POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

Real-time management of water reservoirs with deterministic and ensemble forecasts: the Lake Como case study

Laurea Magistrale in Automation and Control Engineering

Author: Federico Staffa

Advisor: Prof. Andrea Francesco Castelletti

Co-advisor: Dr. Andrea Ficchí

Academic year: 2021-2022

1. Introduction

This thesis investigates the application of advanced on-line control methods, such as Model Predictive Control (MPC) and its more recent stochastic extension, called Tree-Based Model Predictive Control (TB-MPC), for the regulation of Lake Como, in northern Italy. These control schemes can use hydrological forecasts which are increasingly available with better and better accuracy. MPC is a deterministic approach that can employ deterministic predictions to generate an optimal sequence of controls over a receding control horizon, while TB-MPC exploits probabilistic predictions, also called Ensemble Forecasts (EFs). These EFs are a collection of different deterministic trajectories, representing the prediction uncertainty which can be then taken into account in the optimization problem. The real-time knowledge and forecast of lake inflows from the upper catchment is expected to bring short-term improvements to the off-line management policies [1]. However, despite the increasing availability of operational forecasts for the region of Lake Como (and beyond) especially of local deterministic forecasts (but also probabilistic continental forecasts), these have not been tested in any operational real-time framework like MPC yet, that could support the lake regulator in decision-making. Beyond Lake Como, the literature suggests a limited uptake by practitioners for all optimal reservoir management techniques. Thus, there is an urgent need for showing the value of available *operational* forecasts in real-world case studies for the re-operation of existing infrastructure to optimise their benefits.

In this study, the tested on-line approaches, MPC and TB-MPC, will be compared to more extensively studied off-line approaches, namely Deterministic Dynamic Programming (DDP) and Stochastic Dynamic Programming (SDP), and to the current (historical) management by the lake operator (Consorzio dell'Adda). A previous study on Lake Como [2] has analysed the value of exogenous hydro-meteorological information, like inflow forecasts, in an off-line control scheme, but no previous author has investigated the value of real forecasts with on-line control schemes in this case study before.

A few previous studies for other water reservoirs (e.g.[3]) showed that TB-MPC can provide a more adaptive control framework facing uncertainties and outperform deterministic MPC. However, the currently available EFs from continental forecasting systems (e.g. the EU Copernicus - European Flood Awareness System, EFAS) may have large biases and need further model calibration before being potentially used within TB-MPC for Lake Como or other areas. So another interesting research aspect of this study is to generate EFs with a novel machine-learning technique based on the re-calibration of neural networks. A method for this has been recently proposed in the literature [5], but has yet to be applied in a real-world water management problem setting. This method has been receiving growing interest from the scientific community as it could provide skillful EFs at low computational and monetary cost. This could be an interesting locally-calibrated alternative to existing global hydrological EFs generated by big international meteorological data centers, running computationally demanding Numerical Weather Prediction (NWP) models but requiring more efforts for local calibration and bias adjustments.

The main objectives of the thesis are: (i) assessing the skill of available short-term deterministic hydrological predictions (maximum lead time of 3-days) for the Lake Como basin, (ii) investigating in what measure these short-term forecasts can bring benefits in the optimal regulation of the lake for flood control and downstream water supply, with respect to off-line policies and the current management, (iii) assessing the advantages of using TB-MPC with EFs generated via a novel data-driven technique that would be operationally feasible, and represents a computationally-effective alternative to traditional ensembles from NWP.

2. Case study: Lake Como

Lake Como is the third largest lake of Italy and has a total active water storage capacity of $247 Mm^3$. The lake regulation is committed to an authority called Consorzio dell'Adda, controlling the lake levels since 1946 by operating a dam located in Olginate. The lake is part of the Adda River Basin (4.552 Km^2), whose hydro-meteorological regime is typical of the southern Alps, characterized by dry periods in summer and winter, and streamflow peaks in late spring and autumn, fed by snowmelt and rainfall, respectively. Lake Como is regulated mainly to satisfy two primary competing objectives: (i) providing water supply to downstream users, mainly to irrigation districts, and (ii) preventing flooding along the lake shores, especially in the city of Como. To satisfy the summer water demand, the agricultural districts downstream prefer to store the water from snowmelt, but this increases the lake level and the flood risk. The lake is also important in sustaining ecosystems and a range of human activities, often in conflict in terms of objectives and requirements, like hydropower, navigation, tourism, etc.

3. Methods

3.1. Forecast skill and EF generation

The available deterministic inflow forecasts are the operational forecasts produced by PROGEA, an Italian company specialised in hydrological forecasting. Hourly forecasts are produced through their software called EFFORTS, using meteorological forecasts as input of a hydrological model. These are originally provided at 60-h lead time, over 2014 to 2022. Daily forecasts are obtained from these through an aggregation (mean) over 24h. The accuracy and skill of the available deterministic forecasts by PROGEA has been evaluated using a series of overall error-based scores (including RMSE, MAE, and NSE), derived skill scores, and flood event-based scores.

Reliable and bias-corrected hydrological EFs were not available for this case study, as those from EFAS are not calibrated for the region and have known large biases. Also, building local hydrological EFs with physically-based models may have associated monetary costs and require large computational resources. For these reasons, EFs have been generated from the historical observations and deterministic forecasts, with a machine-learning method recently proposed in the literature. This method consists of a data-driven synthetic EF generation through a series of Feed Forward Neural Network (FFNN), as in **Figure 1**. This approach circumvents the complexity and computational burden related to the NWP models used by large data centers, allowing operational users to have access to alternative synthetic ensembles in real-time, potentially more locally effective.

A recent study ([5]) proposed four methods for generating synthetic EFs with the use of neural networks:

- **Random initial perturbations:** It consists in applying a random noise to the training data set of the neural network.
- **Singular Value Decomposition:** It is a technique from linear algebra, used to find and apply minimum input perturbation leading to the maximum output perturbation through the network's Jacobian.
- **Network Retraining:** It consists in carrying out a new training of the network for every ensemble member (inherently a "random" procedure, being a non-linear optimization).
- **Random Dropout:** It consists in a random deactivation of one or more neurons and all their connections, normally applied during training, but applicable during forecasting too, to diversify outputs.

This study will adopt the Network Retraining method, being one of the simplest and the one yielding the best results among them [5].

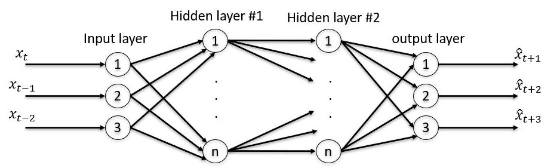

Figure 1: FFNN architecture for daily ensemble forecasting

The number of layers and neurons has been chosen empirically according to the size of the input data available (historical observed inflow) to avoid overfitting. Two hidden layers of $n = 10$ neurons led to good prediction capabilities.
In addition to the main procedure of randomized recalibration, other sources of uncertainty are introduced to differentiate more between output trajectories. These can be activated/deactivated at every training cycle, and they were progressively implemented in order to generate EFs with different associated skills, in order to asses the effect of EF performances on those of the management policy produced by TB-MPC.
These randomness factors are six, and include: changing seed of random number generator and training algorithm at every iteration, using different partition of whole data set for training and validation and adding a Gaussian perturbation to the input data sets and finally random.

3.2. Deterministic MPC

MPC is an advanced control framework that uses a model of the system together with the input prediction to estimate the future state of the system itself over a finite horizon, at every time step. This allows the optimisation of a control sequence with respect to an objective function. Then a receding horizon strategy is adopted, i.e. at each instant, the first control action of the sequence is applied and the horizon is moved towards the future from the next control time step. In our context, such optimization is non linear and constrained, formalized as:

$$\underset{u_t,\dots,u_{t+H}}{minimize} \quad f_{tot} = \sum_{\tau=t}^{t+H} g_\tau(x_\tau, u_\tau) + g_\tau^{end}(x_{\tau+H})$$

$subject\ to:$

$$\begin{aligned} x_{\tau+1} &= x_\tau + \Delta_T(q_\tau - u_\tau) \\ x_{\tau=t} &= x_0 \quad given \end{aligned} \tag{1}$$

$$u_{min} \le u_\tau \le u_{max} \tag{2}$$

Equation (1) represents the model of the system, a mass balance employing the inflow prediction q_τ, the state of the system (lake level) x_τ, and the control action (dam release) u_τ. Equation (2) represents the saturation of the control,dictated by the dam infrastructure characteristics. The control horizon H is set to 3 days (or 60 hours in the hourly case), coinciding with the lead time of the available deterministic hydrological forecasts.

The goal is to obtain an optimal control sequence $u_t, ..., u_{t+H}$ minimizing the total cost f_{tot} over the forecast horizon H. Then, only the first control u_t is actually applied to the real system (receding horizon strategy). The total cost is composed by the step-cost g_τ and the end state penalty g_τ^{end}. The step-cost is a weighted sum of three different components mathematically defining the management objectives:

$$g_\tau^{flood} = (\max_{t...t+H}(x_\tau - h_{flood}, 0))^2$$

$$g_\tau^{wdeficit} = (\max_{t...t+H}(MEF + w - u_\tau, 0))^2$$

$$g_\tau^{lowlevel} = (\max_{t...t+H}(h_{low} - x_\tau, 0))^2$$

g_τ^{flood} is related to the flooding objective, it penalizes solutions that bring the system state over a certain threshold h_{flood} (1.1 m), which delimits the occurrence of a flood along the lake banks in Como. $g_\tau^{wdeficit}$ represents the water deficit objective, it penalizes those solution that do not satisfy the water demand (w) over the horizon; the Minimum Environmental Flow component (MEF) is an additional component representing the minimum release of water necessary for preserving the river ecosystem. Finally, $g_\tau^{lowlevel}$ is the low level objective, which penalizes solutions lowering the level of the lake under a specific threshold h_{low} (-0.2 m). This last objective is not as important as the others, and has been mostly neglected throughout the study by setting its weight to 0.

The end state penalty g_τ^{end} is a critical component of the total cost for MPC to ensure the long-term targets; it is the cost associated to less desirable states of the system. It is of particular importance in water systems as it allows the system to be left in a non-detrimental state once the optimization ends. In this study, it is defined as proposed in [1], using the optimal-cost-to go of a solved off-line management problem (either the DDP or SDP, for MPC with perfect or real predictions, respectively). In other words, g_τ^{end} contains crucial information from the end of the foreseeable future H to the end of the simulation, informing the system of the future cost associated to being in a certain state at the end of the optimization.

3.3. Stochastic Tree-Based MPC

TBMPC is a stochastic modification of MPC that can use EFs and optimises an optimal control tree instead of searching for an optimal control sequence. The inflow EFs are transformed into a tree of inflows. The power of the approach stems from the fact that the optimization takes into account every likely scenario represented by the EF, and the control action can be changed when information about which scenario is actually happening is available, denoting higher flexibility and robustness.
The approach is formalized as:

$$\underset{u_t^{M_\tau},...,u_{t+H}^{M_\tau}}{minimize} \quad f_{tot} = \sum_{z=1}^{Z} p(z) * f_z$$

$$f_z = \left[\sum_{\tau=t}^{t+H} g_\tau(x_{\tau,z}, u_{\tau,z}^{M_\tau}) + g_\tau^{end}(x_{\tau+H,z})\right]$$

subject to

$$\begin{aligned} x_{\tau+1,z} &= x_{\tau,z} + \Delta_T(q_{\tau,z} - u_{\tau,z}^{M_\tau}) \quad \forall z \\ x_{\tau=t,z} &= x_0 \quad given \quad \forall z \end{aligned} \tag{3}$$

$$u_{min} \leq u_{\tau,z}^{M_\tau} \leq u_{max} \quad \forall z \tag{4}$$

Equations (3) and (4) represent the same mass-balance and physical constraint seen before for MPC. The same notation as the original framework of MPC is followed, with the addition of the subscript z, indicating how this optimization is actually spanned across the $z \in Z$ members of the ensemble. Every ensemble has a total cost f_z, coinciding with the one for the single deterministic optimization of MPC. All the different f_z are combined together into the total cost f_{tot} by a weighted sum with their probability of outcome $p(z)$. The simulation aims to minimize the comprehensive cost for every ensemble by devising a set of controls $u_t^{M_\tau}, ..., u_{t+H}^{M_\tau}$ in a tree structure forced by the tree nodal partition matrix M_τ. A tree can be efficiently represented in a mathematical way by means of two sets, $P(\cdot)$ and $B(\cdot)$ defined on every member z of the ensemble, the former being the Parent set of a member (from which member it branches) and the latter the Branches set of a member (which members branch out of it). These have to respect the so called "Non-anticipatory condition" defined as:

$$u_\tau^i = u_\tau^j \quad when \quad \begin{cases} P(i) = j \\ t < B(j) \end{cases} \quad \forall\, i, j \in Z$$

This is an important condition for the tree definition, stating that controls should not depend on the outcome of stochastic variables that have not been extracted yet [4]. Control actions are computed between all the possible branches until uncertainty resolves and the specific control action for the branch happening is selected.

4. Results and discussion

The first comparison between the benchmarks (DDP, SDP and historical management) showed that the inflow knowledge is a valuable information management-wise. DDP significantly outperforms both SDP and the historical management for both objectives (**Figure 2** and **Figure 3**). This is because it has access to perfect knowledge over an infinite horizon; yet the more realistic non-perfect knowledge used by SDP leads to satisfactory results. The deterministic MPC was fed with perfect and real forecasts, with both daily and hourly control frequencies. With perfect forecasts, results proved that indeed an on-line control scheme with short-term forecast has the potential to reach the performance of DDP for flood control (while outperforming SDP, also for the deficit), proving the validity of an on-line management approach. With real forecasts instead, the performances slightly deteriorate, yet still improving indicators with respect to SDP (for floods) and the historical management (both indicators). The good performance is in line with the evaluation of the accuracy of the forecasts, showing that they are quite skillful in reproducing the observations. Overall, an hourly control frequency was found to give slightly better performances than the daily one (even slightly outperforming DDP in terms of flood control). However, as seen in **Figure 2**, the long-term behaviour is mostly the same.

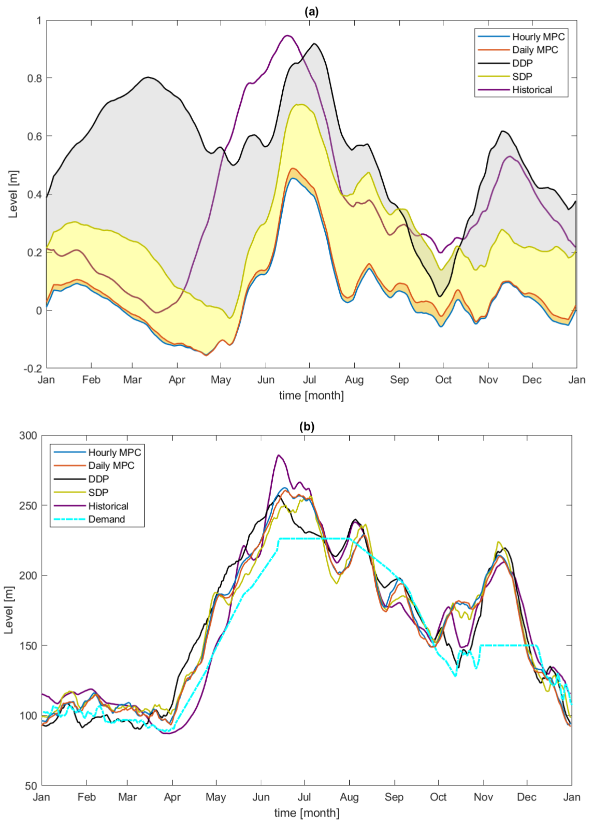

Figure 2: Daily cyclostationary levels and releases over the whole simulation period of 8 years, daily/hourly MPC with real predictions alongside the benchmarks (DDP, SDP) and the historical management.

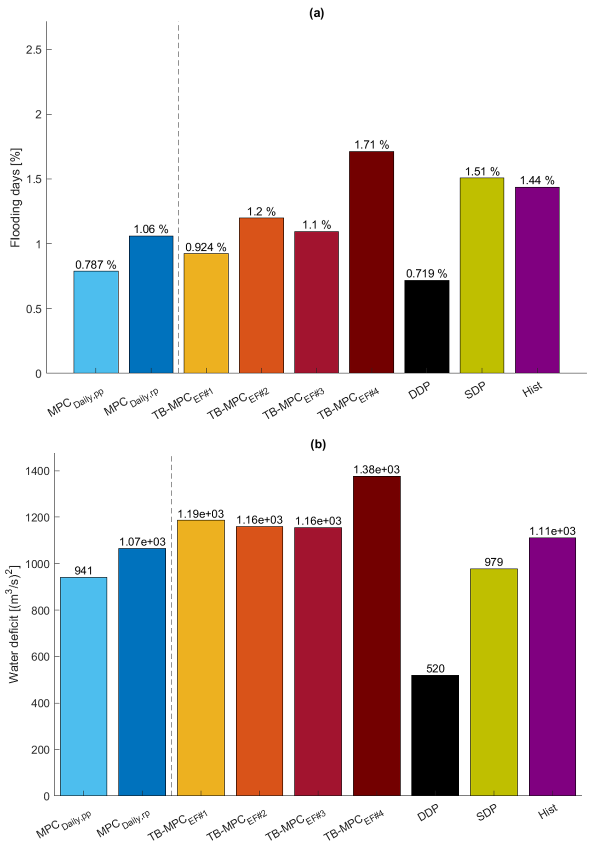

Figure 3: Simulation results: (a) Flooding, and (b) Water deficit indicators, over the 8-year study period, for the four TB-MPC options with different EFs, alongside the daily deterministic MPC with perfect and real forecasts, and benchmarks (DDP, SDP and historical management).

Then, a daily TB-MPC was run with four different synthetic EFs, generated with varying degree of randomness factors in the FFNN, leading to varying degree of skill. They are proposed in order of EF performance in **Figure 3**, from the most skillful (EF#1) to the less skillful (EF#4). No drastic improvement was found, but only a slight one in flood control for the most skillful EF (EF#1). More skilful EFs seem to lead to better control performances. Even if in this simulation, TB-MPC did not yield a clear cut improvement, the key strength of TB-MPC optimizing over inflow trees is to beat deterministic MPC operations across more variable or unexpected scenarios on average. Thus, a stochastic control of this type is a "no-regret" implementation that may increase the control robustness in the long run.

TB-MPC was also tested with a hourly control frequency with a simplified EF generation procedure, though from the results with daily/hourly deterministic MPC (**Figure 2**) only a minor short-term improvement may be expected with the hourly control. Given the high computation time, the hourly TB-MPC was tested only on selected flood and drought periods. The results suggest that the hourly TB-MPC is too influenced by "jumpy" forecast behaviour. So there seems to be no incentive in further developing an operational procedure to generate hourly EFs, when the daily TB-MPC with EFs (from the FFNN procedure) yields better results than the hourly TB-MPC.

5. Conclusions

In conclusion, among the different on-line approaches and resolutions tested, the daily stochastic (TB-MPC) one appeared to be the best option for this case-study, especially for flood control. The EF uncertainty seems to provide more consistent information at the daily scale given the current accuracy and jumpiness of hourly forecasts. The multiple randomized calibration of neural networks method successfully produced skillful ensembles with performance metrics similar to the local operational (PROGEA) deterministic forecasts, proving the potential for the implementation of this method and its use for TB-MPC.

Future work should mainly focus on tackling other practical aspects for the real-world implementation of TB-MPC, aiming particularly at reducing its computational time. TB-MPC has a significantly higher computational time than a standard MPC, due to a higher number of optimization variables, and this requirement would grow even more with larger ensemble sizes and longer lead times, that would probably require more refined optimization routines or to develop efficient ways to distribute computational time. Further work should also assess the link between ensemble forecast skill at different resolutions and the performance of TB-MPC over a larger set of extreme events and other case studies.

6. Acknowledgements

I would like to thank Prof. Andrea Castelletti for giving me the opportunity to work on such an interesting project. I would also like to thank Dr. Andrea Ficchí for guiding me throughout the entire work and helping me fill up my lacks of knowledge in this field of study. I would also like to thank Raffaele Cestari and Matteo Giuliani for their useful advice. Special thanks go also to my family, and to my friends Eric Davide and Marco for being there when I needed it.

References

[1] Castelletti, Pianosi, and Soncini-Sessa. Receding horizon control for water resources management. *Journal of Hydrology*, 204:621–631, 2008.

[2] Denaro, Anghileri, Giuliani, and Castelletti. Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data. *Journal of Hydrology*, 103:51–63, 2017.

[3] Ficchí, Raso, Dorchies, Pianosi, Malaterre, van Overloop, and Jay-Allemande. Optimal operation of the multireservoir system in the seine river basin using deterministic and ensemble forecasts. *Journal of Water Resources Planning and Management*, 142(1):1–12, 2016b.

[4] Raso, van de Giesen, Stive, Schwanenberg, and van Overloop. Tree structure generation from ensemble forecasts for real time control. *Hydrologycal Processes*, 139(27):75–82, 2012.

[5] Scher and Messori. Ensemble methods for neural network-based weather forecasts. *Journal of Advances in Modeling Earth Systems*, 13(2):1–17, 2021.