A stochastic program with tractable time series and affine decision rules for the reservoir management problem

Charles Gauvin[1,2,3] , Erick Delage[2,4], Michel Gendreau[1,3]

[1]École polytechnique de Montréal, C.P. 6079, succursale Centre-ville, Montréal, (Canada), H3C 3A7
[2]GERAD, 3000 chemin de la Côte-Sainte-Catherine, Montréal, (Canada), H3T 2A7
[3]CIRRELT, C.P. 6128, succursale Centre-ville, Montréal, (Canada), H3C 3J7
[4]HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, (Canada), H3T 2A7

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-dimensional hydro-electrical complex. We consider well-studied linear time series model 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 is 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 real river system in Western Québec and perform various numerical experiments based on different inflow generators.

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

---

1 Introduction

The stochastic reservoir management problem consists in designing an optimal release schedule for a set of interconnected reservoirs over a given time horizon subject to uncertainty on factors such as inflows, load, turbine availability and price of electricity while respecting tight operational constraints on reservoir volumes, water flows and electricity generation (see (1)). Reservoir operators must also balance various factors such as irrigation, flood control and electricity generation.

Stochastic reservoir management problems suffer from various computational difficulties. In itself, the sequential decision-making under uncertainty represents a huge theoretical obstacle (see (2)). This problem is exacerbated by the multidimensionality of the random vector as well as complex non-linear physical phenomenon.

In order to solve this problem, this paper proposes a multi-stage stochastic program based on affine decision rules and linear time series model. Our approach leverages techniques both from stochastic programming, stochastic processes and robust optimization.

Starting with the pioneering work of (3), 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 (4) and (5) namely use this framework to maximize the expected electric production for a multi-period and multi-reservoir hydro-electric complex while (6) 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 (7; 8; 9), they only consider very simplified representations of the underlying stochastic process and generally omit serial and spatial correlations. However, the importance of the persistency of inflows has been recognized as a crucial factor in hydrological modelling for stochastic optimization problems. Authors like (10) namely argue that serial correlation of high order is important to consider inflows that are high on many consecutive days and risk producing a flood.

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 recent paper of (11) 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 and to the best of our knowledge has never been studied before. Although we minimize the risk of floods, our work can be extended to electricity maximization and other reservoir management operations. Our approach also has practical implications in various unrelated fields such as financial portfolio management where time series with non-constant variance play an important role.

Our model considers ARMA models of *any* order without increasing the complexity of the

problem. This is a huge improvement over stochastic dynamic programming (SDP) and stochastic dual dynamic programming (SDDP) methods that can only consider serial correlation through autoregressive models of small order. Indeed, higher order models require increasing the state-space which quickly leads to numerical intractability; particularly for multi-reservoir operations (see namely (12; 13; 14)).

To circumvent this curse of dimensionality, authors such as (10; 15; 16; 17) have focused on various 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 distributional 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.

The paper is structured as follows. The model for the stochastic reservoir management problem is presented in section (2). Section (3) discusses inflow representation and general univariate ARMA models. It then presents basic and more elaborate conditional supports based on homoscedastic and heteroscedastic time series. Section (4) explains the solution procedure and simulation framework while section (5) presents an in-depth study of a real river in western Québec. Concluding remarks are drawn in section (6).

1.1 Notation

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

For random variables $X, Y : \Omega \to \mathbb{R}$ with distribution functions F_X, F_Y, we write $F_X(u) = \mathbb{P}(X \leq u), \forall u \in \mathbb{R}$. We also use the notation $\mathrm{E}[Y|\sigma(X)] \equiv \mathrm{E}[Y|X]$ and $\mathbb{P}(Y \leq u|\sigma(X)) \equiv \mathbb{P}(Y \leq u|X) = F_{Y|X}(u), \forall u \in \mathbb{R}$ where $\sigma(X)$ is the σ-algebra generated by X (see (18) for more details).

The support of the distribution of X, which we denote supp F_X, is the smallest closed set on which X takes values with probability one. For any discrete time real valued stochastic process $\{b_t\}_{t\in\mathbb{Z}}$, we denote the $\mathbb{R}^L$ valued random vector $(b_t, ..., b_{t+L-1})^\top \equiv b_{[t,t+L-1]}$ for any $t \in \mathbb{Z}$ and $L \in \mathbb{N}$ with the special notation $b_{[t+L-1]}$ if $t = 1$.

2 The stochastic reservoir management problem

2.1 Minimizing flood risk

Our model aims at establishing the release schedule that minimizes the risk of floods for an interconnected set of reservoirs over a fixed time horizon $\mathbb{T} = \{1, ..., T\}$. As illustrated in section (5), this issue is of prime practical importance for various reservoir systems that are subject to high risk of destructive floods. This is namely the case of sites that are close to human habitations and that are prone to flooding because of the geology of the terrain or sensitivity to seasonal and meteorological patterns.

Because of the interconnected nature of the system and the effect of upstream releases on downstream volumes and flows, the minimization of floods at the different reservoirs may lead to conflicting objectives (see (6)). To address this issue, we consider a convex combination of the various floods where the weights indicate the relative importance of each of the reservoirs. We then define the risk as the expected value of the resulting aggregated flood.

We seek to minimize this risk for the entire hydroelectrical complex subject to various constraints on total volumes, unproductive water discharge, turbined outflow and flow conservation over a time horizon of T periods while considering uncertainty on some of the random factors affecting the reservoir operations.

2.2 Sources of uncertainty

Various factors remain uncertain at the time of developing an initial production plan. Electricity prices, demand, turbine availability and other factors may all have sizeable consequences depending on the particular realized scenario. Nevertheless, 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 adapted and 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 $\mathrm{E}[(\xi_t - \mu_t)^2] = \sigma_t^2$.

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. 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. These decisions must therefore be non-anticipative and $\mathcal{X}_1$ must be a constant function. Once $\xi_{[t-1]}$ is known, $\mathcal{X}_t$ can be implemented to yield the actual decisions $\mathcal{X}_t(\xi) \in \mathbb{R}^{n_t}$ for $\xi \equiv \xi_{[T]}$. This decision process can be visualized in figure 1 (see (19; 20) for reference).

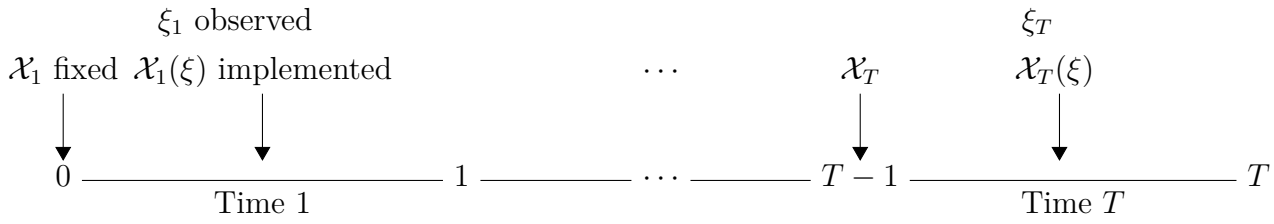

Figure 1: Sequential Dynamic Decision Process

2.4 Mathematical program

At the beginning of time $t \in \mathbb{T}$, we seek a policy $\mathcal{X}_\tau$ which can be implemented at each future time $\tau \in \{t, ..., t+L-1\}$ with $L-1 \leq T-t$ for a given observed past history $\xi_{[t-1]}$. We therefore consider the following mathematical program:

$$
\begin{aligned}
&(\text{RMP}_t) && \min_{\mathcal{X}} \quad \mathrm{E}\left[\sum_{j=1}^{J}\sum_{l=0}^{L-1} \kappa_{j,t+l}\mathcal{E}_{j,t+l}(\xi)|\mathcal{F}_{t-1}\right] && && (1)\\
&(\text{Vol. Bounds}) && \underline{v}_j \leq \mathcal{V}_{j,t+l}(\xi) - \mathcal{E}_{j,t}(\xi) \leq \bar{v}_j && \forall j \in J, l \in \mathbb{L} && (2)\\
&(\text{Water Cons.}) && \mathcal{V}_{j,t+l}(\xi) = \mathcal{V}_{j,t+l-1}(\xi) + && && \\
& && \left(\sum_{i^- \in I^-(j)}\sum_{\bar{l}=\delta_{i^-}^{min}}^{\min\{\delta_{i^-}^{max},t+l-1\}} \lambda_{i^-l}\mathcal{R}_{i^-,t+l-\bar{l}}(\xi) - \sum_{i^+ \in I^+(j)} \mathcal{R}_{i^+,t+l}(\xi) + \alpha_{j,t+l}\xi_{t+l}\right)\beta && && \\
& && && \forall j \in J, l \in \mathbb{L} && (3)\\
&(\text{Flow Bounds}) && \underline{r}_i \leq \mathcal{R}_{i,t+l}(\xi) \leq \bar{r}_i && \forall i \in I, l \in \mathbb{L} && (4)\\
&(\text{Evac. Curve}) && \mathcal{S}_{i,t+l}(\xi) \leq \mathcal{C}_i(\mathcal{V}_{j^-(i),t+l}(\xi)) && \forall i \in I^{evac}, l \in \mathbb{L} && (5)\\
&(\text{Var. Bounds}) && |\mathcal{R}_{i,t+l}(\xi) - \mathcal{R}_{i,t+l-1}(\xi)| \leq \bar{\Delta}_i && \forall i \in I, l \in \mathbb{L} && (6)\\
&(\text{Spill. Bounds}) && \underline{s}_i \leq \mathcal{S}_{i,t+l}(\xi) \leq \bar{s}_i && \forall i \in I, l \in \mathbb{L} && (7)\\
&(\text{Turb. Bounds}) && \underline{q}_i \leq \mathcal{Q}_{i,t+l}(\xi) \leq \bar{q}_i && \forall i \in I, l \in \mathbb{L} && (8)\\
&(\text{Flow Def.}) && \mathcal{R}_{i,t+l}(\xi) = \mathcal{Q}_{i,t+l}(\xi) + \mathcal{S}_{i,t+l}(\xi) && \forall i \in I, l \in \mathbb{L} && (9)\\
&(\text{Floods}) && 0 \leq \mathcal{E}_{j,t+l}(\xi) && \forall j \in J, l \in \mathbb{L} && (10)
\end{aligned}
$$

with $\xi \equiv \xi_{[T]}$, $\mathbb{L} = \{0, ..., L-1\}$ where $L-1 \leq T-t$ and constraints must hold with $\mathbb{P}$ a.s.

The decision rules $\mathcal{V}_t, \mathcal{S}_t$, $\mathcal{Q}_t$, $\mathcal{R}_t$, $\mathcal{E}_t$, respectively represent volumes (hm^3) at the end of period t, average spillage (m^3/s) over time t, average turbined outflow (productive water discharge) (m^3/s) over time t, average total flow (m^3/s) over time t and average floods (hm^3) over time t. The sum of the spillage and the turbined outflow is the average total flow. The aggregate decision vector $\mathcal{X}_t = (\mathcal{V}_t^\top, \mathcal{S}_t^\top, \mathcal{Q}_t^\top, \mathcal{R}_t^\top, \mathcal{E}_t^\top)^\top$ is simply the stacking of each decision at time t. I, I^{evac} and J respectively represent the set of plants, plants with

constraints on evacuation curves and the set of reservoirs.

The decisions pertaining to spillage, flow and discharge represent real implementable decisions used by river operators to control the dynamics of the river system. The decisions to be implemented at some time t must therefore be fixed before observing the future inflows $\xi_{[t,T]}$ and the corresponding subset of random variables $\mathcal{X}_t \circ \xi$ must be $\mathcal{F}_{t-1}$ measurable (refer to figure (1)). On the other hand, volume and flood at time t are only meant to track the evolution of the system and also depend of ξ_t. Hence the associated subset of random variables $\mathcal{X}_t \circ \xi$ must be $\mathcal{F}_t$ measurable.

Constraints (2) ensure that the total water volume remain within tolerable limits $\underline{v}_j$ and $\bar{v}_j$ for all reservoirs $j \in J$. (3) are simply flow conservation constraints ensuring that water released upstream eventually reaches downstream reservoirs where λ_{il} represents the fraction of water released from i to reach the unique downstream reservoir after l days. The constant β allows conversion of $m^3/seconds$ to $hm^3/days$.

At time $t = 1$, we have $\mathcal{V}_{jt-1}(\xi) = v_{j0}$ where v_{j0} represents the fixed known amount of water (in hm^3) in reservoir j at the beginning of the time horizon. We assume a relatively small basin and use the approximation $\xi_{jt} = \alpha_{jt}\xi_t$ to represent the total water inflow in m^3/s from natural precipitations and spring thaw going in reservoir j at time t where α_{jt} is the average fraction of aggregate inflows at time t entering reservoir j.

Constraints (4) ensure that flows are within limits $\underline{r}_{it}, \bar{r}_{it}$ while constraints (6) ensure that the total flow deviation at a given plant i does not exceed a pre-specified threshold Δ_i from one day to the next. Constraints (5) bound the maximum amount of water that can be unproductively spilled for a given volume in the upstream reservoir while (7) ensure respect of absolute upper and lower bounds. These constraints are determined by specific physical characteristic of given plants. Constraints (8) ensure the productive water flows at plant i are within prescribed bounds $\underline{q}_{it}, \bar{q}_{it}$, $\forall t$. These are based on navigation and flood safety thresholds as well as agreements with riparian communities. Finally, (9) defines the total flow as the sum of unproductive spillage and turbined outflow.

Constraints (10) define overflows with respect to the critical water volume levels $\underline{v}_{jt}, \bar{v}_{jt}$ and represent quantities we wish to minimize on average. Since the bounds are taken with respect to a given useful reservoir volume, 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.

The parameters $\kappa_{j,t} > 0$ in (1) 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. For our problem, the dichotomy between critical and non-critical reservoirs is unequivocal, but we could easily adapt this to more intricate cases.

We observe that the optimal policy of problem (1)-(10) at a given time t, implicitly depends on the past decisions as well as the past realized random variables. As reflected in the notation of problem (1) -(10), we therefore consider decision rules to emphasize the fact that at current time t, we must optimize over some functional space.

2.5 Lookahead model with affine decision rules

Even if we ignore uncertainty, solving the multi-stage problem (1) - (10) 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 volumes. For moderate $|J|$ and T, this is already too demanding to be solved directly through classical dynamic programming, which requires state-space discretization.

Thus, we consider a simpler lookahead model based on a restricted class of possibly sub-optimal policies. We specifically consider simple affine functions of the uncertain inflows. These decision rules were popularized in dynamic/adjustable robust optimization models by (3) and have gained considerable attention in the recent years namely in the field of energy (see for instance (6; 4; 11; 21)).

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

$$\mathcal{X}_{k,t+l}(\xi) = \mathcal{X}^0_{k,t+l} + \sum_{l'=0}^{T-t} \mathcal{X}^{l'}_{k,t+l}\xi_{t+l'} \tag{11}$$

where $\mathcal{X}^{l'}_{k,t+l} \in \mathbb{R}$ for $k \in K_{t+l}$, $t \in \mathbb{T}, l \in \mathbb{L}$ and $l' = 0, ..., T-t$. Decisions concerning real implementable decisions at time $t+l$ should only depend on random variables $\xi_{[t+l-1]}$ preceding that time. 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 $t+l$. On the other hands, non-implementable decisions taken at time $t+l$ will also depend on ξ_{t+l} and we must therefore enforce $\mathcal{X}^{l'}_{k,t+l} = 0$, $\forall l' \geq l+1$ and $k \in K_{t+l} \setminus K^{impl}_{t+l}$. The decisions obtained by solving RMP_t at time t will therefore only explicitly involve $\xi_{[t+L-1]}$.

Although (11) and the preceding paragraph only emphasize the dependence on $\xi_{[t,T]}$, the impact of the past observed $\xi_{[t-1]}$ is reflected implicitly in $\mathcal{X}^0_{k,t+l}$. Indeed, since the $\mathcal{X}^{l'}_{k,t+l}$ are decision variables without sign restrictions over which we will optimize, we can always find $\bar{\mathcal{X}}$ such that for a given $\mathcal{X}$ we have $\mathcal{X}^0_{k,t+l} + \sum_{l'=0}^{T-t} \mathcal{X}^{l'}_{k,t+l}\xi_{t+l'} = \bar{\mathcal{X}}^0_{k,t+l} + \sum_{l'=-t+1}^{T-t} \bar{\mathcal{X}}^{l'}_{k,t+l}\xi_{t+l'}$, $\forall k \in K_{t+l}$ with $\mathbb{P}$ a.s. where $\xi_{[t-1]}$ is constant. In sum, at the beginning of time t, problem RMP_t with affine decision rules only requires explicitly considering $\xi_{[t,t+L-1]}$.

This lookahead model 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. Moreover, it becomes much easier to consider constraints such as (3) and (6) that involve decisions in multiple periods. Finally, section (3.7) shows that by leveraging techniques

from robust optimization, we are able to formulate each lookahead problem as a large linear program that can be solved in a single "forward" phase and that makes very little distribution assumptions.

3 Tractable time series models for multi-stage stochastic programs

3.1 General inflow representation

The quality of the solutions returned by solving the lookahead problem RMP_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 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 modelling and tractability purposes, we will also assume that the support of the $\varrho_{[t,t+L-1]}$ is a polyhedron in $\mathbb{R}^L$. More specifically, we assume that with $\mathbb{P}$ a.s. 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 (11), holds:

$$\left\{ \xi : \Omega \to \mathbb{R}^L \;\middle|\; \begin{array}{ll} \exists \zeta, \varrho : \Omega \to \mathbb{R}^L & \text{(12a)} \\ \xi = U_t\zeta + u_t & \text{(12b)} \\ \zeta = V_t\varrho + v_t & \text{(12c)} \\ W_t\varrho \leq w_t & \text{(12d)} \end{array} \right\}$$

We assume that for any ξ, there exists unique ϱ, ζ such that the representation (12) holds. 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. Moreover, it is natural to require this criteria to avoid indeterminate situations. We will further assume that both U_t and V_t as well as their inverse U_t^{-1}, V_t^{-1} 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 prefectly 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 (12c) into (12b). More specifically, we will consider the case where the $\{\zeta_t\}$ follow

well-known autoregressive moving average (ARMA) time series models.

In this context, (12b) 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. 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 $\{r \in \mathbb{R}^L : W_t r \leq w_t\}$ 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 (12) can be updated to more adequately reflect the random environment as we move forward in time and new data is progressively observed.

3.2 Considering general ARMA models

Since ARMA models filter the serial dependency and output white noise, these linear 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 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 $\phi(B)\zeta_t = \theta(B)\varrho_t$ for some $\phi(B) = \sum_{i=0}^{p} \phi_i B^i$ and $\theta(B) = \sum_{i=0}^{q} \theta_i B^i$ with $p, q \in \mathbb{N}$ and $\theta_0 = \phi_0 = 1$ where B represents the backshift operator acting on time indices such that $B^p \zeta_t = \zeta_{t-p}$ for all $t, p \in \mathbb{Z}$ (see (22; 23)). We suppose the ϱ_t are independent identically distributed zero mean and $\mathcal{F}_t$-measurable random variables. In order to guarantee second order stationarity, we require that the process autocovariance function $\gamma(l) = \mathrm{E}\left[\varrho_t \varrho_{t+l}\right]$ depend only on $l \in \mathbb{Z}$ and in particular that the variance $\gamma(0) = \sigma_\varrho^2$ be constant across time.

We also assume that the process $\{\zeta_t\}$ is stable, that is we can find $\psi(B) = \sum_{i=1}^{\infty} \psi_i B^i$ such that $\phi(B)\psi(B) = \theta(B)$ and $\sum_{i=1}^{\infty} |\psi_i| < \infty$. In this case, $\psi(B)\varrho_t = \zeta_t$, we can express ζ_t as an infinite linear combination of past $\{\varrho_\tau\}_{\tau=t,t-1,\ldots}$ and the representation is essentially unique (see (23)). 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 $d \in \mathbb{N}$.

These representations are particularly useful when forecasting the future values of ζ_t given the information available at time $t \in \mathbb{Z}$. Specifically, $\hat{\zeta}_t(l) \equiv \mathrm{E}\left[\zeta_{t+l} | \mathcal{F}_t\right]$ is a natural choice of forecast as it represents the minimum mean squared error linear estimator of ζ_{t+l} given the information up to time $t \in \mathbb{Z}$ for $l \in \mathbb{N}$ (see (23)). We also write $\mathrm{E}\left[\varrho_{t+l} | \mathcal{F}_t\right] \equiv \hat{\varrho}_t(l)$ and observe that $\hat{\varrho}_t(l) = \varrho_{t+l}$ if $l \in \{0, -1, ...\}$ and 0 otherwise.

Given the stable process $\{\zeta_t\}$ and $(t,l)^\top \in \mathbb{Z}^2$ we have $\zeta_{t+l} = \hat{\zeta}_t(l) + \rho_t(l)$ where $\hat{\zeta}_t(l) = \sum_{j=l}^{\infty} \psi_j \varrho_{t+l-j}$ is the forecast and $\rho_t(l) = \sum_{j=0}^{l-1} \psi_j \varrho_{t+l-j}$ is the forecast error. For any pair $(t,l)^\top \in \mathbb{Z}^2$, we observe that $\rho_t(l)$ is $\mathcal{F}_{t+l}$ measurable while $\hat{\zeta}_t(l)$ is $\mathcal{F}_t$ measurable. In particular, for $l \in \{0,-1,...\}$, $\rho_t(l) = 0$ while $\hat{\zeta}_t(l) = \zeta_{t+l}$. If we set $\rho_{t-1,L} \equiv (\rho_{t-1}(1),...,\rho_{t-1}(L))^\top$ for any $t \in \mathbb{T}$, 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 $L \in \{1,...,T-t+1\}$:

$$\rho_{t-1,L} = V_t \varrho_{[t,t+L-1]} \tag{13}$$

where $V_t \in \mathbb{R}^{L \times L}$ is the following invertible and lower triangular square matrix:

$$V_t = \begin{pmatrix} 1 & & \cdots & 0 \\ \psi_1 & 1 & & \vdots \\ \vdots & & \ddots & \\ \psi_{L-1} & \cdots & \psi_1 & 1 \end{pmatrix} \tag{14}$$

We then have the equality:

$$\begin{aligned} \zeta_{[t,t+L-1]} &= \hat{\zeta}_{t-1,L} + \rho_{t-1,L} && (15) \\ &= \hat{\zeta}_{t-1,L} + V_t \varrho_{[t,t+L-1]} && (16) \end{aligned}$$

where $\hat{\zeta}_{t-1,L} \equiv (\hat{\zeta}_{t-1}(1),...,\hat{\zeta}_{t-1}(L))^\top$ corresponds to v_t in the representation (12). The structure of V_t 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 $\hat{\zeta}_{t-1,L}$:

$$\begin{aligned} \xi_{[t,t+L-1]} &= U_t \zeta_{[t,t+L-1]} + u_t && (17) \\ &= U_t(\hat{\zeta}_{t-1,L} + \rho_{t-1,L}) + u_t && (18) \\ &= U_t(\hat{\zeta}_{t-1,L} + V_t \varrho_{[t,t+L-1]}) + u_t && (19) \end{aligned}$$

The affine representation (19) reveals that the $\xi_{[t,t+L-1]}$ vector is completely determined by $\varrho_{[t,t+L-1]}$. We therefore set $\sigma(\varrho_s; s \leq t) = \mathcal{F}_t, \forall t \in \mathbb{T}$ which reflects the fact that observing $\varrho_{[t-1]}$ at the beginning of time $t \in \mathbb{T}$ gives us all the information necessary to apply the real implementable policies at times $1,...,t-1$.

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 $F_{\varrho_t,...,\varrho_{t+L-1}}$, which also corresponds to the set $\{r \in \mathbb{R}^L : W_t r \leq w_t\}$ described in (12), is a bounded polyhedron in $\mathbb{R}^L$ given by the intersection of the following two polyhedrons[1]:

$$\mathscr{B}^{\infty}_{L,\epsilon} = \{r \in \mathbb{R}^L : |r_i|\sigma_{\varrho}^{-1} \leq (L\epsilon)^{1/2}, i = 1, ..., L\} \tag{20a}$$

$$\mathscr{B}^{1}_{L,\epsilon} = \{r \in \mathbb{R}^L : \sum_{i=1}^{L} |r_i|\sigma_{\varrho}^{-1} \leq L\epsilon^{1/2}\} \tag{20b}$$

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^2_{ϱ}, 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 $\Sigma_{\varrho,L} = \sigma^2_{\varrho} I_L$ where I_L is the $L \times L$ identity matrix. Therefore, if tr denotes the (linear) trace operator, for $\tilde{\varrho} : \Omega \to \mathbb{R}^L$, Markov's inequality gives us:

$$\mathbb{P}(\tilde{\varrho}^{\top}\Sigma^{-1}_{\varrho,L}\tilde{\varrho} > L\epsilon) \leq \mathrm{E}\left[\tilde{\varrho}^{\top}\Sigma^{-1}_{\varrho,L}\tilde{\varrho}\right](L\epsilon)^{-1} \tag{21}$$

$$= tr(\Sigma^{-1}_{\varrho,L}\mathrm{E}\left[\tilde{\varrho}\tilde{\varrho}^{\top}\right])(L\epsilon)^{-1} \tag{22}$$

$$= \epsilon^{-1} \tag{23}$$

The polytope $\mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}$ contains the ellipsoid $\mathscr{B}^{2}_{L,\epsilon} = \{r \in \mathbb{R}^L : r^{\top}\Sigma^{-1}_{\varrho,L}r \leq L\epsilon\}$ (see the Appendix for more details). It follows that $\mathbb{P}(\tilde{\varrho} \in \mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}) \geq 1 - \epsilon^{-1}$ for all $t, L \in \mathbb{N}$.

For our particular case, we consider a large ϵ and reasonably assume $\mathbb{P}(\tilde{\varrho} \in \mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}) = 1$. If the $\{\varrho_t\}$ are essentially bounded iid random variables with constant variance σ^2_{ϱ}, 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 (see (24)).

The polyhedron defined by (20a) - (20b) 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.

Before moving on, it is important to acknowledge the time consistency issues with our hypothesis. For instance, considering $L = 1, 2$ reveals that neither $\mathscr{B}^{1}_{1,\epsilon} \times \mathscr{B}^{1}_{1,\epsilon} \not\subset \mathscr{B}^{1}_{2,\epsilon} \cap \mathscr{B}^{\infty}_{2,\epsilon}$ nor $\mathscr{B}^{1}_{1,\epsilon} \times \mathscr{B}^{1}_{1,\epsilon} \not\supset \mathscr{B}^{1}_{2,\epsilon} \cap \mathscr{B}^{\infty}_{2,\epsilon}$. Hence, at time $t \in \mathbb{T}$ for lead time horizon $L \in \mathbb{N}$, we may consider realizations of $\varrho_{[t,t+L-1]}$ that we had not considered possible for $\varrho_{[t+L-1]}$ at the beginning of time 1. Similarly, at time t we will disregard certain realizations of $\varrho_{[t,t+L-1]}$ that we had

[1]The set $\mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}$ is not strictly speaking a polyhedron since (20a)-(20b) involve the non-linear absolute value function. Nonetheless, lifting this set using the commonly used decomposition $r_i = r_i^+ - r_i^-$ and $|r_i| = r_i^+ + r_i^-$ with $r_i^+, r_i^- \geq 0, \forall i$ yields a polyhedron where each projected point lies in the original $\mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}$ (see (24)).

previously considered in the past to determine our previous decisions.

This difficulty is directly related to dynamic update issues raised in section 6 of (19). Resolving the matter could namely be done by using more conservative tractable approximations based on projections. However, this may result in decisions that consider overly pessimistic scenarios. Our simplifying hypotheses are rooted in practical considerations and retain reasonable probabilistic interpretations. Moreover, such inconsistencies are already implicitly present in various problems where uncertainty is progressively revealed (namely (11)).

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

Building on these assumptions, we now consider the conditional support of $\xi_{[t,t+L-1]}$ for $L \in \{1, ..., T-t+1\}$ given the past observations $\xi_{[t-1]}$ and $\varrho_{[t-1]}$, which is all that is required to solve RMP_t at the beginning of time t (see section (2.5)) :

$$\mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1} = \left\{ x \in \mathbb{R}^L \;\middle|\; \begin{array}{ll} \exists r \in \mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon} & \text{(24a)} \\ x = U_t(\hat{\zeta}_{t-1,L} + V_{t-1} r) + u_t & \text{(24b)} \end{array} \right\}$$

Equation (24b) defines the future inflow realizations of $\xi_{[t,t+L-1]}$ for the L day forward horizon through the affine representation (19). Finally, the vector $r \in \mathbb{R}^L$ corresponds to the possible realizations of the future L dimensional random vector $\varrho_{[t,t+L-1]}$ and (24a) ensures that these values reside within the bounded polyhedron defined by (20a)- (20b) described previously. We observe that this set is based on a more compact formulation of (12) using the appropriate substitutions and previous definitions.

The set $\mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1}$ implicitely depends on past $\varrho_{[t-1]}$ through $\hat{\zeta}_{t-1,L}$ 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 $\mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1}$ 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.

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 $m, s \in \mathbb{N}$. In this case, the $\{\varrho_t\}$ are still uncorrelated zero-mean random variables with constant variance σ_a^2. However, they are no longer independent since they respect the following relation:

$$\hat{\sigma}^2_{t-1}(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho^2_{t-i} + \sum_{j=1}^{s} \beta_j \hat{\sigma}^2_{t-1-j}(1), \ \forall t \tag{25}$$

where $\hat{\sigma}^2_t(l) = \text{E}\left[\varrho^2_{t+l}|\mathcal{F}_t\right]^2$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. Furthermore, in order for the $\{\varrho_t\}$ to be second order stationary with

constant variance σ_a^2 we require that:

$$\mathrm{E}\left[\mathrm{E}\left[\varrho_t^2|\mathcal{F}_{t-1}\right]\right] = \mathrm{E}\left[\alpha_0 + \sum_{i=1}^{m}\alpha_i\varrho_{t-i}^2 + \sum_{j=1}^{s}\beta_j\hat{\sigma}_{t-1-j}^2(1)\right] \tag{26}$$

$$\Leftrightarrow \sigma_a^2 = \alpha_0\left(1 - \sum_{j=1}^{\max\{m,s\}}(\alpha_j+\beta_j)\right)^{-1} \tag{27}$$

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 $j > s$. We can also show the squared shocks ϱ_t^2 satisfy the difference equation:

$$\hat{\phi}(B)\varrho_t^2 = \alpha_0 + \hat{\theta}(B)\nu_t \Leftrightarrow \hat{\phi}(B)\left(\varrho_t^2 - \sigma_a^2\right) = \hat{\theta}(B)\nu_t \tag{28}$$

where the $\{\nu_t\}$ are zero mean uncorrelated random variables with $\nu_t = \varrho_t^2 - \hat{\sigma}_{t-1}^2(1)$. $\hat{\theta}(B)$ is an order s polynomial in B and $\hat{\phi}(B)$ is an order $\max\{s,m\}$ polynomial in B and $\hat{\phi}(B)k = k\sum_{i=1}^{\max\{s,m\}}\hat{\phi}_i = k\left(1 - \sum_{j=1}^{\max\{m,s\}}(\alpha_j+\beta_j)\right)$ for any constant $k \in \mathbb{R}$. Given that the $\{\varrho_t\}$ are second order stationary and condition (27) holds, we can find $\hat{\psi}(B)$ such that $\hat{\phi}(B)\hat{\psi}(B) = \hat{\theta}(B)$ and we therefore have:

$$\varrho_t^2 = \sigma_\varrho^2 + \hat{\psi}(B)\nu_t \tag{29}$$

Taking the conditional expectation on both sides of equation (29) at the beginning of time t allows us to obtain an expression for the conditional variance reminiscent of the conditional expectation described in section (3.2). For $l \in \mathbb{N}$, we specifically have:

$$\hat{\sigma}_t^2(l) = \sigma_\varrho^2 + \sum_{j=l}^{\infty}\hat{\psi}_j\nu_{t+l-j} \tag{30}$$

Since σ_ϱ^2 is constant and $\sum_{j=l}^{\infty}\hat{\psi}_j\nu_{t+l-j}$ is $\mathcal{F}_t$ measurable, the difference $\hat{\sigma}_t^2(l) - \sigma_\varrho^2$ is a known value at time t. At the beginning of time t, before observing $\xi_{[t,t+L-1]}$, we therefore use the the same conditional uncertainty set $\mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1}$ as before, but modify constraints (20a)-(20b) to:

$$\mathscr{B}_{L,\epsilon}^{\infty}|\mathcal{F}_{t-1} = \{r \in \mathbb{R}^L : |r_i|\hat{\sigma}_{t-1}(i)^{-1} \leq (L\epsilon)^{1/2}, i = 1,...,L\} \tag{31a}$$

$$\mathscr{B}_{L,\epsilon}^{1}|\mathcal{F}_{t-1} = \{r \in \mathbb{R}^L : \sum_{i=1}^{L}|r_i|\hat{\sigma}_{t-1}(i)^{-1} \leq L\epsilon^{1/2}\} \tag{31b}$$

The use of this set is heuristic and suffers from the same time consistency issues discussed previously. Nonetheless, the conditional Markov inequality yields: $\mathbb{P}(\tilde{\varrho}^\top\Sigma_{\varrho,L,t-1}^{-1}\tilde{\varrho} > L\epsilon|\mathcal{F}_{t-l}) \leq \epsilon^{-1}$ regardless of the realisations of $\tilde{\varrho} = \varrho_{[t,t+L-1]}$ and $\Sigma_{\varrho,L,t-1} = \mathrm{E}[\tilde{\varrho}\tilde{\varrho}^\top|\mathcal{F}_{t-1}]$. The conditional covariance matrix remains diagonal since the ϱ_t are uncorrelated and thus

[2]Although non-standard, we adopt the notation $\hat{\sigma}_t^2(l)$ to maintain the coherence with the past sections and to highlight the similarities with the conditional expectation of ζ_{t+l} for some $t \in \mathbb{T}$ and $l \in \{0,...,T-t\}$ given $\mathcal{F}_t$, which we denoted $\hat{\zeta}_t(l)$.

$\Sigma_{\varrho,L,t-1} = diag(\hat{\sigma}^2_{t-1}(1), ..., \hat{\sigma}^2_{t-1}(L))$.

Using the same arguments as before, the following inclusion holds at time t given knowledge of $\varrho_{[t-1]}$: $\{r \in \mathbb{R}^L : r^\top \Sigma^{-1}_{\varrho,L,t-1} r\} \subset \mathscr{B}^\infty_{L,\epsilon}|\mathcal{F}_{t-1} \cap \mathscr{B}^1_{L,\epsilon}|\mathcal{F}_{t-1}$ and the sets are all deterministic. It follows that $\mathbb{P}(\tilde{\varrho} \in (31a) - (31b)|\mathcal{F}_{t-1}) \geq 1 - \epsilon^{-1}$ for all $t, L \in \mathbb{N}$ and any past observed $\varrho_{[t-1]}$. Using these sets with a large ϵ will therefore capture all possible realizations of the future $\varrho_{[t,t+L-1]}$ at time $t \in \mathbb{T}$ based on past observations with high probability.

In addition, this set is merely a generalization of (20a)-(20b) since we have $\mathrm{E}\left[\varrho_{t+l}\varrho_{t+k}|\mathcal{F}_{t-1}\right] = \mathrm{E}\left[\varrho_{t+l}\varrho_{t+k}\right] = \gamma(|k-l|)$ for any $t, l, k \in 0 \cup \mathbb{N}$ and (31a)-(31b) collapses back to (20a)-(20b) when the ϱ_t are independent. Hence, this set allows us to model the future realizations of the ϱ_t more precisely when the conditional variance is not constant. In any case, this can be considered a basic assumption of our data-driven model.

Under these assumptions, the size of the conditional support depends on past observations as well as the lead time L. This should be contrasted with (20a) - (20b) that only depends on L. In addition, large past errors will boost $\hat{\sigma}_{t-1}(i)$ and hence the size of the polyhedron. These observations suggest that a stochastic process that follows the fitted time series model very closely will 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 $\hat{\sigma}_{t-1}(i)$ and hence to large conditional supports.

3.6 Additional modelling considerations

The representation (24) 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 (19). 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 projected polyhedron, which is perfectly known at the beginning of time $t \in \mathbb{T}$:

$$\mathscr{P}_L|\mathcal{F}_{t-1} = \left\{r \in \mathbb{R}^L \ : U_t V_{t-1} r \geq -(U_t \hat{\zeta}_{t-1,L} + u_t)\right\} \tag{32a}$$

The additional structure imposed by equation (32a) affects the independence 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 $\mathscr{P}_L|\mathcal{F}_{t-1}$. 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.

3.7 Computing the conditional expectation with robust optimization techniques

Consider our affine lookahead model at the beginning of time $t \in \mathbb{T}$, after observing the past $\varrho_{[t-1]}$ and $\xi_{[t-1]}$, given by equations (1)-(10) for a horizon of $L \in \{1, ..., T-t+1\}$ days and where $\xi \equiv \xi_{[t,t+L-1]}$ in a more condensed and abstract form:

$$\min_{\mathcal{X}} \mathrm{E}\left[\sum_{l=0}^{L-1} c_{t+l}^{\top} \mathcal{X}_{t+l}(\xi) | \mathcal{F}_{t-1}\right] \tag{33a}$$

$$\text{s.t.} \quad \sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}_{t+\bar{l}}(\xi) \geq C_{t+l}\xi \qquad \forall l \in \mathbb{L} \qquad \mathbb{P} \text{ a.s.} \tag{33b}$$

for some $A_{t+l,t+\bar{l}} \in \mathbb{R}^{m_{t+l} \times n_{t+\bar{l}}}$, $C_{t+l} \in \mathbb{R}^{m_{t+l} \times L}$, $\mathcal{X}_{t+\bar{l}}(\xi) \in \mathbb{R}^{n_{t+\bar{l}}}$ for $l \in \mathbb{L}$ and $\bar{l} \in \{1, ..., l\}$ for $m_t, n_t \in \mathbb{N}$.

We exploit the decomposition $\xi_{[t,t+L-1]} = U_t(\hat{\zeta}_{t-1,L} + V_t \varrho_{[t,t+L-1]}) + u_t$ presented in section (3.2). We then consider the affine decision rules (11) in a more compact form: $\mathcal{X}_t^0 + \mathcal{X}_t^{\Delta} x$ for $x \in \mathbb{R}^L$, $\mathcal{X}_t^0 \in \mathbb{R}^{n_t}$ and $\mathcal{X}_t^{\Delta} \in \mathbb{R}^{n_t \times L}$ whose structure depends on the non-anticiaptivity of the respective decisions.

Assuming that constraint (32a) is ignored or not binding in most situations, the hypothesis that the ϱ_τ are uncorrelated and $\mathrm{E}[\varrho_{t+l}|\mathcal{F}_{t-1}] = \mathrm{E}[\varrho_{t+l}] = 0, \forall l \in 0 \cup \mathbb{N}$ remain reasonable. Under these assumptions, the objective value then becomes:

$$\mathrm{E}\left[\sum_{l=0}^{L-1} c_{t+l}^{\top} \mathcal{X}_{t+l}(\xi) | \mathcal{F}_{t-1}\right] = \sum_{l=0}^{L-1} c_{t+l}^{\top} (\mathcal{X}_{t+l}^0 + \mathcal{X}_{t+l}^{\Delta} u_t + \mathcal{X}_{t+l}^{\Delta} U_t \hat{\zeta}_{t-1,L}) \tag{34}$$

Using the definition of the conditional joint support of $\xi_{[t,t+L-1]}$ given by (24), we see that for any $f : \mathbb{R}^L \to \mathbb{R}$ and $k \in \mathbb{R}$, $f(x) \geq k, \forall x \in \mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1} \Rightarrow \mathbb{P}(f(\xi_{[t,t+L-1]}) \geq k|\mathcal{F}_{t-1}) = 1$ with $\mathbb{P}$ a.s.. We therefore consider the following problem:

$$\min_{\mathcal{X}^0, \mathcal{X}^{\Delta}} \sum_{l=0}^{L-1} c_{t+l}^{\top} (\mathcal{X}_{t+l}^0 + \mathcal{X}_{t+l}^{\Delta} u + \mathcal{X}_{t+l}^{\Delta} U \hat{\zeta}_{t-1,L}) \tag{35a}$$

$$\text{s.t.} \quad (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}_{t+\bar{l}}^{\Delta} - C_{t+l}) x \geq -\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}_{t+\bar{l}}^0 \quad \forall l \in \mathbb{L}, \qquad \forall x \in \mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1} \tag{35b}$$

Its optimal solution represents an upper bound on (33a) - (33b) because we limit ourselves to affine functions. Since $\mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1}$ is a polyhedron, we can handle the constraints (35b) through robust optimization techniques (see namely (24)). More specifically, we use linear programming duality and write the entire program as a large (minimization) linear program.

From this derivation, we also see that the optimal value of problem (35a)-(35b) will be an upper bound on the conditional expectation of (weighted) floods over the horizon $t, ..., t+L-1$ for *any* distribution of the ϱ_t provided that the true support remains within the polyhedral

support defined by (31), constraint (32a) is not binding and the structure of the ARMA and GARCH models are correct.

4 Solution procedure

4.1 Monte Carlo simulation and rolling horizon framework

Solving the problem (1)-(10) 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 $t, ..., t+L-1$ by considering some horizon $L \in \mathbb{N}$[3]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 (5.2) 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 $\varrho^s_{[T]} \equiv (\varrho_{s,1}, ..., \varrho_{s,T})^\top$ which together with the fixed and deterministic initial inflows ξ_0[4]completely determine inflows $\xi^s_{[T]} \equiv (\xi_{s,1}, ..., \xi_{s,T})^\top$. For a given scenario s at the beginning of time t, after having observed the past history $\xi^s_{[t-1]}$ and $\varrho^s_{[t-1]}$, the initial volume $\mathcal{V}_{j,t-1}(\xi^s_{[T]})$ and the past water releases $\mathcal{R}_{i,t'}(\xi^s_{[T]}), t' \leq t-1, i \in I$, 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 $t, ..., t+L-1$ and by taking the deterministic equivalent when considering affine decision rules with the conditional support $\mathscr{X}_{L,\epsilon}|\mathcal{F}_{t-1}$. We then implement the first day decisions, observe the total random inflow $\xi_{s,t}$ during time t, compute the linear combination of actual floods,

[3]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.

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

update $\mathcal{V}_{jt}(\xi^s_{[T]})$ and the past water releases $\mathcal{R}_{t',i}(\xi^s_{[T]}), t' \leq t;\ i \in I$ and solve RMP_{t+1}. We repeat this step for times $t = 1, ..., T - L + 1$ for each of S sample trajectories.

5 Case study

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 (25). 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 (2)).

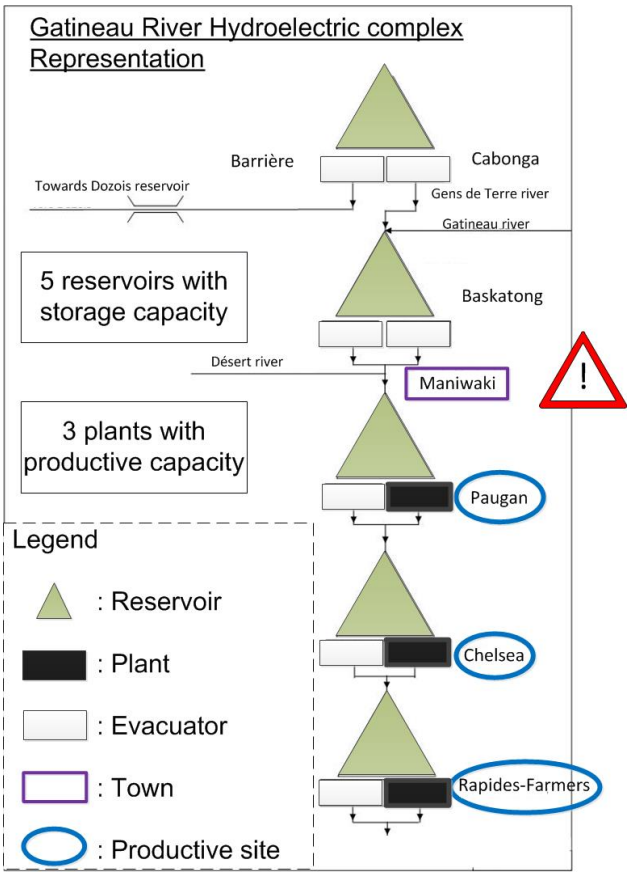

Figure 2: Simplified representation of the Gatineau river system

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 volumes. If the head reservoirs are not sufficiently emptied before the freshet, there is a significant risk of disrupting

normal operating conditions and flooding (see (6)).

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 been used to control baseflow at the greater Montreal region several hundreds of kilometres downstream. As such, respect of minimum and maximum water volume threshold is essential for river operators.

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 (3) illustrates, 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 liquid inflows during the winter months.

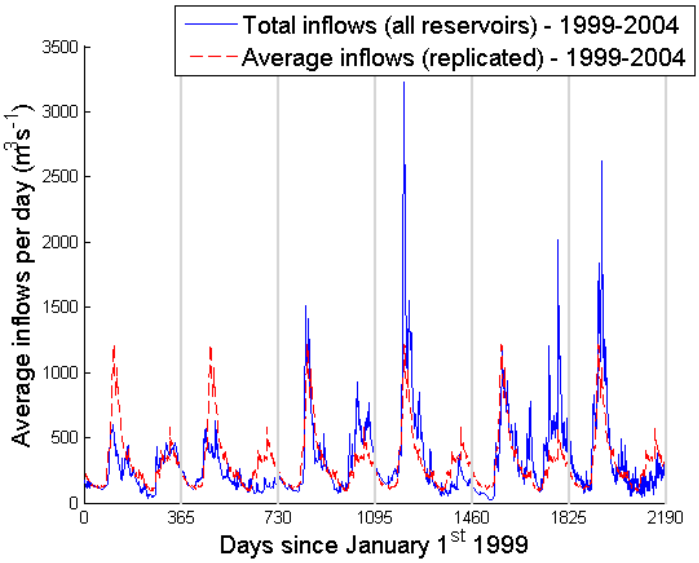

Figure 3: Sample inflows for 1999-2004 (6 years)

5.2 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 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, $u = \mathrm{E}[\xi] = (\mu_1, ..., \mu_T)^\top \in \mathbb{R}^T$ and $U^{-1} = diag(\sigma_1^{-1}, ..., \sigma_T^{-1})$ for our affine representation $\xi = U\zeta + u$ 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 $s \in \mathbb{N}$ (see (26; 22)). These are all compatible with our framework at no additional complexity. We specifically experimented with Fourier analysis, but removing the high frequency terms with smaller power led to patterns that failed to capture

important characteristics of the inflows and ultimately decreased the quality of the forecasts.

The ζ_t still display a significant amount of serial correlation. In order to express them as an infinite linear combination of uncorrelated white noise random variables, we consider Box-Jenkins methodology (see (22)) and find that they approximately follow a ARMA$(1,1)$ process. That is $\phi(B)\zeta_t = \theta(B)\varrho_t$ where $\phi(B) = 1 - \phi B$ and $\theta(B) = 1 + \theta B$. 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 $\phi(B)\psi(B) = \theta(B)$ and $\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$.

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 (4)). However, repeatedly forecasting the future values as new data becomes available in a rolling horizon fashion provides much better predictive power.

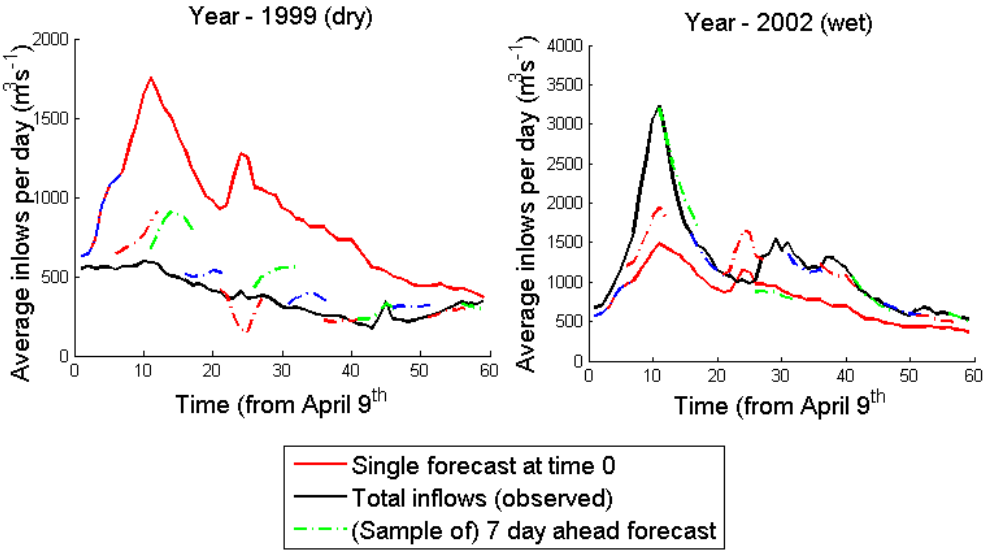

Figure 4: Comparing simple forecasts for 1999 & 2002

5.3 Heteroscedastic inflows

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

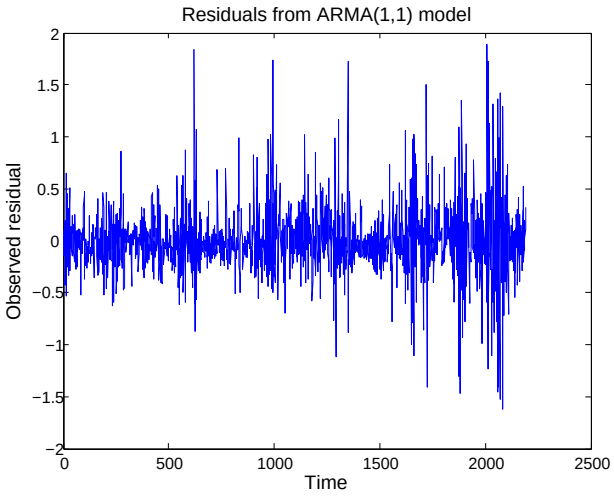

Figure 5: Residuals from ARMA(1,1) model

We find that the residual $\{\varrho_t\}$ approximately follow a GARCH$(1,1)$ model with the following estimates: $\alpha_0 = 0.01, \alpha_1 = 0.14, \beta_1 = 0.84$.

5.4 Comparing forecasts

Sections (3.3) - (3.5) suggests that bad forecasts will lead to poor representation of the future uncertainty in the form of large conditional supports. Before evaluating the effectiveness of our lookahead models at reducing the occurrence and magnitude of floods, we require a formal mechanism to evaluate the relative forecasting abilities of our calibrated ARMA and GARCH models relative to other simpler forecasts.

We specifically consider a statistical measure commonly used by meteorologist and hydrologist known as the *skill* (see (27)). In our case, we wish to evaluate the accuracy of our time series models compared with the naive static forecast consisting of the historical daily mean. More specifically, for a any $t \in \mathbb{T}$ and L dimensional time series $\xi_{[t,t+L-1]}$, we consider the corresponding forecasts $\{\hat{\xi}_{frcst,t+l}\}_{l=0}^{L-1}$ and $\{\hat{\xi}_{naive,t}\}_{l=0}^{L-1}$ where $\hat{\xi}_{naive,t+1} = \mu_{t+l}, \forall l \in \mathbb{L}$.

We use the standard skill score, also known as the Nash-Sutcliffe efficiency which is one of the simplest and most popular variants (see (28; 29)) and 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 $MSE_{naive,t,L} = L^{-1}\sum_{l=0}^{L-1}\mathrm{E}\left[(\xi_{t+l} - \mu_{t+l})^2\right]$ represent the mean square error of both forecasts. 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 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, which reflects the real decision process for river operators during this period. 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 volume 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.

For simulated inflows, we present the sample CVaR_α of floods for $\alpha = 10^{-2}n, n = 0, 1, ..., 100$ 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 $\text{VaR}_\alpha(X) = q_X(1-\alpha) = \inf\{t : \mathbb{P}(X \leq t) \geq 1-\alpha\}$ is the $1 - \alpha$ quantile.

As in (6), 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 (30), 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) reveals that the bound can be relatively tight in certain cases and therefore that our models perform well under certain scenarios.

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

We first assume that 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 $\bar{\phi}$ so that the relationship $\bar{\zeta}_{t+l} = \sum_{i=0}^{\infty}\bar{\psi}_i\varrho_{t+l-i}$ is exact with $\bar{\psi}_0 = 1, \bar{\psi}_i = \bar{\phi}^{i-1}(\bar{\phi}+\bar{\theta}), i \geq 1$, but that the process considered by the model is $\zeta_{t+l} = \sum_{i=0}^{\infty}\psi_i\varrho_{t+l-i}$ with parameters ϕ and θ.

We also consider that the $\{\varrho_t\}$ are zero-mean uncorrelated normal random variables with unconditional variance σ_ϱ^2. Hence these random variables violate our assumption of boundedness made in section (3.3). Choosing random variables with support defined exactly by (20b)-(20a) or (31a)- (31b) leads to virtually no floods.

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:

$$0 \leq 1 - \mathrm{E}\left[MSE_{frcst,t,L}\right]\mathrm{E}\left[MSE_{naive,t,L}\right]^{-1} \tag{36}$$

$$\Leftrightarrow \sum_{l=0}^{L-1}\mathrm{E}\left[(\sigma_{t+l}(\sum_{i=0}^{\infty}(\bar{\psi}_i-\psi_i)\varrho_{t+l-i}))^2\right] \leq \sum_{l=0}^{L-1}\mathrm{E}\left[(\sigma_{t+l}\sum_{i=0}^{\infty}\bar{\psi}_i\varrho_{t+l-i})\right]^2 \tag{37}$$

If $\bar{\phi} = \phi$ and $\bar{\theta} \neq \theta$, then (37) is equivalent to:

$$\sum_{l=0}^{L-1}\sigma_{t+l}^2\frac{(\theta-\bar{\theta})^2}{1-\phi^2}\sigma_\varrho^2 \leq \sum_{l=0}^{L-1}\sigma_{t+l}^2\frac{\bar{\theta}(2\phi+\bar{\theta})+1}{1-\phi^2}\sigma_\varrho^2 \tag{38}$$

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.

As the first 2 graphs of figure (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.

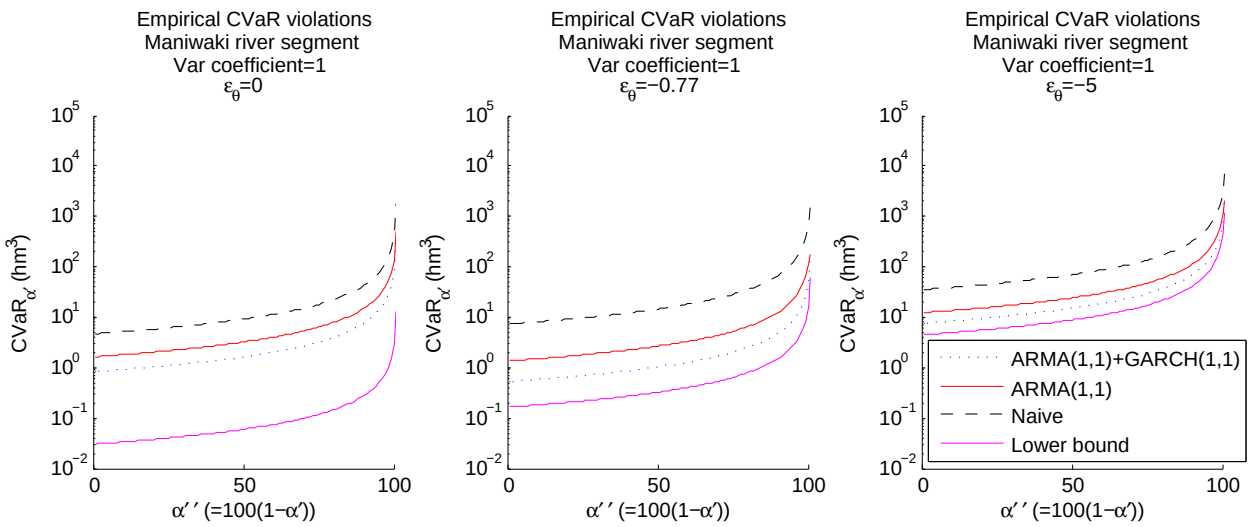

Figure 6: Influence of reduced forecast skill

These conclusions remain valid when we increase the volatility of the true inflow process. Figure (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 $\{\zeta_t\}$.

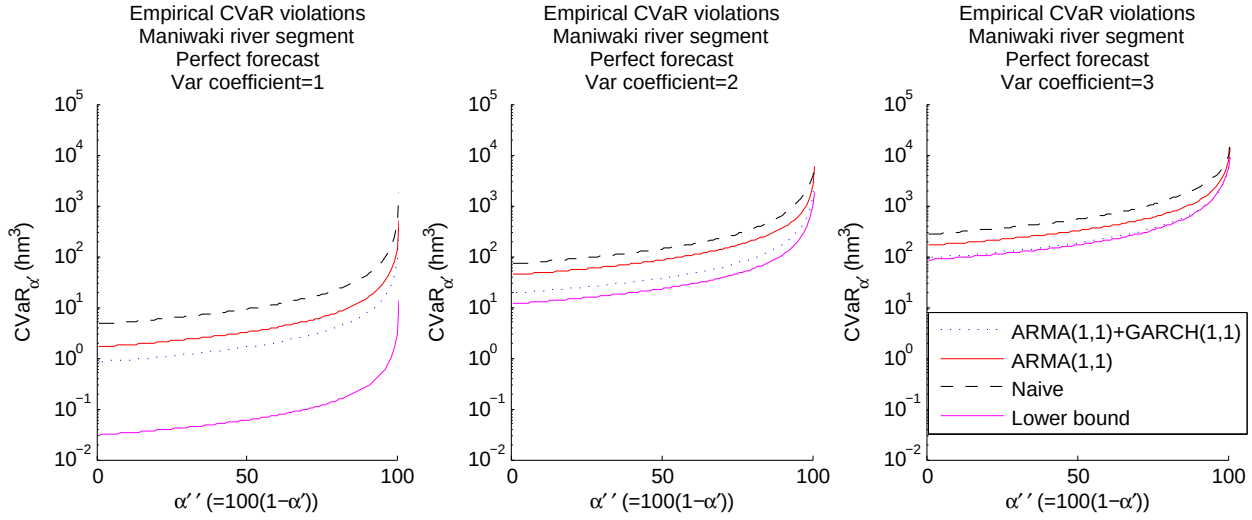

Figure 7: Influence of increased unconditional variance

As illustrated in figures (6) and (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.

5.7 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. Different real scenarios are used as initial inflows to generate varying patterns. This implies that the seasonality of generated inflows may differ from the fixed one considered in the model.

Figure (8) suggests that even in this case, our approach based on dynamic uncertainty sets performs well. It is encouraging to observe such robustness with respect to time series structure.

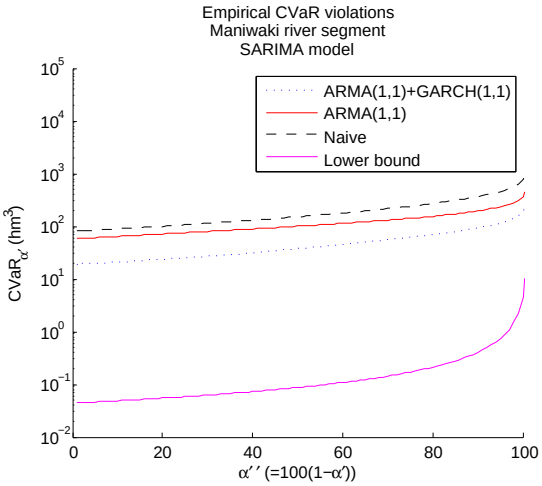

Figure 8: Influence of different time series structure

5.8 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) where used to validate the robustness of our approach.

The real inflows do not perfectly follow an ARMA(1,1) model with GARCH(1,1) error. To evaluate the relative quality of our forecasts based on this time series calibration compared to the naive forecast which is to take the unconditional mean at a given time t, we therefore compute the sample skill $\sum_{s=1}^{\bar{S}}\bar{S}^{-1}\left[1-\frac{\sum_{l=0}^{L-1}(\xi_{t+l}-\hat{\xi}_{frcst,t+l})^2}{\sum_{l=0}^{L-1}(\xi_{t+l}-\hat{\xi}_{naive,t+l})^2}\right]$ at every time $t\in\mathbb{T}$ with $L=7$

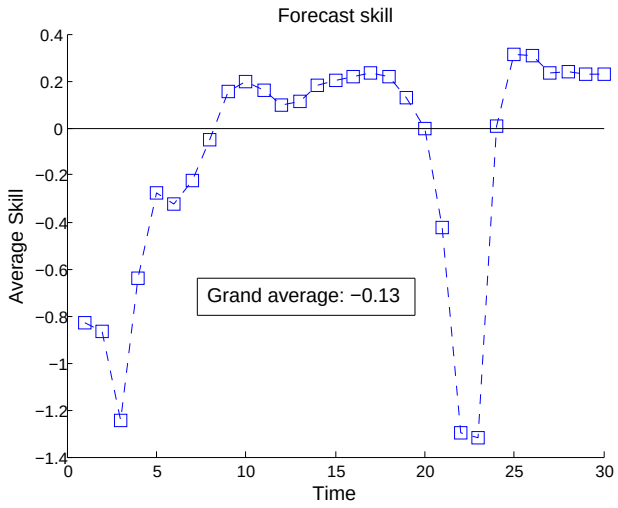

Figure 9: Sample skill of 7 day ahead forecasts for 1999-2004 & 2008-2013 (12 years)

Figure (9) indicates that our forecast does not perform consistently better or worse than the naive forecast, however, on average, it does slightly worse since the average over all time periods is slightly negative. This is reflected in the real behaviour of our model on the 12 years of inflow data (see figure (10)).

The relatively poor forecasting abilities of our calibrated model seem to stem from the fact that the empirical distribution of the ϱ differs significantly from the theoretical distributions used to calibrate the model. This is partly a consequence of limited availability of data as well as the inability to use non-affine transformations to obtain a more satisfactory fit with the theoretical hypothesis of the time series models.

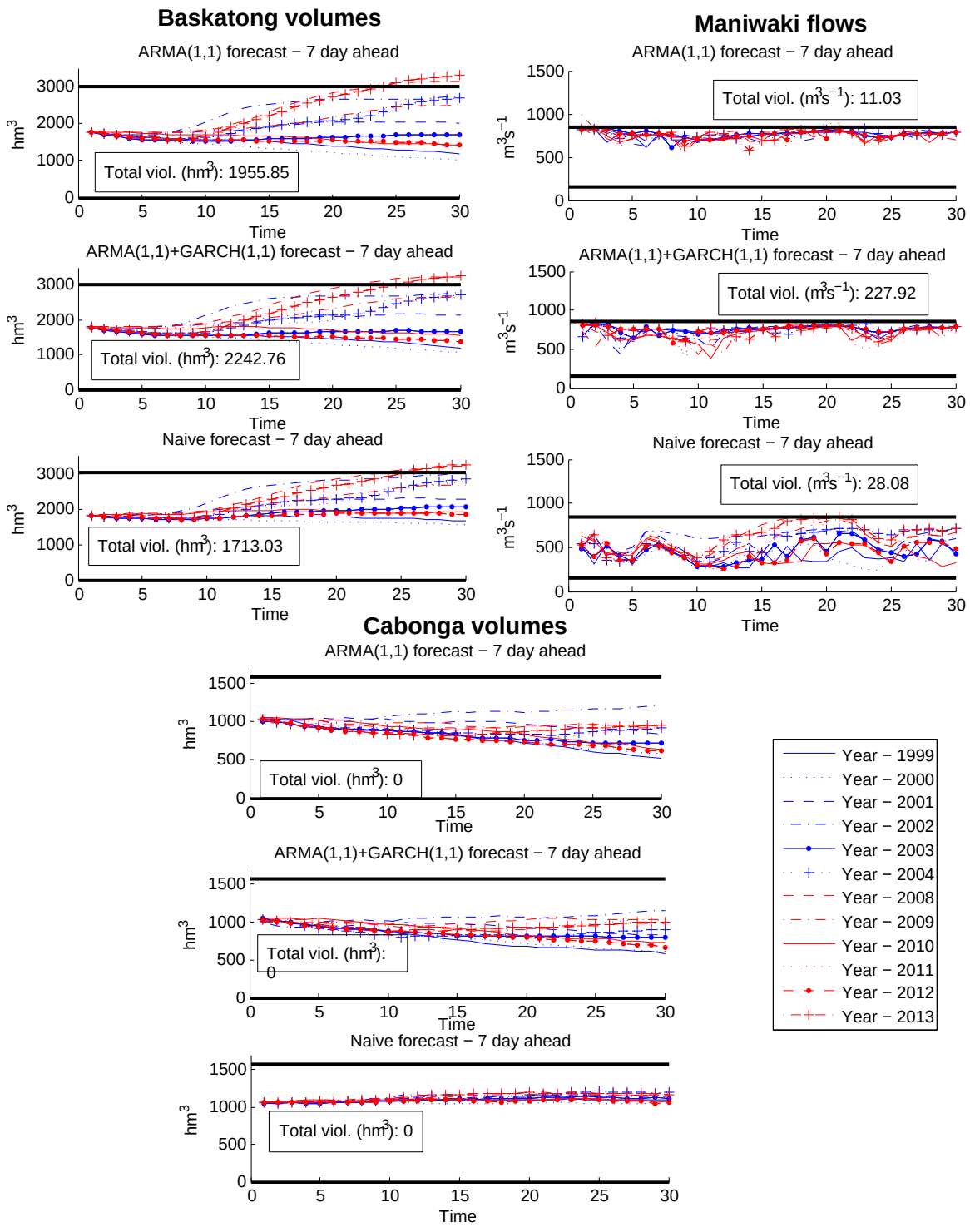

Figure 10: Simulation results for 1999-2004 & 2008-2013 (12 years)

Figure (10) shows violations of volume 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 no method dominates the other two as higher flow violations are associated with lower volume violations and vice-versa.

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 linear ARMA and GARCH time series model provides good quality solutions for a problem that would otherwise be intractable for SDP or SDDP.

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 section (5.6),(5.7) and (5.8) 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 not yield sizeable benefits compared to naive static representations of the random vectors; particularly for out-of-sample inflows.

Our approach based on time series models suffers from weaknesses that are chiefly a consequence of the limited availability of data. The quality of the time series model considered will be significantly influenced by its structure, order and the value of its coefficients. However, the identification of a satisfactory model is not necessarily an easy task which may require more art than science and will likely be hindered by limited data. There might also be an issue of over-fitting due to the parametric nature of the models.

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. This might not be possible for other competing methods such as stochastic programming based on scenario trees or SDP.

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. From a more theoretical perspective, our approach relies on a very limited number of assumptions.

7 Acknowledgments

The authors would like to thank Stein-Erick Fleten 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. This research was supported by the Natural Sciences and Engineering Reasearch Council of Canada (NSERC) and Hydro-Québec.

8 References

References

[1] J. Labadie, Optimal operation of multireservoir systems: state-of-the-art review, Journal of Water Resources Planning and Management 130 (2004) 93–111.

[2] M. Dyer, L. Stougie, Computational complexity of stochastic programming problems, Mathematical Programming Series A.

[3] A. Ben-Tal, E. Goryashko, A. Guslitzer, A. Nemirovski, Adjustable robust solutions of uncertain linear programs, Mathematical Programming 99 (2004) 351–378.

[4] R. Apparigliato, Règles de décision pour la gestion du risque: Application à la gestion hebdomadaire de la production électrique, Ph.D. thesis, École Polytechnique (2008).

[5] L. Pang, M. Housh, P. Liu, X. Chen, Robust stochastic optimization for reservoir operation, Water Resources Research 51.

[6] C. Gauvin, E. Delage, M. Gendreau, A robust optimization model for the risk averse reservoir management problem, Tech. rep., GERAD (2015).

[7] X. Chen, M. Sim, P. Sun, J. Zhang, A linear decision-based approximation approach to stochastic programming, Operations Research 56 (2008) 344–357.

[8] J. Goh, M. Sim, Distributionally robust optimization and its tractable approximations, Operations Research 58 (2010) 902–917.

[9] A. Georghiou, W. Wiesemann, D. Kuhn, Generalized decision rule approximations for stochastic programming via liftings, Mathematical Programming (2014) 1–38.

[10] A. Turgeon, Solving a stochastic reservoir management problem with multilag autocorrelated inflows, Water Resources Research 41 (2005) W12414.

[11] A. Lorca, A. Sun, Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind, IEEE Transactions on Power Systems.

[12] A. Turgeon, Stochastic optimization of multireservoir operation: The optimal reservoir trajectory approach, Water Resources Research 43 (2007) W05420.

[13] C. Cervellera, V. Chen, A. Wen, Optimization of a large-scale water reservoir network by stochastic dynamic programming with efficient state space discretization, European Journal of Operational Research 171 (2006) 1139–1151.

[14] A. Tilmant, R. Kelman, A stochastic approach to analyze trade-offs and risk associated with large-scale water resources systems, Water Resources Research 43 (2007) W06425.

[15] J. R. Stedinger, B. A. Faber, Reservoir optimization using sampling sdp with ensemble streamflow prediction (esp) forecast, Journal of Hydrology 249 (2001) 113–133.

[16] J. A. Tejada-Guibert, S. A. Johnson, J. R. Stedinger, The value of hydrologic information in stochastic dynamic programming models of a multireservoir system, Water Resources Research 31 (10) (1995) 2571–2579.

[17] P. Côté, D. Haguma, R. Leconte, S. Krau, Stochastic optimisation of hydro-quebec hydropower installations: a statistical comparison between sdp and ssdp methods, Revue canadienne de génie civil.

[18] P. Billingsley, Probability and measure, third edition, John Wiley & Sons Inc, 1995.

[19] E. Delage, D. A. Iancu, Robust multi-stage decision making, Tutorials in Operations Research (2015) 20–46.

[20] A. Shapiro, D. Dentcheva, A. Ruszcyński, Lectures on stochastic programming: modeling and theory, MOS-SIAM Series on Optimization, 2009.

[21] P. Rocha, D. Kuhn, Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules, European Journal of Operational Research 216 (2) (2012) 397–408.

[22] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, Time series analysis: forecasting and control, 4th edition, John Wiley & Sons, Inc., Hoboken, New Jersey., 2008.

[23] P. J. Brockwell, R. A. Davis, Time series: theory and methods, Springer-Verlag New York, 1987.

[24] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust Optimization, Princeton University Press, 2009.

[25] Hydro-Québec, Hydro-québec, rapport annuel 2012 (2012).

[26] J. D. Salas, J. W. Delleur, V. Yevjevich, W. L. Lane, Applied modeling of hydrologic time series, Water Resources Publications, Littleton, Colorado, 1980.

[27] P. Roebber, The regime dependence of degree day forecast technique, skill, and value, Weather Forecasting 13 (1998) 783–794.

[28] J. Nash, J. Sutcliffe, River flow forecasting through conceptual models. part i. - a discussion of principles, Journal of Hydrology 10 (1970) 282–290.

[29] H. V. Gupta, H. Kling, K. K. Yilmaz, G. F. Martinez, Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling, Journal of Hydrology 377 (2009) 80–91.

[30] D. Bertsimas, G. J. Lauprete, A. Samarov, Shortfall as a risk measure: properties, optimization and applications, Journal of Economic Dynamics and Control 28 (2004) 1353–1381.

[31] W. Rudin, Real and Complex Analysis, 3rd Ed., McGraw-Hill, Inc., New York, NY, USA, 1987.

[32] M. Sim, D. Bertsimas, The price of robustness, Operations Research 52 (1) (2004) 35–53.

9 Appendix

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

$$\{y \in \mathbb{R}^L : \|y\|_2 \leq \sqrt{L\epsilon}\} \subset \{y \in \mathbb{R}^L : \|y\|_1 \leq L\sqrt{\epsilon}\} \cap \{y \in \mathbb{R}^L : \|y\|_\infty \leq \sqrt{L\epsilon}\} \tag{39}$$

Proof 1. *The Cauchy-Schwartz inequality yields (see (31)):*

$$\|y\|_1 \leq \|y\|_2 L^{1/2} \tag{40}$$

and we see that:

$$|y_i| \leq \left(\sum_{i=1}^{L} |y_i|^2\right)^{1/2}, \; i = 1, ..., L \Rightarrow \max_i |y_i| \leq \left(\sum_{i=1}^{L} |y_i|^2\right)^{1/2} \Leftrightarrow \|y\|_\infty \leq \|y\|_2 \tag{41}$$

If follows that if $\|y\|_2 \leq \sqrt{L\epsilon}$,

$$\|y\|_\infty \leq \|y\|_2 \leq \sqrt{L\epsilon} \tag{42}$$

$$\|y\|_1 \leq \|y\|_2 L^{1/2} \leq L\sqrt{\epsilon} \tag{43}$$

and (39) follows. We also note that the bounds are tight since the vector with $y_i = \epsilon, \forall i$ yields an equality in (43) and the vector with $y_i = \epsilon$ for a single i and 0 otherwise yields an equality in (42). □

Using this fact, we can prove the following theorem:

Theorem 2. *For a fixed $\epsilon > 0$ and $L \in \mathbb{N}$, given the square diagonal invertible matrix $\Sigma = diag(\sigma_1^2, ..., \sigma_L^2) \in \mathbb{R}^{L \times L}$ with $\sigma_i > 0, \forall i$ and $\Sigma^{-1/2} = diag(\sigma_1^{-1}, ..., \sigma_L^{-1})$, the following inclusion holds:*

$$\{x \in \mathbb{R}^L : x^\top \Sigma^{-1} x \leq L\epsilon\} \subset \{x \in \mathbb{R}^L : \sum_i \frac{|x_i|}{\sigma_i} \leq L\sqrt{\epsilon}\} \cap \{x \in \mathbb{R}^L : \frac{|x_i|}{\sigma_i} \leq \sqrt{L\epsilon}, \forall i\} \quad (44)$$

Proof 2. *We first observe that for any $x \in \mathbb{R}^L$, there exists a unique $y \in \mathbb{R}^L$ such that $y = \Sigma^{-1/2} x$ since Σ is invertible and $\Sigma^{-1/2}$ exists.*

We then show that the three sets in (44) can be written in terms of the 1,2 and ∞ norms and apply theorem (1).

1) The ellipsoid $\{x \in \mathbb{R}^L : x^\top \Sigma^{-1} x \leq L\epsilon\}$ can be written in terms of the $\|\cdot\|_2$ norm:

$$\{x \in \mathbb{R}^L : x^\top \Sigma^{-1} x \leq L\epsilon\} = \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2} x, \ \|y\|_2 \leq \sqrt{L\epsilon}\} \quad (45)$$

The equivalence holds since $\left\|\Sigma^{-1/2} x\right\|_2^2 = (\Sigma^{-1/2} x)^\top (\Sigma^{-1/2} x) = x^\top \Sigma^{-1} x$.

2) The set: $\{x \in \mathbb{R}^L : \frac{|x_i|}{\sigma_i} \leq \sqrt{\epsilon}, \forall i\}$ can be written in terms of the $\|\cdot\|_\infty$ norm:

$$\{x \in \mathbb{R}^L : |x_i| \sigma_i^{-1} \leq \sqrt{\epsilon}, \forall i\} = \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2} x, \ \|y\|_\infty \leq \sqrt{L\epsilon}\} \quad (46)$$

3) The set: $\{x \in \mathbb{R}^L : \sum_i |x_i| \sigma_i^{-1} \leq L\sqrt{\epsilon}\}$ can be written in terms of the $\|\cdot\|_1$ norm:

$$\{x \in \mathbb{R}^L : \sum_i |x_i| \sigma_i^{-1} \leq L\sqrt{\epsilon}\} = \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2} x, \ \|y\|_1 \leq L\sqrt{\epsilon}\} \quad (47)$$

By theorem (1), we have:

$$\{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2} x, \ \|y\|_2 \leq \sqrt{L\epsilon}\} \subset \quad (48)$$

$$\{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2} x, \ \|y\|_1 \leq L\sqrt{\epsilon}\} \cap \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2} x, \ \|y\|_\infty \leq \sqrt{L\epsilon}\} \quad (49)$$

which proves theorem (2) and justifies the use of our bounded polyhedral support. □

Observation 1. *We observe that (44) is slightly different from the standard inclusion used in robust optimization (see (24) section 2.3 and (32)). This is due to the fact that the authors in (24) 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 (32) also derive similar uncertainty sets, but with different probabilistic guarantees because they assume independent and bounded random variables with symmetric distribution.*

On the other hand, the bound we derive is based on Markov's/Chebyshev's conditional or unconditional inequality and applies for a larger class of random variables. We do not need to find the maximum realisation of the random variables explicitly and we can weaken the requirements of independence to uncorrelation, which is essential to obtain the probabilistic bounds with the GARCH effects.