Les Cahiers du GERAD ISSN: 0711–2440

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

C. Gauvin, E. Delage,
M. Gendreau

G–2016–24

April 2016
Revised: January 2017

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

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

Charles Gauvin [a]
Erick Delage [b]
Michel Gendreau [c]

[a] *GERAD & Polytechnique Montréal, Montréal (Québec) Canada, H3C 3A7*

[b] *GERAD & HEC Montréal, Montréal (Québec) Canada, H3T 2A7*

[c] *CIRRELT & Polytechnique Montréal, Montréal (Québec) Canada, H3C 3A7*

charles.gauvin@polymtl.ca
erick.delage@hec.ca
michel.gendreau@cirrelt.net

April 2016
Revised: January 2017

Les Cahiers du GERAD

G–2016–24

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.

Key Words: Stochastic programming, stochastic processes, robust optimization, forecasting, OR in energy, risk analysis

Acknowledgments: The authors would like to thank 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. 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.

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 [42, 41, 12, 22].

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 [31]. 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 [22]. Considering uncertainty significantly increases theses difficulties since the sequential decision-making under uncertainty represents a huge theoretical obstacle in itself [18].

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 [4], 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 [2] and [37] use this framework to maximize the expected electric production for a multi-period and multi-reservoir hydro-electric complex while [22] 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 [15, 25, 23], 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 [56, 41] 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 [34] 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 ([45]). 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 [31, 58, 13, 56] and references therein). Although these methods can deal with more realistic nonconvex 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 [57, 14, 55].

Considering heteroscedasticity further increases the state dimension and the resulting computational burden for SDP. As such, we are only aware of the work of [41] 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 [12], sampling SDP [53] and more elaborate discretization schemes [59, 14] 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 [56, 53, 54, 16], 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 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.

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 [55, 51, 46, 24, 38, 40]. Moreover, [35] 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 [51]. To avoid this issue, various studies only consider a limited number of discrete scenarios (<100) [55, 46], but this only leads to approximate solutions of questionable quality [49].

We also mention stochastic programming methods based on decision trees, which are also theoretically capable of explicitly handling highly persistent inflows [10, 26, 19]. 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. [22] 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 [37], 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 2. Section 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 4 explains the solution procedure and simulation framework while Section 5 studies a river in western Québec. Conclusions are drawn in Section 6.

1.1 Notation

The sets $\mathbb{Z}_+ = \{0, 1, 2, ...\}$ and $\mathbb{Z}_- = \{0, -1, -2, ...\}$ 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 $(\Omega, \mathcal{G}, \{\mathcal{G}_t\}, \mathbb{P})$ be a filtered probability space where $\{\mathcal{G}_t\}$ is a collection of σ-algebras representing some information available at time $t \in \mathbb{T}$ where $\mathcal{G}_0 = \{\Omega, \emptyset\}$ and $\mathcal{G}_T = \mathcal{G}$. We let $\mathrm{E}[\cdot]$ denote mathematical expectation while $\mathrm{E}[\cdot|\mathcal{A}]$ represents conditional expectation given any σ-algebra $\mathcal{A} \subseteq \mathcal{G}$. Both expectations are taken with respect to $\mathbb{P}$, the base probability measure on $\mathcal{G}$.

For the real random variables X, $\sigma(X)$ represents the σ-algebra generated by X [6]. 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 $(X_t, ..., X_{t+L-1})^\top \equiv X_{[t,t+L-1]}$ for any $t \in \mathbb{Z}$ and $L \in \mathbb{N}$ with the special notation $X_{[t+L-1]}$ if $t = 1$. For simplicity, we abuse notation and do not distinguish a random variable from a given realisation.

2 The stochastic reservoir management problem

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 [22], 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 4.

At the beginning of time $t \in \mathbb{T}$, we seek a vector of decisions for each future time $\tau \in \{t, ..., t+L-1\}$ 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 (1b)–(1j):

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

$$(\text{Stor. Bounds}) \qquad \underline{s}_j \leq \mathcal{S}_{j,t+l} - \mathcal{E}_{j,t+l} \leq \bar{s}_j \qquad \forall j \in J, l \in \mathbb{L} \tag{1b}$$

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

$$(\text{Flow Bounds}) \qquad \underline{f}_i \leq \mathcal{F}_{i,t+l} \leq \bar{f}_i \qquad \forall i \in I, l \in \mathbb{L} \tag{1d}$$

$$(\text{Evac. Curve}) \qquad \mathcal{L}_{i,t+l} \leq \mathscr{C}_i(\mathcal{S}_{j^-(i),t+l}) \qquad \forall i \in I^{evac}, l \in \mathbb{L} \tag{1e}$$

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

$$(\text{Spill. Bounds}) \qquad \underline{l}_i \leq \mathcal{L}_{i,t+l} \leq \bar{l}_i \qquad \forall i \in I, l \in \mathbb{L} \tag{1g}$$

$$(\text{Rel. Bounds}) \qquad \underline{r}_i \leq \mathcal{R}_{i,t+l} \leq \bar{r}_i \qquad \forall i \in I, l \in \mathbb{L} \tag{1h}$$

$$(\text{Flow Def.}) \qquad \mathcal{F}_{i,t+l} = \mathcal{R}_{i,t+l} + \mathcal{L}_{i,t+l} \qquad \forall i \in I, l \in \mathbb{L} \tag{1i}$$

$$(\text{Floods}) \qquad 0 \leq \mathcal{E}_{j,t+l} \qquad \forall j \in J, l \in \mathbb{L}. \tag{1j}$$

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 (1b) ensure that the total storage remains within tolerable limits $\underline{s}_j$ and $\bar{s}_j$ for all reservoirs $j \in J$. (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.

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 (1d) ensure that flows are within limits $\underline{f}_{it}, \bar{f}_{it}$ while constraints (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 (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 [22], 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 (1g) ensure respect of absolute upper and lower bounds. These constraints are determined by specific physical characteristic of given plants. Constraints (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, (1i) defines the total flow as the sum of unproductive spillage and releases.

Constraints (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 (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 [41] 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 shows it leads to good empirical results.

The parameters $\kappa_{j,t} > 0$ in (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.

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

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 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 $\mathrm{E}\left[(\xi_t - \mu_t)^2\right] = \sigma_t^2$.

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 [17, 50]. 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 1.

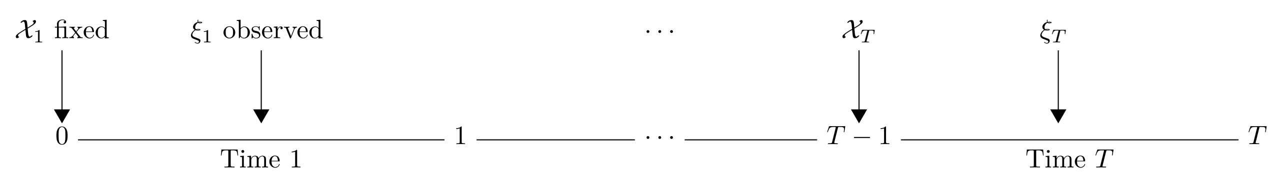

Figure 1: Sequential Dynamic Decision Process

2.4 Affine decision rules

To consider this uncertainty and formulate the stochastic version of model (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 [4] and have gained considerable attention in the recent years, namely in the field of energy [22, 2, 34, 44]. 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 $K_\tau = \{n_{\tau-1} + 1, ..., n_{\tau-1} + n_\tau\}$ 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 future inflow vector $\xi \equiv \xi_{[t,t+L-1]}$ in the form :

$$\mathcal{X}_{k,t+l}(\xi) = \mathcal{X}^0_{k,t+l} + \sum_{l'=0}^{L-1} \mathcal{X}^{l'}_{k,t+l}\xi_{t+l'}, \tag{2}$$

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 (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 (1b) - (1j) at time t requires considering $\min\{\delta^{max}_{i^-}, t-1\}$ 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 $\sum_{j\in J}\sum_{i^-\in I^-(j)}(\delta^{max}_{i^-} - \delta^{min}_{i^-}) = 13$ 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 (1c) and (1f) 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 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.

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 (1a) to :

$$(SP_t) \qquad \min_{\mathcal{X}(\cdot)} \quad \mathrm{E}\left[\sum_{j=1}^{J}\sum_{l=0}^{L-1}\kappa_{jt+l}\mathcal{K}_{jt+l}(\mathcal{E}_{j,t+l}(\xi))|\mathcal{G}_{t-1}\right]. \tag{3}$$

The conditional expectation offers the advantage of being the simplest consistent and coherent dynamic risk measure [43]. 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 (1) at time t as SP_t.

3 Exploiting time series models and linear decision rules

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 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 bounded 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 [34], holds:

$$\Xi_t = \left\{ \xi \in \mathbb{R}^L \;\middle|\; \begin{array}{ll} \exists \zeta, \varrho \in \mathbb{R}^L & \text{(4a)} \\ \xi = U_t \zeta + u_t & \text{(4b)} \\ \zeta = V_t \varrho + v_t & \text{(4c)} \\ W_t \varrho \leq w_t & \text{(4d)} \end{array} \right\}$$

We assume that for any ξ, there exists unique ϱ, ζ such that the representation (4) 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 (4c) into (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, (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 [9, 8]. 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 (4) 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

This section assumes some basic familiarity with linear AR(I)MA time series. For further details, we refer to the classic texts [8, 9]. 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:

$$\phi(B)\zeta_t = \theta(B)\varrho_t, \tag{5}$$

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 $t, p \in \mathbb{Z}$ [8, 9]. We suppose the ϱ_t are independent identically distributed zero mean and $\mathcal{G}_t$-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. For any ARMA process respecting Equation (5), we can equivalently write:

$$\zeta_t = \psi(B)\varrho_t, \tag{6}$$

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 $\{\varrho_\tau\}_{\tau=t,t-1,\ldots}$. 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 (6) is essentially unique [9]. 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 1 *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 (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:

$$\zeta_{t+l} = \underbrace{\sum_{j=l}^{\infty} \psi_j \varrho_{t+l-j}}_{\hat{\zeta}_t(l)} + \underbrace{\sum_{j=0}^{l-1} \psi_j \varrho_{t+l-j}}_{\rho_t(l)}, \tag{7}$$

where $\hat{\zeta}_t(l) = \mathrm{E}[\zeta_{t+l}|\mathcal{G}_t]$ 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 $\mathrm{E}[\varrho_{t+l}|\mathcal{G}_t] = \varrho_{t+l}$ if $l \in \mathbb{Z}_-$ and $\mathrm{E}[\varrho_{t+l}|\mathcal{G}_t] = \mathrm{E}[\varrho_{t+l}] = 0$ otherwise. For any $t \in \mathbb{Z}, l \in \mathbb{Z}_+$, we observe that $\rho_t(l)$ is $\mathcal{G}_{t+l}$ measurable while $\hat{\zeta}_t(l)$ is $\mathcal{G}_t$ 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 $\mathrm{E}[\zeta_{t+l}|\mathcal{G}_t]$ 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 $l \in \mathbb{Z}_+$ [9].

Example 2 *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* $\rho_t(l) = \sum_{i=0}^{l-1} \phi^i \varrho_{t+l-i}$. □

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 $L \in \{1, \cdots, T-t+1\}$:

$$\rho_{t-1,L} = V_t \varrho_{[t,t+L-1]}, \tag{8}$$

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 $t \in \mathbb{Z}$:

$$V = \begin{pmatrix} 1 & & \cdots & 0 \\ \psi_1 & 1 & & \vdots \\ \vdots & & \ddots & \\ \psi_{L-1} & \cdots & \psi_1 & 1 \end{pmatrix} \tag{9}$$

We then have the equality:

$$\begin{aligned} \zeta_{[t,t+L-1]} &= \hat{\zeta}_{t-1,L} + \rho_{t-1,L} && (10) \\ &= \hat{\zeta}_{t-1,L} + V \varrho_{[t,t+L-1]}, && (11) \end{aligned}$$

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 (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 $\hat{\zeta}_{t-1,L}$:

$$\xi_{[t,t+L-1]} = U_t \zeta_{[t,t+L-1]} + u_t \tag{12}$$
$$= U_t(\hat{\zeta}_{t-1,L} + \rho_{t-1,L}) + u_t \tag{13}$$
$$= U_t(\hat{\zeta}_{t-1,L} + V\varrho_{[t,t+L-1]}) + u_t. \tag{14}$$

Example 3 *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:*

$$\begin{pmatrix} \xi_t \\ \xi_{t+1} \\ \vdots \\ \xi_{t+L-1} \end{pmatrix} = \begin{pmatrix} \phi\zeta_{t-1} \\ \phi^2\zeta_{t-1} \\ \vdots \\ \phi^L\zeta_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & & \cdots & 0 \\ \phi^1 & 1 & & \vdots \\ \vdots & & \ddots & \\ \phi^{L-1} & \dots & \phi^1 & 1 \end{pmatrix} \begin{pmatrix} \varrho_t \\ \varrho_{t+1} \\ \vdots \\ \varrho_{t+L-1} \end{pmatrix} + \begin{pmatrix} \mu_t \\ \mu_{t+1} \\ \vdots \\ \mu_{t+L-1} \end{pmatrix} \tag{15}$$

□

The affine representation (14) reveals that the $\xi_{[t,t+L-1]}$ vector is completely determined by $\varrho_{[t,t+L-1]}$. We therefore set $\sigma(\varrho_s; s \leq t) = \mathcal{G}_t, \forall t \in \mathbb{T}$ 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 $1, \cdots, t-1$.

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 $\varrho_{[t,t+l-1]}$, which also corresponds to the set $\{\varrho \in \mathbb{R}^L : W_t\varrho \leq w_t\}$ described in (4), is a bounded polyhedron in $\mathbb{R}^L$ given by the intersection of the following two polyhedrons:[2]

$$\mathscr{B}^{\infty}_{L,\epsilon} = \{\varrho \in \mathbb{R}^L : |\varrho_i|\sigma_\varrho^{-1} \leq (L\epsilon)^{1/2}, i = 1, ..., L\} \tag{16a}$$

$$\mathscr{B}^{1}_{L,\epsilon} = \{\varrho \in \mathbb{R}^L : \sum_{i=1}^{L} |\varrho_i|\sigma_\varrho^{-1} \leq L\epsilon^{1/2}\}. \tag{16b}$$

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:

$$\mathbb{P}(\tilde{\varrho}^\top \Sigma_{\varrho,L}^{-1}\tilde{\varrho} > L\epsilon) \leq \mathrm{E}\left[\tilde{\varrho}^\top \Sigma_{\varrho,L}^{-1}\tilde{\varrho}\right](L\epsilon)^{-1} \tag{17}$$
$$= tr(\Sigma_{\varrho,L}^{-1}\mathrm{E}\left[\tilde{\varrho}\tilde{\varrho}^\top\right])(L\epsilon)^{-1} \tag{18}$$
$$= \epsilon^{-1}. \tag{19}$$

The polytope $\mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}$ contains the ellipsoid $\mathscr{B}^{2}_{L,\epsilon} = \{\varrho \in \mathbb{R}^L : \varrho^\top \Sigma_{\varrho,L}^{-1}\varrho \leq L\epsilon\}$ (see (A) 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 $t, L \in \mathbb{N}$. Although we could use $\mathscr{B}^{2}_{L,\epsilon}$ as the support, preliminary tests demonstrate that it is preferable to consider the exterior polyhedral approximation (16) to speed computations.

[2] The set $\mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}$ is not strictly speaking a polyhedron since (16a)–(16b) involve the non-linear absolute value function. Nonetheless, lifting this set using the commonly used decomposition $\varrho_i = \varrho_i^+ - \varrho_i^-$ 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 $\mathscr{B}^{\infty}_{L,\epsilon} \cap \mathscr{B}^{1}_{L,\epsilon}$ ([3]).

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 [3].

The polyhedron defined by (16a)–(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.

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

Building on these assumptions, it follows that the polyhedron given by (4) represents the conditional support of $\xi_{[t,t+L-1]}$ for $L \in \{1, ..., T-t+1\}$ 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.

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 [21, 32]. As we discuss in Section 5.4, this is also empirically observed in daily inflows. For further details on GARCH processes, see [8, 7].

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

$$\hat{\sigma}^2_{t-1}(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho^2_{t-i} + \sum_{j=1}^{s} \beta_j \hat{\sigma}^2_{t-1-j}(1) \; , \tag{20}$$

where $\hat{\sigma}^2_t(l) = \mathrm{E}\left[\varrho^2_{t+l}|\mathcal{G}_t\right]$[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 (D), we show that under standard stationarity assumptions on the ϱ_t, the squared shocks ϱ^2_t satisfy the difference equation:

$$\hat{\phi}(B)\left(\varrho^2_t - \sigma^2_a\right) = \hat{\theta}(B)\nu_t. \tag{21}$$

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:

$$\varrho^2_t = \sigma^2_{\varrho} + \hat{\psi}(B)\nu_t. \tag{22}$$

[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 $\mathcal{G}_t$, which we denoted $\hat{\zeta}_t(l)$.

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 (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 3.2. For $l \in \mathbb{N}$, we specifically have:

$$\hat{\sigma}_t^2(l) = \sigma_\varrho^2 + \sum_{j=l}^{\infty} \hat{\psi}_j \nu_{t+l-j}. \tag{23}$$

We can then write the conditional covariance matrix as $\Sigma_{\varrho,L,t-1} = diag(\hat{\sigma}_{t-1}^2(1), ..., \hat{\sigma}_{t-1}^2(L))$, where each conditional variance $\hat{\sigma}_t^2(l)$ 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 (D)). In Section 3.7, we show how to obtain a better estimation of the expected flood penalties by exploiting this information.

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.

3.6 Additional modelling considerations

The representation (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 (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 $t \in \mathbb{T}$:

$$\left\{\varrho \in \mathbb{R}^L \ : U_t V \varrho \geq -(U_t \hat{\zeta}_{t-1,L} + u_t)\right\} \tag{24a}$$

The additional structure imposed by Equation (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 (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.

3.7 Optimizing the conditional expected flood penalties with affine decision rules

We suppose that the inflow model (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 $L \in \{1, ..., T-t+1\}$ days and where $\xi \equiv \xi_{[t,t+L-1]}$ in a more condensed and abstract form:

$$\min_{\mathcal{X}} \mathrm{E}\left[\sum_{l=0}^{L-1} \mathcal{X}_{t+l}^\top(\xi) G_{t+l} \mathcal{X}_{t+l}(\xi) | \mathcal{G}_{t-1}\right] \tag{25a}$$

$$\text{s.t.} \quad \sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}_{t+\bar{l}}(\xi) \geq C_{t+l}^{\Delta} \xi + C_{t+l}^0 \qquad \forall l \in \mathbb{L} \qquad \mathbb{P} \text{ a.s.} \tag{25b}$$

$$\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \mathcal{X}_{t+\bar{l}}(\xi) = \hat{E}_{t+l}^{\Delta} \xi + \hat{E}_{t+l}^0 \qquad \forall l \in \mathbb{L} \qquad \mathbb{P} \text{ a.s.}, \tag{25c}$$

for some $A_{t+l,t+\bar{l}} \in \mathbb{R}^{m^{\geq}_{t+l} \times n_{t+\bar{l}}}$, $C^{\Delta}_{t+l} \in \mathbb{R}^{m^{\geq}_{t+l} \times L}, C^{0}_{t+l} \in \mathbb{R}^{m^{\geq}_{t+l}}, D^{\Delta}_{t+l} \in \mathbb{R}^{m^{=}_{t+l} \times L}, E^{0}_{t+l} \in \mathbb{R}^{m^{=}_{t+l}}, \mathcal{X}_{t+l}(\xi) \in \mathbb{R}^{n_{t+l}}$, $G_{t+l} \in \mathbb{S}^{n_{t+l}}_{+}$ for $l \in \mathbb{L}$, for $m^{\geq}_t, m^{=}_t, n_t \in \mathbb{N}$.

We now limit ourselves to affine decision rules and rewrite (2) in the more compact form:

$$\mathcal{X}_t(\xi) = \mathcal{X}^0_t + \mathcal{X}^{\Delta}_t \xi, \tag{26}$$

for $\xi \equiv \xi_{[t,t+L-1]} \in \mathbb{R}^L, \mathcal{X}^0_t \in \mathbb{R}^{n_t}$ and $\mathcal{X}^{\Delta}_t \in \mathbb{R}^{n_t \times L}$ whose structure depends on the non-anticipativity of the respective decisions. Assuming that constraint (24a) does not significantly affect the hypothesis that $\mathrm{E}\left[\varrho_{t+l}|\mathcal{G}_{t-1}\right] = 0, \forall l \in \mathbb{Z}_+$ and setting $\varrho \equiv \varrho_{[t,t+L-1]}$, the objective value (25a) then becomes:

$$\begin{aligned}
\mathrm{E}\left[\sum_{l=0}^{L-1} \mathcal{X}^{\top}_{t+l}(\xi) G_{t+l} \mathcal{X}_{t+l}(\xi) |\mathcal{G}_{t-1}\right] &= \sum_{l=0}^{L-1} \mathrm{E}\left[\hat{\mathcal{X}}^{0,\top}_{t+l} G_{t+l} \hat{\mathcal{X}}^{0}_{t+l} |\mathcal{G}_{t-1}\right] + \sum_{l=0}^{L-1} \mathrm{E}\left[\varrho^{\top} \hat{\mathcal{X}}^{\Delta,\top}_{t+l} G_{t+l} \hat{\mathcal{X}}^{\Delta}_{t+l} \varrho |\mathcal{G}_{t-1}\right] && (27)\\
&= \sum_{l=0}^{L-1} \hat{\mathcal{X}}^{0,\top}_{t+l} G_{t+l} \hat{\mathcal{X}}^{0}_{t+l} + \sum_{l=0}^{L-1} tr(\mathrm{E}\left[\varrho\varrho^{\top}|\mathcal{G}_{t-1}\right] \hat{\mathcal{X}}^{\Delta,\top}_{t+l} G_{t+l} \hat{\mathcal{X}}^{\Delta}_{t+l}) && (28)\\
&= \sum_{l=0}^{L-1} \hat{\mathcal{X}}^{0,\top}_{t+l} G_{t+l} \hat{\mathcal{X}}^{0}_{t+l} + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}^{\Delta}_{t+l} \Sigma^{1/2}_{\varrho,L,t-1})^{\top} G_{t+l} (\hat{\mathcal{X}}^{\Delta}_{t+l} \Sigma^{1/2}_{\varrho,L,t-1}), && (29)
\end{aligned}$$

where:

$$\hat{\mathcal{X}}^{0}_{t+l} = \mathcal{X}^{0}_{t+l} + \mathcal{X}^{\Delta}_{t+l} u_t + \mathcal{X}^{\Delta}_{t+l} U_t \hat{\zeta}_{t-1,L} \tag{30}$$

$$\hat{\mathcal{X}}^{\Delta}_{t+l} = \mathcal{X}^{\Delta}_{t+l} U_t V \tag{31}$$

$$\Sigma^{1/2}_{\varrho,L,t-1} = diag(\hat{\sigma}_{t-1}(1), ..., \hat{\sigma}_{t-1}(L)), \tag{32}$$

and we have used the following decomposition from Section 3.2:

$$\xi_{[t,t+L-1]} = U_t(\hat{\zeta}_{t-1,L} + V \varrho_{[t,t+L-1]}) + u_t. \tag{33}$$

In the case of homoscedasticity, (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 (4) , we see that for any $f : \mathbb{R}^L \to \mathbb{R}$ and $k \in \mathbb{R}$, $f(\xi) \geq k, \forall \xi \in \Xi_t \Rightarrow \mathbb{P}(f(\xi_{[t,t+L-1]}) \geq k|\mathcal{G}_{t-1}) = 1$ with $\mathbb{P}$ a.s.. With (30)–(32), problem (25) can therefore be written as:

$$\min_{\mathcal{X}^0, \mathcal{X}^{\Delta}} \sum_{l=0}^{L-1} \hat{\mathcal{X}}^{0,\top}_{t+l} G_{t+l} \hat{\mathcal{X}}^{0}_{t+l} + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}^{\Delta}_{t+l} \Sigma^{1/2}_{\varrho,L,t-1})^{\top} G_{t+l} (\hat{\mathcal{X}}^{\Delta}_{t+l} \Sigma^{1/2}_{\varrho,L,t-1}) \tag{34a}$$

$$\text{s.t.} \quad (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}^{\Delta}_{t+\bar{l}} - C^{\Delta}_{t+l})\xi \geq -\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \mathcal{X}^{0}_{t+\bar{l}} + C^{0}_{t+l} \qquad \forall l \in \mathbb{L}, \qquad \forall \xi \in \Xi_t \tag{34b}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \mathcal{X}^{\Delta}_{t+\bar{l}} - \hat{E}^{\Delta}_{t+l})\xi = -\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \mathcal{X}^{0}_{t+\bar{l}} + \hat{E}^{0}_{t+l} \qquad \forall l \in \mathbb{L}, \qquad \forall \xi \in \Xi_t. \tag{34c}$$

Its optimal solution represents an upper bound on (25a)–(25c) with arbitrary decision rules because we limit ourselves to affine functions. Since Ξ_t is a polyhedron, we can handle the constraints (34b) through robust optimization techniques and linear programming duality [3]. We also reformulate (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 (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 (B) and (C).

From this derivation, we also see that the optimal value of problem (34a)–(34b) 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 (16a)–(16b), constraint (24a) does not affect the hypothesis of $\mathrm{E}\left[\varrho_{t+l}|\mathcal{G}_{t-1}\right] = 0, \forall l \in \mathbb{Z}_+$ and the structure of the ARMA and GARCH models are correct.

4 Monte Carlo simulation and rolling horizon framework

Solving the stochastic version of problem (1a)–(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 $t, ..., t + L - 1$ 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 5.3 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 $\varrho^s_{[T]} \equiv (\varrho_{s,1}, ..., \varrho_{s,T})^\top$ which together with the fixed and deterministic initial inflows ξ_0[5] completely determine inflows $\xi^s_{[T]} \equiv (\xi_{s,1}, ..., \xi_{s,T})^\top$. 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 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 $t, ..., t + L - 1$ 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 $t = 1, ..., T - L + 1$ 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 [30]. 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).

[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.

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

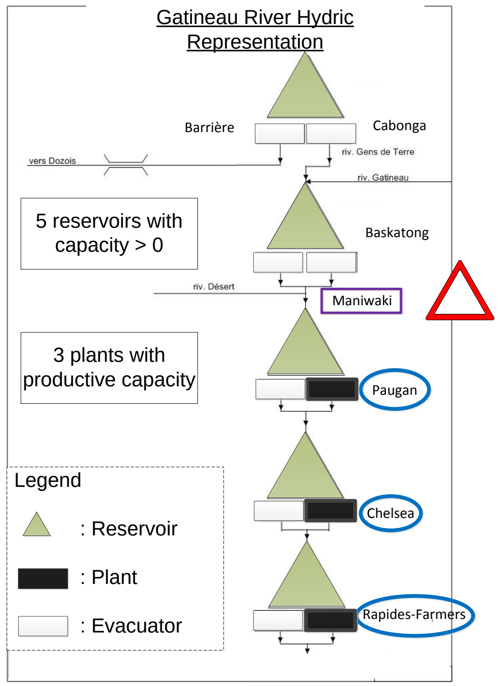

Figure 2: Simplified representation of the Gatineau river system

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 storages. If the head reservoirs are not sufficiently emptied before the freshet, there is a significant risk of disrupting normal operating conditions and flooding [22].

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 [11, 20].

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 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 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.

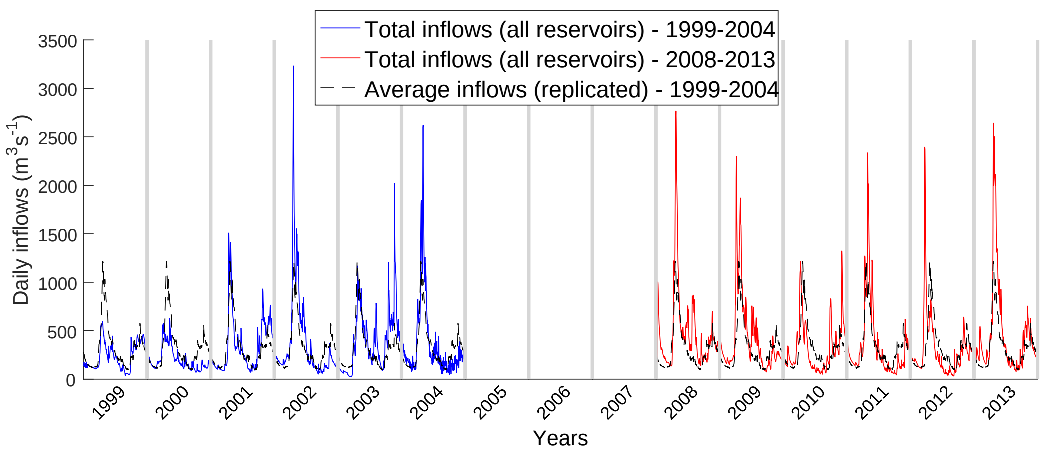

Figure 3: Sample inflows for 1999-2004 and 2008-2013 (12 years) with "in-sample" mean [6]

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 [24, 35] 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, $u_t = \mathrm{E}\left[\xi_{[t,t+L-1]}\right] = (\mu_t, ..., \mu_{t+L-1})^\top \in \mathbb{R}^L$ and $U_t^{-1} = diag(\sigma_t^{-1}, ..., \sigma_{t+L-1}^{-1}) \in \mathbb{R}^{L \times L}$ 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 $s \in \mathbb{N}$ [48, 8]. These are all compatible with our framework at no additional complexity.

We consider Box-Jenkins methodology [8] and find that the 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 provided worst AIC criterion [1]. 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 [8].

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 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.

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

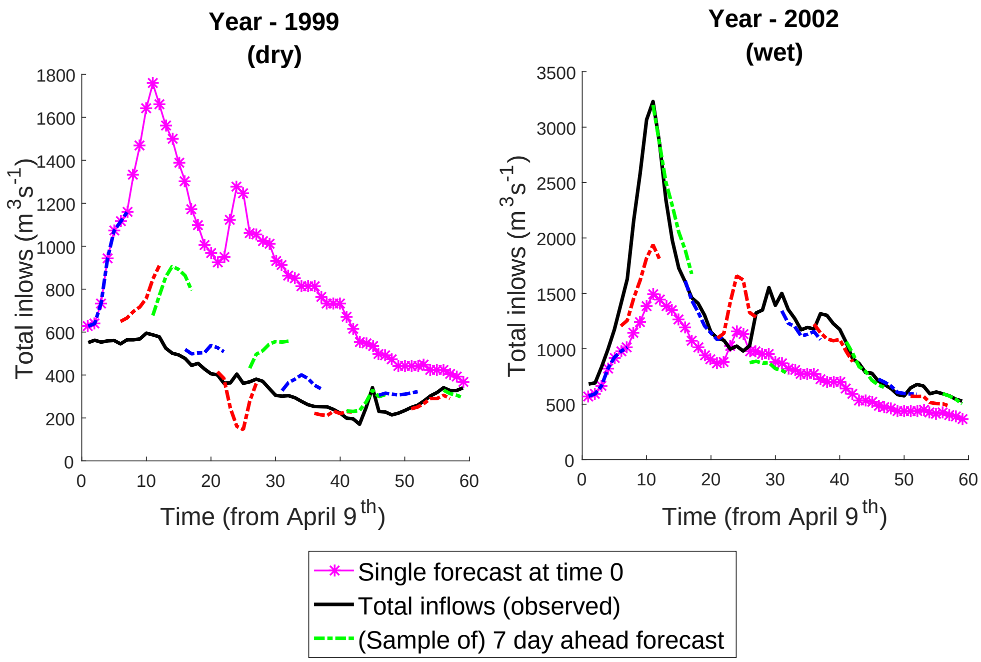

Figure 4: Comparing simple forecasts for 1999 & 2002

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). However, at the 5% level of significance, the Ljung-Box test on the squared residuals reveals the presence of heteroscedasticity [33]. Visual inspection corroborates this conclusion as there are clear signs of volatility clustering.

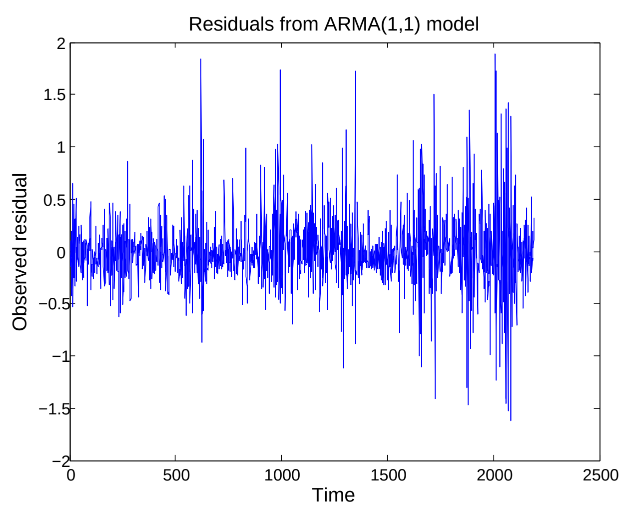

Figure 5: Residuals from ARMA(1,1) model

We find that the residual $\{\varrho_t\}$ approximately follow a GARCH(1, 1) model with the following estimates: $\alpha_0 = 0.01, \alpha_1 = 0.14, \beta_1 = 0.84$. We can show that the coefficients in (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 [32, 28].

5.5 Comparing forecasts

Sections (3.3)–(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 [36, 27]. 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.

5.6 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 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.

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 $\mathrm{CVaR}_\alpha = \mathrm{E}[X|X > \mathrm{VaR}_\alpha(X)]$ and $\mathrm{VaR}_\alpha(X) = q_X(\alpha) = \inf\{t : \mathbb{P}(X \leq t) \geq \alpha\}$ is the α quantile for some $\alpha \in (0, 1)$ [39].

As in [22], 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 [5], 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 5.6 reveals that the bound can be relatively tight in some cases and therefore that our models perform well under some scenarios.

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_ϱ^2. Hence these random variables violate our assumption of boundedness made in Section 3.3. Choosing random variables with support defined exactly by (16b)–(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 (E).

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.

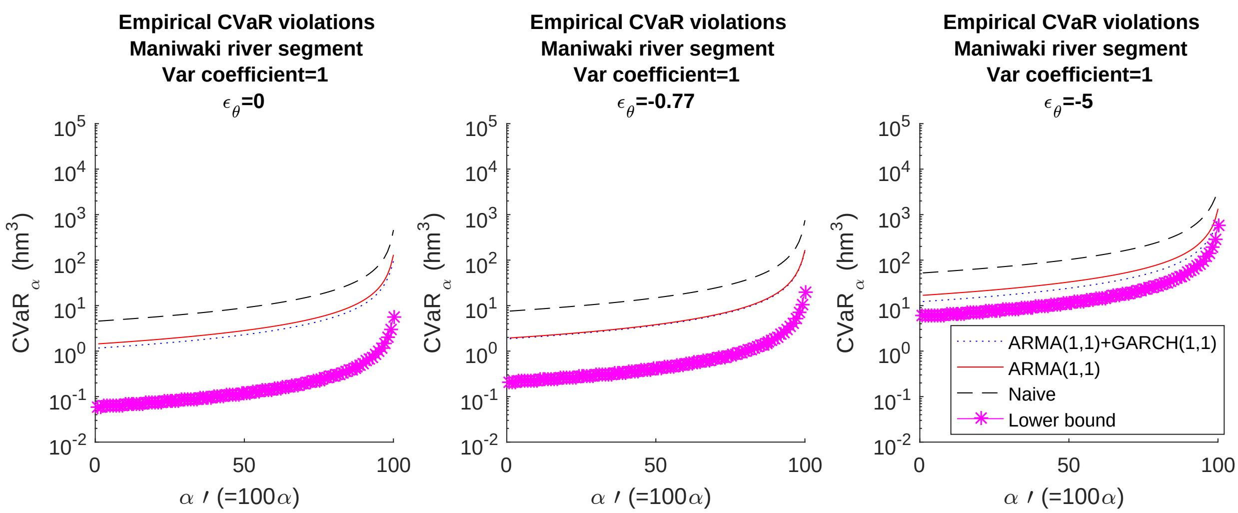

Figure 6: Influence of reduced forecast skill

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 $\{\zeta_t\}$.

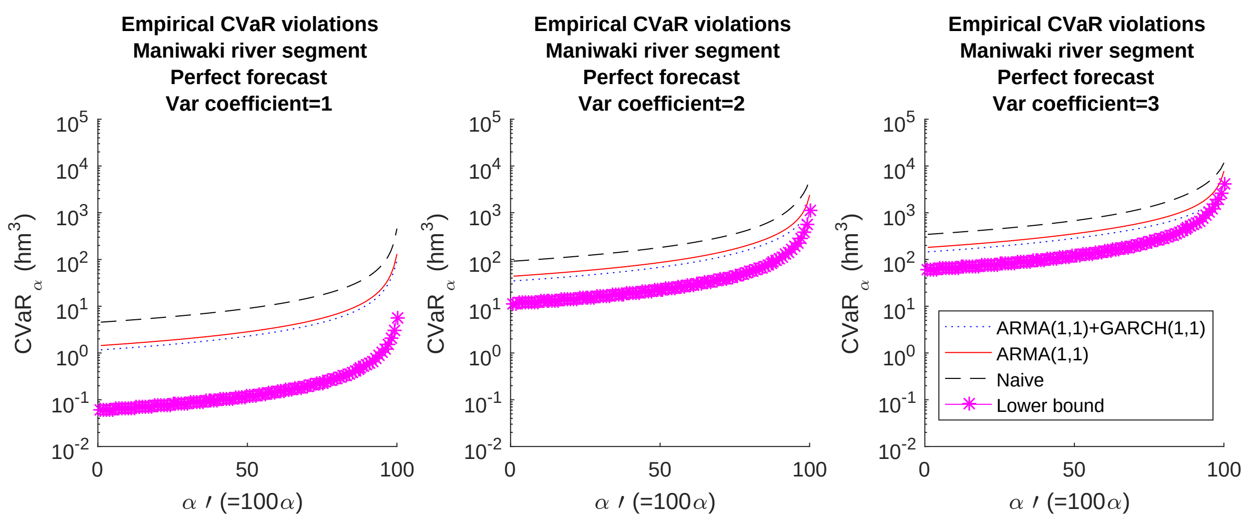

Figure 7: Influence of increased unconditional variance

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.

5.8 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 [8].

Figure 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.

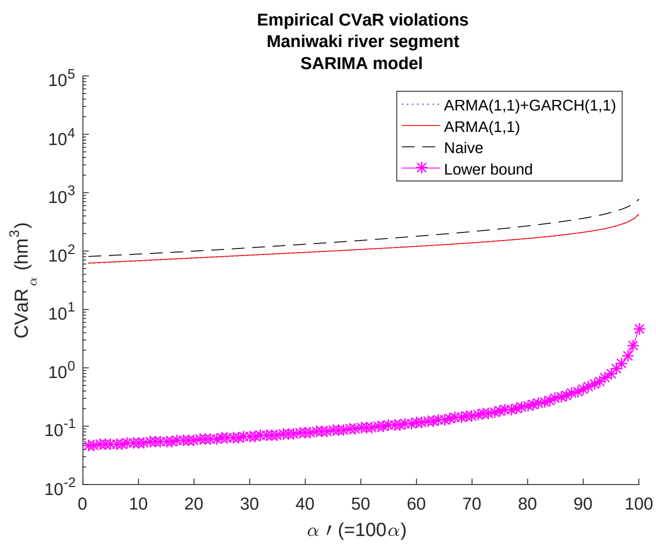

Figure 8: Influence of different time series structure

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 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.

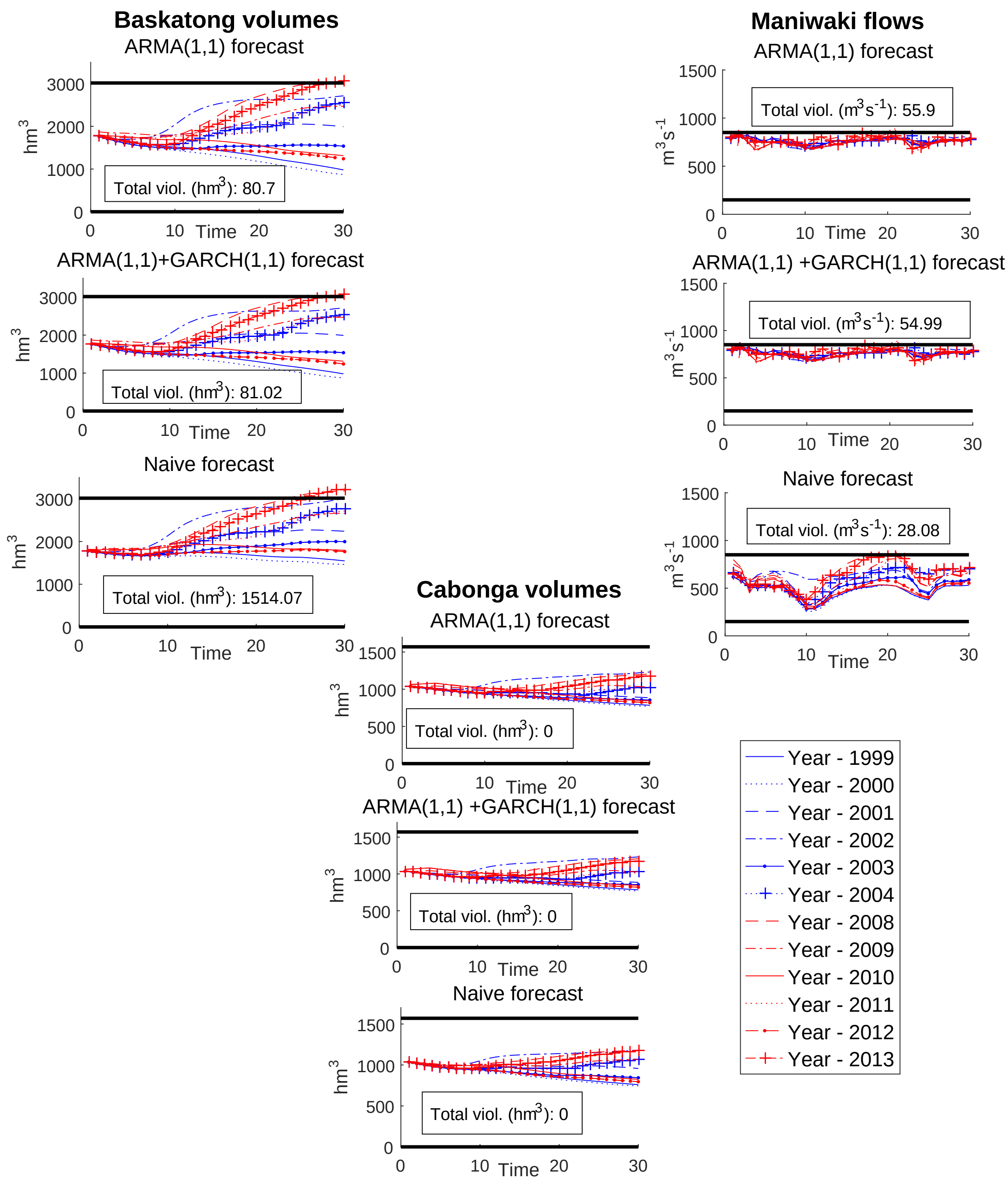

Figure 9: Simulation results for 1999-2004 & 2008-2013 (12 years)

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 [29]. 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 [56].

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 Section 5.7, (5.8) and (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 [22].

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.

Appendix A Joint probabilistic guarantees for the polyhedral support

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

$$\{y \in \mathbb{R}^L : \|y\|_2 \leq \sqrt{L\epsilon}\} \subset \{y \in \mathbb{R}^L : \|y\|_1 \leq L\sqrt{\epsilon}\} \cap \{y \in \mathbb{R}^L : \|y\|_\infty \leq \sqrt{L\epsilon}\} \tag{35}$$

Proof. The Cauchy-Schwartz inequality yields ([47]):

$$\|y\|_1 \leq \|y\|_2 \, L^{1/2} \tag{36}$$

and we see that:

$$|y_i| \leq \left(\sum_{i=1}^{L} |y_i|^2\right)^{1/2}, \; i = 1, ..., L \Rightarrow \max_i |y_i| \leq \left(\sum_{i=1}^{L} |y_i|^2\right)^{1/2} \Leftrightarrow \|y\|_\infty \leq \|y\|_2 \tag{37}$$

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

$$\|y\|_\infty \leq \|y\|_2 \leq \sqrt{L\epsilon} \tag{38}$$

$$\|y\|_1 \leq \|y\|_2 L^{1/2} \leq L\sqrt{\epsilon} \tag{39}$$

and (35) follows. We also note that the bounds are tight since the vector with $y_i = \epsilon, \forall i$ yields an equality in (39) and the vector with $y_i = \epsilon$ for a single i and 0 otherwise yields an equality in (38). □

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:*

$$\{x \in \mathbb{R}^L : x^\top \Sigma^{-1} x \leq L\epsilon\} \subset \{x \in \mathbb{R}^L : \sum_i \frac{|x_i|}{\sigma_i} \leq L\sqrt{\epsilon}\} \cap \{x \in \mathbb{R}^L : \frac{|x_i|}{\sigma_i} \leq \sqrt{L\epsilon}, \forall i\} \tag{40}$$

Proof. 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 (40) 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:

$$\{x \in \mathbb{R}^L : x^\top \Sigma^{-1} x \leq L\epsilon\} = \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2}x,\ \|y\|_2 \leq \sqrt{L\epsilon}\} \tag{41}$$

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:

$$\{x \in \mathbb{R}^L : |x_i|\sigma_i^{-1} \leq \sqrt{\epsilon}, \forall i\} = \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2}x,\ \|y\|_\infty \leq \sqrt{L\epsilon}\} \tag{42}$$

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:

$$\{x \in \mathbb{R}^L : \sum_i |x_i|\sigma_i^{-1} \leq L\sqrt{\epsilon}\} = \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2}x,\ \|y\|_1 \leq L\sqrt{\epsilon}\} \tag{43}$$

By Theorem 1, we have:

$$\{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2}x,\ \|y\|_2 \leq \sqrt{L\epsilon}\} \subset \tag{44}$$

$$\{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2}x,\ \|y\|_1 \leq L\sqrt{\epsilon}\} \cap \{x \in \mathbb{R}^L : \exists y = \Sigma^{-1/2}x,\ \|y\|_\infty \leq \sqrt{L\epsilon}\} \tag{45}$$

which proves Theorem 2 and justifies the use of our bounded polyhedral support. □

Observation 1 *We observe that (40) is slightly different from the standard inclusion used in robust optimization ([3] Section 2.3 and [52]). This is due to the fact that the authors in [3] 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 [52] 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.

Appendix B Equivalent reformulation of the stochastic program according to ϱ

In order to derive the deterministic/robust equivalent of problem (34a)–(34c), it is convenient to formulate the problem only in terms of ϱ. Recall the following uncertainty set for the random inflow vector ξ:

$$\Xi_t = \left\{ \xi \in \mathbb{R}^L \;\middle|\; \begin{array}{l} \exists \varrho \in \mathscr{A}_{L,\epsilon,t} \\ \xi = U_t(\hat{\zeta}_{t-1,L} + V\varrho) + u_t \end{array} \right\} \tag{46a, 46b}$$

where:

$$\mathscr{A}_{L,\epsilon,t} = \left\{ \varrho \in \mathbb{R}^L \;\middle|\; \begin{array}{l} |\varrho_i|\sigma_\varrho^{-1} \leq (L\epsilon)^{1/2}, i = 1, ..., L \\ \sum_{i=1}^{L} |\varrho_i|\sigma_\varrho^{-1} \leq L\epsilon^{1/2} \\ U_t(\hat{\zeta}_{t-1,L} + V\varrho) + u_t \geq 0 \end{array} \right\} \tag{47a–47c}$$

For any $\xi \in \Xi_t$, Equation (46b) 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 $\varrho \in \mathscr{A}_{L,\epsilon,t}$ where:

$$\hat{\mathcal{X}}^0 = \mathcal{X}^0 + \mathcal{X}^\Delta u_t + \mathcal{X}^\Delta U_t \hat{\zeta}_{t-1,L} \tag{48}$$

$$\hat{\mathcal{X}}^\Delta = \mathcal{X}^\Delta U_t V \tag{49}$$

Since $\hat{\mathcal{X}}^0$ and $\hat{\mathcal{X}}^\Delta$ are decision variables unrestricted in sign, we can equivalently write problem (34a)–(34c) as:

$$\min_{\hat{\mathcal{X}}^0, \hat{\mathcal{X}}^\Delta} \sum_{l=0}^{L-1} \hat{\mathcal{X}}_{t+l}^{0,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^0 + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2})^\top G_{t+l} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2}) \tag{50a}$$

$$\text{s.t.} \quad (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^\Delta - \hat{C}_{t+l}^\Delta)\varrho \geq -\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 + \hat{C}_{t+l}^0 \qquad \forall l \in \mathbb{L}, \quad \forall \varrho \in \mathscr{A}_{L,\epsilon,t} \tag{50b}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^\Delta - \hat{E}_{t+l}^\Delta)\varrho = -\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 + \hat{E}_{t+l}^0 \qquad \forall l \in \mathbb{L}, \quad \forall \varrho \in \mathscr{A}_{L,\epsilon,t} \tag{50c}$$

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 $\hat{E}_{t+l}^0 = E_{t+l} U_t \hat{\zeta}_{t-1} + E_{t+l} u_t$.

Appendix C Deriving the robust/deterministic equivalent

In order to robustify constraint (50b), we write $\mathscr{A}_{L,\epsilon,t}$ as a polyhedron in the lifted space $\{(\varrho_1^+, \varrho_1^-, \cdots, \varrho_L^+, \varrho_L^-)^\top \in \mathbb{R}_+^{2L} : \exists \varrho \in \mathbb{R}^L; \varrho_i^+ - \varrho_i^- = \varrho_i; \varrho_i^+ + \varrho_i^- = |\varrho_i|, \forall i = 1, ..., L\}$, 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 $i = 1, ..., L$. We specifically consider:

$$\mathscr{A}_{L,\epsilon}^{lift} | \mathcal{G}_{t-1} = \left\{ \varrho^{lift} = \begin{pmatrix} \varrho_1^+ \\ \varrho_1^- \\ \vdots \\ \varrho_L^+ \\ \varrho_L^- \end{pmatrix} \in \mathbb{R}_+^{2L} \;\middle|\; \begin{array}{l} \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \leq (L\epsilon)^{1/2} \mathbf{1} \\ \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \leq L\epsilon^{1/2} \\ -U_t V R \varrho^{lift} \leq U_t \hat{\zeta}_{t-1,L} + u_t \end{array} \right\} \tag{51a–51c}$$

where

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

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

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

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

Robustifying the $k_l^{th} \in \{1, ..., m_{t+l}\}$ constraint (50b) 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 $\alpha_{k_l}^\top(\hat{\mathcal{X}}) R \varrho^{lift} \le (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}})_{k_l}$ holds for all $\varrho^{lift} \in \mathscr{A}_{L,\epsilon}^{lift} | \mathcal{G}_{t-1}$. Since $\mathscr{A}_{L,\epsilon}^{lift} | \mathcal{G}_{t-1}$ 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 $(\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 - \hat{C}_{t+l}^0)_{k_l}$:

$$(P)\quad \begin{aligned} \max_{\varrho^{lift}\ge 0} \quad & \alpha_{k_l}^\top(\hat{\mathcal{X}}) R \varrho^{lift} \\ \text{s. t:} \quad & \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \le (L\epsilon)^{1/2} \mathbf{1} \\ & \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S \varrho^{lift} \le L\epsilon^{1/2} \\ & -U_t V R \varrho^{lift} \le U_t \hat{\zeta}_{t-1,L} + u_t \end{aligned}$$

$$(D)\quad \begin{aligned} \min_{\pi,\nu\ge 0} \quad & (L\epsilon)^{1/2} \pi^\top \mathbf{1} + \pi^0 L \epsilon^{1/2} + \nu^\top (U_t \hat{\zeta}_{t-1,L} + u_t) \\ \text{s. t:} \quad & \pi^\top \Sigma_{\varrho,L}^{-1/2} S + \pi^0 \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S - \nu^\top (U_t V R) \ge \alpha_{k_l}^\top(\hat{\mathcal{X}}) R \end{aligned}$$

Since $\mathscr{A}_{L,\epsilon,t}$ is full dimensional and contains 0, the linear system (50c) will hold for all $\varrho \in \mathscr{A}_{L,\epsilon,t}$ if and only if:

$$-\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 + \hat{E}_{t+l}^0 = 0 \in \mathbb{R}^{n_{t+l}} \tag{52a}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^\Delta - \hat{E}_{t+l}^\Delta) = 0 \in \mathbb{R}^{m_{t+l}} \tag{52b}$$

It follows that problem (34a)–(34c) is equivalent to:

$$\min_{\substack{\hat{\mathcal{X}}^0, \hat{\mathcal{X}}^\Delta, \\ \pi,\nu\ge 0}} \sum_{l=0}^{L-1} \hat{\mathcal{X}}_{t+l}^{0,\top} G_{t+l} \hat{\mathcal{X}}_{t+l}^0 + \sum_{l=0}^{L-1} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2})^\top G_{t+l} (\hat{\mathcal{X}}_{t+l}^\Delta \Sigma_{\varrho,L,t-1}^{1/2}) \tag{53a}$$

$$\text{s.t.} \quad (L\epsilon)^{1/2} \pi^\top \mathbf{1} + \pi^0 L \epsilon^{1/2} + \nu^\top (U_t \hat{\zeta}_{t-1,L} + u_t) \le (\sum_{\bar{l}=0}^{l} A_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 - \hat{C}_{t+l}^0)_{k_l} \quad \forall k_l \in \{1, ..., m_l\}, l \in \mathbb{L} \tag{53b}$$

$$\pi^\top \Sigma_{\varrho,L}^{-1/2} S + \pi^0 \mathbf{1}^\top \Sigma_{\varrho,L}^{-1/2} S - \nu^\top (U_t V R) \ge \alpha_{k_l}^\top(\hat{\mathcal{X}}) R \quad \forall k_l \in \{1, ..., m_l\}, l \in \mathbb{L} \tag{53c}$$

$$-\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^0 + \hat{E}_{t+l}^0 = 0 \in \mathbb{R}^{n_{t+l}} \tag{53d}$$

$$(\sum_{\bar{l}=0}^{l} D_{t+l,t+\bar{l}} \hat{\mathcal{X}}_{t+\bar{l}}^\Delta - \hat{E}_{t+l}^\Delta) = 0 \in \mathbb{R}^{m_{t+l}} \tag{53e}$$

Appendix D Details on $GARCH(m, s)$ model

Recall the general GARCH(m, s) model:

$$\hat{\sigma}^2_{t-1}(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho^2_{t-i} + \sum_{j=1}^{s} \beta_j \hat{\sigma}^2_{t-1-j}(1) \tag{54}$$

D.1 Stationarity conditions

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

$$\mathrm{E}\left[\mathrm{E}\left[\varrho^2_t | \mathcal{G}_{t-1}\right]\right] = \mathrm{E}\left[\alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho^2_{t-i} + \sum_{j=1}^{s} \beta_j \hat{\sigma}^2_{t-1-j}(1)\right] \tag{55}$$

$$\Leftrightarrow \sigma^2_\varrho = \alpha_0 \left(1 - \sum_{j=1}^{\max\{m,s\}} (\alpha_j + \beta_j)\right)^{-1} \tag{56}$$

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 $j > s$.

D.2 Reformulation of difference equations

We can reformulate (54) by making the substitution $\nu_t = \varrho^2_t - \hat{\sigma}^2_{t-1}(1)$:

$$\hat{\sigma}^2_{t-1}(1) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varrho^2_{t-i} + \sum_{j=1}^{s} \beta_j \hat{\sigma}^2_{t-1-j}(1) \tag{57}$$

$$\Leftrightarrow \varrho^2_{t-i} - \sum_{i=1}^{\max\{m,s\}} (\alpha_i + \beta_i)\varrho^2_{t-i} = \alpha_0 + \nu_t - \sum_{i=1}^{s} \beta_j \nu_{t-j} \tag{58}$$

with $\alpha_i = 0$ if $i > m$ and $\beta_i = 0$ if $i > s$. We can then rewrite (58) in a more compact form similar to the general ARMA model:

$$\hat{\phi}(B)\varrho^2_t = \alpha_0 + \hat{\theta}(B)\nu_t \tag{59}$$

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:

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

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

Given conditions (56):

$$\hat{\phi}(B)\sigma^2_\varrho = \frac{\left(1 - \sum_{j=1}^{\max\{m,s\}}(\alpha_j + \beta_j)\right)}{\left(1 - \sum_{j=1}^{\max\{m,s\}}(\alpha_j + \beta_j)\right)}\alpha_0 \tag{60}$$

Hence (59) is equivalent to:

$$\hat{\phi}(B)(\varrho_t^2 - \sigma_\varrho^2) = \hat{\theta}(B)\nu_t \tag{61}$$

D.3 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" (61). Given any $l \in \mathbb{Z}_+$, we specifically obtain:

$$\varrho_{t+l}^2 = \sigma_\varrho^2 + \sum_{i=0}^{\infty} \hat{\psi}_i \nu_{t+l-i} \tag{62}$$

Taking the conditional expectation $\mathrm{E}\left[\cdot|\mathcal{G}_t\right]$ on both sides of the Equation 62 yields:

$$\hat{\sigma}_t^2(l) = \sigma_\varrho^2 + \sum_{j=l}^{\infty} \hat{\psi}_j \nu_{t+l-j} \tag{63}$$

Equation (63) holds since for $j \leq l-1$ and $l \in \mathbb{Z}_+$:

$$\begin{aligned}
\mathrm{E}\left[\nu_{t+l-j}|\mathcal{G}_t\right] &= \mathrm{E}\left[\varrho_{t+l-j}^2 - \mathrm{E}\left[\varrho_{t+l-j}^2|\mathcal{G}_{t+l-j-1}\right]|\mathcal{G}_t\right] && (64)\\
&= \mathrm{E}\left[\varrho_{t+l-j}^2|\mathcal{G}_t\right] - \mathrm{E}\left[\mathrm{E}\left[\varrho_{t+l-j}^2|\mathcal{G}_{t+l-j-1}\right]|\mathcal{G}_t\right] && (65)\\
&= 0 && (66)
\end{aligned}$$

By the linearity and towering property of the conditional expectation [6].

Appendix E 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:

$$\begin{aligned}
&0 \leq 1 - \mathrm{E}\left[MSE_{frcst,t,L}\right]\mathrm{E}\left[MSE_{naive,t,L}\right]^{-1} && (67)\\
\Leftrightarrow &\sum_{l=0}^{L-1} \mathrm{E}\left[(\sigma_{t+l}(\sum_{i=0}^{\infty}(\bar{\psi}_i - \psi_i)\varrho_{t+l-i}))^2\right] \leq \sum_{l=0}^{L-1} \mathrm{E}\left[(\sigma_{t+l}\sum_{i=0}^{\infty}\bar{\psi}_i \varrho_{t+l-i})\right]^2 && (68)
\end{aligned}$$

If $\bar{\phi} = \phi$ and $\bar{\theta} \neq \theta$, then (68) is equivalent to:

$$\sum_{l=0}^{L-1} \sigma_{t+l}^2 \frac{(\theta - \bar{\theta})^2}{1-\phi^2}\sigma_\varrho^2 \leq \sum_{l=0}^{L-1} \sigma_{t+l}^2 \frac{\bar{\theta}(2\phi + \bar{\theta}) + 1}{1-\phi^2}\sigma_\varrho^2 \tag{69}$$

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.

References

[1] 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, pp. 267–281, Budapest. Akadémiai Kiado.

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

[3] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). Robust Optimization. Princeton University Press.

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

[5] 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.

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

[7] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31:307–327.

[8] 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.

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

[10] Carpentier, P.-L., Gendreau, M., and Bastin, F. (2013). Long-term management of a hydroelectric multireservoir system under uncertainty using the progressive hedging algorithm. Water Resources Research, 49:2812–2827.

[11] 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.

[12] 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.

[13] Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2008). Water reservoir control under economic, social and environmental constraints. Automatica, 44(6):1595–1607.

[14] 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.

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

[16] Côté, P., Haguma, D., Leconte, R., and Krau, S. (2011). Stochastic optimisation of hydro-quebec hydropower installations: a statistical comparison between sdp and ssdp methods. Revue canadienne de génie civil.

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

[18] Dyer, M. and Stougie, L. (2006). Computational complexity of stochastic programming problems. Mathematical Programming Series A, 106(3):423–432.

[19] Fleten, S.-E., K. J. and Näsäkkälä, E. (2011). Handbook of integrated risk management in global supply chains, chapter Risk Management in Electric Utilities, pp. 495–513. John Wiley & Sons, Inc.

[20] Galelli, S., Goedbloed, A., Schwanenberg, D., and van Overloop, P.-J. (2014). Optimal real-time operation of multipurpose urban reservoirs: Case study in singapore. Journal of Water Resources Planning and Management, 140(4):511–523.

[21] 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.

[22] Gauvin, C., Delage, E., and Gendreau, M. (2015). A robust optimization model for the risk averse reservoir management problem. Technical Report G-2015-131, Les cahiers du GERAD.

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

[24] 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, pp. 33–55. Springer Berlin Heidelberg, Berlin, Heidelberg.

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

[26] Gonçalves, R., E.C., Gendreau, M., and Finardi, E. (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, pp. 129–155. Springer US.

[27] 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.

[28] 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.

[29] 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.

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

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

[32] 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.

[33] 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.

[34] 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.

[35] 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, pp. 397–402. Iowa State University.

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

[37] 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.

[38] Pereira, M. and Pinto, L. (1991). Multi-stage stochastic optimization applied to energy planning. Mathematical Programming, 52:359–375.

[39] 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, pp. 272–281. Springer US.

[40] Phillpot, A. and Guan, Z. (2008). On the convergence of stochastic dual dynamic programming and related methods. Operations Research Letters, 36.

[41] 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.

[42] 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, 41(9):839–844.

[43] Riedel, F. (2004). Dynamic coherent risk measures. Stochastic Processes and their Applications, 112:185–200.

[44] 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.

[45] 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):n/a–n/a. W06407.

[46] 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.

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

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

[49] Shapiro, A. (2011). Analysis of stochastic dual dynamic programming method. European Journal of Operational Research, 209:63–72.

[50] Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). Lectures on stochastic programming: modeling and theory. MOS-SIAM Series on Optimization.

[51] 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.

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

[53] 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.

[54] 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.

[55] 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.

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

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

[58] Yeh, W. W.-G. (1985). Reservoir management and operations models: A state-of-the-art review. Water Resources Research, 21(12):1797–1818.

[59] Zéphyr, L., Lang, P., Lamond, B. F., and Côté, P. (2016). Controlled Approximation of the Stochastic Dynamic Programming Value Function for Multi-Reservoir Systems, pp. 31–37. Springer International Publishing, Cham.