Real-time control of water reservoir operations: a learning-based hierarchical approach

Pauline Kergus\textsuperscript{1}, Simone Formentin\textsuperscript{2}, Matteo Giuliani\textsuperscript{2}, and Andrea Castelletti\textsuperscript{2}

Abstract—Hydropower dams represent a way to combat fossil fuel consumption, while also providing irrigation and urban water supply. The optimal control of a water reservoir operations still represents a challenging problem, due to uncertain hydrologic conditions and the need to adapt to changing environment and varying control objectives. In this work, we propose a real-time learning-based control strategy based on a hierarchical predictive control architecture. More specifically, two control loops are considered: the inner loop is aimed to make the overall dynamics similar to an assigned linear one, then the outer economic model-predictive controller compensates for model mismatches, enforces suitable constraints and boosts the tracking performance. The effectiveness of the proposed approach as compared to traditional dynamic programming strategies is illustrated on an accurate simulator of the Hoa Binh reservoir in Vietnam.

Index Terms—water reservoir, data-driven control, learning based predictive control.

I. INTRODUCTION

On a global scale, population and economic growth result in an increasing energy demand. At the same time, climate change and growing populations pressure freshwater resources \cite{1}. Hydropower dams could constitute a response to these big energy challenges since they provide a wide range of benefits such as reduced fossil fuel consumption, irrigation and urban water supply. However, large dams imply substantial financial and environmental costs and, as explained in \cite{2}, many large storage projects worldwide are failing to produce the full range of benefits that would economically justify their development. Therefore, operating existing infrastructures more efficiently, rather than planning new ones, is a critical challenge. New operating policies should be able to adapt release decisions to uncertain hydrologic conditions and to changing objectives such as growing water demands \cite{3}.

This problem has received much attention from different research fields since the 70s but, as explained in \cite{1}, it remains challenging for different reasons: water reservoirs models are highly non-linear and multiple and conflicting interests are at stakes, usually formulated as non-linear and strongly asymmetric objective functions. In addition, the system is affected by strong uncertainties, such as the inflow of water, which cannot be neglected. In the literature, dynamic programming (DP) and its stochastic extension (SDP) are among the most widely used methods for designing optimal operating policies for water reservoirs, see \cite{4}. In practice, the use of SDP is limited by its computational complexity.

In this paper, we propose to control reservoirs operations using a mild knowledge of the system description and a hierarchical learning-based approach using operation data as illustrated in \cite{5}. First, a parametric controller is designed to match some desired closed-loop behavior using the Virtual Reference Feedback Tuning approach of \cite{6}. An outer Model Predictive Controller (MPC) is then used as a reference governor, enabling to enforce constraints on the water release and to compensate for possible tracking deficiencies. It follows that the problem of selecting an achievable reference model becomes less critical than in \cite{6} since low-performance models can be employed, and as such they are easily achieved with a simple control structure like PID (Proportional Integral Derivative)-like blocks.

The above approach shows a number of advantages over traditional SDP: (i) it can be easily implemented online, (ii) being “model-free” (it does not require a full mathematical description of the control reservoir), it can be easily adapted in real-time to any change in the system or operating conditions, (iii) it can handle time-varying constraints within a receding-horizon rationale. Nonetheless, the method presents a number of tuning knobs whose selection will be discussed throughout the paper.

For the sake of completeness, the above learning-based approach has already shown its potential in implicit force control for robotics \cite{7} and position control in mechatronics \cite{8}. However, this is the first time such a rationale is applied to a hydropower system, where the overall control objective is formulated in an economic MPC fashion (see \cite{9}).

An accurate simulator of the Hoa Binh water reservoir system (Vietnam) is used as a case study throughout the whole paper to demonstrate the effectiveness of the proposed approach and to compare its performance with classical SDP.

The remainder of the paper is organized as follows. First, the Hoa Binh case study is presented, followed by a brief summary of the classical dynamic programming approach typically employed for policy search. The proposed hierarchical approach is then described, starting from the data-driven design of the inner-loop based on the VRFT method \cite{6}, and then introducing the outer economic MPC loop, which plays the role of a reference governor. Simulation results are then reported and compared with SDP. Concluding remarks, along with outlooks for future research, are presented in the last section.

This project was partially supported by the Italian Ministry of University and Research under the PRIN’17 project ”Data-driven learning of constrained control systems”, contract no. 2017J89ARP.

\textsuperscript{1} Department of Automatic Control, Lund University, Sweden (e-mail: pauline.kergus@control.lth.se)

\textsuperscript{2} Dipartimento di Elettronica, Informazione e Biomeccanica, Politecnico di Milano, Italy (e-mail: simone.formentin@polimi.it, matteo.giuliani@polimi.it, andrea.castelletti@polimi.it)
II. PRELIMINARIES

A. The Hoa Binh case study

The Hoa Binh reservoir is one of the largest reservoirs in Vietnam characterized by a surface area of about 198 km² and an active storage capacity of about 6 billion m³. The reservoir is located along the Da River (see Fig. 1), which is the main tributary of the Red River, with this latter being the second largest basin of Vietnam accounting for a total area of about 169,000 km². The Hoa Binh reservoir is one of the largest reservoirs in Vietnam being the second largest basin of Vietnam accounting for a total area of about 169,000 km². The reservoir is located along the Da River (see Fig. 1), which is the main tributary of the Red River, with this latter being the second largest basin of Vietnam accounting for a total area of about 169,000 km². The reservoir affects the Red River Basin (see Fig. 2).

The nonlinear dynamics of the reservoir are due to the stochastic Max (MDP) by combining conceptual and data-driven models. More specifically, the Hoa Binh dynamics is represented by a stochastic disturbance \( q_{t+1}^{D} \) (namely, the inflow to the reservoir in the time interval \([t, t+1)\)), i.e.

\[
s_{t+1} = s_{t} + q_{t+1}^{D} - r_{t+1}
\]  

(1)

The nonlinear dynamics of the reservoir are due to the release function, which determines the actual release as

\[
r_{t+1} = f(s_{t}, u_{t}, q_{t+1}^{D})
\]  

(2)

as a function of the control \( u_{t} \), along with the minimum and maximum releases that can be produced in the time interval \([t, t+1)\) starting from \( s_{t} \) with inflows \( q_{t+1}^{D} \) (by keeping all the dam’s gates completely closed and completely open, respectively).

These constraints of the problem are embedded in the model [3], thus guaranteeing the feasibility of the designed solutions. In the adopted notation, the time subscript of a variable indicates the instant when its value is deterministically known.

The reservoir storage \( s_{t} \) is observed at time \( t \), whereas the inflow and release have subscripts \( t + 1 \) as they depend on the realization of the stochastic process in the time interval \([t, t+1)\). The routing of the water released from the reservoir to the delta and the city of Hanoi is simulated by a feedforward neural network providing the water level in Hanoi as a function of the Hoa Binh release along with the natural discharges of the Thao (\( q_{t+1}^{T} \)) and Lo (\( q_{t+1}^{L} \)) rivers [10].

Overall, the modelled system includes two state variables \( x_{t} = [s_{t}, s_{t}^{*}] \), one control \( u_{t} \), and a vector of three stochastic disturbances \( \varepsilon_{t+1} = [q_{t+1}^{D}, q_{t+1}^{T}, q_{t+1}^{L}] \) which are both spatially and temporally correlated. For more details about the model formulation, see [11].

B. Policy design via dynamic programming

The two conflicting interests of hydropower production and flood control that drive the Hoa Binh operation are modeled using the following objective formulations, evaluated over the simulation horizon \( H \):

- **Hydropower production**: daily average energy production (kWh/day) to be maximized, computed as

  \[
  J^{H} = \frac{1}{H} \sum_{t=0}^{H-1} (\eta g \gamma_{w} h_{t} q_{t+1}^{T} \text{urb}) \cdot 10^{-6}
  \]  

  (3)

  where \( \eta \) is the turbine efficiency (which depends on the hydraulic head), \( g = 9.81 \text{ (m/s²)} \) is the gravitational acceleration, \( \gamma_{w} = 1000 \text{ (kg/m³)} \) is the water density, \( h_{t} \) (m) is the net hydraulic head (i.e., reservoir level minus tailwater level), \( q_{t+1}^{T} \text{urb} \) (m³/s) represents the turbined flow;

- **Flood control**: the daily average excess level \( h_{t+1}^{Hanoi} \text{ (cm²/day)} \) in Hanoi with respect to the flooding threshold \( h = 950 \text{ cm} \), to be minimized, computed as

  \[
  J^{F} = \frac{1}{H} \sum_{t=0}^{H-1} max(h_{t+1}^{Hanoi} - h, 0)^{2}
  \]  

  (4)

  where \( h_{t+1}^{Hanoi} \) is the level in Hanoi estimated by the routing model.

The optimal control problem of the Hoa Binh reservoir is formulated as designing the set of Pareto optimal (or
approximate) control policies \( \mathcal{P}^* \) that minimize the objective functions vector \( \mathbf{J} \):

\[
\mathcal{P}^* = \arg \min_{\mathcal{P}} \mathbf{J} = \arg \min_{\mathcal{P}} [-J^H, J^F] \tag{5}
\]

subject to the dynamics of Hoa Binh reservoir (eq. (1)) and the objective functions are defined in eqs. (3)-(4).

The traditional approach to solve the multi-objective optimal control problem (5) is to reformulate it as a series of single-objective problems that can be solved via SDP [4] by computing the Bellman function as

\[
H_t(s_t) = \min_{u_t} E_{\varepsilon_{t+1}} [G_{t+1}(s_t, u_t, \varepsilon_{t+1}) + H_{t+1}(s_{t+1})] \tag{6}
\]

where \( H_t(\cdot) \) estimates the expected long-term cost of a policy over a discrete grid of states (i.e., reservoir storage \( s_t \) and time \( t \)) for the "scalarized" objective, and \( G_{t+1}(\cdot) \) being the corresponding scalarized immediate cost function. The optimal control policy, defined as a periodic sequence of control laws with period \( T \), that minimize a given scalar objective, is then derived as

\[
\mu^*_t(s_t) = \arg \min_{u_t} E_{\varepsilon_{t+1}} [G_{t+1}(s_t, u_t, \varepsilon_{t+1}) + H_{t+1}(s_{t+1})] \tag{7}
\]

In principle, DP can solve problem (7) under relatively mild assumptions [12]. In practice, the application of DP in large-scale control schemes is constrained by the well-known curse of dimensionality [13]. Besides, DP is also constrained by the curse of modeling [14], as any input of the control policy must be explicitly modeled, and the curse of multiple objectives [15] as the generation of the full set of Pareto optimal solutions factorially scales with the growth in the number of the objectives. These three curses, along with the challenges related to the adaptation to variable hydrologic regimes, motivate the search for scalable and more flexible solutions.

### III. Hierarchical Control System Design

An online strategy such as economic Model Predictive Control (eMPC) appears as the best candidate to operate such a reservoir and would also allow us to enforce path constraints. To that extent, the hierarchical approach introduced in [5] looks interesting as the first step consists in reducing the complexity of the system dynamics by closing an inner loop through a model-free feedback controller. Indeed, the design of such a low-level controller \( C \) allows to assimilate the inner-loop to a simpler linear model. This whole hierarchical approach is detailed in this section, starting with the design of the inner-loop controller in III-A. The overall control architecture is shown in Figure 3.

#### A. Data-driven inner-loop design

To start with, a low-level controller \( C \) is designed to reduce the complexity of the plant’s behaviour and to transform it into a known linear inner-loop model. The system is described by the balance equation (1) and by the release nonlinear function \( f \). Since \( f \) does not have a parameterized expression, it motivates the use of a data-driven technique, such as the VRFT [6], to design the inner-loop controller.

1Namely: \( p^* \equiv [\mu^*_0(s_t), \ldots, \mu^*_T(s_{T-1})] \).

![Prediction of \( q^D, q^T, q^L \)]

![ Inner loop ]

Fig. 3: The proposed control architecture: the inner-loop is designed to simplify the model dynamics for the outer reference governor, while the latter consists of an economic model predictive controller generating \( s^{ref} \).

![ Inner loop design using VRFT: \( M \) is the reference model, \( P \) is the system to be controlled and \( C \) represents the controller to be tuned on the basis of the signals \( u^* \) and \( s^* \). The signals \( \tau \) and \( \tilde{\tau} \) are the virtual reference and error respectively. Based on the available time-domain data \( \{u_t, s_t\}_{t=1}^{T} \), a desired closed-loop behaviour \( M \) and a controller structure \( C(\theta) \), the objective of the VRFT is to find the controller parameters \( \theta^* \) such that the resulting closed-loop is as close as possible to the reference model \( M \). The key idea is the computation of the virtual reference signal \( \tau_t = M^{-1}(z)s_t \), as the reference that would feed the loop if the complementary sensitivity function was exactly \( M \). Like that, the optimal controller (i.e., the one achieving \( M \) in closed-loop) can be computed as the system producing \( u_t \) when fed by the virtual error \( \tilde{\tau}_t = \tau_t - s_t \). Since the virtual error and the virtual reference can be computed off-line based on the available dataset, also the controller is retrieved by solving the one-shot optimization problem

\[
\theta^* = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^{N} (u_t - C(z, \theta)\tilde{\tau}_t)^2. \tag{9}
\]

Specifically, in [6], it has been shown that a linear parameterization of the controller \( C(\theta) \) (like a PID) leads to a simple quadratic problem that can be solved via least squares formulas. By suitably prefiltering \( u_t \) and \( s_t \), it can also be shown that the minimum of (9) coincides with the optimal controller, if this belongs to the considered class [16].

In this work, the cyclostationary signals \( u_c \) (release decision) and \( s_c \) (storage), which represents the annual mean behaviour corresponding to the data collected between 1962
and 1969, are used as the dataset for the identification of the controller. The desired closed-loop behavior $M$ is defined as

$$M(z) = \frac{0.2z^{-1}}{1 - 0.8z^{-1}}.$$  \hspace{1cm} (10)

Unlike the classical VRFT framework, here the reference model does not need to represent the desired closed-loop performance: this will be handled by the outer-loop eMPC controller. The reference model only needs to be practically achievable, which is most likely the case when specifying a low-performance closed-loop behavior $M$.

In the present case, the inner-loop controller has a PID structure:

$$C(z, \theta) = \theta_1 + \frac{\theta_2}{1 - z^{-1}} + \theta_3(1 - z^{-1})$$  \hspace{1cm} (11)

and the coefficients obtained through the VRFT procedure are

$$\theta^* = 10^{-6} \left[ -0.4439 \quad -0.1063 \quad -0.1898 \right]^T.$$  \hspace{1cm} (12)

The resulting inner-loop, see Figure 3, contains a nonlinear block $f$ which corresponds to the physical limitations of the dam in terms of water release.

The purpose of the inner-loop controller $C$ is to assimilate the inner-loop to a simple linear model which will be used later for prediction in the outer-loop to simplify the eMPC problem. To that extent, the nonlinear term $f$ is neglected, i.e. $u \approx r$, which leads to the linear model $H$ approximating the inner-loop behaviour:

$$H : \left\{ \begin{array}{l}
s(t + k + 1) = s_{ref}(t + k) + T_s(q_{D}(z) - z^{-1}u(t)) \\
u(t + k + 1) = C_H x(t + k) + D_H \left( \begin{array}{c}s_{ref}(t + k) \\
q_{D}(t + k)\end{array} \right)
\end{array} \right\} \hspace{1cm} (13)$$

Figure 5 compares the water-storage signals that are obtained using the inner-loop linear approximation $H$ or by simulating the real inner-loop visible on Figure 3 both fed with the optimal water storage $s^*$ as reference signal $s^r$. The signal $s^*$ is found through deterministic dynamic programming when solving the optimal control problem for the period 1962-1969. The results indicate that the linear model $H$ can reasonably be used to predict the inner-loop behaviour. Even though the prediction model is not perfect because of the nonlinear part $f$ of the true inner-loop, this aspect will be handled in the outer-loop through constraints enforcement.

Finally, this simulation highlights the fact that the inner-loop alone is not sufficient for this application, for two reasons. First, Figure 5 shows that the inner-loop exhibits reasonable tracking performances when fed with $s^*$ as reference, but may be too slow to produce a sufficient level of benefit in terms of hydropower production and floodings’ damages. Besides that, the reference $s^r$ to be tracked still needs to be defined. The optimal control strategy could not be used in this case, since it would require to optimize over the whole considered period. The best option is therefore to perform an online optimization of the reference signal on a limited prediction horizon. Based on these considerations and on the simple description of the inner-loop obtained through this data-driven control design, an outer-loop model predictive controller is then designed as described in the next paragraph.

B. Model-based outer-loop design using economic MPC

While the inner-loop allows to have a simplified description of the problem, the outer-loop has a double objective: performance optimization and constraints enforcement. To that extent, an economic model predictive controller is implemented, solving the following problem at each time-step $t$:

$$\begin{array}{ll}
\min_{s^r(t)} & -\alpha_{hyd} J_{hyd} + \alpha_{flo} J_{flo} \\
\text{s.t.} & \forall k = 1 \ldots N_p, \end{array} \hspace{1cm} (14)$$

$$\{ x(t + k + 1) = A_H x(t + k) + B_H \left( \begin{array}{c}s_{ref}(t + k) \\
q_{D}(t + k)\end{array} \right) \\
u(t + k + 1) = C_H x(t + k) + D_H \left( \begin{array}{c}s_{ref}(t + k) \\
q_{D}(t + k)\end{array} \right) \}$$

where $r^m(s_t, q^D_{t+1}) \leq u(t) \leq r^\text{max}(s_t, q^D_{t+1})$,

$$s^\text{min} \leq s(t) \leq s^\text{max}$$

First, the outer-loop controller acts as a reference governor for the inner-loop, determining the signal $s^r$ such that good performance is obtained. The performance is determined through the daily average energy production $J_{hyd}$ and the daily average excess level in Hanoi $J_{flo}$ over the prediction horizon $N_p$. Their definition is therefore similar to the ones given in [3] and [4] for the optimal control approach. The major difference is that the performance is here evaluated online on a shorter time window, starting at the current time step $t$ and covering the prediction horizon $N_p$, instead of computing the optimal solution offline using the whole period information (in the present case the 1962-1969 period as in [4]). To evaluate the performances over the prediction horizon, the linear description $H$ of the inner-loop, given in a state-space form $(A_H, B_H, C_H, D_H)$, is used as prediction model (suitably scaled to avoid numerical errors). Its input are the reference water storage $s^r$, which is the optimization variable, and the flow $q^D$ from the Da river. The outputs of the inner-loop are the water storage $s$ and the water release decision $u$.

Secondly, the outer-loop controller should enforce constraints on the release decision $u$ so that the assumption $u \approx r$ made earlier holds. In addition, enforcing the constraints allow to avoid emptying the reservoir, which would happen otherwise since the flood damage objective is mostly equal to zero over time. To that extent, hard constraints are imposed on both outputs. The constraints are constant when it comes to the water storage $s$, with $s_{\text{min}} = 3.8 Gm^3$ and $s_{\text{max}} = 9.9 Gm^3$. The constraints on $u$ are nonlinear and allow to take into account the nonlinearity $f$ of the system: $r_{\text{min}}$ and $r_{\text{max}}$ represent the minimal and maximal water release, respectively. Like $f$, they depend on the current water storage $s_t$ and the incoming flow $q^D$. These functions $r_{\text{min}}$ and $r_{\text{max}}$ are not parameterized: their value is known only for a given set of values of storage and flow.
Fig. 5: Performance of the inner-loop and linear approximation $H$ (13) of the inner-loop. The water release decision $u$ (left) and water storage $s$ (right) are obtained by simulating the true inner-loop and its linear approximation over the period 1962-1969. The signal $s^{ref}$ is taken as a desirable water storage signal, obtained by solving the optimal control problem through DDP over the same period.

Fig. 6: Trajectory of the controlled system when combining the inner-loop controller with the outer-loop economic MPC. The controlled system is simulated over the period 1962-1969 and compared with the Deterministic Dynamic Programming (DDP) approach.

IV. SIMULATION RESULTS

The behaviour of the controlled system, including both the inner-loop and the outer loop controllers, is simulated over the period 1962-1969, using the available data regarding the incoming flows $q^D$, $q^T$ and $q^L$. The prediction horizon is $N_p = 20$ days and the weights in the cost function are selected as $\alpha^{hyd} = 0.05$ and $\alpha^{flo} = 0.95$.

The resulting water release and water storage are visible in Figure 6. Figure 6a highlights that the constraints on the water release decision are satisfied so that $u_t = r_{t+1}$. The corresponding performance can be evaluated through the hydropower production and the excess amount of water in Hanoi, visible in Figure 7. The results of the optimal control approach, solved through Deterministic Dynamic Programming (DDP) over the same period, are also represented in both the figures and can be seen as an "upper" reference. In comparison, the proposed control strategy exhibits a slower reaction to the monsoon when emptying and refilling the reservoir every year and does not reach that high storage for the rest of the time.

The performance of the controlled system can be more easily appreciated when looking at the Pareto front given in Figure 8 which represents the compromise between the average daily hydropower production and the average daily
excess of water in Hanoi. These values are computed over the simulation period 1962-1969 for three different values of the prediction horizon, $N_p = 10$, 15 and 20 days. Shortening the prediction horizon allows to increase the hydropower production, also implies to increase the flooding in Hanoi. Indeed, a prediction horizon of 10 days is not sufficient to predict it.

The performance compromise is also represented for the optimal control approaches DDP and SDP. The DDP approach can be conceived as an upper bound for the achievable performance. Unlike the proposed approach, the optimization is done offline for the whole considered period. Obviously, the performance depends on the weights $\alpha_{hyd}$ and $\alpha_{flo}$ used in the cost function.

According to Figure 8 the proposed approach performs better than SDP, which makes it a good candidate for its online implementation and use in different time periods. In order to demonstrate this, a simulation is run for the period 2007-2008. The results are visible in Figure 9.

V. CONCLUSIONS

In this paper, a hierarchical data-driven control design strategy is proposed for water resources management, with a focus on the Hoa Binh reservoir case study. First, a linear controller is designed from data (with no use of the model of the system) to approximately assign a desired behaviour to the inner-loop. The model of the inner loop is then used for the design of an outer economic model predictive control loop, aimed to handle the performance and satisfy the signals constraints.

Compared to the resolution of the optimal control through traditional SDP tools, the proposed strategy allows to find a better compromise between hydropower production and floodings in Hanoi for a prediction horizon of 20 days. The main strength of the proposed approach is to perform online control and therefore to adapt to new environmental conditions.

Future research will include the study of the prediction of the future input flow from the different involved rivers. Moreover, many reservoirs coexist in the Red River basin, so it would be interesting to investigate their possible coordination.

REFERENCES

[1] R. I. McDonald, P. Green, D. Balk, B. M. Fekete, C. Revenga, M. Todd, and M. Montgomery, “Urban growth, climate change, and freshwater availability,” Proceedings of the National Academy of Sciences, vol. 108, no. 15, pp. 6312–6317, 2011.
Fig. 9: Performance of the controlled system in terms of hydropower production and flood damages over the period 2007-2008, compared with the DDP approach.