules to provide good implementable solutions while maintaining the tractability of the model. Our experiments also suggest that these decision rules offer little optimality loss over more complex piece-wise linear decision rules while providing extremely important computational savings. Affine and piece-wise linear decision rules seem to perform consistently better than decision trees, even when considering a rolling horizon framework.

The results based on historical and synthetic inflows also suggest that the affine decision rules compare favourably with scenario tree models in terms of solution quality. However, these conclusions may not be generalizable to other problems and instances. In addition, more elaborate and representative scenario trees may lead to better results, since these model can exploit probabilities and serial correlation to which the robust model is insensitive. Our intuition is that if the decision makers have very rich and reliable data spanning long time periods, then it might be beneficial to exploit the structure provided by scenario trees. On the other hand, if the decisions makers only have limited data, as was the case in this study, or if they must extrapolate into the future and stationarity is a questionable assumption, then the robust model may provide more reliable results.

Our experiments also showcase the computational advantages of the affine decision rules compared to the tree based model. We feel that this is an important point, notably for operational purposes. This advantage may be less important if the model is only used for strategic long term decisions and long computing times are acceptable. Nonetheless, detailed tree structures will invariably lead to difficulties when using simulation and stress tests. It might be possible to speed up computations by using decomposition techniques such as the ones considered in Carpentier et al. (2013a), however, these will require a significantly amount

of technical work to implement.

Finally, we perform a brief but interesting comparison of box and budgeted uncertainty sets. We show that in the presence of significant serial correlation, decision makers should use caution when choosing the value of Γ. If the model can be calibrated thorough reliable simulation, then it might be preferable to consider the budget uncertainty sets with a properly chosen budget. However, in the context of considerable ambiguity in distributions and possibly non-stationary, the simple box might provide more reliable solutions. In any case, the use of the box uncertainty set should reduce model size and speed up computations compared with the budgeted uncertainty set.

4.8 Acknowledgments

The authors would like to thank everyone at Hydro-Québec and IREQ for their ongoing support, particularly Grégory Émiel, Louis Delorme and Pierre-Marc Rondeau as well as Pierre-Luc Carpentier and Alexia Marchand. We are also grateful to Daniel Kuhn, Pascal Côté and Sara Séguin for valuable discussions. The comments of anonymous referees also helped improve the quality of the paper. This research was supported by the Natural Sciences and Engineering Reasearch Council of Canada (NSERC) and Hydro-Québec through the Industrial Research Chair on the Stochastic Optimization of Electricity Generation and NSERC grant 386416-2010 .

4.9 Appendix

4.9.1 Representative robust constraint under standard affine decision rules and box uncertainty

Consider standard affine decision rules and the maximum turbined flow constraints (4.7) : $\mathcal{R}_{it}(\zeta) \leq \bar{r}_i, \forall \zeta \in \mathcal{Z}_\nu = \bigtimes_{t=1}^{T}[-\min\{\mu_t, \nu\sigma_t\}, \nu\sigma_t]$ for a fixed plant and time. We omit the indices i and t for sake of clarity and formulate these robust constraints as the following linear program :

$$
\begin{aligned}
\max_{\zeta} \quad & \mathcal{R}^0 + \sum_{t'=1}^{T} \mathcal{R}^{t'} \zeta_{t'} \leq \bar{r} \\
\text{s. t.} \quad & (\pi^+_{t'}) \quad \zeta_{t'} \leq \nu\sigma_{t'}, && \forall t' = 1, \cdots, T \\
& (\pi^-_{t'}) \quad -\zeta_{t'} \leq -\min\{\mu_{t'}, \nu\sigma_{t'}\}, && \forall t' = 1, \cdots, T
\end{aligned}
$$

Taking the dual yields :

$$\begin{aligned}
\min_{\pi} \quad & \mathcal{R}^0 + \sum_{t'=1}^{T} \left[\nu\sigma_{t'}\pi_{t'}^{+} - \min\{\mu_{t'}, \nu\sigma_{t'}\}\pi_{t'}^{-}\right] \leq \bar{r} \\
\text{s. t.} \quad & \pi_{t'}^{+} - \pi_{t'}^{+} = \mathcal{R}^{t'}, \quad \forall t' = 1, \cdots, T \\
& \pi_{t'}^{+}, \pi_{t'}^{-} \geq 0, \qquad \forall l' = 1, \cdots, L
\end{aligned}$$

4.9.2 Representative robust constraint under standard affine decision rules and budgeted uncertainty

In contrast, adding a budget constraint with budget $\Gamma > 0$ to our basic box uncertainty yields the following linear program :

$$\begin{aligned}
\max_{\zeta} \quad & \mathcal{R}^0 + \sum_{t'=1}^{T} \mathcal{R}^{t'}(\zeta_{t'}^{+} - \zeta_{t'}^{-}) \leq \bar{r} \\
\text{s. t.} \quad & (\pi_{t'}^{+}) \quad \zeta_{t'}^{+} \leq \nu\sigma_{t'}, && \forall t' = 1, \cdots, T \\
& (\pi_{t'}^{-}) \quad \zeta_{t'}^{-} \leq \min\{\mu_{t'}, \nu\sigma_{t'}\}, && \forall t' = 1, \cdots, T \\
& (\pi^0) \quad \sum_{t'=1}^{T} \frac{\zeta_{t'}^{+}}{\nu\sigma_{t'}} + \frac{\zeta_{t'}^{-}}{\min\{\mu_{t'}, \nu\sigma_{t'}\}} \leq \Gamma \\
& \qquad \zeta_{t'}^{+}, \zeta_{t'}^{-} \geq 0, && \forall t' = 1, \cdots, T
\end{aligned}$$

Taking the dual yields :

$$\begin{aligned}
\min_{\pi} \quad & \mathcal{R}^0 + \sum_{t'=1}^{T} \left[\nu\sigma_{t'}\pi_{t'}^{+} + \min\{\mu_{t'}, \nu\sigma_{t'}\}\pi_{t'}^{-}\right] + \Gamma\pi^0 \leq \bar{r} \\
\text{s. t.} \quad & \pi_{t'}^{+} + \frac{\pi^0}{\nu\sigma_{t'}} \geq \mathcal{R}^{t'}, && \forall t' = 1, \cdots, T \\
& \pi_{t'}^{-} + \frac{\pi^0}{\min\{\mu_{t'}, \nu\sigma_{t'}\}} \geq -\mathcal{Q}^{t'}, && \forall t' = 1, \cdots, T \\
& \pi_{t'}^{+}, \pi_{t'}^{-} \geq 0, && \forall l' = 1, \cdots, L
\end{aligned}$$

Note that since the intervals $[\min\{\mu_t, \nu\sigma_t\}, \nu\sigma_t]$ may be non symmetric for any t, we have to consider the problem on the lifted space with $\zeta_t = \zeta_t^{+} - \zeta_t^{-}$ decomposed into its positive and

negative parts.

4.9.3 Equality constraints under standard affine decision rules

Under standard affine decision rules and the full dimensional $\mathcal{Z}_\nu$ (given either by the intersection of the budget constraint and the box or simply the box), the equality constraints (4.2) will hold if and only if they hold for $\zeta = 0 \in \mathbb{R}^T$:

$$\mathcal{S}^0_{jt} = v_{j0} + \sum_{\tau=1}^{t}[\sum_{l=\delta_{min}}^{\min\{\delta_{max},\tau-1\}} \lambda_l \mathcal{F}^0_{i^+\tau-l} - \mathcal{F}^0_{i^-\tau} + \mu_t \alpha_{jt}]\beta$$

and

$$\mathcal{S}^{t'}_{jt} = \begin{cases} \left(\sum_{\tau=1}^{t}[\sum_{l=\delta_{min}}^{\min\{\delta_{max},\tau-1\}} \lambda_l \mathcal{F}^{t'}_{i^+\tau-l} - \mathcal{F}^{t'}_{i^+\tau-l}]\right)\beta + \alpha_{jt'}\beta & t' \leq t \\ \left(\sum_{\tau=1}^{t}[\sum_{l=\delta_{min}}^{\min\{\delta_{max},\tau-1\}} \lambda_l \mathcal{F}^{t'}_{i^+\tau-l} - \mathcal{F}^{t'}_{i^+\tau-l}]\right)\beta & t' \geq t+1 \end{cases}$$

4.9.4 Details on lifted decision rules with example

We now give a detailed treatment of the lifted decision rules with a particular emphasis on one of the cases considered in our numerical study. The objective of this appendix is to help the reader get better intuition about the use of lifted decision rules. First, Section 4.9.4 describes how a conservative approximation model is obtained when using lifted decision rules and improves on the solutions obtained without the lifting. The following two Subsections of Section 4.9.4 discuss how an equivalent robust reformulation can be obtained for inequality constraints such as the flow conservation constraint (4.2) and equality constraints such as the turbine bounds (4.7) respectively.

Linear mapping F and description of $F\mathcal{Z}_\nu$

Given a bijective (full rank matrix) F and under the hypothesis that $\mathcal{Z}_\nu$ is closed and bounded, we get that $F\mathcal{Z}_\nu \subseteq [l,u]$ where $[l,u] = \bigtimes_{t=1}^{T}[\min_{\zeta\in\mathcal{Z}_\nu} F_t^\top\zeta, \max_{\zeta\in\mathcal{Z}_\nu} F_t^\top\zeta]$ and $F_t^\top$ is the t^{th} row of F. Figure 4.12 illustrates the simple case of a symmetric 2-dimensional uncertainty set given by $\mathcal{Z}_\nu = [-1,1]^2$ and the lower triangular mapping $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ with its inverse

$$F^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

The blue polyhedron on the right corresponds to $\{\zeta' \in \mathbb{R}^2 : \exists \zeta \in \mathcal{Z}_\nu, \zeta' = F\zeta\}$, the image of the box $[-1,1]^2$ under F while the red box $[-1,1] \times [-2,2]$ represents $\bigtimes_{t=1}^{2}[\min_{\zeta\in[-1,1]^2} F_t^\top \zeta, \max_{\zeta\in[-1,1]^2} F_t^\top \zeta]$.

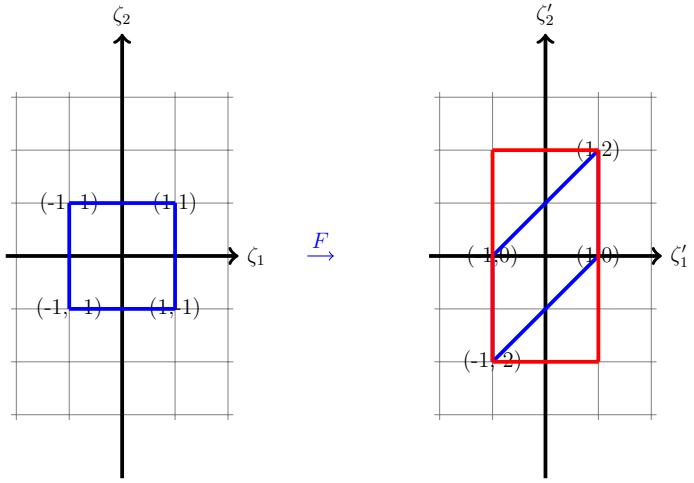

Figure 4.12 Image of polyhedron under a bijective linear mapping and bounding box.

We now describe the lifting operator L and $(L \circ F)(\mathcal{Z}_\nu)$. Recall that for a fixed $r \in \mathbb{N}$ and $t \in \mathbb{T}$ with $z_1^t < z_2^t < \cdots < z_{r-1}^t$:

$$L_{tk}(\zeta^{'}) = \begin{cases} \min\{\zeta_t^{'}, z_1^t\} & \text{if } k = 1 \\ \max\{0, \min\{\zeta_t^{'}, z_k^t\} - z_{k-1}^t\} & \text{if } k = 2, \cdots, r-1 \\ \max\{\zeta_t^{'} - z_{k-1}^t, 0\} & \text{if } k = r \end{cases}$$

and

$$R_t(\zeta^{''}) = \sum_{k=1}^{r} \zeta_{tk}^{''} \equiv \zeta_t^{'} = [F\zeta]_t = F_t^\top \zeta$$

We notice that R_t is linear and R may be represented as a block angular $T \times rT$ matrix :

$$R = \begin{pmatrix} \underbrace{1 \quad \cdots \quad 1}_{r\,terms} & & & & \\ & 1 \quad \cdots \quad 1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \quad \cdots \quad 1 \end{pmatrix}$$

As described in Georghiou et al. (2015), in order to get an accurate description of the uncertainty set for ζ'', we need to characterize the convex hull of $(L \circ F)(\mathcal{Z}_\nu)$. Geoghiou et al. show how this can be done exactly when $F(\mathcal{Z}_\nu)$ is an hyperrectangle. In particular, if one considered the lifted image $L(\bigtimes_t [l_t, u_t])$, then he could exploit the property that $\text{conv}(L(\bigtimes_t [l_t, u_t]))$ $= \bigtimes_t \text{conv} L_t([l_t, u_t])$ and observe that for every t, $L_t([l_t, u_t])$ is a set of connected segments (defined through the choice of breakpoints $z_1^t, \cdots, z_{r-1}^t$) for which $\text{conv} L_t([l_t, u_t])$ simply reduces to the set of convex combinations of the following $r+1$ affinely independent points in $\mathbb{R}^r$:

$$v_1^t = \begin{pmatrix} l_t \\ 0 \\ \cdots \\ 0 \end{pmatrix} v_2^t = \begin{pmatrix} z_1^t \\ 0 \\ \cdots \\ 0 \end{pmatrix} v_3^t = \begin{pmatrix} z_1^t \\ z_2^t - z_1^t \\ \cdots \\ 0 \end{pmatrix} \quad \cdots \quad v_{r+1}^t = \begin{pmatrix} z_1^t \\ z_2^t - z_1^t \\ \cdots \\ u_t - z_{r-1}^t \end{pmatrix}$$

For instance, consider the simple lifting which breaks the components into their positive and negative parts. We have $r = 2$, $z_1^1 = z_2^1 = 0$ and $l_1 = -1, u_1 = 1, l_2 = -2, u_2 = 2$. This is illustrated in Figure 4.13 where the solid green lines indicate $L_t([l_t, u_t])$ and the (light and solid) green represents $\text{conv}(L_t([l_t, u_t]))$.

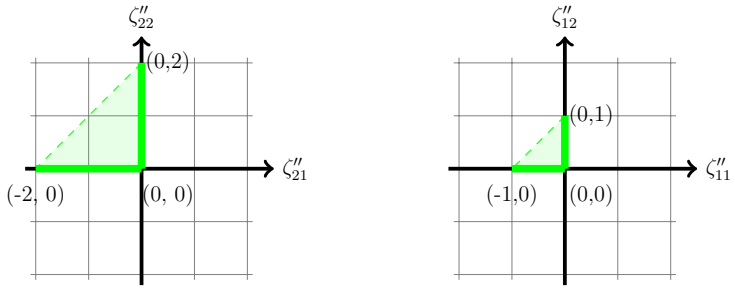

Figure 4.13 $L_t([l_t, u_t])$ and $\text{conv}(L_t([l_t, u_t])), t = 1, 2$.

If we regroup the points $v_1^t,\, v_2^t, \cdots, v_{r+1}^t$ in the following invertible matrix :

$$V_t = \begin{pmatrix} 1 & \cdots & 1 \\ v_1^t & \cdots & v_{r+1}^t \end{pmatrix} \in \mathbb{R}^{(r+1)\times(r+1)}$$

then we have :

$$\begin{aligned} \text{conv}L_t([l_t, u_t]) &= \{\zeta_t^{''} \in \mathbb{R}^r : (1, \zeta_t^{''\top})^\top = V_t\lambda, \lambda \in \mathbb{R}_+^{r+1}\} \\ &= \{\zeta_t^{''} \in \mathbb{R}^r : V_t^{-1}(1, \zeta_t^{''\top})^\top \geq 0\}. \end{aligned}$$

Building on these observations, we obtain :

$$\begin{aligned} \text{conv}(L \circ F)(\mathcal{Z}_\nu) &= \text{conv}((L \circ F)(\mathcal{Z}_\nu) \cap \{\zeta'' \in \mathbb{R}^{rT} : \exists \zeta \in \mathcal{Z}_\nu,\, R\zeta'' = F\zeta\}) && (4.24) \\ &\subseteq \text{conv}((L \circ F)(\mathcal{Z}_\nu)) \cap \text{conv}(\{\zeta'' \in \mathbb{R}^{rT} : \exists \zeta \in \mathcal{Z}_\nu,\, R\zeta'' = F\zeta\}) && (4.25) \\ &\subseteq \text{conv}(L(\bigtimes_t [l_t, u_t])) \cap \text{conv}(\{\zeta'' \in \mathbb{R}^{rT} : \exists \zeta \in \mathcal{Z}_\nu,\, R\zeta'' = F\zeta\}) && (4.26) \\ &= \mathcal{Z}_\nu'' := \{\zeta'' \in \mathbb{R}^{rT} : F^{-1}R\zeta'' \in \mathcal{Z}_\nu, V_t^{-1}\zeta_t'' \geq 0, \forall t\}, && (4.27) \end{aligned}$$

where the first inclusion comes from the property that $\text{conv}(A \cap B) \subseteq \text{conv}(A) \cap \text{conv}(B)$, while the second inclusion comes from the property that $F\mathcal{Z}_\nu \subseteq \bigtimes_t [l_t, u_t]$ which implies that $\text{conv}((L \circ F)(\mathcal{Z}_\nu)) \subseteq \text{conv}(L(\bigtimes_t [l_t, u_t]))$. We will conveniently formulate each robust constraint of the lifted decision rules model in terms of $\mathcal{Z}_\nu''$.

Following our previous example with $r = 2$, we see that $V_t^{-1}\zeta_t'' \geq 0, \forall t$ corresponds to :

$$\underbrace{\begin{pmatrix} 0 & \frac{1}{l_t} & 0 \\ 1 & -\frac{1}{l_t} & -\frac{1}{u_t} \\ 0 & 0 & \frac{1}{u_t} \end{pmatrix}}_{V_t^{-1}} \begin{pmatrix} 1 \\ \zeta_{t1}'' \\ \zeta_{t2}'' \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad t = 1, 2 \qquad (4.28)$$

$$\begin{aligned} \Leftrightarrow \quad & \zeta_{21}'' \leq 0 && \zeta_{11}'' \leq 0 && (4.29) \\ & \frac{-\zeta_{21}''}{2} + \frac{\zeta_{22}''}{2} \leq 1 && -\zeta_{11}'' + \zeta_{12}'' \leq 1 && (4.30) \\ & \zeta_{22}'' \geq 0 && \zeta_{12}'' \geq 0 && (4.31) \end{aligned}$$

with $\zeta'_1 = R_1\zeta'' = \zeta_{11} + \zeta_{12}$ and $\zeta'_2 = R_1\zeta'' = \zeta_{21} + \zeta_{22}$ where we could use the more eloquent labelling $\zeta_{t1} = -\zeta_t^- \leq 0$ and $\zeta_{t2} = \zeta_t^+ \geq 0$ for $t = 1, 2$ to represent the negative and positive part of ζ'_t. At any extreme point of $\mathcal{Z}''_\nu$, at most one of ζ''_{t1} or ζ''_{t2} can be non-zero for $t = 1, 2$. From (4.27), all that is left is to ensure $F^{-1}R\zeta'' \in \mathcal{Z}_\nu$. In our example, this translates to $\mathcal{Z}''_\nu$ being described with the following linear constraints :

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} \leq \underbrace{\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}}_{F^{-1}} \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}}_{R} \begin{pmatrix} \zeta''_{11} \\ \zeta''_{12} \\ \zeta''_{21} \\ \zeta''_{22} \end{pmatrix} \leq \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } (4.29) - (4.31). \tag{4.32}$$

It is worth emphasizing the fact that since $\mathcal{Z}''_\nu$ is an outer approximation of $\text{conv}(L \circ F)(\mathcal{Z}_\nu)$, one should expect that each robust constraint will be replaced by one that is more conservative than needed hence preventing the optimization model from fully exploiting the flexibility provided by the piecewise linear lifting. However, with $\mathcal{Z}''_\nu$, we are at least guaranteed to obtain a better approximation than with the scheme that only uses affine decision rules. This is due to the fact that for any standard and lifted affine decision rules of the form (4.13) and (4.20), respectively :

$$\begin{aligned} \mathcal{X}_{kt}(\zeta) &= \mathcal{X}^0_{kt} + \sum_{t'=1}^{T} \mathcal{X}^{t'}_{kt}\zeta_{t'}, \qquad k \in \mathscr{K}_t \\ \hat{\mathcal{X}}_{kt}(\zeta'') &= \hat{\mathcal{X}}^0_{kt} + \sum_{t'=1}^{T}\sum_{s=1}^{r} \hat{\mathcal{X}}^{t's}_{kt}\zeta''_{t',s} \qquad k \in \mathscr{K}_t \end{aligned}$$

it is always possible to fix the values $\hat{\mathcal{X}}^{t's}_{kt}$ to ensure that the following holds for a fixed affine policy and an arbitrary function $f : \mathbb{R} \times \mathbb{R}^T \to \mathbb{R}$:

$$\begin{aligned} \max_{\zeta \in \mathcal{Z}_\nu} f(\mathcal{X}_{kt}(\zeta), \zeta) &= \max_{\zeta'' \in \mathbb{R}^{rT} : F^{-1}R\zeta'' \in \mathcal{Z}_\nu} f(\hat{\mathcal{X}}_{kt}(\zeta''), F^{-1}R\zeta'') \\ &\geq \max_{\zeta'' \in \mathcal{Z}''_\nu} f(\hat{\mathcal{X}}_{kt}(\zeta''), F^{-1}R\zeta'') \ , \end{aligned}$$

The equality follows from the particular restriction of the lifted decision rules and the fact that the "retracted" $F^{-1}R\zeta''$ lie within $\mathcal{Z}_\nu$. For instance, with the identity mapping $F = I$, it suffices to set $\hat{\mathcal{X}}^{t's}_{kt} = \mathcal{X}^{t'}_{kt}, \forall s \in \{1, \cdots, r\}$, since $\sum_{s=1}^{r} \zeta''_{t',s} = \zeta_{t'}$ for any t'. The inequality

follows from equation (4.27), which implies $\mathcal{Z}''_\nu \subseteq \{\zeta'' \in \mathbb{R}^{rT} : F^{-1}R\zeta'' \in \mathcal{Z}_\nu\}$. We therefore conclude that for any feasible solution based on standard affine decision rules it is possible to find a feasible solution with lifted decision rules with the same objective value.

The following subsections will illustrate how we can derive a reformulation for each type of robust constraint in this lifted decision rules framework under $\mathcal{Z}''_\nu$ as defined in equation (4.32).

Representative robust constraint under lifted decision rules

Under the lifting $(L \circ F)$ where F is linear and bijective, the robust turbined flow constraints (4.7) for a fixed plant i and time t will be satisfied by imposing $\max_{\zeta'' \in \mathcal{Z}''_\nu \supseteq (L\circ F)(\mathcal{Z}_\nu)} \mathcal{R}^0 + \sum_{t=1}^{T}\sum_{j=1}^{r} \mathcal{R}^{tj}\zeta''_{tj} \leq \bar{r}$.

We let a_t represent the 1^{st} column of V_t^{-1} and $b_t \in \mathbb{R}^{(r+1)\times r}$ be the matrix composed of the last r columns of V_t^{-1} from the preceding Section 4.9.4. We can then write this explicitly as the following linear program with block constraint matrix under the hypothesis that $\mathcal{Z}_\nu = \bigtimes_{t=1}^{T}[\min\{\mu_t, \nu\sigma_t\}, \nu\sigma_t]$:

$$
\begin{array}{rrrrrrrrrrrrl}
\mathcal{R}^0 + \max\limits_{\zeta''} & \mathcal{R}^{11}\zeta''_{11}+ & \ldots & +\mathcal{R}^{1r}\zeta''_{1r}+ & \ldots & \mathcal{R}^{T1}\zeta''_{T1}+ & \ldots & +\mathcal{R}^{Tr}\zeta''_{Tr} & \leq \bar{r} \\
\text{s. t.} \quad (\pi_1^+) & F_{11}^{-1}\zeta''_{11} & \ldots & F_{11}^{-1}\zeta''_{1r} & \ldots & F_{1T}^{-1}\zeta''_{T1} & \ldots & F_{1T}^{-1}\zeta''_{Tr} & \leq \nu\sigma_1 \\
(\pi_2^+) & F_{21}^{-1}\zeta''_{11} & \ldots & F_{21}^{-1}\zeta''_{1r} & \ldots & F_{2T}^{-1}\zeta''_{T1} & \ldots & F_{2T}^{-1}\zeta''_{Tr} & \leq \nu\sigma_2 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
(\pi_T^+) & F_{T1}^{-1}\zeta''_{11} & \ldots & F_{T1}^{-1}\zeta''_{1r} & \ldots & F_{TT}^{-1}\zeta''_{T1} & \ldots & F_{TT}^{-1}\zeta''_{Tr} & \leq \nu\sigma_T \\
(\pi_1^-) & -F_{11}^{-1}\zeta''_{11} & \ldots & -F_{11}^{-1}\zeta''_{1r} & \ldots & -F_{1T}^{-1}\zeta''_{T1} & \ldots & F_{1T}^{-1}\zeta''_{Tr} & \leq \min\{\mu_1, \nu\sigma_1\} \\
(\pi_2^-) & -F_{21}^{-1}\zeta''_{11} & \ldots & -F_{21}^{-1}\zeta''_{1r} & \ldots & -F_{2T}^{-1}\zeta''_{T1} & \ldots & -F_{2T}^{-1}\zeta''_{Tr} & \leq \min\{\mu_2, \nu\sigma_2\} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
(\pi_T^-) & -F_{T1}^{-1}\zeta''_{11} & \ldots & -F_{T1}^{-1}\zeta''_{1r} & \ldots & -F_{TT}^{-1}\zeta''_{T1} & \ldots & -F_{TT}^{-1}\zeta''_{Tr} & \leq \min\{\mu_T, \nu\sigma_T\} \\
(\phi_{11}) & -b_{11}^1\zeta''_{11} & \ldots & -b_{1r}^1\zeta''_{1r} & & & & & \leq a_1^1 \\
\ldots & \ldots & \ldots & \ldots & & & & & \\
(\phi_{1(r+1)}) & -b_{(r+1)1}^1\zeta''_{11} & \ldots & -b_{(r+1)r}^1\zeta''_{1r} & & & & & \leq a_{(r+1)}^1 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
(\phi_{T1}) & & & & & -b_{11}^T\zeta''_{T1} & \ldots & -b_{1r}^T\zeta''_{Tr} & \leq a_1^T \\
\ldots & & & & & \ldots & \ldots & \ldots & \\
(\phi_{T(r+1)}) & & & & & -b_{(r+1)1}^T\zeta''_{T1} & \ldots & -b_{(r+1)r}^T\zeta''_{Tr} & \leq a_{r+1}^T
\end{array}
$$

The first set of constraints associated with the π dual variables represent $F^{-1}R\zeta'' \in \mathcal{Z}_\nu$ when $\mathcal{Z}_\nu = \bigtimes_{t=1}^{T}[\max\{-\nu\sigma_t, -\mu_t\}, \nu\sigma_t]$ and the second one associated with the ϕ dual variables represents $V_t^{-1}(1, \zeta''_{t1}, \cdots, \zeta''_{tr})^\top \geq (0, \cdots, 0)^\top, \forall t$. Taking the dual yields :

$$\begin{aligned}
\mathcal{R}^0 + \min_{\pi,\phi} \quad & \sum_{t'=1}^{T}(\nu\sigma_{t'}\pi^+_{t'} + \pi^-_{t'}\min\{\mu_{t'}, \nu\sigma_{t'}\}) + \sum_{t'=1}^{T}\sum_{j'=1}^{r+1}\phi_{t'j'}a^{t'}_{j'} \leq \bar{r} \\
\text{s. t.} \quad & \sum_{t''=1}^{T} F^{-1}_{t''t'}(\pi^+_{t''} - \pi^-_{t''}) - \sum_{j''=1}^{r+1}\phi_{t'j''}b^{t'}_{j''j'} = \mathcal{R}^{t'j'} \qquad \forall t' = 1, \cdots, T, j' = 1, \cdots, r \\
& \pi^-_{t'}, \pi^-_{t'} \geq 0 \qquad \forall t' = 1, \cdots, T \\
& \phi_{t'j''} \geq 0, \qquad \forall t' = 1, \cdots, T, j'' = 1, \cdots, r+1
\end{aligned}$$

Equality constraints under lifted affine decision rules

Under the lifted decision rules, the nominal equality constraints for $\zeta = 0$ remain the same :

$$\mathcal{S}^0_{jt} = v_{j0} + \sum_{\tau=1}^{t}[\sum_{l=\delta_{min}}^{\min\{\delta_{max},\tau-1\}} \lambda_l \mathcal{F}^0_{i^+\tau-l} - \mathcal{F}^0_{i^-\tau} + \mu_\tau\alpha_{j\tau}]\beta$$

However, the other constraints take the form :

$$\mathcal{S}^{j't'}_{jt} = \left(\sum_{\tau=1}^{t}[\sum_{l=\delta_{min}}^{\min\{\delta_{max},\tau-1\}} \lambda_l \mathcal{F}^{t'j'}_{i^+\tau-l} - \mathcal{F}^{t'j'}_{i^+\tau-l} + F^{-1}_{\tau t'}\alpha_{j\tau}]\right)\beta$$

for all $t' = 1, \cdots, T, j' = 1, \cdots, r, t = 1, \cdots, T$ and $j \in J$. In particular, if $F = I$, then $F^{-1}_{\tau t} = 1$ if $\tau = t$ and 0 otherwise, which yields a structure similar to the standard affine case.

CHAPITRE 5 ARTICLE 2: A STOCHASTIC PROGRAM WITH TIME SERIES AND AFFINE DECISION RULES FOR THE RESERVOIR MANAGEMENT PROBLEM

Cet article a été accepté avec révision mineures par *European Journal of Operational Research* en avril 2017. Auteurs : Charles Gauvin, Erick Delage et Michel Gendreau.

Un rapport technique préliminaire est disponible : Gauvin, C., Delage, E. et Gendreau, M., A stochastic program with time series and affine decision rules for the reservoir management problem, *Les Cahiers du GERAD*, Avril 2016, Révisé Janvier 2017, ISSN : 0711–2440 (Gauvin et al., 2016).

Abstract : This paper proposes a multi-stage stochastic programming formulation for the reservoir management problem. Our problem specifically consists in minimizing the risk of floods over a fixed time horizon for a multi-reservoir hydro-electrical complex. We consider well-studied linear time series models and enhance the approach to consider heteroscedasticity. Using these stochastic processes under very general distributional assumptions, we efficiently model the support of the joint conditional distribution of the random inflows and update these sets as new data are assimilated. Using robust optimization techniques and affine decision rules, we embed these time series in a tractable convex program. This allows us to obtain good quality solutions rapidly and test our model in a realistic simulation framework using a rolling horizon approach. Finally, we study a river system in western Québec and perform various numerical experiments based on different inflow generators.

Keywords : Stochastic programming, Stochastic processes, Robust optimization, Forecasting, OR in energy, Risk analysis

5.1 Introduction

This paper considers the problem of minimizing the risk of floods for a multi-reservoir system over a fixed time horizon subject to uncertainty on inflows while respecting tight operational constraints on total storage, spills, releases, water balance and additional physical constraints. This reservoir management problem is of vital importance to various real sites that are close to human habitations and that are prone to flooding (Quentin et al., 2014; Pianosi and Soncini-Sessa, 2009; Castelletti et al., 2010; Gauvin et al., 2017).

The deterministic version of the problem already poses serious challenges since operators must consider complex non-linear phenomena related to the physical nature of the system

(Labadie, 2004). The interconnection of the catchment also complicates decisions as upstream releases affect downstream volumes and flows. This issue is particularly important for large catchments where there can be long water delays (Gauvin et al., 2017). Considering uncertainty significantly increases theses difficulties since the sequential decision-making under uncertainty represents a huge theoretical obstacle in itself (Dyer and Stougie, 2006).

In order to solve this problem, we propose a multi-stage stochastic program based on affine decision rules and well-known time series models. Our approach leverages techniques from stochastic programming, stochastic processes and robust optimization.

Starting with the pioneering work of Ben-Tal et al. (2004), adjustable robust optimization based on affine decision rules has emerged as a viable approach for dynamic problems where uncertainty is progressively revealed. The approach has been shown capable of finding good quality solutions to large multi-stage stochastic problems that would otherwise be unmanageable to traditional methods such as stochastic dynamic programming.

These techniques have been applied to the reservoir management problems with a varying degree of success. The authors of Apparigliato (2008) and Pan et al. (2015) use this framework to maximize the expected electric production for a multi-period and multi-reservoir hydroelectric complex while Gauvin et al. (2017) minimize the risk of floods by adopting a risk averse approach that explicitly considers the multidimensional nature of the problem subject to more realistic operating constraints.

Although some of these studies use elaborate decision rules based on works such as Chen et al. (2008); Goh and Sim (2010) and Georghiou et al. (2015), they only consider very simplified representations of the underlying stochastic process and generally omit serial correlation. However, the importance of the persistence of inflows can play a crucial factor in hydrological modelling for stochastic optimization problems, particularly when daily inflows need to be considered. Authors like Turgeon (2005); Pianosi and Soncini-Sessa (2009) argue that serial correlation of high order is important to consider inflows that are high or low on many consecutive days and risk producing a flood or low baseflow.

This paper addresses the issue by developing a dynamic robust uncertainty set that takes into consideration the dynamic structure and serial correlation of the inflow process. We show that under certain conditions, these sets correspond to the support of the joint conditional distribution of uncorrelated random variables that determine the inflows over a given horizon. Our work shares similarities with the paper of Lorca and Sun (2015) who propose dynamic uncertainty sets based on time series models for a 2-stage economic dispatch problem in the presence of high wind penetration. Like these authors, we take advantage of the dynamic adaptability of the uncertainty sets by incorporating our model in a realistic simulation

framework with rolling horizon.

Nonetheless, we give significantly more details on the construction of these uncertainty sets for general univariate ARMA models and provide key insights which are of value to practitioners and academics alike. We also consider the case of heteroscedasticity which is empirically observed in various inflow time series (Romanowicz et al., 2006). Although we minimize the risk of floods, our work can be directly extended to electricity generation, under the hypothesis that head is constant and that the production function can be modeled as a piecewise linear function.

Our model considers ARMA and GARCH models of *any* order without increasing the complexity of the problem. This is a huge improvement over stochastic dynamic programming (SDP) methods, which have historically been the most popular techniques used for reservoir management both in academia and in practice (see Labadie (2004); Yeh (1985); Castelletti et al. (2008); Turgeon (2005) and references therein). Although these methods can deal with more realistic non-convex optimization problems and provide excellent closed-loop policies, they can usually only consider serial correlation through autoregressive models of small order since higher order models require increasing the state-space, which quickly leads to numerical intractability ; particularly for multi-reservoir operations (Turgeon, 2007; Cervellera et al., 2006; Tilmant and Kelman, 2007).

Considering heteroscedasticity further increases the state dimension and the resulting computational burden for SDP. As such, we are only aware of the work of Pianosi and Soncini-Sessa (2009) that is capable of considering this phenomenon with this methodology. Furthermore, these authors only manage to incorporate heteroscedasticity in a reduced model used for on-line computations.

Numerous refinements of classical SDP have emerged to circumvent some of these difficulties. Neuro-dynamic programming (Castelletti et al., 2010), sampling SDP (Stedinger and Faber, 2001) and more elaborate discretization schemes (Zéphyr et al., 2016; Cervellera et al., 2006) have namely been applied successfully to large reservoir systems while explicitly or implicitely considering high order serial correlation. However, most of these methods still require simplifications of the river dynamics and inflow representation as well as discretization of decisions, which is not the case of our approach.

Other works based on SDP, such as Turgeon (2005); Stedinger and Faber (2001); Tejada-Guibert et al. (1995); Côté et al. (2011), have focused on various low-dimensional hydrological variables such as seasonal forecasts, additional exogenous information like soil moisture and linear combinations of past inflows. Although these aggregate hydrological variables improve the solution quality without excessive computational requirements, they often rely on distri-

butional assumptions such as normality that are not verified in practice or exogenous data that may be difficult to obtain. Our model does not suffer from such limitations.

In recent years, the stochastic dual dynamic programming (SDDP) method has emerged as an effective algorithm capable of successfully tackling multi-dimensional stochastic reservoirs problems (Tilmant and Kelman, 2007; Shapiro et al., 2013; Rougé and Tilmant, 2016; Gjelsvik et al., 2010; Pereira and Pinto, 1991; Phillpot and Guan, 2008). Moreover, Maceira and Damázio (2004) illustrate that this method can consider multiple lag autoregressive processes without excessively increasing the size of the problem for the aggregated 4-state Brazilian hydro-thermal system. However, the algorithm can exhibit relatively slow convergence (Shapiro et al., 2013). To avoid this issue, various studies only consider a limited number of discrete scenarios (<100) (Tilmant and Kelman, 2007; Rougé and Tilmant, 2016), but this only leads to approximate solutions of questionable quality (Shapiro, 2011).

We also mention stochastic programming methods based on decision trees, which are also theoretically capable of explicitly handling highly persistent inflows (Carpentier et al., 2013a; Gonçalves et al., 2013; Fleten et al., 2011). These models are simple to implement when an existing deterministic model already exists and are intuitive to use and interpret. Unfortunately, they display exponential growth in complexity as a function of the time horizon. Therefore, they are usually limited to small decision trees. Gauvin et al. (2017) shows that these methods can be considerably more computationally intensive than corresponding stochastic programs based on decision rules.

The main advantage of our method with respect to SDDP and scenario-tree based stochastic program is its limited distributional assumptions. Whereas these competing methods requires using specific distributions from which it is possible to sample, our approach only demands hypothesis on the first 2 moments of the random variables as well as the correlation between them. As observed by Pan et al. (2015), this is likely to lead to solutions that are more robust when tested on out-of-sample scenarios, which is of prime concern when data availability is limited. Another shortcoming of SDDP and tree-based stochastic programming compared with decision rules and SDP are their inability to provide explicit policies that can be directly used by decision makers. Nonetheless, this is usually not a critical point, particularly when rolling horizon simulations are considered.

The paper is structured as follows. The model for the deterministic reservoir management problem and the stochastic version based on affine decision rules are presented in Section 5.2. Section 5.3 discusses inflow representation and general univariate ARMA models and the resulting conditional supports. It then studies heteroscedastic time series and their impact on the model formulation. Section 5.4 explains the solution procedure and simulation framework

while Section 5.5 studies a river in western Québec. Conclusions are drawn in Section 5.6.

5.1.1 Notation

The sets $\mathbb{Z}_+ = \{0, 1, 2, \cdots\}$ and $\mathbb{Z}_- = \{0, -1, -2, \cdots\}$ represent the non-negative and non-positive integers while $\mathbb{S}_+^n$ is the cone of square $n \times n$ semi-definite matrices. The set $\mathbb{T} = \{1, \cdots, T\}$ represents the entire horizon of T periods and $\mathbb{L} = \{0, \cdots, L-1\}$ denotes a limited look-ahead horizon for some $L < T$.

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbb{P})$ be a filtered probability space where $\{\mathscr{F}_t\}$ is a collection of σ-algebras representing some information available at time $t \in \mathbb{T}$ where $\mathscr{F}_0 = \{\Omega, \emptyset\}$ and $\mathscr{F}_T = \mathscr{F}$. We let $\mathrm{E}[\cdot]$ denote mathematical expectation while $\mathrm{E}[\cdot|\mathscr{A}]$ represents conditional expectation given any σ-algebra $\mathscr{A} \subseteq \mathscr{F}$. Both expectations are taken with respect to $\mathbb{P}$, the base probability measure on $\mathscr{F}$.

For the real random variables X, $\sigma(X)$ represents the σ-algebra generated by X (Billingsley, 1995). We abuse language and refer to the support of X as the smallest closed set on which X takes values with probability one. For any discrete time real valued stochastic process $\{X_t\}_{t\in\mathbb{Z}}$, we denote the $\mathbb{R}^L$ valued random vector $(X_t, \cdots, X_{t+L-1})^\top \equiv X_{[t,t+L-1]}$ for any $t \in \mathbb{Z}$ and $L \in \mathbb{N}$ with the special notation $X_{[L]}$ if $t = 1$. For simplicity, we abuse notation and do not distinguish a random variable from a given realisation.

5.2 The stochastic reservoir management problem

5.2.1 Deterministic look-ahead model for flood minimization

Before describing our stochastic reservoir management problem with affine decision rules, we describe the deterministic version. A similar model is described in Gauvin et al. (2017), but we present it here to make the paper self-contained. We write the problem in a "look-ahead" form to facilitate its integration in the rolling horizon framework presented in Section 5.4.

At the beginning of time $t \in \mathbb{T}$, we seek a vector of decisions for each future time $\tau \in \{t, \cdots, t+L-1\}$ that will minimize a coarse measure of flood damages over a limited horizon of L periods, where $L - 1 \leq T - t$. These decisions must respect the operational constraints (5.1b)-(5.1j) :

$$(P_t) \qquad \min_{\mathcal{X}} \quad \sum_{j=1}^{J} \sum_{l=0}^{L-1} \kappa_{jt+l} \mathcal{K}_{jt+l}(\mathcal{E}_{j,t+l}) \tag{5.1a}$$

$$\text{(Stor. Bounds)} \qquad \underline{s}_j \leq \mathcal{S}_{j,t+l} - \mathcal{E}_{j,t+l} \leq \bar{s}_j \qquad \forall j \in J, l \in \mathbb{L} \tag{5.1b}$$

$$\text{(Water Bal.)} \qquad \mathcal{S}_{j,t+l} = \mathcal{S}_{j,t+l-1} + \left(\sum_{i^- \in I^-(j)} \sum_{\bar{l}=\delta_{i^-}^{min}}^{\min\{\delta_{i^-}^{max}, t+l-1\}} \lambda_{i^-\bar{l}} \mathcal{F}_{i^-,t+l-\bar{l}} - \sum_{i^+ \in I^+(j)} \mathcal{F}_{i^+,t+l} + \alpha_{j,t+l} \xi_{t+l} \right)$$

$$\forall j \in J, l \in \mathbb{L} \tag{5.1c}$$

$$\text{(Flow Bounds)} \qquad \underline{f}_i \leq \mathcal{F}_{i,t+l} \leq \bar{f}_i \qquad \forall i \in I, l \in \mathbb{L} \tag{5.1d}$$

$$\text{(Evac. Curve)} \qquad \mathcal{L}_{i,t+l} \leq \mathscr{C}_i(\mathcal{S}_{j^-(i),t+l}) \qquad \forall i \in I^{evac}, l \in \mathbb{L} \tag{5.1e}$$

$$\text{(Var. Bounds)} \qquad |\mathcal{F}_{i,t+l} - \mathcal{F}_{i,t+l-1}| \leq \bar{\Delta}_i \qquad \forall i \in I, l \in \mathbb{L} \tag{5.1f}$$

$$\text{(Spill. Bounds)} \qquad \underline{l}_i \leq \mathcal{L}_{i,t+l} \leq \bar{l}_i \qquad \forall i \in I, l \in \mathbb{L} \tag{5.1g}$$

$$\text{(Rel. Bounds)} \qquad \underline{r}_i \leq \mathcal{R}_{i,t+l} \leq \bar{r}_i \qquad \forall i \in I, l \in \mathbb{L} \tag{5.1h}$$

$$\text{(Flow Def.)} \qquad \mathcal{F}_{i,t+l} = \mathcal{R}_{i,t+l} + \mathcal{L}_{i,t+l} \qquad \forall i \in I, l \in \mathbb{L} \tag{5.1i}$$

$$\text{(Floods)} \qquad 0 \leq \mathcal{E}_{j,t+l} \qquad \forall j \in J, l \in \mathbb{L}. \tag{5.1j}$$

For $\tau = t + l$ with $l \in \mathbb{L}$, the decisions $\mathcal{S}_\tau, \mathcal{L}_\tau$, $\mathcal{R}_\tau$, $\mathcal{F}_\tau$, $\mathcal{E}_\tau$, respectively represent storage (hm^3) at the end of period τ, average spillage (hm^3/period) over time τ, average releases (productive water discharge) (hm^3/period) over time τ, average total flow (hm^3/period) over time τ and average floods (hm^3) over time τ [1]. The sum of the spillage and the releases is the total flow. The aggregate decision vector $\mathcal{X}_\tau = (\mathcal{S}_\tau^\top, \mathcal{L}_\tau^\top, \mathcal{R}_\tau^\top, \mathcal{F}_\tau^\top, \mathcal{E}_\tau^\top)^\top$ is simply the stacking of each decision at time τ. I and J respectively represent the set of plants and the set of reservoirs.

Constraints (5.1b) ensure that the total storage remains within tolerable limits $\underline{s}_j$ and $\bar{s}_j$ for all reservoirs $j \in J$. (5.1c) are simply flow conservation (water balance) constraints ensuring that water released upstream eventually reaches downstream reservoirs where λ_{il} represents the fraction of water released from plant i to reach the unique downstream reservoir after l periods. The constant δ_i^{max} represents the number of periods required for 100% of the water released to reach the downstream reservoir. Similarly, δ_i^{min} represents the number of periods before any water released upstream reaches the downstream reservoir. We consider $\sum_{l=\delta_i^{min}}^{\delta_i^{max}} \lambda_{il} = 1, \forall i$, but we could also use $\sum_{l=\delta_i^{min}}^{\delta_i^{max}} \lambda_{il} < 1$ to model evaporation or other losses.

1. Throughout the text, we refer interchangeably to storage as volumes, releases as discharge or turbined outflow and plants as powerhouses.

At time $t = 1$, we have $\mathcal{S}_{jt-1}(\xi) = s_{j0}$ where s_{j0} represents the fixed known amount of water (in hm^3) in reservoir j at the beginning of the time horizon. We use the approximation $\xi_{jt} = \alpha_{jt}\xi_t$ where ξ_t and ξ_{jt} represent the total inflows over all reservoirs and the inflows at a specific reservoir j at time t where α_{jt} is the average fraction of total inflows at time t entering reservoir j. Inflows are expressed in hm^3/period and namely come from natural precipitations, run-off and spring thaw.

Constraints (5.1d) ensure that flows are within limits $\underline{f}_{it}, \bar{f}_{it}$ while constraints (5.1f) ensure that the total flow deviation at a given plant i does not exceed a pre-specified threshold Δ_i from one period to the next. Constraints (5.1e) bound the maximum amount of water that can be unproductively spilled at plant $i \in I^{evac} \subset I$ for a given storage in the upstream reservoir, where I^{evac} is the set of plants with such constraints. This relationship is given by the non-linear function $\mathscr{C}(\cdot)$ called an evacuation curve. Following Gauvin et al. (2017), we approximate it by an affine function in our model, but maintain the true structure in our simulations. We successfully tested more precise representations by introducing binary decisions, but found that our simple affine approximation has negligible impact on the final results.

Equations (5.1g) ensure respect of absolute upper and lower bounds. These constraints are determined by specific physical characteristic of given plants. Constraints (5.1h) ensure the releases at plant i are within prescribed bounds $\underline{r}_{it}, \bar{r}_{it}$, $\forall t$. These are based on navigation and flood safety thresholds as well as agreements with riparian communities. Finally, (5.1i) defines the total flow as the sum of unproductive spillage and releases.

We explicitly model spills and releases separately for illustrative purposes and to facilitate the extension to the case where hydroelectrical generation and spills should also be optimized. However, this is not necessary for flood-minimizing purposes, and it would be possible to reduce the dimensionality of the problem by aggregating both decisions to consider only total flows. In this situation, the upper bounds on total flows become a function of the upstream storage due to the evacuation curves.

Constraints (5.1j) define overflows (floods) with respect to the critical storage levels $\underline{s}_{jt}, \bar{s}_{jt}$ and represent quantities we wish to minimize. Since the bounds are taken with respect to a given useful reservoir storage, underflows (droughts) can theoretically exist at a reservoir j, but are physically bounded by a small constant $0 < \epsilon_j$ and are highly undesirable. We therefore chose to forbid them, even if they can be added to our model very straightforwardly.

We consider a convex quadratic penalization function : $\mathcal{K}_{jt}(\mathcal{E}_{jt}) = a\mathcal{E}_{jt}^2$ with $a > 0$ in the objective (5.3) to reflect the fact that larger floods have increasingly disastrous consequences. It is possible to extend our formulation to more general functions such as the ones considered

in Pianosi and Soncini-Sessa (2009) by introducing binary decisions to model discontinuities and non-convexities. However, this would lead to a mixed integer program and would affect the tractability of our model and ultimately its use by decision makers. Indeed, our model may be solved frequently to analyse what-if scenarios and therefore needs to be rapidly solvable by commercial off-the-shelf software. Moreover, Section 5.5 shows it leads to good empirical results.

The parameters $\kappa_{j,t} > 0$ in (5.3) represent the relative weight of each reservoir at a given time. We define the set J^{crit} representing reservoirs located near riparian populations and high risks of floods as well as those with critical importance. We then fix $\kappa_{jt} = W\kappa$, $\forall j \in J^{crit}, t$ for some large $W \in \mathbb{N}$ and some fixed $\kappa > 0$ and impose that the sum of the weights equal one so that we have a convex combination. For our problem, the dichotomy between critical and non-critical reservoirs is unequivocal, but we could easily adapt this to more intricate cases.

5.2.2 Sources of uncertainty

We focus on the stochasticity surrounding inflows which is one of the main factors of the risks of floods and droughts. We therefore study the discrete time stochastic process $\{\xi_t\}$ representing total inflows over the river system at time $t \in \mathbb{Z}$ where the $\xi_t : \Omega \to \mathbb{R}$ are real valued random variables bounded and non-negative with probability one. Although we focus on minimizing the risk of floods for a finite time interval $\mathbb{T}$, the process $\{\xi_t\}_{t\in\mathbb{Z}}$ extends infinitely far in the past and the future. We denote the mean and variance at time $t \in \mathbb{Z}$ as $\mathrm{E}[\xi_t] = \mu_t$ and $\mathrm{E}[(\xi_t - \mu_t)^2] = \sigma_t^2$.

5.2.3 General framework

We consider a dynamic setting were the true realization of the random process $\{\xi_t\}$ is gradually revealed as time unfolds over the horizon of T days (Delage and Iancu; Shapiro et al., 2009). A sequence of controls $\{\mathcal{X}_t\}$ must be fixed at each stage $t \in \mathbb{T}$ after observing the realized history $\xi_{[t-1]}$, but before knowing the future random variables. Once $\xi_{[t-1]}$ is known, $\mathcal{X}_t$ can be implemented to yield the actual decisions $\mathcal{X}_t(\xi_{[t-1]}) \in \mathbb{R}^{n_t}$ where $n_t \in \mathbb{N}$ represents the number of decisions to be taken at time t. This decision process can be visualized in Figure 5.1.

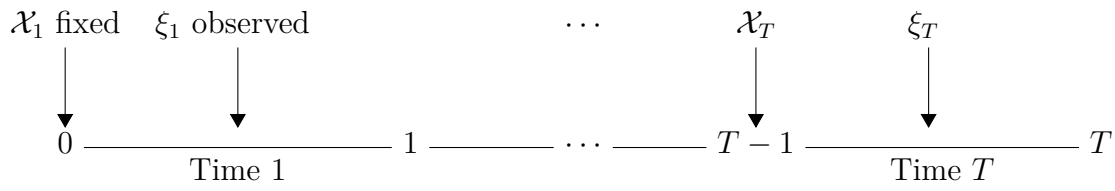

Figure 5.1 Sequential Dynamic Decision Process

5.2.4 Affine decision rules

To consider this uncertainty and formulate the stochastic version of model (5.1), we change the decision variables $\mathcal{X}$ to functions $\mathcal{X}(\cdot)$ of the underlying stochastic process. We specifically consider simple affine functions of the uncertain inflows, which are a restricted class of possibly suboptimal policies. These decision rules were popularized in dynamic/adjustable robust optimization models by Ben-Tal et al. (2004) and have gained considerable attention in the recent years, namely in the field of energy (Gauvin et al., 2017; Apparigliato, 2008; Lorca and Sun, 2015; Rocha and Kuhn, 2012). Although they do not guarantee optimality in general, they often lead to good-quality solutions that can be obtained very efficiently.

At the beginning of period $t \in \mathbb{T}$, we let $K_\tau = \{n_{\tau-1}+1, \cdots, n_{\tau-1}+n_\tau\}$ represent the indices associated with decisions at time $\tau = t+l$ for lead times $l \in \mathbb{L}$ and horizon $L \in \{0, \cdots, T-t+1\}$. We can then express affine functions of the future inflow vector $\xi \equiv \xi_{[t,t+L-1]}$ in the form :

$$\mathcal{X}_{k,t+l}(\xi) = \mathcal{X}^0_{k,t+l} + \sum_{l'=0}^{L-1} \mathcal{X}^{l'}_{k,t+l}\xi_{t+l'}, \tag{5.2}$$

where $\mathcal{X}^{l'}_{k,t+l} \in \mathbb{R}$ for $k \in K_{t+l}$, $t \in \mathbb{T}$ and $l, l' \in \mathbb{L}$. The decisions pertaining to spillage, flow and discharge represent real implementable decisions used by river operators to control the dynamics of the river system. Decisions that must be implemented at some future time $\tau = t+l$ for $l \in \mathbb{L}$ can therefore only depend on information up to time $\tau - 1$. We therefore require that : $\mathcal{X}^{l'}_{k,t+l} = 0$, $\forall l' \geq l$ and $k \in K^{impl}_{t+l} \subset K_{t+l}$ where the K^{impl}_{t+l} represents the set of indices associated to such decisions at time $t+l$.

However, storage and flood are analysis variables that are only meant to track the evolution of the system. As such, decisions at time τ can also depend of information up to time τ and we enforce $\mathcal{X}^{l'}_{k,t+l} = 0$, $\forall l' \geq l+1$ and $k \in K_{t+l} \setminus K^{impl}_{t+l}$. In both cases, the impact of the past observed $\xi_{[t-1]}$ is reflected implicitly in $\mathcal{X}^0_{k,t+l}$.

The stochastic lookahead model (5.1) based on affine decision rules provides the important advantage of avoiding the curse of dimensionality as well as the discretization of the random variables and decisions required to solve the problem through its dynamic programming recursions. This observation is important since finding a feasible solution to the deterministic multi-stage problem (5.1b) - (5.1j) at time t requires considering $\min\{\delta_{i^-}^{max}, t-1\}$ past water releases for all $i^- \in I^-(j)$ and all $j \in J$ as well as $|J|$ initial storages. For the values $|J| = 5$ and $\sum_{j\in J}\sum_{i^-\in I^-(j)}(\delta_{i^-}^{max} - \delta_{i^-}^{min}) = 13$ used in our case study, this leads to a state of dimension 18, which is already extremely demanding for classical dynamic programming. Considering the uncertainty and persistence of inflows further increases the state dimension. For instance, an autoregressive model of order $p \in \mathbb{N}$ would require an additional p states.

Moreover, it becomes much easier to consider constraints such as (5.1c) and (5.1f) that involve decisions in multiple periods. Finally, Section 5.3.7 shows that by leveraging techniques from robust optimization, we are able to formulate each lookahead problem as a tractable convex program that can be solved in a single "forward" phase and that makes limited distribution assumptions compared with SDP or stochastic programming based on scenario trees.

5.2.5 Minimizing flood risk

We define the risk of floods at the beginning of time t as the conditional expected value of the penalized aggregated flood $\sum_{j=1}^{J}\sum_{l=0}^{L-1}\kappa_{jt+l}\mathcal{K}_{jt+l}(\mathcal{E}_{j,t+l}(\xi))$ over the L period look-ahead horizon given the information up to time $t-1$ where $\xi \equiv \xi_{[t,t+L-1]}$. We therefore modify the objective (5.1a) to :

$$(SP_t) \qquad \min_{\mathcal{X}(\cdot)} \quad \mathrm{E}\left[\sum_{j=1}^{J}\sum_{l=0}^{L-1}\kappa_{jt+l}\mathcal{K}_{jt+l}(\mathcal{E}_{j,t+l}(\xi))|\mathscr{F}_{t-1}\right]. \tag{5.3}$$

The conditional expectation offers the advantage of being the simplest consistent and coherent dynamic risk measure (Riedel, 2004). It also enjoys various interesting properties such as linearity and has been extensively studied in stochastic processes as well as reservoir management applications. We refer to the stochastic version of problem (5.1) at time t as SP_t.

5.3 Exploiting time series models and linear decision rules

5.3.1 General inflow representation

The quality of the solutions returned by solving the lookahead problem SP_t crucially depends on the representation of the underlying stochastic process $\{\xi_t\}$. Assuming simple independent time series will likely lead to poor quality solutions in the presence of significant serial correlation. However, we also want to maintain the tractability of the overall linear program considering affine decision rules.

In order to achieve these conflicting objectives, we assume that at the beginning of each time $t \in \mathbb{T}$, the future inflows $\xi \equiv \xi_{[t,t+L-1]}$ over the next L days lie within the set Ξ_t with probability a.s.. We also assume that this random vector can be represented as an affine function of some vector $\varrho \equiv \varrho_{[t,t+L-1]}$ of real valued, uncorrelated, zero mean, second order stationary random variables. This affine representation will allow us to construct the serial dependence empirically observed in the $\xi_{[t,t+L-1]}$ while exploiting the convenient statistical properties of the $\varrho_{[t,t+L-1]}$.

For modeling and tractability purposes, we will also assume that the support of the $\varrho_{[t,t+L-1]}$ is a bounded polyhedron in $\mathbb{R}^L$. More specifically, we assume that there exists $U_t, V_t \in \mathbb{R}^{L\times L}, W_t \in \mathbb{R}^{c\times L}$ and $u_t, v_t \in \mathbb{R}^L, w_t \in \mathbb{R}^c$ for some $c \in \mathbb{N}$ such that the following representation, which shares important similarities with the one presented in Lorca and Sun (2015), holds :

$$\Xi_t = \left\{ \xi \in \mathbb{R}^L \;\middle|\; \begin{array}{ll} \exists \zeta, \varrho \in \mathbb{R}^L & \text{(5.4a)} \\ \xi = U_t\zeta + u_t & \text{(5.4b)} \\ \zeta = V_t\varrho + v_t & \text{(5.4c)} \\ W_t\varrho \leq w_t & \text{(5.4d)} \end{array} \right\}$$

We assume that for any ξ, there exists ϱ, ζ such that the representation (5.4) holds. Although this is not required in general, we naturally assume for the rest of the paper that the ϱ, ζ are unique. This is enforced by requiring that both U_t and V_t be of full rank L. Although this condition may seem strong, we will see that it arises automatically in important contexts. As detailed in Section (5.3.7), a possible advantage of this unicity is the possibility to define a correspondence between affine functions of ξ and affine functions of the uncorrelated ϱ.

We will further assume that both U_t and V_t, as well as their inverse U_t^{-1}, V_t^{-1} when these exists, are lower triangular. This requirement is related to the concept of non-anticipativity

discussed previously and intuitively ensures that each ζ_t and ξ_t is only a function of the past $\varrho_{[t]}$. We assume that ζ_t, ξ_t and ϱ_t are perfectly known and observable at each time t.

As will become clear in the next Section, the relationship between $\zeta \equiv \zeta_{[t,t+L-1]}$ and ϱ as well as the structure of V_t and v_t play a very important role in our analysis. We therefore explicitly consider the intermediary $\mathbb{R}^L$ dimensional vector random ζ even if we could directly substitute (5.4c) into (5.4b). More specifically, we will consider the case where the $\{\zeta_t\}$ follow well-known autoregressive moving average (ARMA) time series models.

In this context, (5.4b) can be naturally interpreted as a way to remove a deterministic trend, seasonal component or perform other preprocessing as is commonly done in time series analysis (Brockwell and Davis, 1987; Box et al., 2008). The ϱ can then also be seen as the residuals obtained after fitting a specific ARMA model to the ζ. We assume the random vector ϱ lies within the polyhedron $\{\varrho \in \mathbb{R}^L : W_t\varrho \ \leq w_t\}$ and show there exists systematic and sound probabilistic methods to construct these polyhedral sets. We begin by assuming that the ϱ_t are serially independent, but then generalize the approach by considering generalized autoregressive conditional heteroscedastic (GARCH) time series models.

Finally, using the theory of ARMA and GARCH models, we will show how the representation (5.4) can be updated to more adequately reflect the random environment as we move forward in time and new data is progressively observed.

5.3.2 Considering general ARMA models

This Section assumes some basic familiarity with linear AR(I)MA time series. For further details, we refer to the classic texts (Box et al., 2008; Brockwell and Davis, 1987). ARMA models are simple linear and discrete time series that filter the serial dependency and output white noise. These time series model allow us to express future random variables as an affine function of independent random variables. Furthermore, their parsimonious representation, practical importance, successful utilization in past hydrological models for stochastic reservoir optimization and linear structure make them as invaluable stochastic model that can be incorporated directly in our multi-stage stochastic problem. We assume that at each time $t \in \mathbb{Z}$, the real valued ζ_t satisfy the equation :

$$\phi(B)\zeta_t = \theta(B)\varrho_t, \tag{5.5}$$

for some $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$ with $\phi_i, \theta_i \in \mathbb{R}$ for $p, q \in \mathbb{N}$ where

B represents the backshift operator acting on time indices such that $B^p\zeta_t = \zeta_{t-p}$ for all $t, p \in \mathbb{Z}$ (Box et al., 2008; Brockwell and Davis, 1987). We suppose the ϱ_t are independent identically distributed zero mean and $\mathscr{F}_t$-measurable random variables. In order to guarantee second order stationarity, we require that the process autocovariance function $\gamma(l) = \mathrm{E}[\varrho_t \varrho_{t+l}]$ depend only on $l \in \mathbb{Z}$ and in particular that the variance $\gamma(0) = \sigma_\varrho^2$ be constant across time. For any ARMA process respecting equation (5.5), we can equivalently write :

$$\zeta_t = \psi(B)\varrho_t, \tag{5.6}$$

where $\phi(B)\psi(B) = \theta(B)$ for some $\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$. We can therefore express ζ_t as an infinite linear combination of past $\{\varrho_\tau\}_{\tau=t,t-1,\dots}$. We specifically consider the case where $\sum_{i=0}^{\infty} |\psi_i| < \infty$. In this case, we say that the process $\{\zeta_t\}$ is stable. Since we assume the ϱ are bounded with probability 1, the representation (5.6) is essentially unique (Brockwell and Davis, 1987). We can relax the assumption that the original ζ_t are stable if $(1-B)^d\zeta_t = \zeta'_t$ for some $d \in \mathbb{N}$ such that the ζ'_t are stable. It follows that our framework also applies to ARIMA models of any integer integration order $d \in \mathbb{N}$.

Example *Consider the simple AR(1) model where $\zeta_t = \phi\zeta_{t-1} + \varrho_t$ holds $\forall t \in \mathbb{Z}$. In this case $(1-\phi B)\sum_{i=0}^{\infty} \psi_i B^i = 1$ if and only if $\psi_i = \phi^i, i \in \mathbb{Z}_+$. It follows the process is stable if and only if $|\phi| < 1$. If $\phi = 1$, then taking $\zeta'_t = \zeta_t - \zeta_{t-1}$ implies that $\zeta'_t = \varrho_t, \forall t \in \mathbb{Z}$ are iid random variables that satisfy our assumptions.* □

Representation (5.6) is particularly useful when forecasting the future values of ζ_t given the information available at time $t \in \mathbb{Z}$. For any $t \in \mathbb{Z}, l \in \mathbb{Z}_+$ we have :

$$\zeta_{t+l} = \underbrace{\sum_{j=l}^{\infty} \psi_j \varrho_{t+l-j}}_{\hat{\zeta}_t(l)} + \underbrace{\sum_{j=0}^{l-1} \psi_j \varrho_{t+l-j}}_{\rho_t(l)}, \tag{5.7}$$

where $\hat{\zeta}_t(l) = \mathrm{E}[\zeta_{t+l}|\mathscr{F}_t]$ is the forecast and $\rho_t(l) = \zeta_{t+l} - \hat{\zeta}_t(l)$ is the forecast error. This follows by the linearity of conditional expectation together with $\mathrm{E}[\varrho_{t+l}|\mathscr{F}_t] = \varrho_{t+l}$ if $l \in \mathbb{Z}_-$ and $\mathrm{E}[\varrho_{t+l}|\mathscr{F}_t] = \mathrm{E}[\varrho_{t+l}] = 0$ otherwise. For any $t \in \mathbb{Z}, l \in \mathbb{Z}_+$, we observe that $\rho_t(l)$ is $\mathscr{F}_{t+l}$ measurable while $\hat{\zeta}_t(l)$ is $\mathscr{F}_t$ measurable. For $l \in \mathbb{Z}_-$, the forecast coincides with the actual observed random variable and therefore $\rho_t(l) = 0$ while $\hat{\zeta}_t(l) = \zeta_{t+l}$. The conditional

expectation $\mathrm{E}\left[\zeta_{t+l}|\mathscr{F}_t\right]$ is a natural choice of forecast as it represents the minimum mean squared error estimator of ζ_{t+l} given the information up to time $t \in \mathbb{Z}$ for $l \in \mathbb{Z}_+$ (Brockwell and Davis, 1987).

Example (continued) *For the stable AR(1) model and any $t \in \mathbb{Z}, l \in \mathbb{Z}_+$, we have $\hat{\zeta}_{t+l} = \phi^l \zeta_t$ and $\rho_t(l) = \sum_{i=0}^{l-1} \phi^i \varrho_{t+l-i}$.* $\square$

If we set $\rho_{t-1,L} \equiv (\rho_{t-1}(1), \cdots, \rho_{t-1}(L))^\top$ for any $t \in \mathbb{Z}$, we can then express the forecast error vector $\rho_{t-1,L}$ as a linear function of the independent $\varrho_{[t,t+L-1]}$. More specifically, the following holds for all $L \in \{1, \cdots, T-t+1\}$:

$$\rho_{t-1,L} = V_t \varrho_{[t,t+L-1]}, \tag{5.8}$$

where $V_t \equiv V \in \mathbb{R}^{L \times L}$ is the following invertible and lower triangular square matrix, which is constant across all $t \in \mathbb{Z}$:

$$V = \begin{pmatrix} 1 & & \cdots & 0 \\ \psi_1 & 1 & & \vdots \\ \vdots & & \ddots & \\ \psi_{L-1} & \cdots & \psi_1 & 1 \end{pmatrix} \tag{5.9}$$

We then have the equality :

$$\begin{aligned} \zeta_{[t,t+L-1]} &= \hat{\zeta}_{t-1,L} + \rho_{t-1,L} && (5.10) \\ &= \hat{\zeta}_{t-1,L} + V \varrho_{[t,t+L-1]}, && (5.11) \end{aligned}$$

where $\hat{\zeta}_{t-1,L} \equiv (\hat{\zeta}_{t-1}(1), \cdots, \hat{\zeta}_{t-1}(L))^\top$ corresponds to v_t in the representation (5.4). The structure of V as well as the definition of $\hat{\zeta}_{t-1,L}$ and $\rho_{t-1,L}$ ensures that the representation is unique. Putting all these together yields a crisp representation of the inflows $\xi_{[t,t+L-1]}$ as an affine function of $\varrho_{[t,t+L-1]}$ whose structure depends on the past observations through $\hat{\zeta}_{t-1,L}$:

$$\xi_{[t,t+L-1]} = U_t \zeta_{[t,t+L-1]} + u_t \tag{5.12}$$
$$= U_t(\hat{\zeta}_{t-1,L} + \rho_{t-1,L}) + u_t \tag{5.13}$$
$$= U_t(\hat{\zeta}_{t-1,L} + V\varrho_{[t,t+L-1]}) + u_t. \tag{5.14}$$

Example (continued) *Consider the discrete time process* $\{\xi_t\}$ *with mean* $E[\xi_t] = \mu_t, \forall t$ *such that* $U_t = I$, $u_t = \mu_t$ *and therefore* $\xi_t = \zeta_t + \mu_t$ *holds* $\forall t$. *Also suppose that* $\{\zeta_t\}$ *follows a stable AR(1) model. For any* $t \in \mathbb{Z}$, $l \in \mathbb{Z}_+$ *and* $L \in \mathbb{N}$, *we can represent* $\xi_{[t,t+L-1]}$ *as :*

$$\begin{pmatrix} \xi_t \\ \xi_{t+1} \\ \vdots \\ \xi_{t+L-1} \end{pmatrix} = \begin{pmatrix} \phi\zeta_{t-1} \\ \phi^2\zeta_{t-1} \\ \vdots \\ \phi^L\zeta_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & & \cdots & 0 \\ \phi^1 & 1 & & \vdots \\ \vdots & & \ddots & \\ \phi^{L-1} & \cdots & \phi^1 & 1 \end{pmatrix} \begin{pmatrix} \varrho_t \\ \varrho_{t+1} \\ \vdots \\ \varrho_{t+L-1} \end{pmatrix} + \begin{pmatrix} \mu_t \\ \mu_{t+1} \\ \vdots \\ \mu_{t+L-1} \end{pmatrix}. \tag{5.15}$$

□

The affine representation (5.14) reveals that the $\xi_{[t,t+L-1]}$ vector is completely determined by $\varrho_{[t,t+L-1]}$. We therefore set $\sigma(\varrho_s; s \leq t) = \mathscr{F}_t, \forall t \in \mathbb{T}$ which reflects the fact that observing $\varrho_{[t-1]}$ at the beginning of time $t \in \mathbb{Z}$ gives us all the information necessary to apply the real implementable policies at times $1, 2, \cdots, t-1$.

5.3.3 Support of the joint distribution of the $\{\varrho_t\}$

Having defined the relationship between ϱ, ζ and ξ, we now study the support hypothesis for the ϱ vector. For any $L \in \mathbb{N}$, we specifically assume that the support of $\varrho_{[t,t+l-1]}$, which also corresponds to the set $\{\varrho \in \mathbb{R}^L : W_t\varrho \leq w_t\}$ described in (5.4), is a bounded polyhedron in $\mathbb{R}^L$ given by the intersection of the following two polyhedrons [2] :

$$\mathscr{B}^{\infty}_{L,\epsilon} = \{\varrho \ \in \mathbb{R}^L : |\varrho_i|\sigma_\varrho^{-1} \leq (L\epsilon)^{1/2}, i = 1, \cdots, L\} \tag{5.16a}$$
$$\mathscr{B}^{1}_{L,\epsilon} = \{\varrho \ \in \mathbb{R}^L : \sum_{i=1}^{L} |\varrho_i|\sigma_\varrho^{-1} \leq L\epsilon^{1/2}\}. \tag{5.16b}$$

2. The set $\mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}$ is not strictly speaking a polyhedron since (5.16a)-(5.16b) involve the non-linear absolute value function. Nonetheless, lifting this set using the commonly used decomposition $\varrho_i = \varrho_i^+ - \varrho_i^-$

Although limiting, the use of this polyhedron is motivated by a sound probabilistic interpretation. If the $\{\varrho_t\}$ are (possibly unbounded) iid random variables with constant variance σ_ϱ^2, then for $t, L \in \mathbb{N}$, the covariance matrix of $\tilde{\varrho} \equiv \varrho_{[t,t+L-1]}$ is simply the positive definite matrix $\Sigma_{\varrho,L} = \sigma_\varrho^2 I_L$ where I_L is the $L \times L$ identity matrix. Therefore, if tr denotes the (linear) trace operator and $\tilde{\varrho}$ is a $\mathbb{R}^L$ valued random vector, Markov's inequality gives us :

$$\mathbb{P}(\tilde{\varrho}^\top \Sigma_{\varrho,L}^{-1} \tilde{\varrho} > L\epsilon) \leq \mathrm{E}\left[\tilde{\varrho}^\top \Sigma_{\varrho,L}^{-1} \tilde{\varrho}\right] (L\epsilon)^{-1} \tag{5.17}$$

$$= tr(\Sigma_{\varrho,L}^{-1} \mathrm{E}\left[\tilde{\varrho}\tilde{\varrho}^\top\right])(L\epsilon)^{-1} \tag{5.18}$$

$$= \epsilon^{-1}. \tag{5.19}$$

The polytope $\mathscr{B}_{L,\epsilon}^\infty \cap \mathscr{B}_{L,\epsilon}^1$ contains the ellipsoid $\mathscr{B}_{L,\epsilon}^2 = \{\varrho \in \mathbb{R}^L : \varrho^\top \Sigma_{\varrho,L}^{-1} \varrho \leq L\epsilon\}$ (see 5.8.1 for more details). It follows that $\mathbb{P}(\tilde{\varrho} \in \mathscr{B}_{L,\epsilon}^\infty \cap \mathscr{B}_{L,\epsilon}^1) \geq 1 - \epsilon^{-1}$ for all $t, L \in \mathbb{N}$. Although we could use $\mathscr{B}_{L,\epsilon}^2$ as the support, preliminary tests demonstrate that it is preferable to consider the exterior polyhedral approximation (5.16) to speed computations.

For our particular case, we consider a large ϵ and reasonably assume $\mathbb{P}(\tilde{\varrho} \in \mathscr{B}_{L,\epsilon}^\infty \cap \mathscr{B}_{L,\epsilon}^1) = 1$. If the $\{\varrho_t\}$ are essentially bounded iid random variables with constant variance σ_ϱ^2, then this assumption is not restrictive, as we can always find an ϵ that respects this hypotheses. Since inflows can only take a finite value with $\mathbb{P}$ a.s., the essential boundedness assumption is realistic.

In a robust optimization context, this polyhedral support would be referred to as an "uncertainty set" since it represents the set of possible values that the random variables can take. Our approach can be straightforwardly extended to more complex polytopes and it will retain polynomial complexity if it is extended to the intersection of polytopes and second order or semi-definite cones (Ben-Tal et al., 2009).

The polyhedron defined by (5.16a) - (5.16b) is namely influenced by the lead time L and extending far into the future intuitively leads to a larger set. We also note that if the calibrated time series model fits the in-sample data poorly, then the estimated σ_ϱ will be large and hence the size of the support will increase.

This distribution-free uncertainty set offers important advantages in the presence of limited data availability. Indeed, since it provides valid bounds for any distribution, the ellipsoid $\{\tilde{\varrho}^\top \Sigma_{\varrho,L}^{-1} \tilde{\varrho} \leq L\epsilon\}$ and the enclosing polyhedral set we consider are rather large compared with

and $|\varrho_i| = \varrho_i^+ + \varrho_i^-$ with $\varrho_i^+, \varrho_i^- \geq 0, \forall i$ yields a polyhedron where each projected point lies in the original $\mathscr{B}_{L,\epsilon}^\infty \cap \mathscr{B}_{L,\epsilon}^1$ (Ben-Tal et al., 2009).

similar sets derived from assuming specific distributions.

For instance, if we assume $\tilde{\varrho}$ is jointly normal, then $\mathbb{P}(\tilde{\varrho}^\top \Sigma^{-1} \tilde{\varrho} \leq L\epsilon) = F_\chi(L\epsilon, L)$, where $F_\chi(x, k)$ is the cdf of a chi-square distribution with k degrees of freedom evaluated at x. For the values of ϵ, L used in our numerical study, $F_\chi(L\epsilon, L)$ can be considerably larger than $1 - \epsilon^{-1}$. For instance, taking $\epsilon = 16$ and $L = 7$, we get : $F_\chi(L\epsilon, L) \approx 1 > 0.9375 = 1 - \epsilon^{-1}$.

This additional conservativeness proves beneficial with limited data as we do not construct overly small and optimistic uncertainty sets. We also observe empirically that the distribution-free set captures a much larger proportion of the out-of-sample data than a similar set with a fixed distribution.

5.3.4 Support of the (conditional) joint distribution of the $\{\xi_t\}$

Building on these assumptions, it follows that the polyhedron given by (5.4) represents the conditional support of $\xi_{[t,t+L-1]}$ for $L \in \{1, \cdots, T-t+1\}$ given the past observed $\varrho_{[t-1]}$. The set Ξ_t implicitly depends on past $\varrho_{[t-1]}$ through $v_t \equiv \hat{\zeta}_{t-1,L}$ and is therefore perfectly known at the beginning of time t. Given our past hypothesis on the support of ϱ as well as knowledge of $\varrho_{[t-1]}$, the future inflows $\xi_{[t,t+L-1]}$ reside within Ξ_t with probability 1. In robust optimization terminology, this polytope can be seen as a *dynamic* uncertainty set determining the possible realizations of the random vector $\xi_{[t,t+L-1]}$ based on past observations.

5.3.5 Considering heteroscedasticity

We now relax the assumption that the $\{\varrho_t\}$ are independent and consider the case when the residual $\{\varrho_t\}$ follow a GARCH(m, s) model where $m, s \in \mathbb{N}$. GARCH processes have proven useful for numerous applications, namely in the field of finance where the assumption of constant conditional variance is not always verified and large shocks tend to be followed by periods of increased volatility (Garcia et al., 2005; Lamoureux and Lastrapes, 1990). As we discuss in Section 5.5.4, this is also empirically observed in daily inflows. For further details on GARCH processes, see Box et al. (2008); Bollerslev (1986).

In this case, we still assume that for any $t \in \mathbb{Z}$, $\mathrm{E}[\varrho_{t+l}|\mathscr{F}_t] = 0, \forall l \in \mathbb{Z}_+$, $\mathrm{E}[\varrho_t] = 0$ and $\mathrm{E}[\varrho_t^2] = \sigma_\varrho^2$. We also have $\mathrm{E}[\varrho_{t+k}\varrho_{t+l}|\mathscr{F}_{t-1}] = \mathrm{E}[\varrho_{t+l}\mathrm{E}[\varrho_{t+k}|\mathscr{F}_{t+l}]\,|\mathscr{F}_{t-1}] = 0$ for $l < k \in \mathbb{L}$. In other words, the $\{\varrho_t\}$ are *uncorrelated* zero-mean random variables with constant variance σ_a^2. However, they are not necessarily *independent* since they are linked through the following relation for all $t \in \mathbb{Z}$:

$$\hat{\sigma}^2_{t-1}(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho^2_{t-i} + \sum_{j=1}^{s} \beta_j \hat{\sigma}^2_{t-1-j}(1) \ , \tag{5.20}$$

where $\hat{\sigma}^2_t(l) = \mathrm{E}\left[\varrho^2_{t+l} | \mathscr{F}_t\right]$ [3]for $l \in \mathbb{N}$ and $\alpha_0, \alpha_i, \beta_j \geq 0, \forall i, j$ to ensure non-negativity of the conditional variance. In 5.8.4, we show that under standard stationarity assumptions on the ϱ_t, the squared shocks ϱ^2_t satisfy the difference equation :

$$\hat{\phi}(B)\left(\varrho^2_t - \sigma^2_a\right) = \hat{\theta}(B)\nu_t. \tag{5.21}$$

where the $\{\nu_t\}$ are zero mean uncorrelated random variables with $\nu_t = \varrho^2_t - \hat{\sigma}^2_{t-1}(1)$. The polynomials are given by $\hat{\theta}(B) = 1 + \sum_{j=1}^{s} \hat{\theta}_j B^j$ and $\hat{\phi}(B) = 1 - \sum_{i=1}^{\max\{s,m\}} \hat{\phi}_i B^i$ for $\hat{\theta}_i, \hat{\phi}_i \in \mathbb{R}$. We can then find $\hat{\psi}(B) = \sum_{i=0}^{\infty} \hat{\psi}_i B^i$ such that $\hat{\phi}(B)\hat{\psi}(B) = \hat{\theta}(B)$ and we therefore have :

$$\varrho^2_t = \sigma^2_\varrho + \hat{\psi}(B)\nu_t. \tag{5.22}$$

Similarly to the ARMA model, assuming $\sum_{i=0}^{\infty} |\hat{\psi}_i| < \infty$ with ν_t essentially bounded ensures that the representation is unique. Taking the conditional expectation on both sides of equation (5.22) at the beginning of time t allows us to obtain an expression for the conditional variance reminiscent of the conditional expectation described in Section 5.3.2. For $l \in \mathbb{N}$, we specifically have :

$$\hat{\sigma}^2_t(l) = \sigma^2_\varrho + \sum_{j=l}^{\infty} \hat{\psi}_j \nu_{t+l-j}. \tag{5.23}$$

We can then write the conditional covariance matrix as $\Sigma_{\varrho,L,t-1} = diag(\hat{\sigma}^2_{t-1}(1), \cdots, \hat{\sigma}^2_{t-1}(L))$, where each conditional variance $\hat{\sigma}^2_t(l)$ is a known value at time t for $l \in \{1, \cdots, L\}$. The matrix remains diagonal since the $\{\varrho_t\}$ are uncorrelated (details are provided in 5.8.4). In Section 5.3.7, we show how to obtain a better estimation of the expected flood penalties by exploiting this information. Note that we keep the same support described in the preceding Section and

3. Although non-standard, we adopt the notation $\hat{\sigma}^2_t(l)$ to maintain the coherence with the past sections and to highlight the similarities with the conditional expectation of ζ_{t+l} for some $t \in \mathbb{Z}$ and $l \in \mathbb{Z}_+$ given $\mathscr{F}_t$, which we denoted $\hat{\zeta}_t(l)$.

that only the objective changes compared to the initial ARMA model.

When the residuals follow a GARCH process, large past errors will boost $\hat{\sigma}_{t-1}(i)$ for all $i \in \{1, \cdots, L\}$ over the look-ahead horizon. On one hand, a stochastic process that follows the fitted time series model very closely will therefore generate small conditional and unconditional variances. On the other hand, a poor time series model will not only lead to imprecise forecasts and a large unconditional σ_ϱ, but also to extremely large $\hat{\sigma}_{t-1}(i)$, which may make it more difficult to find solutions without any storage violations. However considering a variable conditional variance may help adapt to inflows that deviate from the forecast.

5.3.6 Additional modelling considerations

The representation (5.4) assumes that $\xi_t \in \mathbb{R}, \forall t \in \mathbb{T}$, which is not physically meaningful since inflows must always be non-negative. We can correct this by using the affine representation (5.14). More precisely, we can impose $\xi_{[t,t+L-1]} \in \mathbb{R}^L_+$ by requiring that with $\mathbb{P}$ a.s., the future random vector $\varrho_{[t,t+L-1]}$ reside within the following polyhedron, which is perfectly known at the beginning of time $t \in \mathbb{T}$:

$$\left\{\varrho \in \mathbb{R}^L \ : U_t V \varrho \geq -(U_t \hat{\zeta}_{t-1,L} + u_t)\right\} \tag{5.24a}$$

The additional structure imposed by equation (5.24a) affects the independence and the uncorrelation of the ϱ_τ, but this hypothesis may not be severely violated if the constraint is not binding "very often", which is the case in our numerical experiments.

If the violation of independence seems severely violated and this negatively impacts performance, our modelling approach can still be used by ignoring (5.24a). This simply leads to a more conservative modelling of the uncertainty and may be considered an exterior polyhedral uncertain set on the "true" uncertainty set representing the support.

5.3.7 Optimizing the conditional expected flood penalties with affine decision rules

We suppose that the inflow model (5.4) is correct and that the $\{\varrho\}$ are possibly heteroscedastic. Next, we consider our lookahead model SP_t at the beginning of time $t \in \mathbb{T}$ with arbitrary decision rules, after observing the past $\varrho_{[t-1]}$ and $\xi_{[t-1]}$ for a horizon of $L \in \{1, \cdots, T-t+1\}$ days and where $\xi \equiv \xi_{[t,t+L-1]}$ in a more condensed and abstract form :

$$\min_{\mathcal{X}} \mathrm{E}\left[\sum_{l=0}^{L-1} \mathcal{X}_{t+l}^{\top}(\xi) G_{t+l} \mathcal{X}_{t+l}(\xi) | \mathscr{F}_{t-1}\right] \tag{5.25a}$$

$$\text{s.t.} \quad \sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}_{t+\bar{l}}(\xi) \geq C_{t+l}^{\Delta} \xi + C_{t+l}^{0} \qquad \forall l \in \mathbb{L} \qquad \mathbb{P} \text{ a.s.} \tag{5.25b}$$

$$\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \mathcal{X}_{t+\bar{l}}(\xi) = \hat{E}_{t+l}^{\Delta} \xi + \hat{E}_{t+l}^{0} \qquad \forall l \in \mathbb{L} \qquad \mathbb{P} \text{ a.s.}, \tag{5.25c}$$

for some $A_{t+l,t+\bar{l}} \in \mathbb{R}^{m_{t+l}^{\geqslant} \times n_{t+\bar{l}}}$, $C_{t+l}^{\Delta} \in \mathbb{R}^{m_{t+l}^{\geqslant} \times L}, C_{t+l}^{0} \in \mathbb{R}^{m_{t+l}^{\geqslant}}, D_{t+l}^{\Delta} \in \mathbb{R}^{m_{t+l}^{=} \times L}, E_{t+l}^{0} \in \mathbb{R}^{m_{t+l}^{=}}, \mathcal{X}_{t+l}(\xi) \in \mathbb{R}^{n_{t+l}}$, $G_{t+l} \in \mathbb{S}_{+}^{n_{t+l}}$ for $l \in \mathbb{L}$, for $m_t^{\geqslant}, m_t^{=}, n_t \in \mathbb{N}$.

We now limit ourselves to affine decision rules and rewrite (5.2) in the more compact form :

$$\mathcal{X}_t(\xi) = \mathcal{X}_t^0 + \mathcal{X}_t^{\Delta} \xi, \tag{5.26}$$

for $\xi \equiv \xi_{[t,t+L-1]} \in \mathbb{R}^L, \mathcal{X}_t^0 \in \mathbb{R}^{n_t}$ and $\mathcal{X}_t^{\Delta} \in \mathbb{R}^{n_t \times L}$ whose structure depends on the non-anticipativity of the respective decisions. Assuming that constraint (5.24a) does not significantly affect the hypothesis that $\mathrm{E}\left[\varrho_{t+l} | \mathscr{F}_{t-1}\right] = 0, \forall l \in \mathbb{Z}_+$ and setting $\varrho \equiv \varrho_{[t,t+L-1]}$, the objective value (5.25a) then becomes :

$$\mathrm{E}\left[\sum_{l=0}^{L-1} \mathcal{X}_{t+l}^{\top}(\xi) G_{t+l} \mathcal{X}_{t+l}(\xi) | \mathscr{F}_{t-1}\right] = \sum_{l=0}^{L-1} \mathrm{E}\left[\hat{\mathcal{X}}_{t+l}^{0,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^{0} | \mathscr{F}_{t-1}\right] + \sum_{l=0}^{L-1} \mathrm{E}\left[\varrho^{\top} \hat{\mathcal{X}}_{t+l}^{\Delta,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^{\Delta} \varrho | \mathscr{F}_{t-1}\right] \tag{5.27}$$

$$= \sum_{l=0}^{L-1} \hat{\mathcal{X}}_{t+l}^{0,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^{0} + \sum_{l=0}^{L-1} tr(\mathrm{E}\left[\varrho \varrho^{\top} | \mathscr{F}_{t-1}\right] \hat{\mathcal{X}}_{t+l}^{\Delta,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^{\Delta}) \tag{5.28}$$

$$= \sum_{l=0}^{L-1} \hat{\mathcal{X}}_{t+l}^{0,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^{0} + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}_{t+l}^{\Delta} \Sigma_{\varrho,L,t-1}^{1/2})^{\top} G_{t+l} (\hat{\mathcal{X}}_{t+l}^{\Delta} \Sigma_{\varrho,L,t-1}^{1/2}), \tag{5.29}$$

where :

$$\hat{\mathcal{X}}^0_{t+l} = \mathcal{X}^0_{t+l} + \mathcal{X}^{\Delta}_{t+l} u_t + \mathcal{X}^{\Delta}_{t+l} U_t \hat{\zeta}_{t-1,L} \tag{5.30}$$

$$\hat{\mathcal{X}}^{\Delta}_{t+l} = \mathcal{X}^{\Delta}_{t+l} U_t V \tag{5.31}$$

$$\Sigma^{1/2}_{\varrho,L,t-1} = diag(\hat{\sigma}_{t-1}(1), \cdots, \hat{\sigma}_{t-1}(L)), \tag{5.32}$$

and we have used the following decomposition from Section 5.3.2 :

$$\xi_{[t,t+L-1]} = U_t(\hat{\zeta}_{t-1,L} + V\varrho_{[t,t+L-1]}) + u_t. \tag{5.33}$$

In the case of homoscedasticity, (5.32) is simply replaced by its unconditional version $\Sigma^{1/2}_{\varrho,L} = \sigma_\varrho I_L$. Using the definition of the conditional joint support of $\xi_{[t,t+L-1]}$ given by (5.4) , we see that for any $f : \mathbb{R}^L \to \mathbb{R}$ and $k \in \mathbb{R}$, $f(\xi) \geq k, \forall \xi \in \Xi_t \Rightarrow \mathbb{P}(f(\xi_{[t,t+L-1]}) \geq k | \mathscr{F}_{t-1}) = 1$ with $\mathbb{P}$ a.s.. With (5.30)-(5.32), problem (5.25) can therefore be written as :

$$\min_{\mathcal{X}^0, \mathcal{X}^{\Delta}} \sum_{l=0}^{L-1} \hat{\mathcal{X}}^{0,\top}_{t+l} G_{t+l} \hat{\mathcal{X}}^0_{t+l} + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}^{\Delta}_{t+l} \Sigma^{1/2}_{\varrho,L,t-1})^\top G_{t+l} (\hat{\mathcal{X}}^{\Delta}_{t+l} \Sigma^{1/2}_{\varrho,L,t-1}) \tag{5.34a}$$

$$\text{s.t. } (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}^{\Delta}_{t+\bar{l}} - C^{\Delta}_{t+l})\xi \geq -\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}^0_{t+\bar{l}} + C^0_{t+l} \qquad \forall l \in \mathbb{L}, \qquad \forall \xi \in \Xi_t \tag{5.34b}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \mathcal{X}^{\Delta}_{t+\bar{l}} - \hat{E}^{\Delta}_{t+l})\xi = -\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \mathcal{X}^0_{t+\bar{l}} + \hat{E}^0_{t+l} \qquad \forall l \in \mathbb{L}, \qquad \forall \xi \in \Xi_t. \tag{5.34c}$$

Its optimal solution represents an upper bound on (5.25a) - (5.25c) with arbitrary decision rules because we limit ourselves to affine functions. Since Ξ_t is a polyhedron, we can handle the constraints (5.34b) through robust optimization techniques and linear programming duality (Ben-Tal et al., 2009). We also reformulate (5.34c) without the ξ by exploiting the fact that Ξ_t is full dimensional and contains 0. The resulting feasible domain of the robust equivalent is therefore polyhedral. Since equation (5.34a) is of the convex quadratic type, the deterministic equivalent is a large (minimization) second-order cone program (SOCP), which can be solved very efficiently with interior point solvers. We give more details in 5.8.2 and 5.8.3.

From this derivation, we also see that the optimal value of problem (5.34a)-(5.34b) will be an upper bound on the conditional expectation of (weighted) floods over the horizon $t, \cdots, t+L-1$ for *any* distribution of the ϱ_t provided that the true support remains within

the polyhedral support defined by (5.16a) - (5.16b), constraint (5.24a) does not affect the hypothesis of $\mathrm{E}\left[\varrho_{t+l}|\mathscr{F}_{t-1}\right] = 0, \forall l \in \mathbb{Z}_+$ and the structure of the ARMA and GARCH models are correct.

As can be seen from (5.30) - (5.31) and (5.32), given $\hat{\mathcal{X}}^0_{t+l}, \hat{\mathcal{X}}^\Delta_{t+l}$ that solve problem (5.34), it is straightforward to obtain an affine mapping in terms of the ξ when U_t and V are invertible. Indeed, we only need to compute $\hat{\mathcal{X}}^\Delta_{t+l} V^{-1} U_t^{-1}$ and then subtract the constant $\mathcal{X}^\Delta_{t+l} u_t + \mathcal{X}^\Delta_{t+l} U_t \hat{\zeta}_{t-1,L}$ from $\hat{\mathcal{X}}^0_{t+l}$. This might be beneficial if river operators prefer to have decisions expressed directly in terms of the inflows rather than the residuals ϱ.

5.4 Monte Carlo simulation and rolling horizon framework

Solving the stochastic version of problem (5.1a)-(5.1j) with affine decision rules at the beginning of time 1 for $L = T$ provides an upper bound on the value of the "true" problem over the horizon $\mathbb{T}$ when various hypotheses on ϱ_t and ξ_t are verified. However, simulating the behaviour of the system with a given distribution can give a better assessment of the real performance of these decisions. Using random variables that violate the support assumptions also provides interesting robustness tests.

Furthermore, the full potential of ARMA and GARCH models crucially depends on the ability to assimilate new data as it is progressively revealed. Using time series model to construct a single forecast at time 1 for the entire horizon may lead to more realistic uncertainty modelling than considering an uncertainty set that completely ignores the serial correlation. However, computing new forecasts as inflows are progressively revealed will increase the precision of our model. We capture this fact by considering a rolling horizon framework.

A rolling horizon framework also reflects the true behaviour of river operators who must take decisions now at the beginning of time t for each future time $t, \cdots, t+L-1$ by considering some horizon $L \in \mathbb{N}$ [4]and will update the parameters of the model as the time horizon progresses and new information on inflows and other random variables is revealed. Section 5.5.3 also illustrates that the consequences of bad forecasts can be mitigated by adapting past previsions.

The rolling horizon simulation works as follows. We simulate a T dimensional trajectory of zero-mean, constant unconditional variance and uncorrelated random variables $\varrho^s_{[T]} \equiv (\varrho_{s,1}, \cdots, \varrho_{s,T})^\top$ which together with the fixed and deterministic initial inflows ξ_0 [5]completely determine inflows $\xi^s_{[T]} \equiv (\xi_{s,1}, \cdots, \xi_{s,T})^\top$. For a given scenario s at the beginning of time t,

4. We consider a rolling rather than receding horizon approach. More specifically, the future time horizon $L \in \mathbb{N}$ is held constant at each optimization and does not decrease. This reflects the true approach used by river operators.

after having observed the past history $\xi^s_{[t-1]}$ and $\varrho^s_{[t-1]}$, the initial storage $\mathcal{S}_{j,t-1}(\xi^s_{[T]})$ and the past water releases $\mathcal{F}_{i,t'}(\xi^s_{[T]}), t' \leq t-1, i \in I$, but before knowing the future inflows $\xi^s_{[t,T]}$, we compute the conditional expectation and variance of the inflows using the ARMA and GARCH models.

We then solve the affine problem at time t by considering the (future) time horizon $t, \cdots, t+L-1$ and by taking the deterministic equivalent when considering affine decision rules with the conditional support Ξ_t. We then implement the first period decisions, observe the total random inflow $\xi_{s,t}$ during time t, compute the linear combination of actual floods, update $\mathcal{S}_{jt}(\xi^s_{[T]})$ and the past water releases $\mathcal{F}_{t',i}(\xi^s_{[T]}), t' \leq t$; $i \in I$ and solve SP_{t+1}. We repeat this step for times $t = 1, \cdots, T-L+1$ for each of S sample trajectories.

5.5 Case study

5.5.1 The river system

We apply our methodology to the Gatineau river in Québec. This hydro electrical complex is part of the larger Outaouais river basin and is managed by Hydro-Québec, the largest hydroelectricity producer in Canada (Hydro-Québec, 2012). It is composed of 3 run-of-the-river plants with relatively small productive capacity and 5 reservoirs, of which only Baskatong and Cabonga have significant capacity (see figure 5.2).

5. We fix $\xi_0 = \mathrm{E}[\xi_0]$ as the unconditional mean inflow at time 0.

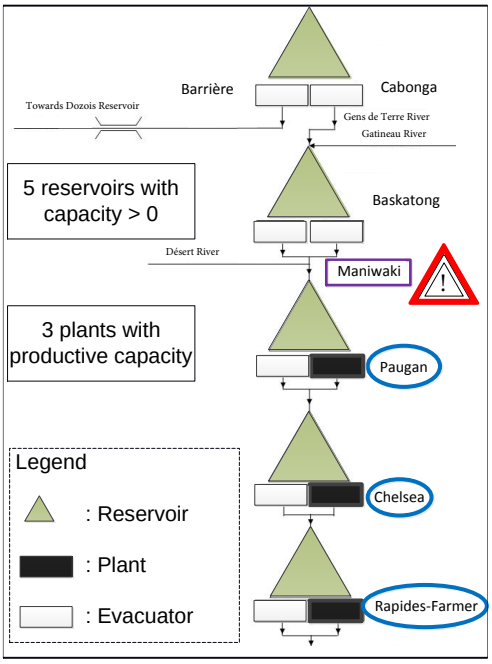

Figure 5.2 Simplified representation of the Gatineau river system . Figure taken from Gauvin et al. (2017).

The Gatineau represents an excellent case study as it runs near the small town of Maniwaki which is subject to high risks of flooding, particularly during the spring freshet. Indeed, the city has suffered 4 significant floods in 1929, 1936, 1947 and 1974. Moreover, the reservoir system has relatively tight operational constraints on flows and storages. If the head reservoirs are not sufficiently emptied before the freshet, there is a significant risk of disrupting normal operating conditions and flooding (Gauvin et al., 2017).

The Baskatong reservoir is the largest of the broader Outaouais-Gatineau catchment and plays a critical role in the management of the river. It is used to manage risk of floods during the freshet period as well as droughts during the summer months. It has even been used to control baseflow at the greater Montreal region several hundreds of kilometres downstream. As such, respect of minimum and maximum storage threshold is essential for river operators. During certain months of the year including the freshet period, flood management decision are taken at daily time steps and we therefore consider daily periods for the rest of the numerical study.

Although the Gatineau also serves recreational, ecological and hydro-electicity generation purposes, flood management remains the most important consideration due to the proximity of human settlements. Nonetheless, it would be useful to adopt a more holistic and integrated

approach in future work in the same vein as works such as Castelletti et al. (2012) and Galelli et al. (2014).

5.5.2 Historical daily inflows

Statistical properties of the total inflows process over the entire river $\{\xi_t\}_t$ also provide an interesting application of our general framework. As figure 5.3 illustrates, water inflows are particularly important and volatile during the months of March through April (freshet) as snow melts. There is a second surge during fall caused by greater precipitations and finally there are very little liquid inflows during the winter months.

Figure 5.3 also emphasizes the differences between the 6 in-sample years 1999-2004 used to calibrate our model with the 6 out-of-sample years 2008-2013 used for validation. Due to the dryer years 1999-2000, the average in-sample inflows underestimate the true average. As exemplified by the large deviation at the beginning of the year 2008, the actual inflows can significantly differ from the historical mean.

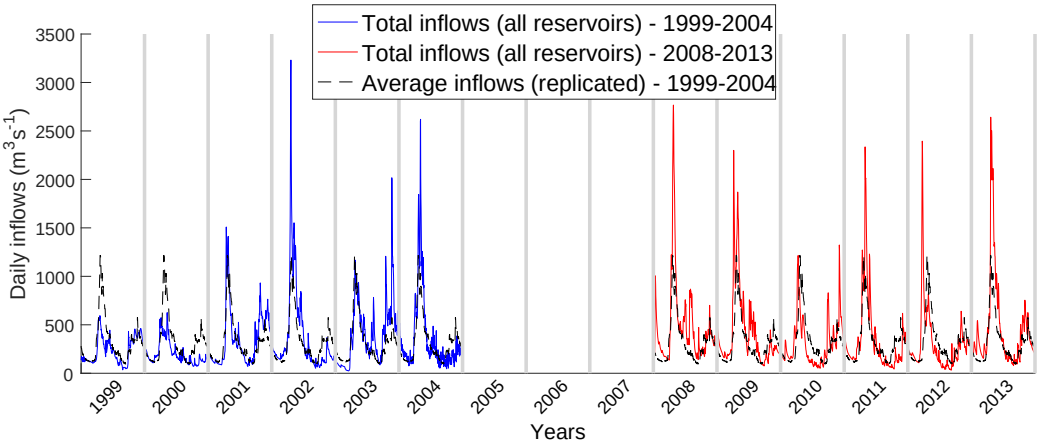

Figure 5.3 Sample inflows for 1999-2004 and 2008-2013 (12 years) with "in-sample" mean [6]

Due to the limited data available to fit our time-series model, we believe it is particularly useful to consider the distribution-free uncertainty set (5.16). Indeed, it's conservativeness compared with a similar set derived with specific distributional assumptions proves useful to capture new out-of-sample observations. Nonetheless, considering more data points in our calibration would surely improve the skill of our forecast and in turn the solutions found by our stochastic model.

6. The years 2005-2007 were not provided by Hydro-Québec.

5.5.3 Forecasting daily inflows

We estimate μ_t and σ_t^2 for the inflows $\{\xi_t\}_t$ at time $t \in \mathbb{T}$ by using the sample mean and variance at that time. We then follow Gjelsvik et al. (2010) and Maceira and Damázio (2004) and fix $\zeta_t = \frac{\xi_t - \mu_t}{\sigma_t}$, which makes sense as raw inflows can be assumed to have constant mean and variance at the same time of the year. In this case, $u_t = \mathrm{E}\left[\xi_{[t,t+L-1]}\right] = (\mu_t, \cdots, \mu_{t+L-1})^\top \in \mathbb{R}^L$ and $U_t^{-1} = diag(\sigma_t^{-1}, \cdots, \sigma_{t+L-1}^{-1}) \in \mathbb{R}^{L\times L}$ for our affine representation $\xi_{[t,t+L-1]} = U_t\zeta_{[t,t+L-1]} + u_t$ with $\mathbb{P}$ a.s..

Alternative ways to deal with the seasonal component of the time series include Fourier analysis to identify a deterministic trend and the use of seasonal difference operators $\Delta^s\xi_t = \xi_t - \xi_{t-s}$ for some seasonal offset $s \in \mathbb{N}$ (Salas et al., 1980; Box et al., 2008). These are all compatible with our framework at no additional complexity, although they do not all lead to invertible affine mappings $\xi_{[t,t+L-1]} = U_t\zeta_{[t,t+L-1]} + u_t$.

We next consider Box-Jenkins methodology (Box et al., 2008). More specifically, we identify a limited subset of candidate ARMA(p, q) models for the ζ_t, based on the empirical partial autocorrelation (PACF) and autocorrelation (ACF) functions. We then fit these candidate models using the econometrics toolbox from Matlab and select the best based on the AIC criterion (Akaike, 1973).

Results suggest that ζ_t approximately follow a ARMA$(1, 1)$ process, that is $\phi(B)\zeta_t = \theta(B)\varrho_t$ where $\phi(B) = 1 - \phi B$ and $\theta(B) = 1 + \theta B$. We point out that this is the best forecast as suggested by the data. Although our method can handle arbitrary time series models, tests with ARMA(p, q) models of order $p, q \in \{1, \cdots, 4\}$ namely provide worst AIC criterion. We also point out that the moving average term implicitly considers arbitrarily long delays since a finite an stable pure moving average model can be equivalently expressed as an infinite order autoregressive process (Box et al., 2008).

The residuals resemble zero-mean independent white noise. The Ljung-Box Q-test also indicates that at the 5% significance level, there is *not* enough evidence to reject the null hypothesis that the residuals are *not* autocorrelated. Based on the data sample, we obtain the following estimates : $\sigma_\varrho = 0.30, \phi = 0.96, \theta = -0.13$. Since $|\phi| < 1$, we can express $\zeta_t = \psi(B)\varrho_t$ with $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

Although the initial forecast made at time 0 provides a much better estimate than the historical expected value for small lead times, it does not perform very well for medium lead times (see figure 5.4). Indeed, low and high inflows in the beginning of the freshet resulted in very small and large forecasted inflows compared with the actual inflows for the rest of the period for the years 2002 and 1999, respectively. However, as the dotted lines reveal,

repeatedly forecasting the future values as new data becomes available in a rolling horizon fashion provides much better predictive power. [7].

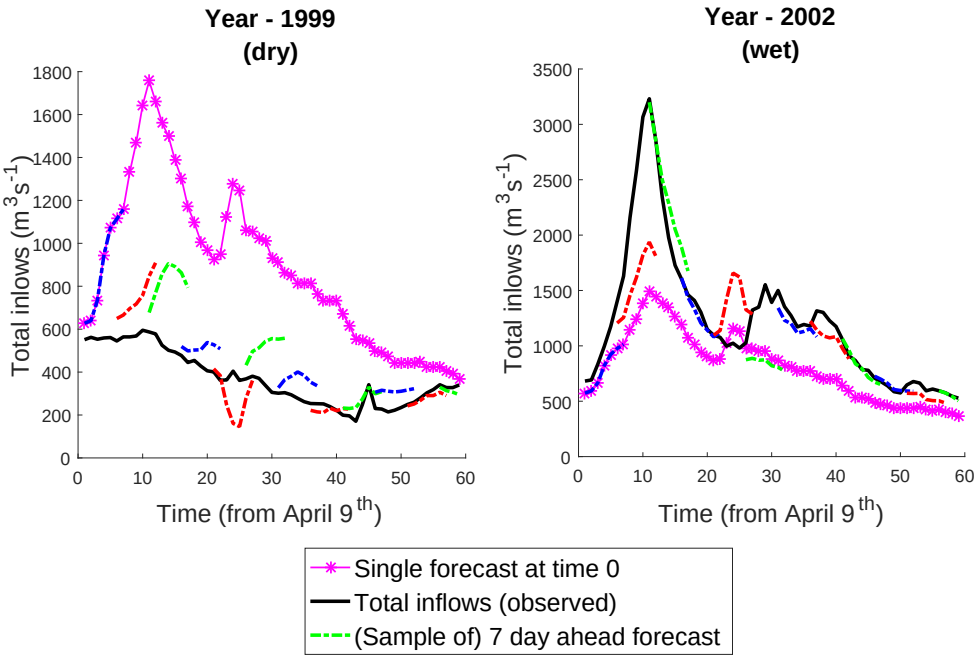

Figure 5.4 Comparing simple forecasts for 1999 & 2002

5.5.4 Heteroscedastic inflows

After fitting the ARMA(1,1) model, the residual $\{\varrho_t\}$ do not seem to display any serial correlations (see figure 5.5). However, at the 5% level of significance, the Ljung-Box test on the squared residuals reveals the presence of heteroscedasticity (Ljung and Box, 1978). Visual inspection corroborates this conclusion as there are clear signs of volatility clustering.

7. The blue, red and green broken lines represent 7-day lookahead forecasts made at times $t = 0, 5, 10, \cdots, 55$ during a 60 day period starting at May 9th. Only the first 5 forecasted values are shown for the last forecast made at time 55 since we plot a horizon of 60 days. We did not plot all 60 7-day lookahead forecasts because there are too many and we chose to use 3 colors since this makes it easier to identify different forecasts. The magenta dotted lines represent the 60 day lookahead forecast made at time 0. The full black line represents the actual observed inflows over the 60 day period.

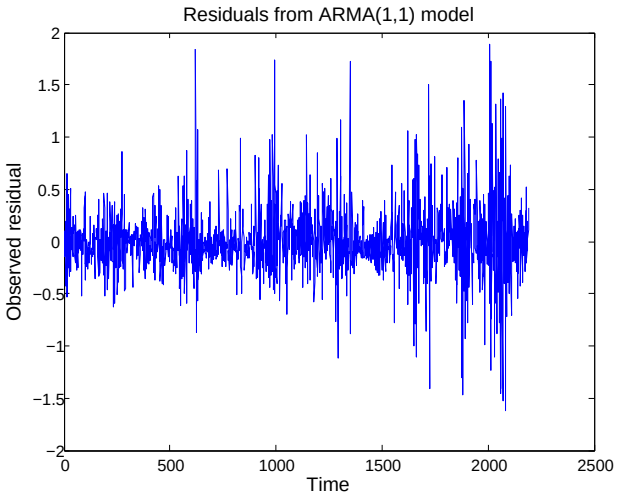

Figure 5.5 Residuals from ARMA(1,1) model

We find that the residual $\{\varrho_t\}$ approximately follow a GARCH$(1,1)$ model with the following estimates : $\alpha_0 = 0.01, \alpha_1 = 0.14, \beta_1 = 0.84$. We can show that the coefficients in (5.23) are given by $\hat{\psi}_i = \hat{\phi}_1^{i-1}(\hat{\phi} + \hat{\theta}) = (\alpha_1 + \beta_1)^{i-1}\alpha_1$ since $\hat{\phi}_i = (\alpha_1 + \beta_1)$ and $\hat{\theta}_1 = -\beta_1$. It follows that $\sum_{i=0}^{\infty} |\hat{\psi}_i| < \infty$. We point out that this parsimonious GARCH process has proven very effective to forecast financial time-series data (Lamoureux and Lastrapes, 1990; Hansen and Lunde, 2005).

5.5.5 Comparing forecasts

Sections 5.3.3 - 5.3.5 suggests that bad forecasts will lead to poor quality solutions. In order to evaluate the forecasting accuracy of our time series model $\{\hat{\xi}_{frcst,t+l}\}_{l=0}^{L-1}$, we therefore compare it with the naive static forecast consisting of the historical daily mean $\{\hat{\xi}_{naive,t}\}_{l=0}^{L-1}$ where $\hat{\xi}_{naive,t+1} = \mu_{t+l}, \forall l \in \mathbb{L}$.. We then use the standard skill score, also known as the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970; Gupta et al., 2009). This statistical measure can be expressed as : $1 - \mathrm{E}\left[MSE_{frcst,t,L}\right]\mathrm{E}\left[MSE_{naive,t,L}\right]^{-1}$ where $MSE_{frcst,t,L} = L^{-1}\sum_{l=0}^{L-1}\mathrm{E}\left[(\hat{\xi}_{frcst,t+l} - \xi_{t+l})^2\right]$ and $MSE_{naive,t,L} = L^{-1}\sum_{l=0}^{L-1}\mathrm{E}\left[(\xi_{t+l} - \mu_{t+l})^2\right]$. A skill score of 1 indicates a perfect forecast with zero mean square error while a skill score of $-\infty$ indicates a forecast doing infinitely worse than the reference forecast. Positive, null and negative skill score respectively indicate superior, identical and inferior performance relative to the reference forecast.

5.5.6 Numerical experiments

To validate the practical importance of our multi-stage stochastic program based on ARMA and GARCH time series and affine decision rules, we perform a series of tests based on different inflow generators for the Gatineau river. We consider a total horizon of $T = 59$ days beginning at the start of the spring freshet and use daily time steps. We chose this granularity because it corresponds to the approach used by river operators to manage the river during the freshet. We concentrate on the freshet as it represents the most difficult and interesting case for our problem. However, we only report results for the first 30 days, which are also the most volatile and wet.

All experiments are performed by solving the problem in a rolling horizon fashion. Each optimization problem uses a lookahead period of $L = 30$ days and we perform 30 model resolutions for each simulated inflow trajectory. For each resolution, we only consider uncertainty on a limited time horizon of 7 days and use the deterministic mean inflows for the remaining 23 days. This reduces the impact of poor quality forecasts and speeds up computations.

To test the robustness of the different methods and evaluate if our approach could be used to avoid emptying the head reservoirs before spring as is currently done, we always consider an initial storage that is considerably higher than the normal operating conditions for this period. Results were obtained by assuming no past water releases at time 0.

We used a $\epsilon^{1/2} = 4$ which generates relatively large supports. Simulations were run on 2000 randomly generated scenarios and took several hours (> 3 hours) to complete although most individual problems are solved in less than 5 seconds. Problems were solved using AMPL with solver CPLEX 12.5 on computers with 16.0 GB RAM and i7 CPU's @ 3.4 GHz.

Unfortunately, this large ϵ combined with very wet or dry conditions as well as the tight flow bounds at Maniwaki can make our model infeasible. When this is the case, we allow flow bounds violations in addition to floods to get complete recourse.

For simulated inflows, we present the sample CVaR_α of floods for $\alpha = 10^{-2}n, n = 0, 1, \cdots, 100$ over the entire time horizon of $T = 30$ days where for the continuous random variable X, we define $\text{CVaR}_\alpha = \text{E}[X|X > \text{VaR}_\alpha(X)]$ and $\text{VaR}_\alpha(X) = q_X(\alpha) = \inf\{t : \mathbb{P}(X \leq t) \geq \alpha\}$ is the α quantile for some $\alpha \in (0, 1)$ (Pflug, 2000).

As in Gauvin et al. (2017), we choose to represent the empirical CVaR rather than the empirical distributions since it allows rapid graphical comparison of the expected value ($\alpha = 0$) and worst case ($\alpha = 1$). Moreover, as mentioned namely in Bertsimas et al. (2004), CVaR is consistent with second order stochastic dominance which is of prime concern for risk averse decision makers.

To evaluate the suboptimality of our policies, we also plot the expected value of the 'wait-and-see' solution with perfect foresight. Although very simple, Section 5.5.6 reveals that the bound can be relatively tight in some cases and therefore that our models perform well under some scenarios.

5.5.7 Simulations with ARMA(1,1) & GARCH(1,1) generator

We first assume that the $\{\varrho_t\}$ are zero-mean uncorrelated normal variables with unconditional variance σ^2_ϱ. Hence these random variables violate our assumption of boundedness made in Section 5.3.3. Choosing random variables with support defined exactly by (5.16b)-(5.16a) leads to virtually no floods.

We then suppose the true inflow process $\{\bar{\xi}_t\}$ is given by $\bar{\xi}_t = \sigma_t \bar{\zeta}_t + \mu_t$ where the $\{\bar{\zeta}_t\}$ follow an ARMA(1,1) model with parameters $\bar{\theta}$ and $\bar{\phi}$. We specifically fix $\bar{\phi} = \phi$ and $\theta = \bar{\theta} + \epsilon_\theta$ where ϕ and θ are the values used by our model through prior calibration.

Under these hypothesis, the skill of the ARMA forecast relative to the naive forecast will be non-negative if and only if $-0.87 \leq \epsilon_\theta \leq 0.87$. This is independent of GARCH effects. Details are provided in 5.8.5.

As the first 2 graphs of figure 5.6 illustrate, taking $-0.87 \leq \epsilon_\theta \leq 0.87$ with ARMA and GARCH models unsurprisingly leads to the greatest flood reductions while only considering an ARMA model still improves the solution quality compared to the naive forecast. The last graph with $\epsilon_\theta = -5$ reveals that even with negative skill, it may pay off to consider ARMA or the combined GARCH and ARMA models when the true process follows the same structure as those used by the model.

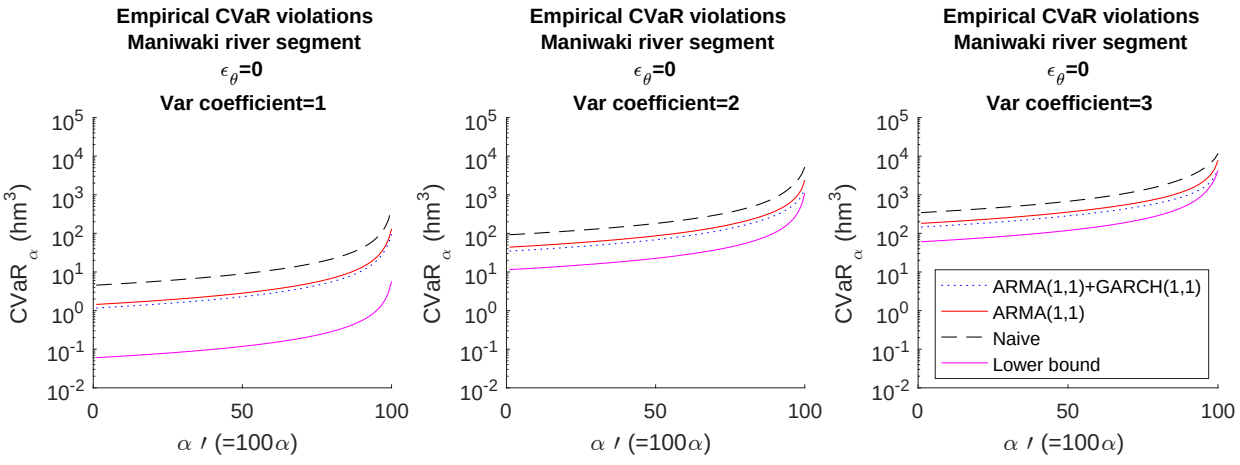

Figure 5.6 Influence of reduced forecast skill

These conclusions remain valid when we increase the volatility of the true inflow process. Figure 5.7 namely illustrates the impact of taking $\bar{\xi}_t = \bar{\epsilon}^{1/2}\sigma_t\bar{\zeta}_t + \mu_t$ with the variance coefficient $\bar{\epsilon}^{1/2} = 1, 2, 3$ on the floods when the time series model used by our multi-stage stochastic problem used to represent $\{\bar{\zeta}_t\}$ is exactly the same as the one used by the inflow generator to represent $\{\zeta_t\}$.

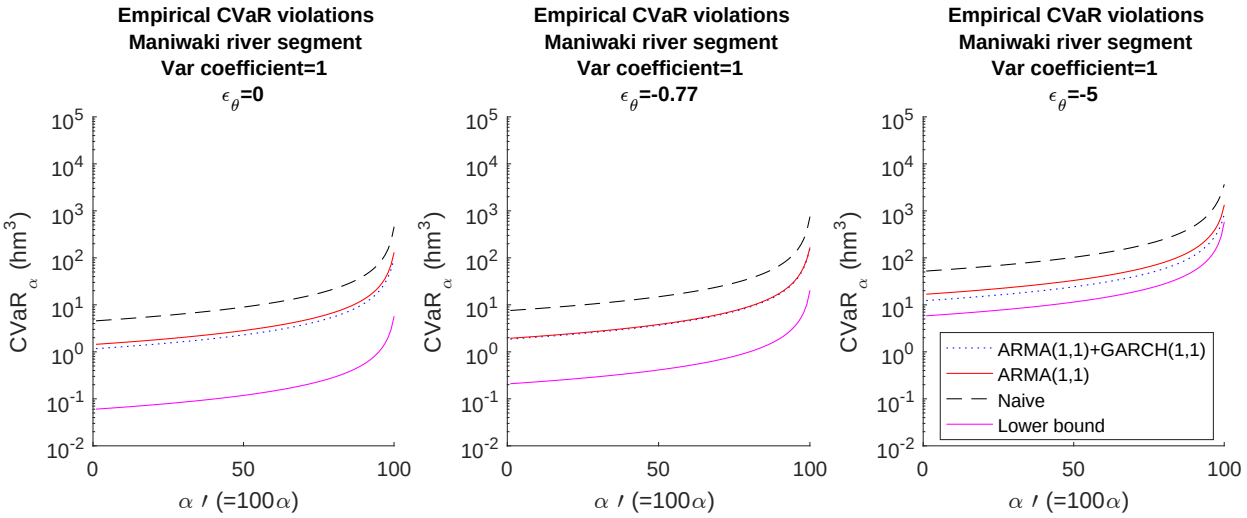

Figure 5.7 Influence of increased unconditional variance

As illustrated in figures 5.6 and 5.7, the gap between the different suboptimal policies and the lower bounds provided by the wait-and-see solution decreases as inflows become more volatile and extreme. This is likely a consequence of the fact that persistently high inflows will invariably lead to high floods, even with perfect foresight. In this case, the model therefore has little manoeuvrability left.

5.5.8 Simulation with different time series model

To test the robustness of our model with dynamic uncertainty sets, we also consider a more complicated SARIMA$(2, 0, 1) \times (0, 1, 1)$ generator which does *not* rely on the affine decomposition $\frac{\xi_t - \mu_t}{\sigma_t} = \zeta_t, \forall t$ assumed by our optimization problem (Box et al., 2008).

Figure 5.8 suggests that even in this case, our approach based on dynamic uncertainty sets performs better than the one based on the naive forecast. It is encouraging to observe such robustness with respect to time series structure. In this case, it makes virtually no difference whether we add GARCH effects or not. Although the ARMA+GARCH model dominates the ARMA model, the curves are nearly confounded.

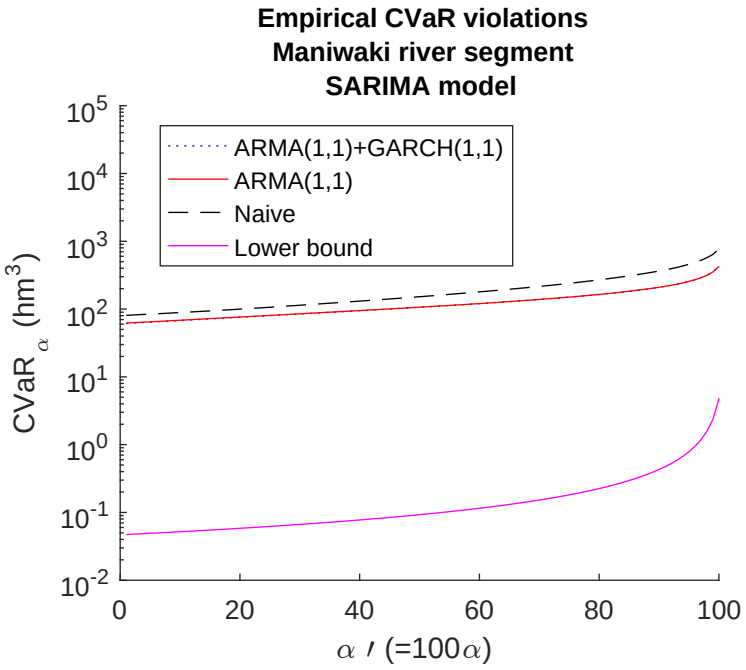

Figure 5.8 Influence of different time series structure

5.5.9 Real Scenarios

We now consider $\bar{S} = 12$ real historical inflows provided by Hydro-Québec. The first 6 years (1999-2004) were used as in-sample scenarios to determine sample moments and calibrate the time series-model. The remaining 6 years (2008-2013) were used to validate the robustness of our approach.

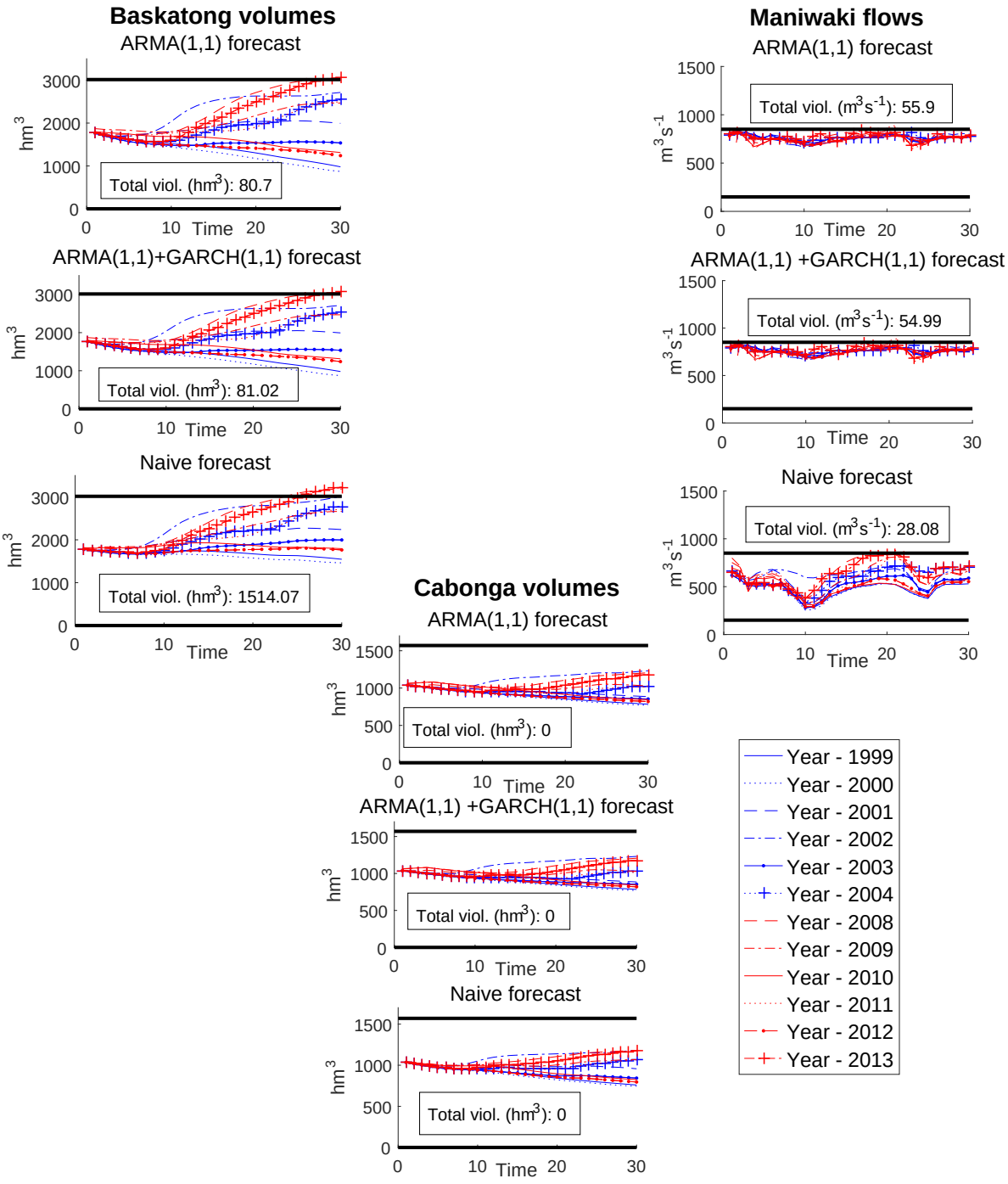

Figure 5.9 Simulation results for 1999-2004 & 2008-2013 (12 years)

Figure 5.9 shows violations of storage for the two large head reservoirs Baskatong and Cabonga as well as flow bounds violations for the town of Maniwaki. Upper and lower bounds are indicated by solid black lines. The figure indicates that violations occur at out-of-sample

years (in red) while in-sample years (in blue) respect all constraints. The two wet years 2008 and 2013 are particularly problematic. The plots show that the models with ARMA and GARCH forecast usually yield overall superior performance compared with the naive forecast. Indeed, these policies produce significantly less storage violations at the expense of only slightly increased flow violations. The GARCH + ARMA and ARMA models give qualitatively similar policies.

Our results also highlight the value of parsimony and the use of criteria such as AIC that penalize models with numerous parameters. Complex models with higher order lags tended to produce very bad results when tested on the out-of-sample historical inflows since the effect of bad forecasts had lasting consequences. For instance, the abnormally high inflows at the beginning of the freshet in 2008 resulted in residuals that were more than 10 times higher than the (in-sample) standard deviation at the beginning of the period. Models based on ARMA(p,q) with large p or q produced overly wet forecasts over the next days, which led to model infeasibility.

This issue is directly related to over-fitting problems, which have attracted considerable attention in statistics, namely in machine learning (Hastie et al., 2001). We believe that these results provide a counterexample to Turgeon's claim that we should always try to optimize the (in-sample) performance of our model and completely disregard parsimony (Turgeon, 2005).

5.6 Conclusion

In conclusion, we illustrate the importance and value of considering the persistence of inflows for the reservoir management problem. Repeatedly solving a simple data-driven stochastic lookahead model using affine decision rules and ARMA and GARCH time series model provides good quality solutions for a problem that would otherwise be intractable for classical SDP and that would require considerably more distributional assumptions for SDDP or tree-based stochastic programming.

We give detailed explanations on the construction and update of the forecasts as well as the conditional distribution of the inflows. We also generalize the approach to consider heteroscedasticity. Although our method is applied to the reservoir management problem, various insights and results can be easily exported to other stochastic problems where serial correlation plays an important role.

As the results from Sections 5.5.7, 5.5.8 and 5.5.9 seem to suggest, it is beneficial to consider ARMA and GARCH models when theses models describe the real inflow process sufficiently

well. When this is not the case, the stochastic models based on these forecasts may only yield modest benefits compared to naive static representations of the random vectors.

Nonetheless, our method offers numerous advantages that in our opinion outweigh its drawbacks. From a practical point of view, our method can be easily incorporated into an existing stochastic programming formulation based on affine decision rules. When the time series model is parsimonious, it is rather straightforward to compute the forecasts and update the stochastic model. The computational overhead is negligible and our approach can be used for *any* time series model of any structure and any order.

Combined with affine decision rules, our lookahead model is not only tractable from a theoretical point of view, it is also extremely fast to solve. This allows us to embed our optimization in a heavier rolling horizon framework and to perform extensive simulations with various inflow generators. Although this would also be theoretically possible for other competing methods such as stochastic programming based on scenario trees, the computation requirements could quickly become excessive for practical purposes (Gauvin et al., 2017).

In addition, results indicate that even when the true process differs from the model considered the dynamic uncertainty sets can achieve superior performance. It is likely that further performance gain can be obtained by easily incorporating additional exogenous information such as soil moisture at no complexity cost.

A possible avenue for further research is to extend the framework to multivariate time-series. Although this is feasible in theory, we believe it may be harder to correctly identify a multivariate model (Tsay, 2005). With limited data availability, the increased degrees of freedom provided by the multivariate model may also overfit the data and in turn produce a model of questionable forecasting skill. The increased dimensionality of the problem may also pose numerical difficulties for multi-reservoir systems.

We also believe that it might be possible to consider ARIMA models with exogeneous variables such as soil moisture, temperature and weather forecasts. Although this may improve the forecasting skill, this approach is considerably more data intensive and will make model calibration more difficult.

It might also be possible to combine our approach with hydrologic models by performing statistical analysis on the forecast errors of models such as Morin and Paquet (2007) or other ensemble streamflow predictions (ESP) made by weather agencies. This might help correct the biases of ESP forecast mentioned in Wood and Schaake (2008).

5.7 Acknowledgments

The authors would like to thank Charles Audet, Fabian Bastin, Angelos Georghiou, Stein-Erick Fleten and Amaury Tilmant for valuable discussion as well as everyone at Hydro-Québec and IREQ for their ongoing support, particularly Grégory Émiel, Louis Delorme, Laura Fagherazzi and Pierre-Marc Rondeau. The comments of two anonymous referee also helped improved the initial version of our paper. This research was supported by the Natural Sciences and Engineering Reasearch Council of Canada (NSERC) and Hydro-Québec through the Industrial Research Chair on the Stochastic Optimization of Electricity Generation and NSERC grant 386416-2010.

5.8 Appendix

5.8.1 Joint probabilistic guarantees for the polyhedral support

Theorem 1. *For any $\epsilon > 0$ and $L \in \mathbb{N}$, we have :*

$$\frac{|x_i|}{\sigma_i} \leq \sqrt{L\epsilon}, \quad \forall i = 1, 2, \cdots, L \quad \text{and} \quad \sum_{i=1}^{L} \frac{|x_i|}{\sigma_i} \leq L\sqrt{\epsilon} \tag{5.35}$$

Proof 1. *The matrix Σ satisfies $\Sigma^{-1} = AA$ where $A = diag(\sigma_1^{-1}, \cdots, \sigma_L^{-1})$. Classical inequalities between norms (see Rudin (1987) for example) ensure that :*

$$\frac{|x_i|}{\sigma_i} \leq \|Ax\|_\infty \leq \|Ax\|_2, \quad \forall i = 1, 2, \cdots, L \tag{5.36}$$

$$\sum_{i=1}^{L} \frac{|x_i|}{\sigma_i} = \|Ax\|_1 \leq \sqrt{L}\,\|Ax\|_2. \tag{5.37}$$

The results follow by observing that $\|Ax\|_2 = \sqrt{x^\top \Sigma^{-1} x}$ for any $x \in \mathbb{R}^L$ and justifies the use of our bounded polyhedral support. □

Observation 1. *We observe that* (5.35) *is slightly different from the standard inclusion used in robust optimization (Ben-Tal et al. (2009) Section 2.3 and Sim and Bertsimas (2004)). This is due to the fact that the authors in Ben-Tal et al. (2009) intersect the ellipsoid with a box (norm infinity ball) of radius 1, assume independent random variables and use Berstein's inequality to obtain probabilistic guarantees. The authors in Sim and Bertsimas (2004) also derive similar uncertainty sets, but with different probabilistic guarantees because they assume*

independent and bounded random variables with symmetric distribution.

However, the bound we derive is based on Markov's/Chebyshev's conditional or unconditional inequality and does not require finding the maximum realisation of the random variables explicitly. Our bounds also hold when we weaken the requirements of independence to uncorrelation, which is required when considering heteroscedasticity and the GARCH model.

5.8.2 Equivalent reformulation of the stochastic program according to ϱ

In order to derive the deterministic/robust equivalent of problem (5.34a)-(5.34c), it is convenient to formulate the problem only in terms of ϱ. Recall the following uncertainty set for the random inflow vector ξ :

$$\Xi_t = \left\{ \xi \in \mathbb{R}^L \;\middle|\; \begin{array}{ll} \exists \varrho \in \Upsilon_{L,\epsilon,t} & (5.38a) \\ \xi = U_t(\hat{\zeta}_{t-1,L} + V\varrho) + u_t & (5.38b) \end{array} \right\}$$

where :

$$\Upsilon_{L,\epsilon,t} = \left\{ \varrho \in \mathbb{R}^L \;\middle|\; \begin{array}{ll} |\varrho_i|\sigma_\varrho^{-1} \leq (L\epsilon)^{1/2}, i = 1, \cdots, L & (5.39a) \\ \sum_{i=1}^{L} |\varrho_i|\sigma_\varrho^{-1} \leq L\epsilon^{1/2} & (5.39b) \\ U_t(\hat{\zeta}_{t-1,L} + V\varrho) + u_t \geq 0 & (5.39c) \end{array} \right\}$$

For any $\xi \in \Xi_t$, equation (5.38b) implies that the affine decision rules $\mathcal{X}(\xi) = \mathcal{X}^0 + \mathcal{X}^\Delta \xi$ can be written as $\hat{\mathcal{X}}^0 + \hat{\mathcal{X}}^\Delta \varrho$ for some $\varrho \in \Upsilon_{L,\epsilon,t}$ where :

$$\hat{\mathcal{X}}^0 = \mathcal{X}^0 + \mathcal{X}^\Delta u_t + \mathcal{X}^\Delta U_t \hat{\zeta}_{t-1,L} \tag{5.40}$$

$$\hat{\mathcal{X}}^\Delta = \mathcal{X}^\Delta U_t V \tag{5.41}$$

Since $\hat{\mathcal{X}}^0$ and $\hat{\mathcal{X}}^\Delta$ are decision variables unrestricted in sign, we can equivalently write problem

(5.34a)-(5.34c) as :

$$\min_{\hat{\mathcal{X}}^0,\hat{\mathcal{X}}^\Delta} \sum_{l=0}^{L-1} \hat{\mathcal{X}}_{t+l}^{0,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^0 + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2})^\top G_{t+l} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2}) \tag{5.42a}$$

$$\text{s.t. } (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^\Delta - \hat{C}_{t+l}^\Delta)\varrho \geq -\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 + \hat{C}_{t+l}^0 \qquad \forall l \in \mathbb{L}, \qquad \forall \varrho \in \Upsilon_{L,\epsilon,t} \tag{5.42b}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^\Delta - \hat{E}_{t+l}^\Delta)\varrho = -\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 + \hat{E}_{t+l}^0 \qquad \forall l \in \mathbb{L}, \qquad \forall \varrho \in \Upsilon_{L,\epsilon,t} \tag{5.42c}$$

where $\hat{C}_{t+l}^\Delta = C_{t+l} U_t V$, $\hat{C}_{t+l}^0 = C_{t+l} U_t \hat{\zeta}_{t-1} + C_{t+l} u_t$, $\hat{E}_{t+l}^\Delta = E_{t+l} U_t V$ and $\hat{E}_{t+l}^0 = E_{t+l} U_t \hat{\zeta}_{t-1} + E_{t+l} u_t$.

5.8.3 Deriving the robust/deterministic equivalent

In order to robustify constraint (5.42b), we write $\Upsilon_{L,\epsilon,t}$ as a polyhedron in the lifted space $\{(\varrho_1^+, \varrho_1^-, \cdots, \varrho_L^+, \varrho_L^-)^\top \in \mathbb{R}_+^{2L} : \exists \varrho \in \mathbb{R}^L; \varrho_i^+ - \varrho_i^- = \varrho_i; \varrho_i^+ + \varrho_i^- = |\varrho_i|, \forall i = 1, \cdots, L\}$, which is the image of $\mathbb{R}^L$ under the coordinate-wise lifting $\mathcal{L}_i(\varrho) = (-\min\{\varrho_i, 0\}, \max\{\varrho_i, 0\})^\top$ for $i = 1, \cdots, L$. We specifically consider :

$$\Upsilon_{L,\epsilon}^{lift} | \mathscr{F}_{t-1} = \left\{ \varrho^{lift} = \begin{pmatrix} \varrho_1^+ \\ \varrho_1^- \\ \vdots \\ \varrho_L^+ \\ \varrho_L^- \end{pmatrix} \in \mathbb{R}_+^{2L} \;\middle|\; \begin{array}{ll} \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \leq (L\epsilon)^{1/2} \mathbf{1} & (5.43a) \\ \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \leq L\epsilon^{1/2} & (5.43b) \\ -U_t V R \varrho^{lift} \leq U_t \hat{\zeta}_{t-1,L} + u_t & (5.43c) \end{array} \right\}$$

where

$$R = \begin{pmatrix} 1 & -1 & & & & \\ & & 1 & -1 & & \\ & & & & \ddots & \\ & & & & & 1 \quad -1 \end{pmatrix} \in \mathbb{R}^{L \times 2L}$$

$$S = \begin{pmatrix} 1 & 1 & & & & \\ & & 1 & 1 & & \\ & & & & \ddots & \\ & & & & & 1 & 1 \end{pmatrix} \in \mathbb{R}^{L \times 2L}$$

$$\mathbf{1}^\top = (1, \cdots, 1)^\top \in \mathbb{R}^L$$

with $S_i^\top \varrho^{lift} = |\varrho_i|$ and $R_i^\top \varrho^{lift} = \varrho_i$ for any $i = 1, \cdots, L$.

Robustifying the $k_l^{th} \in \{1, \cdots, m_{t+l}\}$ constraint (5.42b) for a fixed $l \in \mathbb{L}$ can therefore be achieved by setting $\alpha_{k_l}^\top(\hat{\mathcal{X}}) = -(\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^{\Delta} - \hat{C}_{t+l}^{\Delta})_{k_l}^\top \in \mathbb{R}^{1 \times L}$ as the k_l^{th} row of $-(\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^{\Delta} - \hat{C}_{t+l}^{\Delta})$ and ensuring that $\alpha_{k_l}^\top(\hat{\mathcal{X}}) R \varrho^{lift} \leq (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}})_{k_l}$ holds for all $\varrho^{lift} \in \Upsilon_{L,\epsilon}^{lift} | \mathscr{F}_{t-1}$. Since $\Upsilon_{L,\epsilon}^{lift} | \mathscr{F}_{t-1}$ is a closed, non-empty and bounded polyhedron, this can be achieved by ensuring that the maximum value of the following pair of linear programs is smaller or equal than $(\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^{0} - \hat{C}_{t+l}^{0})_{k_l}$:

$$\begin{array}{lll} \max\limits_{\varrho^{lift} \geq 0} & \alpha_{k_l}^\top(\hat{\mathcal{X}}) R \varrho^{lift} & \\ \text{s. t :} & \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \leq (L\epsilon)^{1/2} \mathbf{1} & \\ (P) & \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \leq L\epsilon^{1/2} & \\ & -U_t V R \varrho^{lift} \leq U_t \hat{\zeta}_{t-1,L} + u_t & \end{array} \quad \left| \quad \begin{array}{lll} \min\limits_{\pi,\nu \geq 0} & (L\epsilon)^{1/2} \pi^\top \mathbf{1} + \pi^0 L \epsilon^{1/2} + \nu^\top (U_t \hat{\zeta}_{t-1,L} + u_t) & \\ \text{s. t :} & \pi^\top \Sigma_{\varrho,L}^{-1/2} S + \pi^0 \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S & \\ (D) & -\nu^\top (U_t V R) \geq \alpha_{k_l}^\top(\hat{\mathcal{X}}) R & \end{array} \right.$$

Since $\Upsilon_{L,\epsilon,t}$ is full dimensional and contains 0, the linear system (5.42c) will hold for all $\varrho \in \Upsilon_{L,\epsilon,t}$ if and only if :

$$-\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^{0} + \hat{E}_{t+l}^{0} = 0 \in \mathbb{R}^{n_{t+l}} \tag{5.44a}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^{\Delta} - \hat{E}_{t+l}^{\Delta}) = 0 \in \mathbb{R}^{m_{t+l}} \tag{5.44b}$$

It follows that problem (5.34a)-(5.34c) is equivalent to :

$$\min_{\substack{\hat{\mathcal{X}}^0,\hat{\mathcal{X}}^\Delta,\\ \pi,\nu\geq 0}} \sum_{l=0}^{L-1} \hat{\mathcal{X}}_{t+l}^{0,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^0 + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2})^\top G_{t+l} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2}) \tag{5.45a}$$

$$\text{s.t. } (L\epsilon)^{1/2}\pi^\top \mathbf{1} + \pi^0 L\epsilon^{1/2} + \nu^\top (U_t \hat{\zeta}_{t-1,L} + u_t) \leq (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 - \hat{C}_{t+l}^0)_{k_l}$$

$$\forall k_l \in \{1,\cdots,m_l\}, l \in \mathbb{L} \tag{5.45b}$$

$$\pi^\top \Sigma_{\varrho,L}^{-1/2} S + \pi^0 \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S - \nu^\top (U_t V R) \geq \alpha_{k_l}^\top (\hat{\mathcal{X}}) R$$

$$\forall k_l \in \{1,\cdots,m_l\}, l \in \mathbb{L} \tag{5.45c}$$

$$-\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 + \hat{E}_{t+l}^0 = 0 \in \mathbb{R}^{n_{t+l}} \tag{5.45d}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^\Delta - \hat{E}_{t+l}^\Delta) = 0 \in \mathbb{R}^{m_{t+l}} \tag{5.45e}$$

5.8.4 Details on GARCH(m,s) model

Recall the general GARCH(m,s) model :

$$\hat{\sigma}_{t-1}^2(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho_{t-i}^2 + \sum_{j=1}^{s} \beta_j \hat{\sigma}_{t-1-j}^2(1) \tag{5.46}$$

Stationarity conditions

In order for the $\{\varrho_t\}$ to be second order stationary with constant variance σ_a^2 we require that for any $t \in \mathbb{Z}$:

$$\mathrm{E}\left[\mathrm{E}\left[\varrho_t^2 | \mathscr{F}_{t-1}\right]\right] = \mathrm{E}\left[\alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho_{t-i}^2 + \sum_{j=1}^{s} \beta_j \hat{\sigma}_{t-1-j}^2(1)\right] \tag{5.47}$$

$$\Leftrightarrow \sigma_\varrho^2 = \alpha_0 \left(1 - \sum_{j=1}^{\max\{m,s\}} (\alpha_j + \beta_j)\right)^{-1} \tag{5.48}$$

which is finite and positive if and only if $\sum_{j=1}^{\max\{m,s\}} (\alpha_j + \beta_j) < 1$ with $\alpha_j = 0$ if $j > m$ and $\beta_j = 0$ if $j > s$.

Reformulation of difference equations

We can reformulate (5.46) by making the substitution $\nu_t = \varrho_t^2 - \hat{\sigma}_{t-1}^2(1)$:

$$\hat{\sigma}_{t-1}^2(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho_{t-i}^2 + \sum_{j=1}^{s} \beta_j \hat{\sigma}_{t-1-j}^2(1) \tag{5.49}$$

$$\Leftrightarrow \varrho_{t-i}^2 - \sum_{i=1}^{\max\{m,s\}} (\alpha_i + \beta_i)\varrho_{t-i}^2 = \alpha_0 + \nu_t - \sum_{i=1}^{s} \beta_j \nu_{t-j} \tag{5.50}$$

with $\alpha_i = 0$ if $i > m$ and $\beta_i = 0$ if $i > s$. We can then rewrite (5.50) in a more compact form similar to the general ARMA model :

$$\hat{\phi}(B)\varrho_t^2 = \alpha_0 + \hat{\theta}(B)\nu_t \tag{5.51}$$

where $\hat{\phi}(B) = (1 - \sum_{i=1}^{\max\{m,s\}} \hat{\phi}_i B^i)$ and $\hat{\theta}(B) = (1 + \sum_{i=1}^{s} \hat{\theta}_i B^i)$. The coefficients of the polynomials are given by :

$$\hat{\phi}_i = \begin{cases} 1 & \text{if } i = 0 \\ (\alpha_i + \beta_i) & \text{if } i = 1, \cdots, \min\{s, m\} \\ \beta_i & \text{if } \min\{s, m\} < i \leq \max\{s, m\} \text{ and } s > m \\ \alpha_i & \text{if } \min\{s, m\} < i \leq \max\{s, m\} \text{ and } s < m \\ 0 & \text{otherwise} \end{cases}$$

and

$$\hat{\theta}_i = \begin{cases} 1 & \text{if } i = 0 \\ -\beta_i & \text{if } i = 1, \cdots, m \\ 0 & \text{otherwise.} \end{cases}$$

Given conditions (5.48) :

$$\hat{\phi}(B)\sigma_{\varrho}^2 = \frac{\left(1 - \sum_{j=1}^{\max\{m,s\}} (\alpha_j + \beta_j)\right)}{\left(1 - \sum_{j=1}^{\max\{m,s\}} (\alpha_j + \beta_j)\right)} \alpha_0 \tag{5.52}$$

Hence (5.51) is equivalent to :

$$\hat{\phi}(B)(\varrho_t^2 - \sigma_\varrho^2) = \hat{\theta}(B)\nu_t. \tag{5.53}$$

Computing the conditional variance for arbitrary lead times

We can find $\hat{\psi}(B)$ such that $\hat{\phi}(B)\hat{\psi}(B) = \hat{\theta}(B)$ and we can therefore "invert" (5.53). Given any $l \in \mathbb{Z}_+$, we specifically obtain :

$$\varrho_{t+l}^2 = \sigma_\varrho^2 + \sum_{i=0}^{\infty} \hat{\psi}_i \nu_{t+l-i}. \tag{5.54}$$

Taking the conditional expectation $\mathrm{E}\left[\cdot|\mathscr{F}_t\right]$ on both sides of the equation (5.54) yields :

$$\hat{\sigma}_t^2(l) = \sigma_\varrho^2 + \sum_{j=l}^{\infty} \hat{\psi}_j \nu_{t+l-j} \tag{5.55}$$

Equation (5.55) holds since for $j \leq l-1$ and $l \in \mathbb{Z}_+$:

$$\begin{aligned} \mathrm{E}\left[\nu_{t+l-j}|\mathscr{F}_t\right] &= \mathrm{E}\left[\varrho_{t+l-j}^2 - \mathrm{E}\left[\varrho_{t+l-j}^2|\mathscr{F}_{t+l-j-1}\right]|\mathscr{F}_t\right] && (5.56)\\ &= \mathrm{E}\left[\varrho_{t+l-j}^2|\mathscr{F}_t\right] - \mathrm{E}\left[\mathrm{E}\left[\varrho_{t+l-j}^2|\mathscr{F}_{t+l-j-1}\right]|\mathscr{F}_t\right] && (5.57)\\ &= 0 && (5.58) \end{aligned}$$

and by the linearity and towering property of the conditional expectation (Billingsley, 1995).

5.8.5 Forecast skill with synthetic ARMA(1,1) time series

If the true process follows an ARMA(1,1) model with parameters $\bar{\theta}$ and $\bar{\phi}$, the quality of a L day ahead forecast made at any time t will be superior to that of the naive forecast on average when the skill of that forecast is non-negative :

$$0 \leq 1 - \mathrm{E}\left[MSE_{frcst,t,L}\right]\mathrm{E}\left[MSE_{naive,t,L}\right]^{-1} \tag{5.59}$$

$$\Leftrightarrow \sum_{l=0}^{L-1} \mathrm{E}\left[(\sigma_{t+l}(\sum_{i=0}^{\infty}(\bar{\psi}_i - \psi_i)\varrho_{t+l-i}))^2\right] \leq \sum_{l=0}^{L-1} \mathrm{E}\left[(\sigma_{t+l}\sum_{i=0}^{\infty}\bar{\psi}_i\varrho_{t+l-i})\right]^2. \tag{5.60}$$

If $\bar{\phi} = \phi$ and $\bar{\theta} \neq \theta$, then (5.60) is equivalent to :

$$\sum_{l=0}^{L-1}\sigma_{t+l}^2\frac{(\theta-\bar{\theta})^2}{1-\phi^2}\sigma_{\varrho}^2 \leq \sum_{l=0}^{L-1}\sigma_{t+l}^2\frac{\bar{\theta}(2\phi+\bar{\theta})+1}{1-\phi^2}\sigma_{\varrho}^2 \tag{5.61}$$

which for values $\bar{\phi} = -0.96$ and $\bar{\theta} = -0.13$ is satisfied if and only if $-1 \leq \theta \leq 0.74$. Hence if we have the perturbed model $\theta = \bar{\theta} + \epsilon_\theta$, then the skill of the forecast will be non-negative if and only if $-0.87 \leq \epsilon_\theta \leq 0.87$. This is independent of GARCH effects since we consider the expected mean square error and not the conditional expected mean square error.

CHAPITRE 6 ARTICLE 3: A SUCCESSIVE LINEAR PROGRAMMING ALGORITHM WITH NON-LINEAR TIME SERIES FOR THE RESERVOIR MANAGEMENT PROBLEM

Cet article a été soumis à *IEEE Transactions on Power Systems* en mars 2017. Auteurs : Charles Gauvin, Erick Delage et Michel Gendreau.

Abstract : This paper proposes a multi-stage stochastic programming formulation based on affine decision rules for the reservoir management problem. Our approach seeks to find a release schedule that balances flood control and power generation objectives while considering realistic operating conditions as well as variable water head. To deal with the non-convexity introduced by the variable water head, we implement a simple, yet effective, successive linear programming algorithm. We also introduce a novel non-linear inflow representation that captures serial correlation of arbitrary order. We test our method on a real river system and discuss policy implications. Our results namely show that our method can decrease flood risk compared to historical decisions, albeit at the cost of reduced final storages.

Keywords : Mathematical programming, Stochastic processes, Forecasting, Risk analysis

6.1 Nomenclature

6.1.1 Sets and parameters

T	Total number of periods
L	Length of look-ahead period ($L \leq T$)
$\mathbb{T}_{t,L}$	Look-ahead period of L periods starting at time t
$\mathbb{L}$	Lead times
$\mathbb{K}^{prod.\,ref.}$	Piecewise linear segments approximating the reference production curve
$\mathbb{K}^{flood\,dmg.}$	Piecewise linear segments approximating the flood damage curve
$\mathbb{K}^{f.val.}$	Piecewise linear segments approximating the storage value function curve
I	Set of plants
I^{evac}	Set of plants with evacuation curve constraints
J	Set of reservoirs
$j^+(i)$, $j^-(i)$	Unique reservoir upstream or downstream of plant i, respectively
$\mathcal{H}_i^{ref}$	Reference water head at plant i
$\mathcal{N}_{jt}^{ref}$	Reference water level at reservoir j
λ_{il}	Fraction of the releases from plant i reaching the unique downstream reservoir after l periods
δ_i^{min}, δ_i^{max}	Minimum and maximum water delays for plant i
κ_{jt}	Relative importance of flood damages at reservoir j at time t
ϕ	Relative importance of power generation objectives compared to flood control
α_{jt}	Average proportion of the total inflows entering reservoir j at time t
$\underline{l}_i$, $\overline{l}_i$	Upper and lower bounds on total spillage at plant i
$\underline{r}_i$, $\overline{r}_i$	Upper and lower bounds on releases at plant i
$\underline{f}_i$, $\overline{f}_i$	Upper and lower bounds on total releases/total flow at plant i
$\overline{\Delta}_i$	Maximum flow variation between periods at plant i
$\underline{s}_j$, $\overline{s}_j$	Upper and lower bounds on storage at reservoir j
$\mathscr{C}_i^{\Delta}$, $\mathscr{C}_i^0$	Parameters for the linear approximation of the evacuation curve at plant i
p_{ik}^{Δ}, p_{ik}^0	Parameters for the k^{th} segment of piecewise linear approximation of the reference production function for plant i
n_j^{Δ}, n_j^0	Parameters for the linear approximation of the water level curve for reservoir j
e_{jk}^{Δ}, e_{jk}^0	Parameters for the k^{th} segment of piecewise linear approximation of the flood penalty at reservoir j
s_{jk}^{Δ}, s_{jk}^0	Parameters for the k^{th} segment of piecewise linear approximation of the final water value function at reservoir j
ν_t	Price of 1 MWh during time t
η	Number of hours per period

6.1.2 Decision variables

$\mathcal{R}_{it}$	Releases at plant i during time t
$\mathcal{L}_{it}$	Spills at plant i during time t
$\mathcal{F}_{it}$	Total flow at plant i during time t
$\mathcal{S}_{jt}, \bar{\mathcal{S}}_{jt}$	Storage and value of storage at reservoir j at the end of time t
$\mathcal{E}_{jt}, \bar{\mathcal{E}}_{jt}$	Floods and flood damages at reservoir j at the end of time t
$\mathcal{H}_{it}$	Relative water head at plant i and time t
$\mathcal{P}_{it}, \mathcal{P}_{it}^{total}$	Reference and real production at plant i and time t
$\mathcal{N}_{jt}$	Water level at reservoir j and time t
$\mathcal{O}_t^{floods}$	Cost associated with floods at time t
$\mathcal{O}_t^{prod}$	Cost associated with electricity generation and final value of storage at time t

6.1.3 Random variables

ξ_t	Total inflows over all reservoirs during time t
ξ_{jt}	Inflows at reservoir j during time t
ζ_t	Logged (total) inflows during time t
ϱ_t	Residuals at time t
$\mathscr{F}_t$	Information known at time t
Ξ_t	Dynamic uncertainty set for inflows given $\mathscr{F}_t$
$\hat{\Xi}_t$	Exterior polyhedral approximation of Ξ_t
Υ	Polyhedral uncertainty set for residuals

6.2 Introduction

This paper considers the optimal operation of a multi-reservoir hydro-electrical system over a given time horizon subject to uncertainty on inflows. We seek to balance hydro-electrical production and flood control while respecting a host of operating constraints on storages, spills, releases and water balance. Our model explicitly considers water delays as well as variable water head. In addition, we consider inflow persistence through ARIMA time series of *arbitrary* order.

As mentioned in Labadie (2004), this stochastic multi-stage non-convex problem is extremely challenging to solve. The variable water head and the associated non-convexity make it impractical to find a globally optimal solution while the multi-dimensional nature of the problem often requires various simplifications that may lead to policies of limited practical use (Castelletti et al., 2008).

To solve this problem, we propose a stochastic program based on affine decision rules, non-linear ARIMA time series and a successive linear programming (SLP) algorithm. We incorporate our lightweight model in a rolling horizon framework and test it with historical inflows. Our approach builds on previous work (Gauvin et al., 2017, 2016), but makes important improvements in terms of modeling and empirical results, namely regarding flood reduction.

Following the work of Ben-Tal et al. (2004); Kuhn et al. (2011), affine decision rules have gained some attention for reservoir management problems and other energy-related problems (Gauvin et al., 2017; Lorca and Sun, 2015). Although they do not guarantee optimality in general, these simple parametric policies often lead to good-quality solutions that can be obtained very efficiently.

Unlike competing methods such as traditional stochastic dynamic programming (SDP) (Turgeon, 2005) and various other extensions such as sampling stochastic dynamic programming (Stedinger and Faber, 2001) and SDP based on improved discretization schemes (Zéphyr et al., 2016), affine decision rules do not require discretizing the state-space, which results in important computational gains.

As illustrated in Gauvin et al. (2016), affine decision rules also make it possible to consider long water delays as well as highly persistent inflows without increasing the computational burden as would be the case of algorithms belonging to the family of SDP. Considering these long inflow delays can be essential to avoid overly optimistic solutions in the face of floods (Gauvin et al., 2016; Turgeon, 2005; Pianosi and Soncini-Sessa, 2009).

Contrary to scenario-tree based stochastic programming and stochastic dual dynamic programming (SDDP), these decision rules also provide easily implementable policies that can be computed off-line and used by operators for simulations.

This paper also extends Gauvin et al. (2016) and Lorca and Sun (2015) by introducing a new inflow representation based on a simple non-linear time series model. Although this leads to a non-convex support, the resulting model has nice statistical interpretation and good forecasting ability.

By exploiting this representation along with affine decision rules, we are able to obtain a compact bilinear expression for the objective function. In order to tackle the non-convexity, we implement a simple SLP algorithm. These algorithms have been successfully utilized in hydro-electrical reservoir management problems in the context of stochastic programs based on scenario-trees (De Ladurantaye et al., 2009), distributionally robust optimization (Pan et al., 2015) as well as at the operational level for deterministic and stochastic optimization, namely at Hydro-Québec. These experimental results suggest that this simple algorithm can

provide good quality solutions since the objective is bilinear and therefore well approximated by a linear first-order Taylor approximation in a restricted neighborhood.

Considering variable water head is important since it captures the fact that a better production schedule can produce more electricity by using less water than the corresponding optimal plan with fixed head (Séguin et al., 2016a). This is also an important consideration from an operational perspective in various regions such as Québec where cascaded river systems are widespread. Evaluating water head is an important advantage of our method compared with SDDP since this method relies on convexity assumptions to obtain hyperplanes bounding the optimal cost-to-go functions (Rougé and Tilmant, 2016; Pereira and Pinto, 1991; Shapiro et al., 2013).

The paper is structured as follows. We present the deterministic reservoir management problem in Section 6.3. We then discuss our non-linear inflow representation and the associated non-convex uncertainty set in Section 6.4. Section 6.5 presents the multi-stage stochastic program based on affine decision rules as well as the SLP algorithm used to solve it approximately. We then present the river system used for our numerical experiments in Section 6.6. Section 6.7 discusses results and Section 6.8 draws concluding comments.

6.3 Deterministic formulation

6.3.1 Daily river operations

River operators must determine the average water release that will flow through the turbines of the powerhouse i during day $\tau = t + l$ where $\tau \in \mathbb{T}_{t,L} = \{t, \cdots, t + L - 1\}$. For the level of granularity considered by our model (daily decisions), the different turbines are aggregated and it is assumed that no maintenance occurs so that all turbines are available for power generation. Because of plant specificities and other physical limitations, the total releases are required to reside within some closed interval :

$$\underline{r}_i \leq \mathcal{R}_{i\tau} \leq \bar{r}_i \qquad \forall i \in I, \tau \in \mathbb{T}_{t,L}. \tag{6.1}$$

Based on physical characteristics of the sites, a certain amount of water can also be unproductively spilled :

$$\underline{l}_i \leq \mathcal{L}_{i\tau} \leq \bar{l}_i \qquad \forall i \in I, \tau \in \mathbb{T}_{t,L}. \tag{6.2}$$

Certain sites $i \in I^{evac} \subset I$ can also evacuate water through a separate spillway. Due to the physical structure of these evacuators, the maximal amount of water that can be physically spilled is bounded by a function of $\mathcal{S}_{j\tau}$, the storage (hm^3) in the unique upstream reservoir $j^-(i)$ at the end of time τ. We approximate these functions with affine functions parametrized by $\mathscr{C}_i^{\Delta}, \mathscr{C}_i^0 \in \mathbb{R}$ to obtain :

$$\mathcal{L}_{i\tau} \leq \mathscr{C}_i^{\Delta} \mathcal{S}_{j^-(i)\tau} + \mathscr{C}_i^0, \qquad \forall i \in I^{evac}, \tau \in \mathbb{T}_{t,L}. \tag{6.3}$$

We define $\mathcal{F}_{i\tau} = \mathcal{R}_{i\tau} + \mathcal{L}_{i\tau}$ as the total flow (hm^3/day) $\forall i \in I, \tau \in \mathbb{T}_{t,L}$, which is simply the sum of unproductive spillage and releases. These quantities are bounded for navigation and recreation purposes, environmental needs as well as to ensure smooth ice formation during the winter months :

$$\underline{f}_i \leq \mathcal{F}_{i\tau} \leq \bar{f}_i \qquad \forall i \in I, \tau \in \mathbb{T}_{t,L}. \tag{6.4}$$

To avoid plant damages, we also ensure that the total flow does not change excessively from one day to the next :

$$|\mathcal{F}_{i\tau} - \mathcal{F}_{i\tau-1}| \leq \bar{\Delta}_i \qquad \forall i \in I, \tau \in \mathbb{T}_{t,L}, \tag{6.5}$$

where $\mathcal{F}_{i,t+l}$ are already known at time t for any $l \leq 0$.

We require that the final water storage at the end of time τ at reservoir j lie within fixed bounds :

$$\underline{s}_j \leq \mathcal{S}_{j\tau} \leq \bar{s}_j \qquad \forall j \in J, \tau \in \mathbb{T}_{t,L}. \tag{6.6}$$

Given very wet conditions, it is possible that the upper bound constraint (6.6) is violated.

We therefore introduce the non-negative variable $\mathcal{E}_{j\tau} \geq 0$ representing floods and modify the second inequality to : $\mathcal{S}_{j\tau} - \mathcal{E}_{j\tau} \leq \bar{s}_j, \forall j \in J, \tau \in \mathbb{T}_{t,L}$. For sake of simplicity, it is assumed that floods remain in the reservoir. Although physically possible in some cases when either $\underline{s}_j > 0$ or reservoirs represent "useful" storages, we do not consider lower bounds violations.

We model the flood severity (\$) as a piecewise linear increasing convex function of floods, which reflects the assumption that greater floods have increasing marginal consequences. We specifically introduce the decision variable $\bar{\mathcal{E}}_{j\tau} \geq 0$ and impose :

$$\bar{\mathcal{E}}_{j\tau} \geq e^{\Delta}_{jk}\mathcal{E}_{j\tau} + e^{0}_{jk}, \forall k \in \mathbb{K}^{flood\,dmg.}, \tag{6.7}$$

for all j, τ where $\mathbb{K}^{flood\,dmg.} = \{1, \cdots, K^{flood\,dmg.}\}$ and $e^{\Delta}_{jk}, e^{0}_{jk}$ define the hyperplanes. In reality, the impact of floods are often non-convex due to threshold effects and other complex physical phenomenon Pianosi and Soncini-Sessa (2009). However, we believe our model is satisfactory since our policies yield very limited floods when tested on historical scenarios.

Water balance constraints ensure that the current storage is given as the previous storage plus the net inflow. The net inflow is itself defined as the sum, over all upstream reservoirs, of past releases (hm^3/day) weighted by the flow delay coefficients plus the daily inflows ξ_{jt} (hm^3/day) during time τ less the releases from all plants during the day (hm^3/day) :

$$\begin{aligned}\mathcal{S}_{j,\tau} &= \mathcal{S}_{j,\tau-1}{}^{1} + \\ &\sum_{i^- \in I^-(j)} \sum_{\bar{l}=\delta^{min}_{i^-}}^{\min\{\delta^{max}_{i^-},\tau-1\}} \lambda_{i^-\bar{l}}\mathcal{F}_{i^-,\tau-\bar{l}} - \sum_{i^+ \in I^+(j)} \mathcal{F}_{i^+,\tau+l} + \xi_{jt},\end{aligned} \tag{6.8}$$

where $\mathcal{S}_{j,t+l}$ are already known at time t for any $l \leq 0$.

for all $j \in J, \tau \in \mathbb{T}_{t,L}$ where $\delta^{min}_{i^-}$ and $\delta^{max}_{i^-}$ represent the minimum and maximum time taken for a drop of water released from plant $i^-(j) \equiv i^-$ to reach reservoir j. We make the simplifying assumptions that $\xi_{jt} = \alpha_{jt}\xi_t$ where α_{jt} is the average proportion of inflows entering reservoir j at time t and ξ_t are the inflows over the entire catchment at time t.

As is customarily done, we model the final water value function with a piecewise linear increasing concave function (Carpentier et al., 2013a). More precisely, we introduce the decision variable $\bar{\mathcal{S}}_{j\tau} \geq 0$ and impose :

$$\bar{\mathcal{S}}_{j\tau} \leq s^{\Delta}_{jk}\mathcal{S}_{j\tau} + s^{0}_{jk}, \forall k \in \mathbb{K}^{f.val.}, \tag{6.9}$$

for all j, τ where $\mathbb{K}^{f.val.} = \{1, \cdots, K^{f.val.}\}$ and $s^{\Delta}_{jk}, s^{0}_{jk}$ define the hyperplanes, which are based on the output of other long-term scheduling tools.

The total power produced (in MW) during time τ for plant i is taken as the product of the reference production $\mathcal{P}_{i\tau}$ at some fixed reference water head times the relative water head $\mathcal{H}_{i\tau}$ (De Ladurantaye et al., 2007; Gjelsvik et al., 2010; De Ladurantaye et al., 2009) :

$$\mathcal{P}^{total}_{i\tau} = \mathcal{P}_{i\tau}\mathcal{H}_{i\tau}. \qquad \forall i \in I, \tau \in \mathbb{T}_{t,L}. \tag{6.10}$$

The relative water head is given by the net water head divided by the reference water head $\mathcal{H}^{ref}_{i}$:

$$\mathcal{H}_{i\tau} = (\mathcal{N}_{j^-\tau} - \mathcal{N}_{j^+\tau})/\mathcal{H}^{ref}_{i} \quad \forall i \in I, \tau \in \mathbb{T}_{t,L}. \tag{6.11}$$

The net water head is simply the difference between forebay $\mathcal{N}_{j^+\tau}$ and tailrace $\mathcal{N}_{j^-\tau}$ elevations during time τ, where $j^+ \equiv j^+(i)$ and $j^- \equiv j^-(i)$ respectively represent the reservoir upstream and downstream of plant i.

The forebay and tailrace elevation (m) at a given plant i are concave increasing functions of the average storage in the upstream and downstream reservoir during time τ respectively. These functions are sufficiently well approximated by affine functions of the water storage in the associated reservoir :

$$\mathcal{N}_{j^-\tau} = n^{0}_{j^-} + n^{\Delta}_{j^-}\mathcal{S}_{j^-\tau} \qquad \forall j \in J, \tau \in \mathbb{T}_{t,L} \tag{6.12a}$$

$$\mathcal{N}_{j^+\tau} = n^{0}_{j^+} + n^{\Delta}_{j^+}\mathcal{S}_{j^+\tau} \qquad \forall j \in J, \tau \in \mathbb{T}_{t,L}, \tag{6.12b}$$

where $n^{0}_{j}, n^{\Delta}_{j} \in \mathbb{R}, \forall j$. For plants where the downstream reservoir is sufficiently far, the tailrace water level is taken as a constant $\mathcal{N}_{j^-} = \mathcal{N}^{ref}_{j^+}$.

The reference production itself depends on the total flows (hm^3/day). For every i and τ, we

use the following approximation for the reference production function by using $K^{prod.\,ref.} \in \mathbb{N}$ hyperplanes :

$$\mathcal{P}_{i\tau} \leq p^{\Delta}_{ik}\mathcal{F}_{i\tau} + p^{0}_{ik}, \forall k \in \mathbb{K}^{prod.\,ref.}, \tag{6.13}$$

where $\mathbb{K}^{prod.\,ref.} = \{1, \cdots, K^{prod.\,ref.}\}$ and $p^{\Delta}_{ik}, p^{0}_{ik} \in \mathbb{R}$. Each piecewise function is increasing only on a subset of its domain to reflect negative tailrace effects caused by excessive flows. Indeed, for each plant, we have $p^{\Delta}_{ik} > 0, k \in \mathbb{K} \setminus K^{prod.\,ref.}$ and $p^{\Delta}_{iK^{prod.\,ref.}} < 0$. Section 6.6 provides concrete examples and additional details.

6.3.2 Biobjective problem

We wish to find a production plan that balances some systemic measure of flood occurrence while maximizing electricity production and avoiding completely emptying the reservoirs at the end of the horizon.

Based on previous work (Gauvin et al., 2017), we let $\kappa_{jt} \geq 0$ reflect the relative importance of floods at reservoir j at time t where $\sum_{j\in J} \kappa_{jt} = 1, \forall t$. We then consider $\mathcal{O}^{floods}_{\tau}$, the weighted flood damages over the entire catchment at time $\tau \in \mathbb{T}_{t,L}$ ($), which is given by the following simple relationship :

$$\mathcal{O}^{floods}_{\tau} = \sum_{j\in J} \kappa_{j\tau}\bar{\mathcal{E}}_{j\tau}. \tag{6.14}$$

We evaluate the hydro-electric productive benefits associated to a reservoir release schedule at time $\tau \in \mathbb{T}_{t,L}$ by considering the (negative) total hydroelectric production value and the value of the final storage of reservoirs ($) :

$$\mathcal{O}^{prod}_{\tau} = \begin{cases} -\sum_{i\in I^{prod}} \mathcal{P}^{total,}_{i,\tau}\eta\nu_{\tau} - \sum_{j\in J} \bar{\mathcal{S}}_{j\tau} & \text{if } \tau = t + L - 1 \\ -\sum_{i\in I^{prod}} \mathcal{P}^{total}_{i,\tau}\eta\nu_{\tau} & \text{else,} \end{cases} \tag{6.15}$$

where η is the number of hours per period, ν_{τ} is the price of 1 MWh and $\bar{\mathcal{S}}_{j\tau}$ indicates the final value of this water storage. Electricity generation is assumed constant throughout each period τ. We also assume that electricity prices are known constants, which makes sense in

the case of Hydro-Québec.

Finally, at the beginning of time $t \in \mathbb{T}$, we seek to minimize

$$\sum_{\tau=t}^{t+L-1} \left((1-\varphi)\mathcal{O}_\tau^{floods} + \varphi \mathcal{O}_\tau^{prod} \right), \tag{6.16}$$

for some $\varphi \in [0,1]$ reflecting the relative importance of the aggregate production measure $\mathcal{O}_\tau^{prod}$ compared to the aggregated flood measure $\mathcal{O}_\tau^{floods}$.

We acknowledge that this approach has important theoretical limitations. Since we consider a non-convex objective, we have no guarantee of being able to explore the set of Pareto-optimal solutions of the true problem by varying φ (Ehrgott, 2005). However, in practice, this simple parametrization is sufficient to gain important insights on the solution space when considering simulated or historical inflows. We briefly explore these issues in Section 6.7.4.

In any case, it would be easily possible to consider a lexicographic optimization approach and minimize floods, then maximize production objectives subject to flood restrictions depending on the optimal solution of the first problem. However, this will slightly increase computing times since 2 models must be successively solved.

6.4 Incorporating uncertainty

We want to model $\{\xi_t\}_{t\in\mathbb{Z}}$, the discrete time stochastic process representing total inflows. At each time $t \in \mathbb{T} = \{1, \cdots, T\}$ for the lead times $\mathbb{L} = \{0, \cdots, L-1\}$, we assume that the $\xi \equiv \xi_{[t,t+L-1]}$ lie within the following set with probability 1 :

$$\Xi_t = \left\{ \xi \in \mathbb{R}^L \;\middle|\; \begin{array}{ll} \exists \zeta, \varrho \in \mathbb{R}^L & (6.17a) \\ \xi_{t+l} = e^{\zeta_{t+l}}, \forall l \in \mathbb{L} & (6.17b) \\ \zeta = V_t \varrho + v_t & (6.17c) \\ W\varrho \leq w & (6.17d) \end{array} \right\}$$

where $\zeta \equiv \zeta_{[t,t+L-1]}$ and $\varrho \equiv \varrho_{[t,t+L-1]}$ represent random variables for the next future L days. We fix $V_t \in \mathbb{R}^{L\times L}$, $W \in \mathbb{R}^{c\times L}$, $v_t \in \mathbb{R}^L$ and $w \in \mathbb{R}^c$ for some $c \in \mathbb{N}$.

Representation (6.17) essentially states that the logarithms of inflows follow some ARIMA process, whose structure is codified by the matrix V_t and the vector v_t, and where the residuals ϱ are random variables taking values in the polyhedral set $\Upsilon = \{\varrho \in \mathbb{R}^L : W\varrho \leq w\}$.

This representation captures commonly used inflow representations as a particular case. For instance, considering large Υ and assuming $\varrho_{[t,t+L-1]}$ follow a multivariate normal distribution implies that the inflows are correlated and approximately follow the popular log-normal distribution, often used in reservoir management problems (Castelletti et al., 2010; Shapiro et al., 2013; Turgeon and Charbonneau, 1998).

The exponential function used in equation (6.17b) also possesses various intuitive and statistically appealing properties. Indeed, the function ensures non-negativeness of inflows and belongs to the family of Box-Cox transforms commonly used to obtain a more adequate fit with a theoretical model (Box and Cox, 1964).

6.5 Stochastic multi-stage formulation

We consider a dynamic setting where the inflows are progressively revealed as time unfolds over the horizon of T days. At each time $t \in \mathbb{T}$, the decision maker fixes a sequence of policies to be implemented at future times $\tau = t, \cdots, t+L-1$ under the assumptions that the future inflows $\xi_{[t,t+L-1]}$ belong to Ξ_t. Decisions must be non-anticipative in the sense that each $\mathcal{X}_\tau$ must only be a function of the past random variables $\xi_{[\tau-1]}$.

6.5.1 Affine decision rules

Solving the true multistage problem to optimality requires considering arbitrary functions, which will lead to an intractable problem. We therefore limit ourselves to affine policies, which are suboptimal but might still capture well the general structure of future decisions. These functions take the following form at time $\tau = t + l$ for $t \in \mathbb{T}$ and $l \in \mathbb{L}$:

$$\mathcal{X}_{k\tau}(\xi) = \mathcal{X}^0_{k\tau} + \sum_{\tau'=t}^{t+L-1} \mathcal{X}^{\tau'}_{k\tau}\xi_{\tau'}. \tag{6.18}$$

In this case, $k \in \{n_{\tau-1}+1, \cdots, n_{\tau-1}+n_\tau\}$ and $n_\tau \in \mathbb{N}$ represents the number of decisions at time τ. $\mathcal{X}^0_{k\tau}, \mathcal{X}^{\tau'}_{k\tau} \in \mathbb{R}, \forall k, \tau$ become the decision variables in our deterministic equivalent formulation. Non-anticipativity is respected by imposing $\mathcal{X}^{\tau'}_{k\tau} = 0, \forall \tau' \geq \tau$.

As mentioned in Gauvin et al. (2017) and Egging et al. (2017), we replace each decision variable in the deterministic formulation by its associated affine decision in the stochastic model. To ensure that the inequalities hold with probability 1, we solve an optimization problem. For instance, constraint (6.1) is respected with probability 1 if $\max_{\xi\in\Xi_t} R^0_{it} + \sum_{t'=t}^{t+L-1} \mathcal{R}^{t'}_{it}\xi_{t'} \leq \bar{r}$.

However, this optimization problem is intractable given our choice of Ξ_t, since this set is not generally convex for an arbitrary $t \in \mathbb{T}$ (see the Appendix 6.9.1). Fortunately, it is possible to obtain the following exterior polyhedral conservative approximation :

$$\hat{\Xi}_t = \bigtimes_{l=1}^{L} \left[e^{\min\limits_{W\varrho \leq w} V_l^\top \varrho + v_{t,l}}, e^{\max\limits_{W\varrho \leq w} V_l^\top \varrho + v_{t,l}} \right]. \tag{6.19}$$

Indeed, using properties of the exponential and Υ, we can check that $\Xi_t \subseteq \hat{\Xi}_t$. The set $\hat{\Xi}_t$ is simply the Cartesian product of intervals (*i.e.* a box) and can therefore be represented with only $2L$ hyperplanes.

We then use strong linear programming duality to write the deterministic equivalent as a a large linear program (see Gauvin et al. (2017) for details).

6.5.2 Objective function and composite risk

We define the (composite) risk of the production plan at the beginning of time t as :

$$\mathrm{E}\left[\sum_{l=0}^{L-1} \left((1-\varphi)\mathcal{O}_{t+l}^{floods}(\xi) + \varphi \mathcal{O}_{t+l}^{prod}(\xi) \right) | \mathscr{F}_{t-1} \right], \tag{6.20}$$

where $\mathscr{F}_t = \sigma(\varrho_\tau : \tau \leq t)$ is the σ-algebra generated by the past $\{\varrho_\tau\}$ and that we can simply interpret as the information known at time t (Billingsley, 1995).

By linearity of the conditional expectation, it follows that this composite risk is simply a convex combination of the flood risk and the production risk. Appendix 6.9.2 details the analytical expression of these terms.

6.5.3 Using successive linear programming to approximate the true problem

Due to (6.10), (6.20) contains a bilinear term (see Appendix 6.9.2 for details). In order to deal with this non-convexity, we consider a commonly used first order Taylor approximation, which is detailed in the Appendix 6.9.3, as well as a SLP algorithm (see Palacios-Gomez et al. (1982); Castillo et al. (2016) for a similar algorithm based on a first-order approximation and Nocedal and Wright (2006); Yuan (2015) for additional details on the more general class of trust-region algorithms). Our algorithm begins by considering the linear problem of minimizing flood damages and builds on this solution.

The SLP algorithm displayed very rapid convergence and after only 4 iterations, the flood and production risk remained roughly constant. We obtained similar results with different stopping criteria, period of the year and weights φ in the objective function (6.20).

6.5.4 Rolling horizon simulation

We embed the SLP algorithm within a broader rolling horizon simulation that reflects the real behavior of river operators (see Figure 6.1).

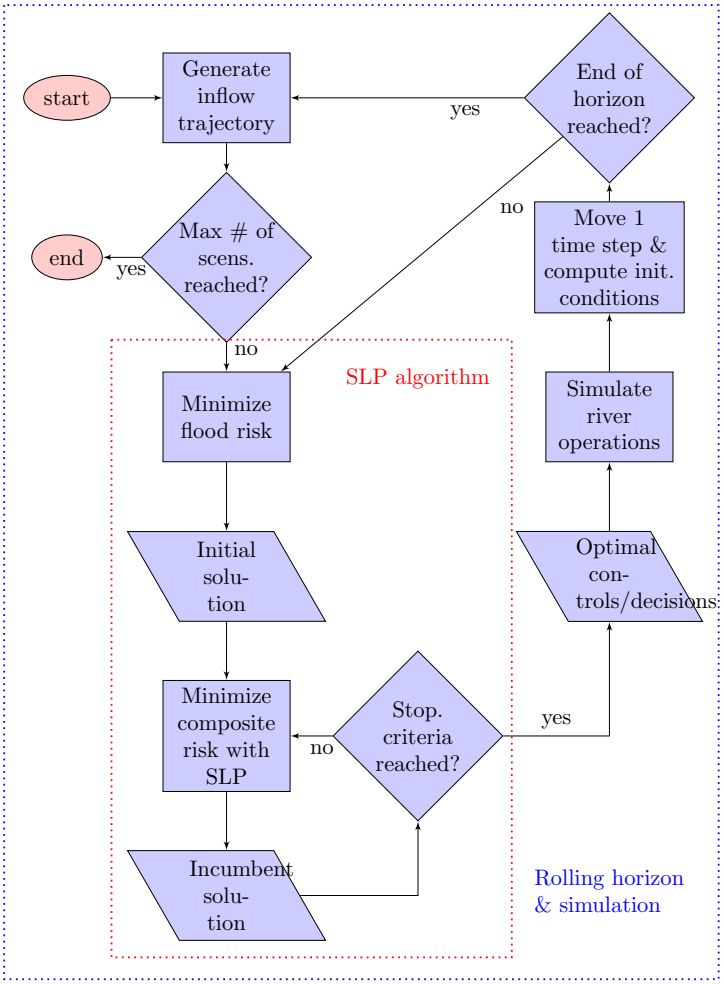

Figure 6.1 Flow chart for the rolling horizon simulation algorithm.

6.6 The Gatineau river system

We apply our framework to the Gatineau river in Western Québec. This hydro-electrical complex is part of the larger Outaouais catchment and is managed by Hydro-Québec. It is

composed of 5 reservoirs and 3 run-of-the river plants.

The first 2 head reservoirs, Baskatong and Cabonga have storage capacity of 1563 hm^3 and 3049 hm^3, respectively, while the remaining downstream reservoirs Paugan, Chelsea and Rapides-Farmer have virtually none.

The Baskatong reservoir is the largest of the larger Outaouais-Gatineau catchment and plays a critical role in the management of the river. It is used to manage flood risk during the freshet as well as baseflow during summer. Furthermore, it is located upstream of the town of Maniwaki, which has witnessed 4 significant floods in the past and therefore imposes extremely tight operating constraints on water flows at this segment.

The three run-of-the-river plants Paugan, Chelsea and Rapides-Farmer have relatively small installed capacity. As illustrated in Figure 6.2, up to some critical threshold, increasing flow will increase the reference production. On the other hand, excessive releases and spills above this value will have negative effect on water head, which is captured with an additional linear segment of negative slope.

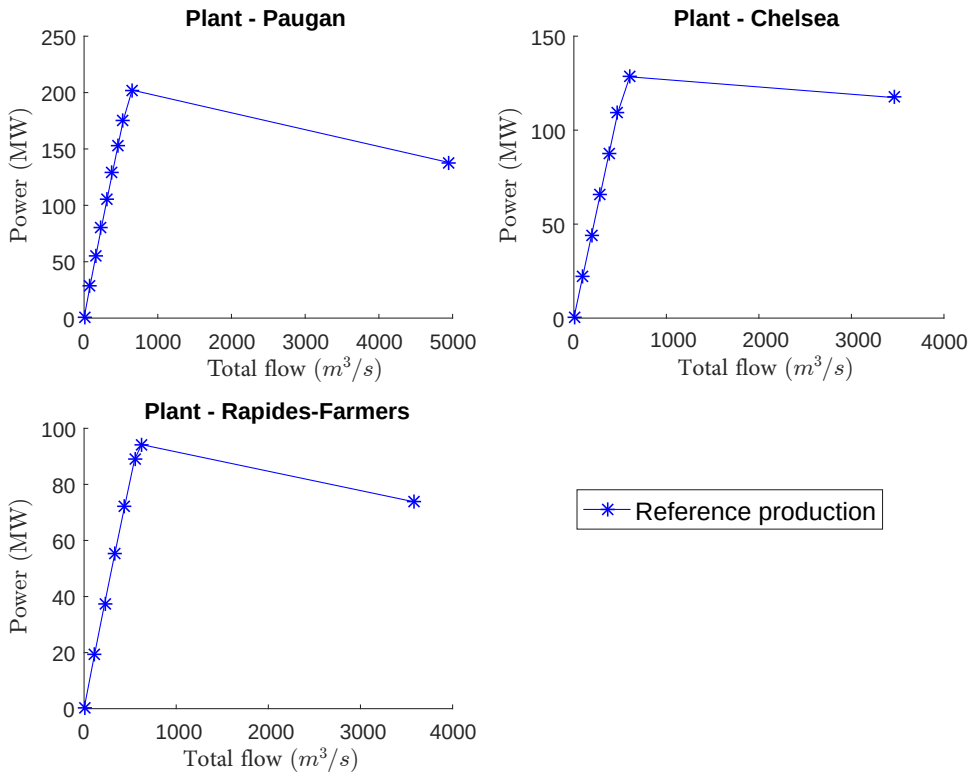

Figure 6.2 Production functions at reference water head.

6.7 Numerical experiments

6.7.1 Daily inflows

We consider 12 years of historical daily inflows provided by Hydro-Québec. We calibrated our model on 6 years of in-sample data 1999-2004, but we also tested our framework on 6 additional out-of-sample years 2008-2013, which are on average wetter than the in-sample years.

6.7.2 Model performance under hard synthetic scenario compared with alternative inflow representations

The non-linear ARMA representation (6.17) offers significant advantages over other alternative inflow representations. This is illustrated in Table 6.1, which shows the results of the flood minimizing solutions with three different inflow representations. Results were obtained by considering the 12 historical scenarios during the freshet under the same hard stress test where the initial storages are fixed at very high levels. We set $\varphi = 0$ in the objective (6.16), which reflects the real behaviour of operators in such a critical situation.

The first naive representation is a static uncertainty set that consists of intervals around the historical sample mean and does not explicitly consider serial correlation. Details can be found in Gauvin et al. (2017). The second one considers the linear ARMA(1,1) model described in Gauvin et al. (2016), which is similar to (6.17), but does not take into account the non-linear log transformation. The third model is based on the non-linear ARMA model presented in this paper where $\xi_t = e^{\zeta_t + \hat{\mu}_t}$, $\hat{\mu}_t$ is the sample log average and the AIC criterion suggests that the ζ_t follow an AR(4) model.

Tableau 6.1 Impact of different uncertainty representations. Values in the table represent the sum over 12 scenarios.

	Baskatong storage violations (hm^3)	Maniwaki flow violations (m^3/s)	Energy (GWh)	Final Volume ($10^6\, hm^3$)
Non-linear ARMA	0	0	140.44	22.22
Linear ARMA	1605.07	27.63	147.28	40.33
Naive forecast	1768.26	28.08	136.42	43.56

The absence of violations in the case of these extremely challenging stress tests is very encouraging. Moreover, these flood reduction objectives are achieved without sacrificing much energy generation compared to the best model. These results suggest that Hydro-Québec can maintain higher storages without violating operating constraints nor social obligations. However, Table 6.1 reveals that maintaining higher storage is of questionable value as a significant amount of water needs to be evacuated out of the system to avoid floods.

6.7.3 Model performance under realistic historical conditions compared with historical decisions

Flood reduction

Next, we consider the very wet historical scenario of fall 2003. Figure 6.3 illustrates that these meteorological conditions led to very high storages at the Baskatong and Cabonga reservoirs around early November and in turn to upper flow bounds violations of 55 m^3/s at the downstream town of Maniwaki on 3 occasions during the period November 14 to December 13.

We consider November 14 as the starting period, which is the day the Baskatong storage reached a level very close to maximal normal operating limits. We therefore naturally fix $\varphi = 0$ and only focus on minimizing flood risk for the catchment. We fixed the initial storages at their historical levels and used the realized inflows over this 30 day period.

Figure 6.3 reveals that our model out-performed the historical decisions since the production plan leads to solutions that respect all operating constraints. We are able to respect flow bounds at Maniwaki by emptying the Baskatong reservoir faster and better anticipating the very large inflows at times 5 and 6. Our release schedule maintains approximately the same final storages in Baskatong as the historical decisions, which illustrates the importance of properly timing the decisions. However, as in section 6.7.2, a significant amount of water is released from Cabonga out of the system. This result suggests that under hard scenarios, it is very hard to simultaneously reduce floods and maintain the same level of storage.

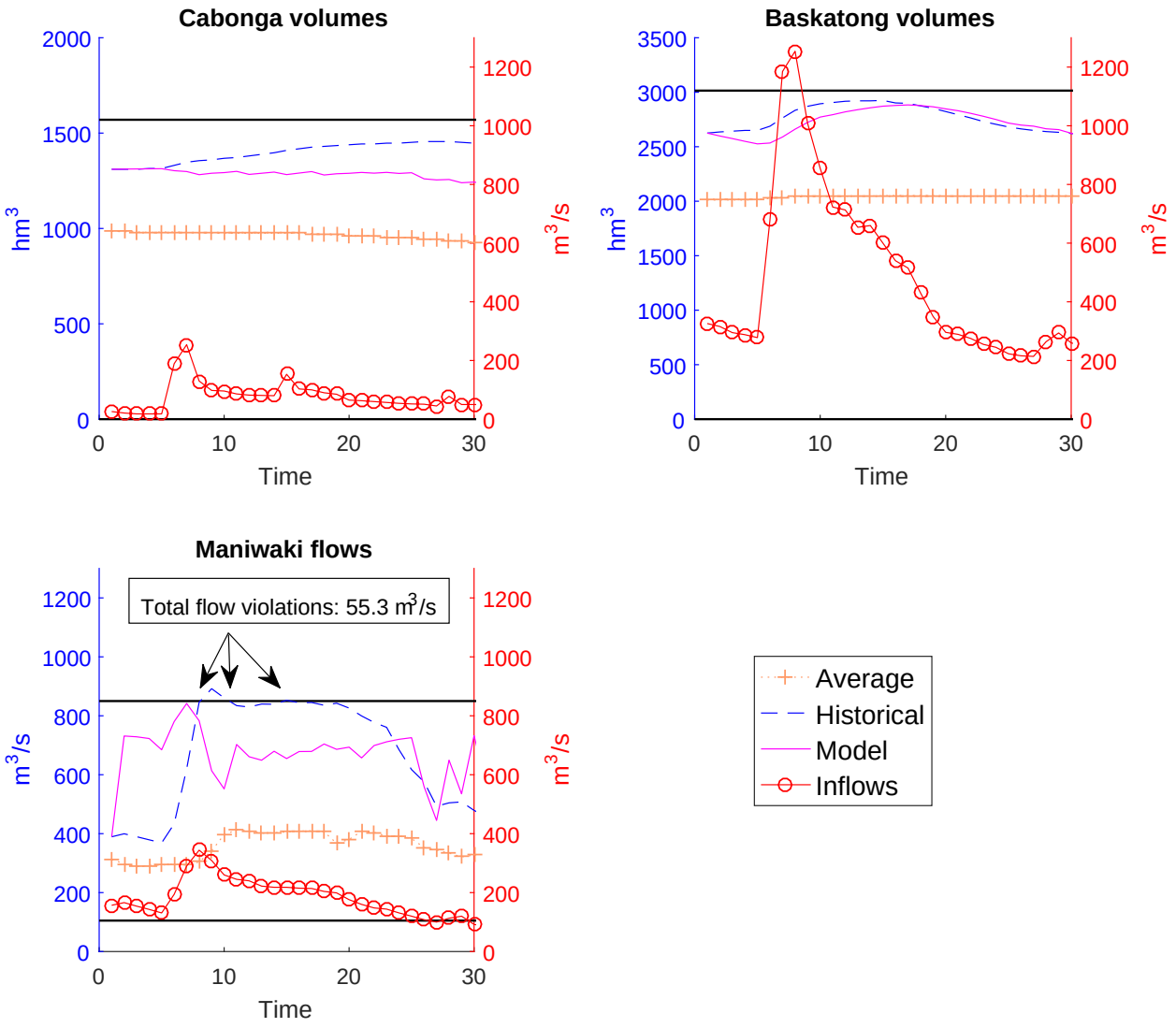

Figure 6.3 Model results on hard historical scenario. The time series 'Average' represents the daily average storages at Cabonga, daily average storages at Baskatong and the daily average flows at Maniwaki, respectively. These averages are computed over the 5 years 1999-2003 using the historical data. The time series 'Average', 'Historical' and 'Model' correspond to the left axis, while the right axis corresponds to the 'Inflows' time series. Upper and lower bounds are indicated by solid black lines.

Energy generation

To validate the utility of our approach for electricity generation purposes, we tested our model in the summer for the years 1999-2003. The summer period corresponds to the days 200-229. We only considered the years 1999-2003, because the data for the historical production plan for the years 2008-2013 was not available. We considered this period, since it represents the

moment where the reservoirs are relatively full, there are no pressing flood risks and it is possible to exploit the variable water head to derive operating efficiencies. Table 6.2 reveals that our model yields an average energy increase of 0.3 % over a 30 day horizon relative to the historical production plan. These figures are obtained by using the same amount of water as the historical plan and avoiding spillage and floods.

Tableau 6.2 Energy generation increase (%) with model water head. Value in this table where obtained by setting $\varphi = 1$ and focusing on energy production.

Year	Paugan	Chelsea	Rapides-Farmer	Total
1999	0.5	-6.5	10.1	0.2
2000	0.4	-5.4	8.6	0.4
2001	0.5	-5.2	8.0	0.4
2002	0.4	-3.5	5.4	0.2
2003	0.1	-3.5	5.1	0.0
Average	0.4	-4.8	7.4	0.3

6.7.4 Exploring the solution space

Figure 6.4 displays the realized performance when considering the low winter regime (in blue), the high freshet regime with normal conditions (in green) and the high freshet regime with high initial volumes (in red) for different values of $\varphi \in \{0, \frac{1}{11}, \frac{1}{2}, \frac{10}{11}, 1\}$. With $\varphi = \frac{1}{11}$ and $\varphi = \frac{10}{11}$, flood control and production are 10 times more important than the other factor, respectively. If $\varphi = \frac{1}{2}$, both factors have the same weight. Finally, $\varphi = 0$ and $\varphi = 1$ represents the solution when only floods and production are considered, respectively.

The plot suggests that for the winter period, it is possible to focus exclusively on production since all 5 points display no floods. As the φ increases to 1, the solutions yield more energy during the period while increasing the final value. This is a consequence of a more efficient use of the variable water head and avoidance of negative tailrace effects causes by excess flow, since the flow rarely exceed the optimal operating threshold during the period.

However, the graph also shows that there is no free lunch in the freshet period. Indeed, all solutions with weight $\varphi > 0$ produce floods. Although the influence of φ on the performance of the solutions is not as clear as in the winter case, it seems that maintaining larger final volumes/storage can only be achieved at a cost of degraded flood control. Considering $\varphi = \frac{1}{11}$

reveals that even policies that place a small weight on production can produce large floods. These results reflect the temporal nature of the operational procedures already in place. During the freshet, the Gatineau is monitored more closely than during the winter months and the productive goals are often set aside to focus on flood control.

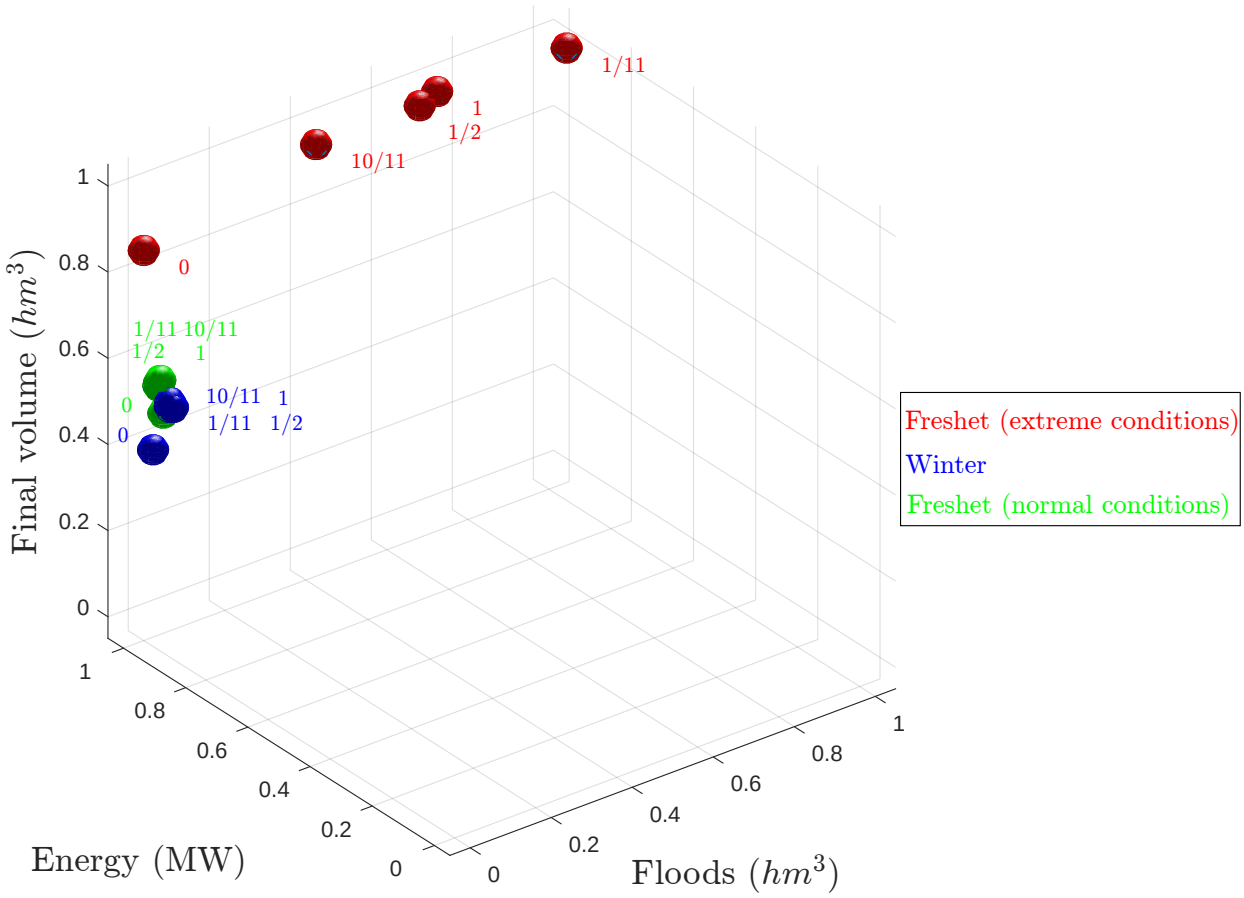

Figure 6.4 Exploring the solution space. The green and red scatter plots were obtained by considering real historical inflows from the Freshet period. While the green plots are obtained by taking the average historical storage at the initial time, the red plots consider hard initial conditions. The blue plots correspond to historical winter scenarios with high initial conditions. The floods and energy are aggregated over all 30 days and 12 historical series while the final volumes are aggregated over all 12 series only. All axis are normalized to $[0, 1]$. The labels indicate the value of φ.

6.8 Conclusion

This paper presents encouraging evidence that stochastic programming models based on affine decision rules can improve short-term river management operations. Our framework based on non-linear time series and a SLP algorithm can successfully consider important phenomenon such as water delays, variable water head and inflow persistence of arbitrary order.

More importantly, our model can yield sizable improvements when compared with historical decisions. We also show that it is possible to find feasible release schedules with storage considerably larger than average historical levels, even with high inflows. This suggests it is possible to revise the drawdown-refill cycle currently used in practice to maintain higher storage and limit unproductive spills.

Finally, the simple convex combination of flood control and production objectives we consider can allow river operators to easily explore the solution space. It is namely interesting to assess possible trade-offs using different inflow simulations. Since our model is extremely fast to solve, it is highly feasible to perform such sensitivity analysis in a realistic setting.

Acknowledgment

The authors would like to thank Charles Audet, Fabian Bastin, Angelos Georghiou, Grégory Émiel, Louis Delorme, Pierre-Marc Rondeau, Sara Séguin and Pierre-Luc Carpentier. This research was supported by the Natural Sciences and Engineering Reasearch Council of Canada (NSERC) and Hydro-Québec through the Industrial Research Chair on the Stochastic Optimization of Electricity Generation and NSERC grant 386416-2010 .

6.9 Appendix

6.9.1 Details on Ξ_t

Time series model linking ζ and ϱ

We assume that at each time $t \in \mathbb{Z}$, there exists some finite constant $\bar{v}_t$ such that $\zeta_t = \ln \xi_t - \bar{v}_t{}^2$ are zero mean random variables that satisfy the equation :

$$\phi(B)\zeta_t = \theta(B)\varrho_t, \tag{6.21}$$

where B represents the backshift operator acting on time indices such that $B^p\zeta_t = \zeta_{t-p}$ for all $t, p \in \mathbb{Z}$ Box et al. (2008); Brockwell and Davis (1987), $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$ with $\phi_i, \theta_i \in \mathbb{R}$ for $p, q \in \mathbb{N}$.

We also suppose $\{\varrho_t\}$ are second order stationary zero-mean i.i.d. random variables (Box et al., 2008) and that we can find $\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$ where $\phi(B)\psi(B) = \theta(B)$ such that $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Under suitable conditions on $\phi(B)$ and $\theta(B)$, the representation $\zeta_t = \psi(B)\varrho_t$ holds and is essentially unique (Brockwell and Davis, 1987).

We then exploit the following linear decomposition :

$$\hat{\zeta}_t(l) \equiv \mathrm{E}\left[\zeta_{t+l} | \mathscr{F}_t\right] = \sum_{i=0}^{l-1} \psi_i \varrho_{t+l-i} \tag{6.22}$$

$$\rho_t(l) \equiv \zeta_{t+l} - \hat{\zeta}_t(l) = \sum_{i=0}^{l-1} \psi_i \varrho_{t+l-i}, \tag{6.23}$$

for any $t \in \mathbb{T}$ and $l \in \mathbb{L}$. $\hat{\zeta}_t(l)$ can be naturally interpreted as a forecast of ζ_{t+l} given information up to time t and $\rho_t(l)$ as the forecast error.

If we set $\rho_{t-1,L} \equiv (\rho_{t-1}(1), \cdots, \rho_{t-1}(L))^\top$ for any $t \in \mathbb{Z}$, we can then express the forecast error vector $\rho_{t-1,L}$ as a linear function of the independent $\varrho_{[t,t+L-1]}$. More specifically, $\rho_{t-1,L} = V_t \varrho_{[t,t+L-1]}$ holds for all $L \in \{1, \cdots, T-t+1\}$, where $V_t \equiv V \in \mathbb{R}^{L \times L}$ is the following invertible and lower triangular square matrix, which is constant across all $t \in \mathbb{Z}$:

$$V = \begin{pmatrix} 1 & & \cdots & 0 \\ \psi_1 & 1 & & \vdots \\ \vdots & & \ddots & \\ \psi_{L-1} & \cdots & \psi_1 & 1 \end{pmatrix} \tag{6.24}$$

We then have the system of equalities :

$$\zeta_{[t,t+L-1]} = \hat{\zeta}_{t-1,L} + V \varrho_{[t,t+L-1]} \tag{6.25}$$

2. Given S years of data where $\xi_{t,i}$ represents the inflows on the t^{th} period of the i^{th} series, we can namely consider $\bar{v}_t = \sum_{i=1}^{S} \frac{1}{S} \ln \xi_{t,i}, \forall t$, the sample daily logged mean, which is also used in Shapiro et al. (2013).

and it follows that $\xi_{t+l} = e^{\bar{v}_{t+l}+\hat{\zeta}_{t-1}(l+1)+\rho_{t-1}(l+1)}, \forall l \in \mathbb{L}$ and $\hat{\zeta}_{t-1,L} + \bar{v}_{[t,t+L-1]} \equiv v_t \in \mathbb{R}^L$ in representation (6.17) where $\hat{\zeta}_{t-1,L} \equiv (\hat{\zeta}_{t-1}(1), \cdots, \hat{\zeta}_{t-1}(L))^\top$ and $\bar{v}_{[t,t+L-1]} = (\bar{v}_t, \cdots, \bar{v}_{t+L-1})^\top$. For more details, see Gauvin et al. (2016).

Polyhedral support of the ϱ_t

We fix some $\epsilon > 0$ and consider the following polyhedron :

$$\Upsilon = \{\varrho \in \mathbb{R}^L : \sum_{i=1}^{L} |\varrho_i| \leq L\sqrt{\epsilon\sigma_\varrho^2}; |\varrho_i| \leq \sqrt{L\epsilon\sigma_\varrho^2}, \forall i\}. \tag{6.26}$$

This set is motivated by Markov's inequality. Indeed, for the independent random variables $\tilde{\varrho} \equiv \varrho_{[t,t+L-1]}$, we know that $\mathbb{P}(\tilde{\varrho} \in \Upsilon) \geq 1 - \epsilon^{-1}$. For more details, see Gauvin et al. (2016).

Counterexample showing that Ξ_t is generally non-convex

We show that in general, Ξ_t is not convex for an arbitrary $t \in \mathbb{T}$. Consider the following instance :

$$\begin{aligned}
&V = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \\
&v_t = (0,0)^\top \\
&\Upsilon = \{\varrho \in \mathbb{R}^L : -1 \leq \varrho_l \leq 1, \forall l = 1, \cdots, L; \sum_{i=1}^{L} |\varrho_i| \leq \sqrt{L}\} \\
&L = 2.
\end{aligned}$$

Figure (6.5) displays $\Xi_t = \{\xi \in \mathbb{R}^L : \exists \varrho \in \Upsilon, \xi_l = \exp(V_l^\top \varrho), \forall l = 1, \cdots, L\}$ and illustrates that Ξ_t is in general not convex.

For a slightly more formal demonstration, it is possible to show that given the two points $\hat{\varrho}_1 = (1-\sqrt{2}, 1)^\top \in \Upsilon$ and $\hat{\varrho}_2 = (\sqrt{2}-1, 1)^\top \in \Upsilon$ illustrated in Figure (6.5) as well as $\lambda = \frac{1}{2}$, there exists no $\varrho \in \Upsilon$ such that :

$$\lambda \begin{pmatrix} e^{V_1^\top \hat{\varrho}_1} \\ e^{V_2^\top \hat{\varrho}_1} \end{pmatrix} + (1-\lambda) \begin{pmatrix} e^{V_1^\top \hat{\varrho}_2} \\ e^{V_2^\top \hat{\varrho}_2} \end{pmatrix} = \begin{pmatrix} e^{V_1^\top \varrho} \\ e^{V_2^\top \varrho} \end{pmatrix}. \tag{6.27}$$

Equivalently, we can show that $\forall \varrho \in \Upsilon$,

$$\left\| \begin{pmatrix} e^{V_1^\top \varrho} \\ e^{V_2^\top \varrho} \end{pmatrix} - \lambda \begin{pmatrix} e^{V_1^\top \hat{\varrho}_1} \\ e^{V_2^\top \hat{\varrho}_1} \end{pmatrix} + (1-\lambda) \begin{pmatrix} e^{V_1^\top \hat{\varrho}_2} \\ e^{V_2^\top \hat{\varrho}_2} \end{pmatrix} \right\|_\infty > 0.$$

This can be shown by solving the following linear program and observing that its optimal value is strictly larger than 0 :

$$\begin{aligned} \min_{\varrho^+ \geq 0, \varrho^- \geq 0, t \geq 0} \quad & t \\ \text{s. t.} \quad & \sum_{i=1}^{2} (\varrho_i^+ + \varrho_i^-) \leq \sqrt{2} \\ & \varrho_i^+ + \varrho_i^- \leq 1, \quad \forall i = 1, 2 \\ & V_i^\top (\varrho^+ - \varrho^-) - l_i^\lambda \leq t, \quad i = 1, 2 \\ & l_i^\lambda - V_i^\top (\varrho^+ - \varrho^-) \leq t, \quad i = 1, 2 \end{aligned}$$

where $l_i^\lambda = ln(\lambda e^{V_i^\top \hat{\varrho}_1} + (1-\lambda) e^{V_i^\top \hat{\varrho}_2})$ is a known constant.

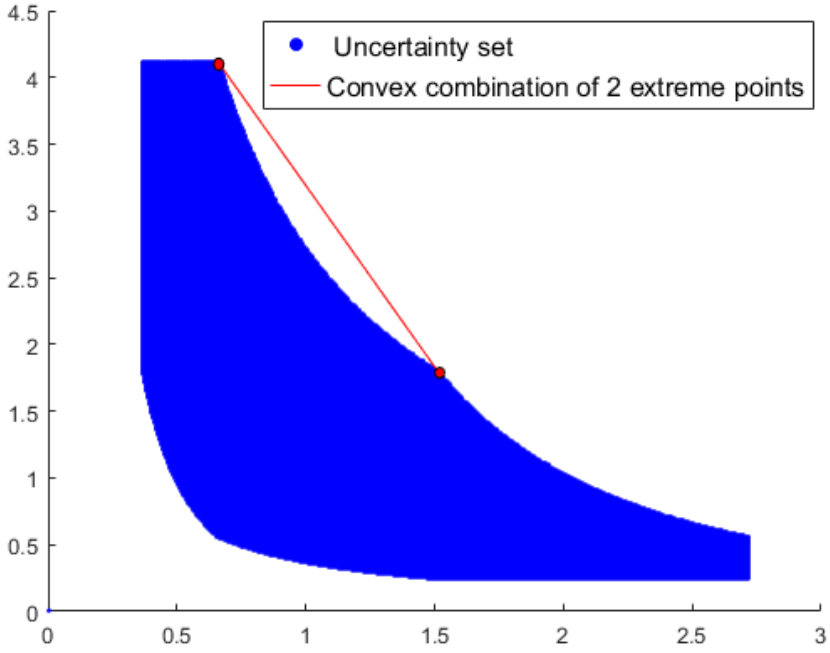

Figure 6.5 Non convex uncertainty set.

6.9.2 Analytical expression for the composite risk

Conditional expected value of $\sum_{l=0}^{L-1} \mathcal{O}_{t+l}^{floods}(\xi)$ and $\sum_{l=0}^{L-1} \mathcal{O}_{t+l}^{prod}(\xi)$

By considering affine decision rules, deriving an analytical expression for the flood and production risk becomes straightforward. Indeed,

$$\mathrm{E}\left[\sum_{l=0}^{L-1} \mathcal{O}_{t+l}^{floods}(\xi)|\mathscr{F}_{t-1}\right] = \sum_{l=0}^{L-1}\sum_{j\in J} \kappa_{j\tau}\mathrm{E}\left[\bar{\mathcal{E}}_{j,t+l}(\xi)|\mathscr{F}_{t-1}\right] \tag{6.28}$$

where

$$\mathrm{E}\left[\bar{\mathcal{E}}_{j,t+l}(\xi)|\mathscr{F}_{t}\right] = \bar{\mathcal{E}}_{j,t+l}^{0} + \sum_{l'=0}^{L-1} \bar{\mathcal{E}}_{j,t+l}^{t+l'}\mathrm{E}\left[\xi_{t+l'}|\mathscr{F}_{t}\right] \tag{6.29}$$

The production risk is defined as :

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{l=0}^{L-1}\mathcal{O}_{t+l}^{prod}(\xi)|\mathscr{F}_{t-1}\right] \\
&\quad = -\sum_{l=0}^{L-1}\sum_{i\in I^{prod}}\mathrm{E}\left[\mathcal{P}_{i,t+l}(\xi)\mathcal{H}_{i,t+l}(\xi)|\mathscr{F}_{t-1}\right] \\
&\qquad -\sum_{j\in J}\mathrm{E}\left[\bar{\mathcal{S}}_{j,t+L-1}(\xi)|\mathscr{F}_{t-1}\right].
\end{aligned} \tag{6.30}
$$

The conditional expected value of the final value of storages $\mathrm{E}\left[\bar{\mathcal{S}}_{j,t+L-1}(\xi)|\mathscr{F}_{t-1}\right]$ is defined similarly to (6.29) while $\mathrm{E}\left[\mathcal{P}_{i,t+l}(\xi)\mathcal{H}_{i,t+l}(\xi)|\mathscr{F}_{t-1}\right]$, which involves the product of two affine functions, is defined as :

$$
\begin{aligned}
\mathrm{E}\left[\mathcal{P}_{i,t+l}(\xi)\mathcal{H}_{i,t+l}(\xi)|\mathscr{F}_{t-1}\right] = &\quad \mathcal{P}_{i,t+l}^{0}\mathcal{H}_{i,t+l}^{0} \\
&+\sum_{k=0}^{L-1}(\mathcal{P}_{i,t+l}^{0}\mathcal{H}_{i,t+l}^{k}+\mathcal{H}_{i,t+l}^{0}\mathcal{P}_{i,t+l}^{k})\mathrm{E}\left[\xi_{t+k}|\mathscr{F}_{t-1}\right] \\
&+\sum_{m=0}^{L-1}\sum_{k=0}^{L-1}\mathcal{P}_{i,t+l}^{m}\mathcal{H}_{i,t+l}^{k}\mathrm{E}\left[\xi_{t+m}\xi_{t+k}|\mathscr{F}_{t-1}\right].
\end{aligned} \tag{6.31}
$$

The conditional expectations $\mathrm{E}\left[\xi_{t+k}|\mathscr{F}_{t-1}\right]$ and $\mathrm{E}\left[\xi_{t+k}\xi_{t+m}|\mathscr{F}_{t-1}\right]$ in (6.29) and (6.31) are detailed in the following Section 6.9.2.

Conditional expected value of ξ_{t+l}

Theorem 2. *For any $t\in\mathbb{Z}$ and $l\in\mathbb{N}$:*

$$
E\left[\xi_{t+l}|\mathscr{F}_t\right] = e^{\hat{\zeta}_t(l)}\cdot E\left[e^{\rho_t(l)}\right] \tag{6.32}
$$

and for $k\geq l\geq 1$:

$$
\begin{aligned}
E\left[\xi_{t+l}\xi_{t+k}|\mathscr{F}_t\right] = e^{\hat{\zeta}_t(l)+\hat{\zeta}_t(k)}\cdot E\left[e^{\sum_{i=0}^{k-l-1}\psi_i\varrho_{t+k-i}}\right] \\
\cdot E\left[e^{\sum_{i=k-l}^{k-1}(\psi_i+\psi_{l-k+i})\varrho_{t+k-i}}\right].
\end{aligned} \tag{6.33}
$$

Proof 2. *This is a direct application of Theorem 34.3 of Billingsley (1995) together with the observation that the* $\{\varrho_t\}$ *are independent and hence* $e^{\rho_t(l)} = e^{\sum_{i=0}^{l-1}\psi_i\varrho_{t+l-i}}$ *is independent of* $\{\varrho_\tau\}_{\tau=-\infty}^{t}$*, which in turn implies that* $E\left[e^{\rho_t(l)}|\mathscr{F}_t\right] = E\left[e^{\rho_t(l)}\right]$. $\square$

At time t, $e^{\hat{\zeta}_t(l)}$ is a known deterministic value for any $l \in \mathbb{N}$ that depends on the past observed $\{\varrho_\tau\}_{\tau=-\infty}^{t}$. To compute $\mathrm{E}\left[e^{\sum_{i=0}^{k}\chi_i\varrho_{t+i}}\right]$ for any $k \in \{0,\cdots,L\}$ and $\chi_i \in \mathbb{R}, \forall i \in \{0,\cdots,L\}$, we perform Monte Carlo simulation.

6.9.3 First order Taylor approximation of the composite risk

We first fix $\mathrm{E}\left[\mathcal{P}_{i,t+l}(\xi)\mathcal{H}_{i,t+l}(\xi)|\mathscr{F}_{t-1}\right] \equiv F_{i,t+l}(\mathcal{H}_{i,t+l},\mathcal{P}_{i,t+l})$ where $\mathcal{H}_{i,t+l} = (\mathcal{H}^0_{i,t+l},\mathcal{H}^{t+1}_{i,t+l},\cdots,\mathcal{H}^{t+L-1}_{i,t+l})^\top \in \mathbb{R}^L$ and $\mathcal{P}_{i,t+l} = (\mathcal{P}^0_{i,t+l},\mathcal{P}^{t+1}_{i,t+l},\cdots,\mathcal{P}^{t+L-1}_{i,t+l})^\top \in \mathbb{R}^L$. Given the point $(\hat{\mathcal{H}}^\top_{i,t+l},\hat{\mathcal{P}}^\top_{i,t+l})^\top \in \mathbb{R}^{2L}$, we then obtain :

$$
\begin{aligned}
&F_{i,t+l}(\mathcal{H}_{i,t+l},\mathcal{P}_{i,t+l}) \\
&\quad\approx F_{i,t+l}(\hat{\mathcal{H}}_{i,t+l},\hat{\mathcal{P}}_{i,t+l}) \\
&\qquad+\hat{\mathcal{H}}^0_{i,t+l}(\mathcal{P}^0_{i,t+l}-\hat{\mathcal{P}}^0_{i,t+l}) \\
&\qquad+\sum_{k=0}^{L-1}\hat{\mathcal{H}}^k_{i,t+l}(\mathcal{P}^0_{i,t+l}-\hat{\mathcal{P}}^0_{i,t+l})\mathrm{E}\left[\xi_{t+k}|\mathscr{F}_{t-1}\right] \\
&\qquad+\hat{\mathcal{P}}^0_{i,t+l}(\mathcal{H}^0_{i,t+l}-\hat{\mathcal{H}}^0_{i,t+l}) \\
&\qquad+\sum_{k=0}^{L-1}\hat{\mathcal{P}}^k_{i,t+l}(\mathcal{H}^0_{i,t+l}-\hat{\mathcal{H}}^0_{i,t+l})\mathrm{E}\left[\xi_{t+k}|\mathscr{F}_{t-1}\right] \\
&\qquad+\sum_{k=0}^{L-1}\hat{\mathcal{P}}^0_{i,t+l}(\mathcal{H}^{t+k}_{i,t+l}-\hat{\mathcal{H}}^{t+k}_{i,t+l})\mathrm{E}\left[\xi_{t+k}|\mathscr{F}_{t-1}\right] \\
&\qquad+\sum_{m=0}^{L-1}\sum_{k=0}^{L-1}\hat{\mathcal{P}}^{t+m}_{i,t+l}(\mathcal{H}^{t+k}_{i,t+l}-\hat{\mathcal{H}}^{t+k}_{i,t+l})\mathrm{E}\left[\xi_{t+m}\xi_{t+k}|\mathscr{F}_{t-1}\right] \\
&\qquad+\sum_{k=0}^{L-1}\hat{\mathcal{H}}^0_{i,t+l}(\mathcal{P}^{t+k}_{i,t+l}-\hat{\mathcal{P}}^{t+k}_{i,t+l})\mathrm{E}\left[\xi_{t+k}|\mathscr{F}_{t-1}\right] \\
&\qquad+\sum_{m=0}^{L-1}\sum_{k=0}^{L-1}\hat{\mathcal{H}}^{t+m}_{i,t+l}(\mathcal{P}^{t+k}_{i,t+l}-\hat{\mathcal{P}}^{t+k}_{i,t+l})\mathrm{E}\left[\xi_{t+m}\xi_{t+k}|\mathscr{F}_{t-1}\right]. \qquad (6.34)
\end{aligned}
$$

CHAPITRE 7 DISCUSSION GÉNÉRALE

7.1 Synthèse des travaux

Les chapitres précédents présentent trois travaux sur la gestion mensuelle de la rivière Gatineau en présence d'incertitude sur les apports. Les résultats montrent que sous certaines conditions hydrologiques, il est difficile d'établir un plan de production qui respecte toujours toutes les contraintes. Comme l'indique le chapitre 4, même en éliminant complètement l'incertitude, il est impossible de trouver un plan de production réalisable si les niveaux d'eau et les apports sont suffisamment importants.

Comme en témoignent les résultats des chapitres 4 et 5 basés sur les règles de décision, les modèles déterministes ainsi que ceux basés sur la programmation stochastique sur arbre, l'ajout de l'incertitude complique significativement la prise de décisions et rend la minimisation des inondations très difficile. Il est donc encourageant d'obtenir une solution sans inondations au chapitre 6.

Bien que le chapitre 6 suggère que le modèle développé puisse améliorer certaines décisions historiques et qu'il soit possible de réduire le risque d'inondations, il semble prudent d'affirmer que le modèle ne permettra pas de révolutionner la gestion de la rivière du jour au lendemain.

Effectivement, il semble difficile, voire impossible de réduire le risque d'inondations tout en maintenant le même stock d'eau dans tous les réservoirs. Notre modèle permet toutefois d'explorer l'espace des solutions grâce à une expression paramétrique simple de l'objectif à optimiser.

Cependant, les résultats des chapitres 4, 5 et 6 demeurent encourageants, car ils suggèrent que le modèle peut être utilisé efficacement et à très faibles coûts en termes de temps de calcul. L'approche par règles de décision linéaires offre assurément des avantages importants par rapport à la prise de décision actuelle, basée sur la résolution de plusieurs modèles déterministes sous différents scénarios d'apports.

Le chapitre 4 suggère que l'approche par règles de décision linéaires est supérieure à celle basée sur les arbres de scénarios au niveau des qualités fournies avec des temps de calcul similaires. Les règles de décision permettent également de considérer des problèmes beaucoup plus grands que ceux en programmation stochastique dynamique tout en faisant moins d'hypothèses sur les apports. Par ailleurs, le chapitre 6 indique que contrairement à l'algorithme SDDP, il est possible de considérer la hauteur de chute variable avec notre méthodologie.

Malgré ces résultats prometteurs, plusieurs points demeurent en suspens ou sujets à amélioration. La section suivante fait un survol rapide des améliorations potentielles pouvant être étudiées lors de travaux futurs.

7.2 Limitations de la solution proposée

Les trois articles présentés aux Sections 4, 5 et 6 exploitent tous une représentation basée sur :

$$\xi_{jt} = \alpha_{jt}\xi_t, \quad \forall j, t \tag{7.1}$$

où α_{jt} représente la proportion moyenne des apports totaux entrant dans le réservoir j au temps t. Bien qu'elle soit compacte et réduise la dimension du problème, cette formulation suppose implicitement que les apports sont parfaitement corrélés spatialement. Cette hypothèse est assez forte considérant que la distance entre certains réservoirs est importante. [1] Heureusement, la figure 7.1 suggère que l'hypothèse n'est pas complètement aberrante.

En effet, comme Baskatong reçoit systématiquement une grande proportion des apports totaux, la représentation (7.1) a l'avantage d'expliquer une grande partie de l'incertitude. Par ailleurs, comme Maniwaki se situe immédiatement en aval de Baskatong et qu'il existe une forte corrélation entre Baskatong et Maniwaki (voir la Figure 7.2), cette représentation offre également l'avantage de bien prédire les apports au point névralgique de Maniwaki, ce qui se révèle utile pour éviter les inondations.

La faible corrélation entre les apports totaux et les apports aux deux derniers réservoirs n'est pas dramatique, car ces derniers reçoivent le moins d'eau de la rivière et il n'y a historiquement jamais d'inondations à ces centrales. La forme en "V" des données à ces 2 réservoirs suggère néanmoins la présence de deux régimes hydrologiques très différents ainsi qu'une différence structurelle importante entre certaines années. Une analyse multivariée plus poussée semble de mise.

1. Dans la Section 1, on mentionne que la rivière Gatineau couvre une distance d'environ 400 km. Cependant, il s'agit de la distance entre la source de la rivière et son exutoire non loin de la centrale Rapides-Farmer. En fait, la distance entre le barrage de Cabonga et la centrale Rapides-Farmer se situe aux alentours de 250-300 km alors que la distance entre Baskatong et Rapides-Farmer est environ de 150-200 km.

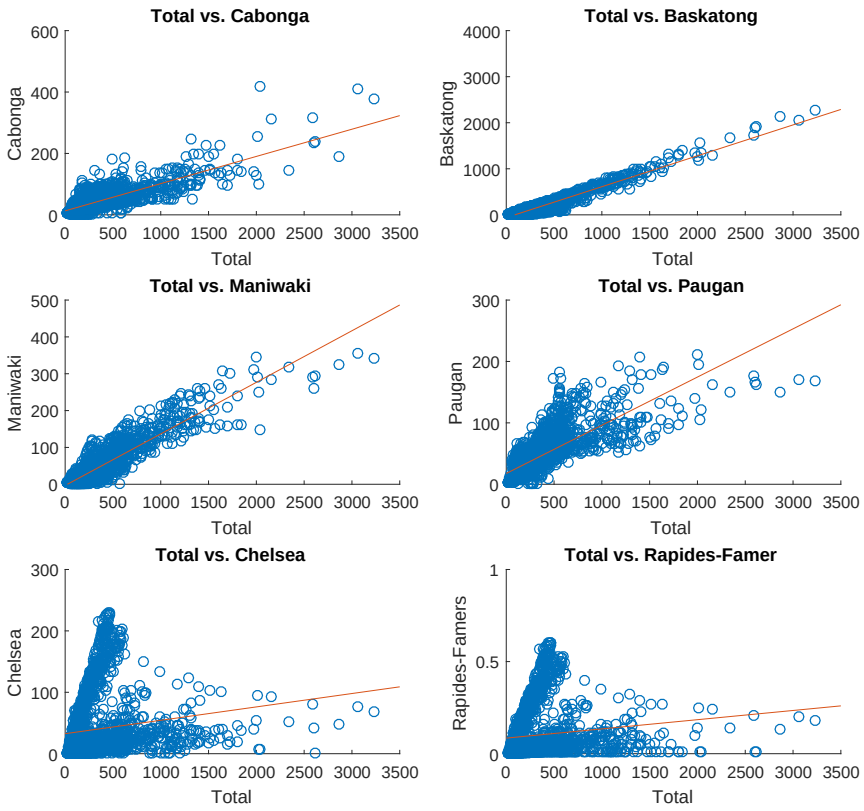

Figure 7.1 Nuage de points pour les apports historiques journaliers (en m^3/s) de 1999-2006. Relation entre les apports totaux et les apports aux différents réservoirs. La droite en rouge indique la droite des moindres carrés.

7.3 Améliorations futures

7.3.1 Représentation des apports pour le premier article

Un façon de modifier la représentation des apports proposée à la Section 4 pour tenir compte de la corrélation spatiale consiste à utiliser $\xi_{jt} = \mu_{jt} + \zeta_{jt}$ où pour chaque t, $\zeta_t \equiv (\zeta_{1t}, \cdots, \zeta_{Jt})^\top$ àà l'ellipse suivante :

$$\ddot{\mathcal{Z}}_{t,\nu} = \left\{ \zeta_t \in \mathbb{R}^{|J|} \mid \zeta_t^\top \Sigma_t^{-1} \zeta_t \leq |J| \nu^2 \right\} \tag{7.2}$$

où $\Sigma_t \in \mathbb{R}^{|J| \times |J|}$ est la matrice de covariance au temps t pouvant être estimée à partir des données de l'échantillon. Cette ellipse nous permet d'obtenir des garanties probabilistes similaires à la Section 5.8 pour un t fixé : $P(\zeta_t^\top \Sigma_t^{-1} \zeta_t \leq |J| \epsilon) \geq 1 - \epsilon^{-1} = 1 - \nu^{-2}$ si $\nu^2 = \epsilon$.

Néanmoins, on pourrait accuser cette représentation d'être "trop conservatrice", car il est possible qu'elle contienne des déviations ζ_{jt} pouvant mener à des apports ξ_{jt} négatifs. On peut donc prendre l'intersection avec l'orthant non-négatif : $\ddot{\mathcal{Z}}_{t,\nu} \cap \{\zeta_{j,t} \geq 0, \forall j \in J\}$ qui nous donne encore la garantie probabiliste $1 - \epsilon^{-1}$, car les apports sont non-négatifs avec probabilité 1. L'usage de cet ensemble implique que notre équivalent déterministe deviendra un *second order cone problem* (SOCP).

Si on choisit d'ignorer la corrélation complètement, il serait également possible d'utiliser la représentation polyhédrale suivante, plus près de celle utilisée dans la Section 4 : $\xi_{jt} = \mu_{jt} + \zeta_{jt}$ où pour chaque t, $\zeta_t \equiv (\zeta_{1t}, \cdots, \zeta_{Jt})^\top$ appartient à :

$$\breve{\mathcal{Z}}_{t,\nu} = \left\{ \zeta_t \in \mathbb{R}^{|J|} \mid -\min\{\mu_{jt}, \nu\sigma_{jt}\} \leq \zeta_{jt} \leq \nu\sigma_{jt}, \forall j \right\}. \tag{7.3}$$

Les σ_{jt}, μ_{jt} peuvent encore une fois être estimés à partir des données d'échantillons recueillies.

7.3.2 Représentation des apports pour le deux derniers articles

Il pourrait également être bénéfique de raffiner la représentation utilisée dans le chapitre 5. En effet, comme l'indique les Figures 7.1 et 7.2, il semble exister une corrélation spatiale plus complexe que celle supposée dans cette thèse. Afin d'avoir une meilleure compréhension des relations entre les apports aux différents réservoirs, une analyse par séries chronologiques multivariées semble prometteuse.

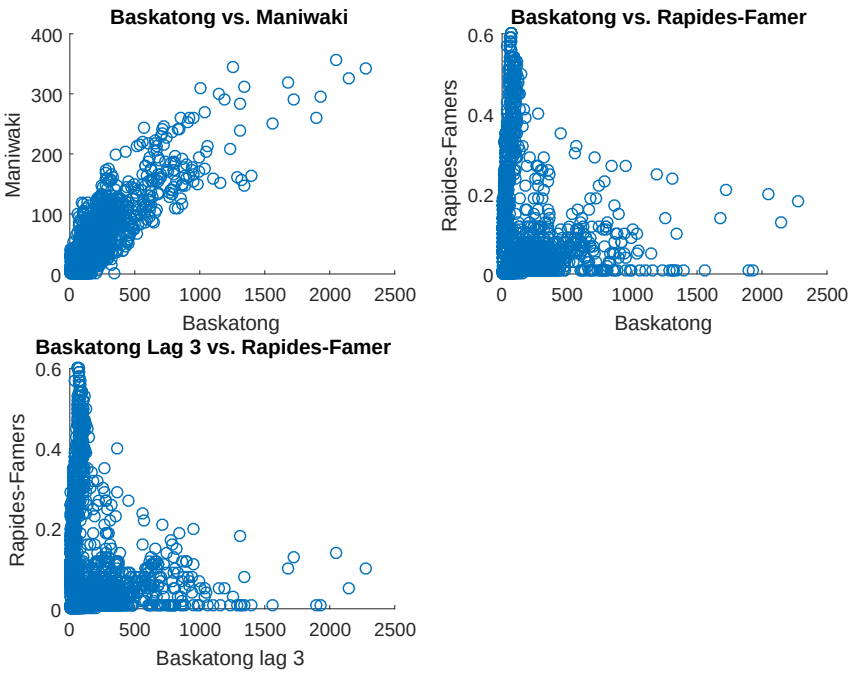

Figure 7.2 Nuage de points pour les apports historiques journaliers (en m^3/s) de 1999-2006. Relation entre Baskatong, Maniwaki et Rapides-Farmer [3]

Afin d'étendre les résultats du chapitre 5 au cas multivarié, il serait possible de considérer l'ensemble d'incertitude suivant :

$$\ddot{\xi} = (\ddot{\xi}_t^\top, \cdots, \ddot{\xi}_{t+L-1}^\top)^\top \in \mathbb{R}^{L|J|} \tag{7.4}$$

$$\ddot{\zeta} = (\ddot{\zeta}_t^\top, \cdots, \ddot{\zeta}_{t+L-1}^\top)^\top \in \mathbb{R}^{L|J|} \tag{7.5}$$

$$\ddot{\varrho} = (\ddot{\varrho}_t^\top, \cdots, \ddot{\varrho}_{t+L-1}^\top)^\top \in \mathbb{R}^{L|J|}, \tag{7.6}$$

où $\ddot{\xi}_t = (\xi_{t,1}, \cdots, \xi_{t,|J|})^\top \in \mathbb{R}^L$, $\ddot{\zeta}_t = (\zeta_{t,1}, \cdots, \zeta_{t,|J|})^\top \in \mathbb{R}^L$ et $\ddot{\varrho}_t = (\varrho_{t,1}, \cdots, \varrho_{t,|J|})^\top \in \mathbb{R}^L$ pour tout $t \in \mathbb{Z}$. Due à la corrélation temporelle des $\{\varrho_{t,j}\}$, on obtient :

3. La figure révèle que certaines relations pouvant sembler intuitives ne sont pas toujours significatives empiriquement. Par exemple, comme un m^3 d'eau prend jusqu'à 3 jours pour s'écouler de Baskatong à Rapides-Farmer, on pourrait croire qu'il existe une corrélation entre les apports de Baskatong il y a 3 jours et ceux de Rapides-Farmer aujourd'hui. Cependant, cette hypothèse ne semble pas corroborée.

$$\mathrm{E}\left[\ddot{\varrho}_l \ddot{\varrho}_k^\top\right] = \begin{cases} \Sigma \in \mathbb{R}^{|J|\times|J|} & \text{if } k = l \\ 0 \in \mathbb{R}^{|J|\times|J|} & \text{if } k \neq l \end{cases} \tag{7.7}$$

et donc :

$$\mathrm{E}\left[\ddot{\varrho}\ddot{\varrho}^\top\right] = \ddot{\Sigma} \tag{7.8}$$

$$= \begin{pmatrix} \Sigma & & \cdots & 0 \\ & \Sigma & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \Sigma \end{pmatrix} \in \mathbb{R}^{L|J|\times L|J|}. \tag{7.9}$$

Il devient ensuite possible de considérer l'ensemble d'incertitude suivant :

$$\ddot{\Xi}_t = \left\{ \ddot{\xi} \in \mathbb{R}^{L|J|} \;\middle|\; \begin{array}{ll} \exists \ddot{\zeta}, \ddot{\varrho} \in \mathbb{R}^{L|J|} & \text{(7.10a)} \\ \ddot{\xi} = \ddot{U}_t \ddot{\zeta} + \ddot{u}_t & \text{(7.10b)} \\ \ddot{\zeta} = \ddot{V}_t \ddot{\varrho} + \ddot{v}_t & \text{(7.10c)} \\ \ddot{\varrho}^\top \ddot{\Sigma}^{-1} \ddot{\varrho} \leq L|J|\epsilon, & \text{(7.10d)} \end{array} \right\}$$

où $\ddot{U}_t, \ddot{V}_t \in \mathbb{R}^{L|J|\times L|J|}$ sont des matrices creuses bloc angulaires ayant une structure similaire à $\ddot{\Sigma}$ et jouant un rôle semblable à celle dans la représentation univariée du chapitre 5. On pourrait notamment considérer l'application $U_t = diag(\sigma_{t,1}, \cdots, \sigma_{t,|J|}, \cdots, \sigma_{t+L-1,1}, \cdots, \sigma_{t+L-1,|J|}) \in \mathbb{R}^{L|J|\times L|J|}$ et $u_t = (\mu_{t,1}, \cdots \mu_{t,|J|}, \cdots, \mu_{t+L-1,1}, \cdots \mu_{t+L-1,|J|})^\top \in \mathbb{R}^{L|J|}$, où $\sigma_{t,j}^2 = \mathrm{E}\left[(\xi_{t,j} - \mu_{t,j})^2\right]$ et $\mu_{t,j} = \mathrm{E}\left[\xi_{t,j}\right]$, qui est simplement la version multivariée de la transformation affine utilisée au chapitre 5.

Cependant, il n'est pas garanti que la matrice $\ddot{\Sigma}^{-1}$ existe, car $\ddot{\Sigma}$ et Σ ne sont pas nécessairement inversibles. il s'agit d'un exemple de difficulté additionnelle engendrée pas la nature multidimensionnelle des séries chronologiques qui n'apparaissait pas dans le cas univarié. Dans le cas où Σ est de rang $k < |J|$ ou presque singulière (ce qu'on peut notamment évaluer à l'aide d'une décomposition SVD et du nombre de conditionnement de la matrice), il est possible de considérer k combinaisons linéaires indépendantes des $\ddot{\zeta}_t, \forall t$. On pourrait notamment utiliser l'analyse par composantes principales pour identifier les k combinaisons linéaires indépendants qui expliquent la plus grande variabilité des données. On peut ensuite effectuer

l'analyse par séries chronologiques sur les vecteurs aléatoires de dimension k appropriés.

On considère également la matrice suivante :

$$\ddot{V}_t = \begin{pmatrix} I & & \cdots & 0 \\ \Psi_1 & I & & \vdots \\ \vdots & & \ddots & \\ \Psi_{L-1} & \cdots & \Psi_1 & I \end{pmatrix} \in \mathbb{R}^{L|J|\times L|J|}. \tag{7.11}$$

Par exemple, dans le cas d'un modèle VAR(1)(*vector autoregressive*), on obtiendrait $\Psi_i = \Phi^i,\ i = 0, 1, \cdots, L$ où $\Phi \in \mathbb{R}^{|J|\times|J|}$ représente la matrice à calibrer en utilisant les données.

L'ellipse (7.10d) nous donne les mêmes garanties probabilistes qu'aux chapitres 5 et 6 :

$$P(\ddot{\varrho}^\top \ddot{\Sigma}^{-1} \ddot{\varrho} > L|J|\epsilon) \leq tr(\ddot{\Sigma}^{-1} \mathrm{E}\left[\ddot{\varrho}\ddot{\varrho}^\top\right])(L|J|\epsilon)^{-1} \tag{7.12}$$

$$= \sum_{l=1}^{L} tr(\Sigma^{-1} \mathrm{E}\left[\ddot{\varrho}_l \ddot{\varrho}_l^\top\right])(L|J|\epsilon)^{-1} \tag{7.13}$$

$$= \epsilon^{-1}. \tag{7.14}$$

Il serait possible d'obtenir une approximation polyhédrale conservatrice en utilisant les même techniques que celles présentées aux chapitres 5 et 6. On pourrait plus spécifiquement utiliser l'inclusion :

$$\{y : \left\|\ddot{\Sigma}^{-1/2} y\right\|_2 \leq \sqrt{L|J|\epsilon}\} \subset \{y : \left\|\ddot{\Sigma}^{-1/2} y\right\|_1 \leq L|J|\sqrt{\epsilon}\} \cap \{y : \left\|\ddot{\Sigma}^{-1/2} y\right\|_\infty \leq \sqrt{L|J|\epsilon}\} \tag{7.15}$$

présentée à l'annexe 5.8.1 en fixant $y = \ddot{\Sigma}^{1/2}\varrho \Leftrightarrow \ddot{\Sigma}^{-1/2} y = \varrho$ [4] et en décomposant y en sa partie positive et négative : $y = y^+ - y^-$ où $y^+, y^- \in \mathbb{R}_+^{|J|L}$. On obtient ensuite le polyhèdre lifté suivant, qui est une approximation conservatrice de l'ellipse (7.10d) :

4. En faisant l'hypothèse que $\ddot{\Sigma}^{-1/2}$ existe.

$$\ddot{\Upsilon} = \left\{ \ddot{\varrho} \in \mathbb{R}_+^{L|J|} \;\middle|\; \begin{array}{ll} \exists y^+, y^- \in \mathbb{R}_+^{|J|L} & \text{(7.16a)} \\ \ddot{\varrho}_i = \ddot{\Sigma}_i^{-1/2}(y^+ - y^-),\ \forall i = 1, \cdots, L|J| & \text{(7.16b)} \\ \sum_{i=1}^{L|J|}(y_i^+ + y_i^-) \leq L|J|\epsilon^{1/2} & \text{(7.16c)} \\ (y_i^+ + y_i^-) \leq (L|J|\epsilon)^{1/2},\ \forall i = 1, \cdots, L|J| & \text{(7.16d)} \end{array} \right\}.$$

Bien qu'il existe peu de différences conceptuelles entre les ensembles d'incertitude pour les séries chronologiques univariées et multivariées, on constate rapidement qu'il existe des différences pratiques non-négligeables entre les deux approches. Outre le besoin d'éliminer la colinéarité dans les $\ddot{\zeta}_t$, il faut également considérer le fait que l'analyse multivariée est habituellement beaucoup plus complexe que celle dans le cas univarié (Tsay, 2005). L'identification d'un modèle adéquat, sa calibration ainsi que sa validation seront effectivement plus difficiles.

Par ailleurs, il faut prendre en compte les impacts de l'augmentation de la dimension de l'incertitude. La sous-section suivante illustre que bien que l'augmentation de la taille de l'équivalent déterministe soit polynomiale, car on passera de $O(L)$ à $O(L|J|)$ variables et contraintes pour chaque contrainte dans le problème déterministe, il est fort probable que l'impact sur les temps de calcul sera assez important. Une seule résolution du modèle pourrait demander quelques minutes, ce qui représente une augmentation significative par rapport aux quelques secondes nécessaires dans le cas univarié.

Il semble également possible d'utiliser un modèle multivarié semblable pour étendre les travaux du chapitre 6 en considérant le logarithme des apports à chaque réservoir et chaque période. Toutefois, cette nouvelle représentation demanderait également un travail non négligeable.

7.3.3 Conséquence additionnelle d'une représentation plus fine des apports

Les représentations désagrégées proposées à la section 7.3 auront une conséquence majeure sur la tractabilité et les temps de calcul des problèmes. Par exemple, si on fait l'hypothèse que les règles de décision prennent la forme :

$$\mathcal{X}_{kt}(\zeta) = \mathcal{X}_{kt}^0 + \sum_{j'=1}^{J}\sum_{t'=1}^{T} \mathcal{X}_{kt}^{j't'}\zeta_{t'}, \quad k \in \mathscr{K}_t, \tag{7.17}$$

il faut résoudre le problème suivant pour s'assurer que $\mathcal{R}(\zeta) \leq \bar{r}, \forall \zeta_t \in \breve{\mathcal{Z}}_{t,\nu}, \forall t$, où $\breve{\mathcal{Z}}_{t,\nu}$ est

décrit par l'équation (7.3) :

$$\begin{aligned}
\max_{\zeta} \quad & \mathcal{R}^0 + \sum_{t'=1}^{T} \sum_{j'=1}^{|J|} \mathcal{R}^{j',t'} \zeta_{j',t'} \leq \bar{q} \\
\text{s. t.} \quad & (\pi^+_{t'}) \quad \zeta_{j',t'} \leq \nu \sigma_{j',t'}, && \forall j' \in J, t' = 1, \cdots, T \\
& (\pi^-_{t'}) \quad -\zeta_{j',t'} \leq -\min\{\mu_{j',t'}, \nu\sigma_{j',t'}\}, && \forall j' \in J, t' = 1, \cdots, T.
\end{aligned}$$

En prenant le dual on constate qu'on multiplie le nombre de variables et de contraintes par $|J|$ par rapport au problème formulé à la section 4.9.1.

Or, il est essentiel de passer à la représentation (7.17) si on considère un ensemble de dimension $|J|$ comme $\breve{\mathcal{Z}}_{t,\nu}$. En effet, les contraintes de conservation de masse ne peuvent tenir si la granularité des règles de décision est trop faible par rapport à celle de la représentation des apports.

Supposons que l'incertitude prenne la forme $\xi_{jt} = \mu_{jt} + \zeta_{jt}$ et qu'elle soit donc représentée par un vecteur de dimension $|J|$ à chaque t. C'est notamment le cas si on considère $\breve{\mathcal{Z}}_{t,\nu}, \forall t$ décrit par (7.3). Supposons également pour les besoins de la contradiction qu'on conserve la représentation agrégée originale pour les règles de décision :

$$\mathcal{X}_{kt}(\zeta) = \mathcal{X}^0_{kt} + \sum_{t'=1}^{T} \mathcal{X}^{t'}_{kt}(\sum_{j' \in J} \zeta_{j't'}), \quad k \in \mathscr{K}_t. \tag{7.18}$$

Pour que les contraintes de conservation de masse (4.2) soient respectées pour tout $\zeta_t \in \breve{\mathcal{Z}}_{t,\nu}$, on doit s'assurer que la somme des coefficients des $\zeta_{j',t'}$ soit nulle et donc pour chaque j et t :

$$\mathcal{S}^{t'}_{jt} = \left(\sum_{\tau=1}^{t} [\sum_{l=\delta_{min}}^{\min\{\delta_{max}, \tau-1\}} \lambda_l \mathcal{F}^{t'}_{i^+\tau-l} - \mathcal{F}^{t'}_{i^+\tau-l}] \right) \beta + \beta \, \mathbb{I}_{t' \leq t} \mathbb{I}_{j'=j}. \tag{7.19}$$

On a donc un système linéaire à $|J||T|$ contraintes et seulement $3|T|$ variables qui n'a pas de solutions si $|J| > 3$. On conclut que si l'incertitude a la dimension $|J|$ à chaque t, alors on ne peut conserver de règles de décision qui dépendent seulement de la variable aléatoire agrégée $\sum_j \zeta_{jt}$ à chaque t.

7.3.4 Utilisation d'un modèle hydrologique

Le domaine de la gestion hydrique bénéficie du fait que l'incertitude associée aux apports est extrinsèque aux décisions prises. La fonte de neige et les précipitations sont effectivement indépendantes de la production hydroélectrique et des débits soutirés. Il existe par ailleurs un couplage généralement assez faible entre les modèles d'optimisation et les représentations de l'incertitude. Il est donc possible d'exploiter les nombreuses années de travail des hydrologues en modélisation hydrologique. Il est notamment possible d'incorporer des modèles hydrologiques comme CEQUEAU ou HYDROTEL (Morin and Paquet, 2007; Fortin et al., 1995) développés par l'INRS-ETE au processus de prévision et d'optimisation d'un bassin versant.

Contrairement aux processus ARIMA, ces modèles permettent d'expliquer la relation entre les précipitations et les débits selon des équations physiques en s'appuyant sur le topologie du terrain, la composition du sol, certains facteurs comme l'humidité ainsi que divers indicateurs météorologiques. Ils offrent donc une capacité prévisionnelle supérieure aux processus ARIMA pour prévoir les débits lorsque les conditions hydrologiques sont différentes du passé (Turcotte et al., 2004). De tels modèles sont notamment utilisés de façon opérationnelle chez Rio Tinto pour l'optimisation stochastique des centrales hydroélectriques (Côté and Leconte, 2015).

De tels modèles n'ont pas été utilisés dans cette thèse par souci de simplicité, par manque de connaissance de ces ressources et par manque de données. Il est fort à parier que ces modèles permettraient d'améliorer la qualité prévisionnelles des apports de manière appréciable. Comme la prévision des apports représente un aspect central de la gestion des rivières, notamment pour le contrôle des inondations, ceci semble être un domaine de recherche prometteur.

Il serait peut-être possible de coupler les travaux effectués sur les processus ARIMA dans cette thèse aux modèles hydrologiques. En effet, il est possible que les erreur de prévision associées au modèle hydrologique démontrent une certaine corrélation temporelle. Par exemple, il est possible que le modèle prévoit systématiquement plus d'eau en été ou qu'ils ait tendance à en prévoir plus à un temps donné s'il en a déjà prévu durant la semaine passée.

7.3.5 Modélisation des courbes d'évacuation

Dans les trois chapitres 4 5 et 6, on utilise la représentation suivante pour les contraintes reliées aux courbes d'évacuation :

$$\mathcal{L}_{it}(\zeta) \leq \mathscr{C}_i^0 + \mathscr{C}_i^{\Delta}\mathcal{S}_{j^-(i)t}(\zeta), \qquad \forall i \in I^{evac}, t. \tag{7.20}$$

Utiliser une telle représentation affine des volumes pour représenter les courbes d'évacuation est une approche assez naïve. Afin d'avoir une représentation plus réaliste, on pourrait modéliser le problème en utilisant des variables binaires. Pour chaque centrale $i \in I^{evac}$ et temps t, on peut considérer K_{it} fonctions affines (hyperplans) et utiliser :

$$\mathcal{L}_{it}(\zeta) \leq \mathscr{C}_i^0 + \mathscr{C}_{i,k}^{\Delta} \mathcal{S}_{j^-(i)t,k}(\zeta) + M(1 - b_{it,k}), \qquad \forall k \in \{1, \cdots, K_{it}\} \tag{7.21a}$$

$$\sum_{k=1}^{K_{it}} b_{it,k} = 1 \tag{7.21b}$$

$$b_{it,k} \in \{0, 1\}, \qquad \forall k \in \{1, \cdots, K_{it}\} \tag{7.21c}$$

où M est une grande constante, par exemple le déversé maximal associé au plus grand volume historique. Avec les courbes présentées à la section 4.6, il semble suffisant de considérer $K_{it} = 2, \forall i \in I^{evac}, t \in \mathbb{T}$ et on obtient ainsi un problème avec $|I^{evac}|T$ variables entières. Dans notre étude de cas, $|I^{evac}| = 4$ et il est envisageable d'utiliser cette modélisation.

Des résultats numériques préliminaires on montré que les temps de calcul augmentaient de façon modérée, mais notable, en considérant uniquement 2 segments à la centrale de Rapides-Farmer. Néanmoins, les temps de calcul demeurent raisonnables si on ne procède pas à la simulation de la politique avec cette nouvelle modélisation.

7.3.6 Résolution en horizon roulant sur un plus long horizon et réplication de la politique de gestion

Tel que suggéré par Grégory Émiel et Pierre-Marc Rondeau d'Hydro-Québec Production, il est intéressant d'évaluer le résultat de l'application de l'algorithme 6.1 sur un horizon long terme. En effet, notre modèle court terme dispose uniquement d'une vision assez myope du futur à travers la fonction de valeur de l'eau. Il est donc possible que la politique de gestion suggérée par notre modèle produisent des effets indésirables à long terme, par exemple le vidage complet du réservoir pour éviter une inondation durant la période considérée par l'algorithme.

Des simulations ont été effectuées sur un horizon d'un an avec des apports historiques dans le but d'évaluer s'il était possible d'obtenir une évolution périodique des volumes similaires à celle engendrée historiquement et illustrée par la figure 1.5. En fixant φ dans l'équation $\sum_{\tau=t}^{t+L-1} \left((1-\varphi)\mathcal{O}_\tau^{floods} + \varphi \mathcal{O}_\tau^{prod}\right)$ pour un certain $\varphi \in [0, 1]$, on observe que les réservoirs se vident ou se remplissent très rapidement, ce qui mène à plusieurs violations des bornes

supérieures ou inférieures sur les volumes sur un horizon annuel. Ces résultats semblaient indépendants des stocks d'eau initiaux.

Des modifications heuristiques de $\varphi_t \in [0, 1]$ à chaque t selon la saison ou selon un processus de moyenne mobile (*moving average*) combinés à des variations des fonctions de valeur d'eau n'ont pas permis d'observer les comportements escomptés. Il semble donc qu'il faille passer plus de temps afin de créer un algorithme totalement automatique ne nécessitant aucune intervention humaine. L'interaction entre les modèles court et moyen termes demeure une question ouverte d'une grande pertinence.

CHAPITRE 8 CONCLUSION

Cette thèse considère la gestion mensuelle de la rivière Gatineau en présence d'incertitude sur les apports. Elle démontre que les règles de décision linéaires possèdent plusieurs avantages importants comme la facilité d'implémentation numérique, la rapidité de calcul et l'obtention directe d'une règle de gestion. Par ailleurs, cette technique permet de mieux modéliser certains phénomènes comme les longs délais d'écoulement et la persistance des apports que d'autres techniques alternatives populaires comme la programmation dynamique stochastique classique.

Les résultats des chapitres 4, 5 et 6 indiquent que les règles de décision linéaires peuvent être utilisées avec succès pour faciliter la gestion de la rivière Gatineau lorsque combinées avec un modèle de l'incertitude ayant une bonne capacité prédictive.

La section 7 illustre plusieurs pistes de solutions permettant de raffiner l'approche et de corriger certaines de ses limitations. Cependant, ces extensions auront probablement des conséquences assez importantes sur la clarté et la formulation du modèle, en plus d'entraîner des augmentations de temps de calcul conséquentes.

RÉFÉRENCES

Aasgård, E., Andersen, G., Fleten, S.-F., and Haugstvedt, D. (2014). Evaluating a stochastic-programming-based bidding model for a multireservoir system. *IEEE Transactions on Power Systems*, 29(4) :1748–1757.

Acciacio, B. and Penner, I. (2011). *Advanced mathematical methods for finance*, chapter Dynamic Risk Measures, pages 1–34. Springer.

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26 :1487–1503.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csaki, F., editors, *Second International Symposium on Information Theory*, pages 267–281, Budapest. Akadémiai Kiado.

Apparigliato, R. (2008). *Règles de décision pour la gestion du risque : Application à la gestion hebdomadaire de la production électrique*. PhD thesis, École Polytechnique.

Ardestani-Jaafari, A. and Delage, E. (2014). The value of flexibility in robust location transportation problem. Technical Report G201483, Les cahiers du GERAD.

Armbruster, B. and Delage, E. (2015). Decision making under uncertainty when preference information is incomplete. *Management Science*, 61(1) :111–128.

Artus, V., . Durlofsky, L., & Onwunalu, J., and Aziz, K. (2006). Optimization of non-conventional wells under uncertainty using statistical proxies. *Computational Geosciences*, 10(4) :389–404.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3) :203–228.

Babaei, M., Pan, I., and Alkhatib, A. (2015). Robust optimization of well location to enhance hysteretical trapping of co2 : Assessment of various uncertainty quantification methods and utilization of mixed response surface surrogates. *Water Resources Research*, 51(2015WR017418) :9402–9424.

Bana e Costa, C. A. and Vansnick, J.-C. (1997). *A Theoretical Framework for Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH)*, pages 15–24. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press.

Ben-Tal, A., Goryashko, E., Guslitzer, A., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99 :351–378.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23 :769–805.

Bertsimas, D., Brown, D., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53 :464–501.

Bertsimas, D. and Georghiou, A. (2015). Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research*, 63(3) :610–627.

Bertsimas, D., Iancu, D. A., and Parrilo, P. A. (2010). Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2) :363–394.

Bertsimas, D., Lauprete, G. J., and Samarov, A. (2004). Shortfall as a risk measure : properties, optimization and applications. *Journal of Economic Dynamics and Control*, 28 :1353–1381.

Bertsimas, D., Litvinov, E., Sun, X. A., Zhao, J., and Zheng, T. (2013). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1) :52–63.

Billingsley, P. (1995). *Probability and measure, third edition*. John Wiley & Sons Inc.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31 :307–327.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, 2 :211–252.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time series analysis : forecasting and control, 4th edition*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Braaten, V. S., Gønnes, O., and Hjertvik, J. K. (2015). Seasonal hydropower planning using linear decision rules. Master's thesis, Norwegian University of Science and Technology - NTNU Trondheim.

Brockwell, P. J. and Davis, R. A. (1987). *Time series : theory and methods*. Springer-Verlag New York.

Brown, B. D. (2007). Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters*, 35(6) :722–730.

Brown, D. and Bertsimas, D. (2009). Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57 :1483–1495.

Carpentier, P.-L., Gendreau, M., and Bastin, F. (2013a). Long-term management of a hydroelectric multireservoir system under uncertainty using the progressive hedging algorithm. *Water Resources Research*, 49 :2812–2827.

Carpentier, P.-L., Gendreau, M., and Bastin, F. (2013b). Optimal scenario set partitioning for multistage stochastic programming with the progressive hedging algorithm. Technical Report CIRRELT-2013-55, CIRRELT.

Carpentier, P.-L., Gendreau, M., and Bastin, F. (2015). Managing hydroelectric reservoirs over an extended horizon using benders decomposition with a memory loss assumption. *IEEE Transactions on Power Systems*, 30(2) :563–572.

Castelletti, A., Galelli, S., Restelli, M., and Soncini-Sessa, R. (2012). Data-driven dynamic emulation modelling for the optimal management of environmental systems. *Environmental Modelling & Software*, 34 :30 – 43. Emulation techniques for the reduction and sensitivity analysis of complex environmental models.

Castelletti, A., Galetti, S., Restelli, M., and Soncini-Sessa, R. (2010). Tree-based reinforcement learning for optimal water reservoir operation. *Water Resources Research*, 46 :W09507.

Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2008). Water reservoir control under economic, social and environmental constraints. *Automatica*, 44(6) :1595–1607.

Castillo, A., Lipka, P., Watson, J. P., Oren, S. S., and O'Neill, R. P. (2016). A successive linear programming approach to solving the iv-acopf. *IEEE Transactions on Power Systems*, 31(4) :2752–2763.

Cervellera, C., Chen, V., and Wen, A. (2006). Optimization of a large-scale water reservoir network by stochastic dynamic programming with efficient state space discretization. *European Journal of Operational Research*, 171 :1139–1151.

Chen, W., Sim, M., Sun, J., and Teo, C.-P. (2010). From cvar to uncertainty set : implications in joint chance-constrained optimization. *Operations Research*, 58 :470–485.

Chen, X., Sim, M., Sun, P., and Zhang, J. (2008). A linear decision-based approximation approach to stochastic programming. *Operations Research*, 56 :344–357.

Commission de toponymie : Gouvernement du Québec (2017). Rivière gatineau.

Côté, P. (2010). *Gestion d'un ensemble d'installations hydroélèctriques soumis á une contrainte de demande*. PhD thesis, École Polytechnique de Montréal.

Côté, P., Haguma, D., Leconte, R., and Krau, S. (2011). Stochastic optimisation of hydroquebec hydropower installations : a statistical comparison between sdp and ssdp methods. *Revue canadienne de génie civil*, 38(12) :1427–1434.

Côté, P. and Leconte, R. (2015). Comparison of stochastic optimization algorithms for hydropower reservoir operation with ensemble streamflow prediction. *Journal of Water Resources Planning and Management*, 142(2).

De Ladurantaye, D., Gendreau, M., and Potvin, J.-Y. (2007). Strategic bidding for price-taker hydroelectricity producers. *IEEE Transactions on Power Systems*, 22(4) :2187–2203.

De Ladurantaye, D., Gendreau, M., and Potvin, J.-Y. (2009). Optimizing profits from hydroelectricity production. *Computers and Operations Research*, 36 :499–529.

Delage, E. and Iancu, D. A. *Robust Multistage Decision Making*, chapter 2, pages 20–46.

Dyer, M. and Stougie, L. (2006). Computational complexity of stochastic programming problems. *Mathematical Programming Series A*, 106(3) :423–432.

Egging, R., Fleten, S. E., Grønvik, I., Hadziomerovic, A., and Ingvoldstad, N. (2017). Linear decision rules for hydropower scheduling under uncertainty. *IEEE Transactions on Power Systems*, 32(1) :103–113.

Ehrgott, M. (2005). *Multicriteria optimization, 2nd edition*. Springer.

El Ghaoui, L. and Lebret, H. (1997). Robust solution to least-squares problems with uncertain data. *SIAM Journal of Matrix Analysis and Applications*, 18 :1035–1064.

El Ghaoui, L., Oustry, F., and Lebret, H. (1998). Robust solutions to uncertain semidefinite programs. *SIAM Journal of Optimization*, 9 :33–52.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 75(4) :643–669.

Feng, Y. and Ryan, S. M. (2014). Scenario reduction for stochastic unit commitment with wind penetration. In *2014 IEEE PES General Meeting | Conference Exposition*, pages 1–5.

Fleten, S.-E., Keppo, J., and Näsäkkälä, E. (2011). *Handbook of integrated risk management in global supply chains*, chapter Risk Management in Electric Utilities, pages 495–513. John Wiley & Sons, Inc.

Fleten, S.-E. and Kristoffersen, T. K. (2008). Short-term hydropower production planning by stochastic programming. *Computers and Operations Research*, 35(8) :2656 – 2671.

Föllmer, H. and Knispel, T. (2013). Convex risk measures : basic facts, law-invariance and beyond, asymptotics for large portfolios. In MacLean, L. and Ziemba, W., editors, *Handbook of the Fundamentals of Financial Decision Making, Part II*, pages 507 – 554. World Scientific.

Föllmer, H. and Schied, A. (2011). *Stochastic finance : an introduction in discrete time*. De Gruyter graduate. De Gruyter, Berlin, New York.

Fortin, J.-P., Moussa, R., Bocquillon, C., and Villeneuve, J.-P. (1995). Hydrotel, un modèle hydrologique distribué pouvant bénéficier des données fourniers par la télédétection et les systèmes d'information géographiques. *Revues des sciences de l'eau*, 8 :97–124.

Galelli, S., Goedbloed, A., Schwanenberg, D., and Overloop, P.-J. V. (2014). Optimal real-time operation of multipurpose urban reservoirs : Case study in singapore. *Journal of Water Resources Planning and Management*, 140(4) :511–523.

Garcia, R. C., Contreras, J., van Akkeren, M., and Garcia, J. B. C. (2005). A garch forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, 20(2) :867–874.

Gauvin, C., Delage, E., and Gendreau, M. (2016). A stochastic program with time series and affine decision rules for the reservoir management problem. Technical Report G-2016-24, Les cahiers du GERAD. Revised January 2017.

Gauvin, C., Delage, E., and Gendreau, M. (2017). Decision rule approximations for the risk averse reservoir management problem. *European Journal of Operational Research*, 261(1) :317–336.

Georghiou, A., Wiesemann, W., and Kuhn, D. (2015). Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming*, 152(1) :301–338.

Gjelsvik, A., Mo, B., and Haugstad, A. (2010). *Long- and Medium-term Operations Planning and Stochastic Modelling in Hydro-dominated Power Systems Based on Stochastic Dual Dynamic Programming*, pages 33–55. Springer Berlin Heidelberg, Berlin, Heidelberg.

Goh, J. and Sim, M. (2010). Distributionally robust optimization and its tractable approximations. *Operations Research*, 58 :902–917.

Goh, J. and Sim, M. (2011). Robust optimization made easy with rome. *Operations Research*, 59 :973–985.

Gonçalves, R., Gendreau, M., and Finardi, E. C. (2013). Medium-term operational planning for hydrothermal systems. In Kovacevic, R. M., Pflug, G. C., and Vespucci, M. T., editors, *Handbook of Risk Management in Energy Production and Trading*, volume 199 of *International Series in Operations Research & Management Science*, pages 129–155. Springer US.

Gonçalves, R. E., Finardi, E. C., and da Silva, E. L. (2012). Applying different decomposition schemes using the progressive hedging algorithm to the operation planning problem of a hydrothermal system. *Electric Power Systems Research*, 83(1) :19 – 27.

Growe-Kuska, N., Heitsch, H., and Romisch, W. (2003). Scenario reduction and scenario tree construction for power management problems. In *Power Tech Conference Proceedings, 2003 IEEE Bologna*, volume 3.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and nse performance criteria : Implications for improving hydrological modelling. *Journal of Hydrology*, 377 :80–91.

Hajkowicz, S. and Higgins, A. (2008). A comparison of multiple criteria analysis techniques for water resource management. *European Journal of Operational Research*, 184(1) :255–265.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models : does anything beat a GARCH(1,1) ? *Journal of Applied Econometrics*, 20(7) :873–889.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Heitsch, H. and Römisch, W. (2009). Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118 :371–406.

Hydro-Québec (2012). Hydro-Québec, rapport annuel 2012.

Hydro-Québec (2015). Hydro-Québec, rapport annuel 2015.

Hydro-Québec (2016). Centrales hydro-électriques (au 1er janvier 2016).

Kelman, J., Stedinger, J., Cooper, L. A. Hsu, E., and Yuan, S.-Q. (1990). Sampling stochastic dynamic programming applied to reservoir operation. *Water Resources Research*, 26 :447–454.

Kuhn, D., Wiesemann, W., and Georghiou, A. (2011). Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming Series A*, 130 :177–209.

Labadie, J. (2004). Optimal operation of multireservoir systems : state-of-the-art review. *Journal of Water Resources Planning and Management*, 130 :93–111.

Lamoureux, C. G. and Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, 8(2) :225–234.

Leitner, J. (2005). A short note on second-order stochastic dominance preserving coherent risk measures. *Mathematical Finance*, 15(4) :649–651.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2) :297–303.

Lorca, A. and Sun, A. (2015). Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. *IEEE Transactions on Power Systems*, 30(4) :1702–1713.

Lorca, A., Sun, A., Litvinov, E., and Zheng, T. (2016). Multistage adaptive robust optimization for the unit commitment problem. *Operations Research*, 64(1) :32–51.

Loucks, D. P. and Dorfman, P. J. (1975). An evaluation of some linear decision rules in chance-constrained models for reservoir planning and operation. *Water Resources Research*, 11 :777–785.

Maceira, M. and Damázio, J. (2004). The use of par(p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the brazilian hydropower system. 8th International Conference on Probabilistic Methods Applied to Power Systems, pages 397–402. Iowa State University.

Morin, G. and Paquet, P. (2007). Modèle hydrologique cequeau. Technical report, INRS-ETE.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models. part i. - a discussion of principles. *Journal of Hydrology*, 10 :282–290.

Natarajan, N., Pachamanova, D., and Sim, M. (2009). Constructing risk measures from uncertainty sets. *Operations Research*, 57 :1129–1141.

Neumann, J. v. and Morgenstern, O. (1953). *Theory of Games and Economic Behavior (Third ed.)*. Princeton, NJ : Princeton University Press.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization, 2nd Edition*. Springer Series in Operations Research. Springer.

Palacios-Gomez, F., Lasdon, L., and Engquist, M. (1982). Nonlinear optimization by successive linear programming. *Management Science*, 28(10) :1106–1120.

Pan, L., Housh, M., Liu, P., Cai, X., and Chen, X. (2015). Robust stochastic optimization for reservoir operation. *Water Resources Research*, 51(1) :409–429.

Pereira, M. and Pinto, L. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52 :359–375.

Pflug, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In Uryasev, S., editor, *Probabilistic Constrained Optimization*, volume 49 of *Nonconvex Optimization and Its Applications*, pages 272–281. Springer US.

Pflug, G. and Pichler, A. (2016). From empirical observations to tree models for stochastic optimization : Convergence properties. *SIAM Journal on Optimization*, 26(3) :1715–1740.

Phillpot, A. and Guan, Z. (2008). On the convergence of stochastic dual dynamic programming and related methods. *Operations Research Letters*, 36.

Pianosi, F. and Soncini-Sessa, R. (2009). Real-time management of a multipurpose water reservoir with a heteroscedastic inflow model. *Water Resources Research*, 45(10) :1–12. W10430.

Pina, J., Tilmant, A., and Anctil, F. (2016). Horizontal approach to assess the impact of climate change on water resources systems. *Journal of Water Resources Planning and Management*.

Powell, W. B. (2007). *Approximate dynamic programming : solving the curses of dimensionality*. Wiley.

Quentin, D., Côté, P. P., and Robert, L. (2014). Role of hydrologic information in stochastic dynamic programming : a case study of the kemano hydropower system in british columbia. *Canadian Journal of Civil Engineering/Revue canadienne de génie civil*, 41(9) :839–844.

ReVelle, C. S. and Kirby, W. (1970). Linear decision rule in reservoir management and design, 2, performance optimization. *Water Resources Research*, 6 :1033–1044.

Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112 :185–200.

Rocha, P. and Kuhn, D. (2012). Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. *European Journal of Operational Research*, 216(2) :397–408.

Rockafellar, R. and Wets, R. (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16 :119–147.

Rockafellar, T. R. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26 :1443–1471.

Romanowicz, R. J., Young, P. C., and Beven, K. J. (2006). Data assimilation and adaptive forecasting of water levels in the river severn catchment, united kingdom. *Water Resources Research*, 42(6). W06407.

Rönnqvist, M. (2003). Optimization in forestry. *Mathematical Programming*, 97(1) :267–284.

Rougé, C. and Tilmant, A. (2016). Using stochastic dual dynamic programming in problems with multiple near-optimal solutions. *Water Resources Research*, 52(5) :4151–4163.

Rudin, W. (1987). *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., New York, NY, USA.

Saad, M., Bigras, P., and Turgeon, A. (1996). Fuzzy learning decomposition for the scheduling of hydroelectric power system. *Water Resources Research*, 32 :179–186.

Saad, M. and Turgeon, A. (1988). Application of principal component anbalysis to long-term reservoir management. *Water Resources Research*, 24 :907–912.

Saddoune, M. (2010). *Optimisation simultanée des rotations et des blocs mensuels des équipages aériens*. PhD thesis, École Polytechnique de Montréal.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (1980). *Applied modeling of hydrologic time series*. Water Resources Publications, Littleton, Colorado.

Séguin, S. (2016). *Optimisation stochastique de la répartition spatio-temporelle d'un volume d'eau aux groupes turbo-alternateurs d'un système de production hydroélectrique*. PhD thesis, École Polytechnique de Montréal.

Séguin, S., Côté, P., and Audet, C. (2016a). Self-scheduling short-term unit commitment and loading problem. *IEEE Transactions on Power Systems*, 31(1) :133–142.

Séguin, S., Fleten, S.-E., Côté, P., Pichler, A., and Audet, C. (2016b). Stochastic short-term hydropower planning with inflow scenario trees. *European Journal of Operational Research*, 259(3) :156–1168.

Shapiro, A. (2009). On a time consistency concept in risk averse multistage stochastic programming. *Operations Research Letters*, 37(3) :143 – 147.

Shapiro, A. (2011). Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209 :63–72.

Shapiro, A., Dentcheva, D., and Ruszcyński, A. (2009). *Lectures on stochastic programming : modeling and theory*. MOS-SIAM Series on Optimization.

Shapiro, A., Tekaya, W., da Costa, J. P., and Soares, M. P. (2013). Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research*, 224(2) :375 – 391.

Sim, M. and Bertsimas, D. (2004). The price of robustness. *Operations Research*, 52(1) :35–53.

Soyster, A. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5) :1154–1157.

Stedinger, J. (1984). The performance of ldr models for preliminary design and reservoir operation. *Water Resources Research*, 20 :215–224.

Stedinger, J. R. and Faber, B. A. (2001). Reservoir optimization using sampling sdp with ensemble streamflow prediction (esp) forecast. *Journal of Hydrology*, 249 :113–133.

Tandberg, C. and Vefring, E. S. (2012). The linear decision rule approach applied to the hydrothermal generation planning problem. Master's thesis, Norwegian University of Science and Technology - NTNU Trondheim.

Tejada-Guibert, J. A., Johnson, S. A., and Stedinger, J. R. (1995). The value of hydrologic information in stochastic dynamic programming models of a multireservoir system. *Water Resources Research*, 31(10) :2571–2579.

Tilmant, A. and Kelman, R. (2007). A stochastic approach to analyze trade-offs and risk associated with large-scale water resources systems. *Water Resources Research*, 43 :W06425.

Tsay, R. S. (2005). *Analysis of Financial Time Series, Second Edition*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Turcotte, R., Lacombe, P., Dimnik, C., and Villeneuve, J.-P. (2004). Prévision hydrologique distribuée pour la gestion des barrages publics du québec. *Canadian Journal of Civil Engineering/Revue canadienne de génie civil*, 31(2) :308–320.

Turgeon, A. (2005). Solving a stochastic reservoir management problem with multilag autocorrelated inflows. *Water Resources Research*, 41 :W12414.

Turgeon, A. (2007). Stochastic optimization of multireservoir operation : The optimal reservoir trajectory approach. *Water Resources Research*, 43 :W05420.

Turgeon, A. and Charbonneau, R. (1998). An aggregation-disaggregation approach to long-term reservoir management. *Water Resources Research*, 34 :3585–3594.

Wood, A. W. and Schaake, J. C. (2008). Correcting errors in streamflow forecast ensemble mean and spread. *Journal of hydrometeorology*, 9 :132–174.

Yeh, W. W.-G. (1985). Reservoir management and operations models : A state-of-the-art review. *Water Resources Research*, 21(12) :1797–1818.

Yuan, Y.-x. (2015). Recent advances in trust region algorithms. *Mathematical Programming Series B*, 151(1) :249–281.

Zéphyr, L., Lang, P., Lamond, B. F., and Côté, P. (2016). *Controlled Approximation of the Stochastic Dynamic Programming Value Function for Multi-Reservoir Systems*, pages 31–37. Springer International Publishing, Cham.