POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

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

Tesi di Laurea Magistrale in
Automation and Control Engineering
- Ingegneria dell'Automazione

Author: **Federico Staffa**

Student ID: 945700
Advisor: Prof. Andrea Francesco Castelletti
Co-advisors: Dr. Andrea Ficchí
Academic Year: 2021-2022

Abstract

Lake Como, located in northern Italy, is an important regulated water reservoir supporting a wide range of human activities, and is regulated to satisfy two primary competing objectives: (i) providing water supply to downstream users, mainly farmers and hydropower plants, and (ii) preventing flooding along the lake shores, especially in the city of Como. This work is aimed at assessing the performance of different on-line control approaches for the management of such an important water resource, using both deterministic and probabilistic hydrological forecasts that are increasingly available with better and better accuracy. These are deterministic Model Predictive Control (MPC) and the relatively new stochastic modification called Tree Based MPC (TB-MPC), which has attracted growing interest for its increased robustness and flexibility shown in few previous applications. Their performance is compared to the historical management and to two benchmarks known as Deterministic Dynamic Programming (DDP) and Stochastic Dynamic Programming (SDP), both off-line management approaches. Both MPC and TB-MPC make use of short-term hydrological predictions of the lake inflow: while the MPC uses only deterministic forecasts (here from an operational product), the TB-MPC exploits probabilistic information from an Ensemble Forecast (EF), a set of multiple deterministic predictions that inherently represents the forecast uncertainty. TB-MPC could be a key step towards improving water management to help contrasting the rise of climate-related crises, providing a more adaptive control framework facing uncertainties. However, EFs are not always operationally and readily available at the local scale for different reasons (e.g. high computational cost, lack of local calibration, etc.). So this study has made one of the first attempts in the literature to feed TB-MPC with synthetic EFs from a recently-developed machine learning algorithm that could be used operationally. The simulation results suggest that the lake regulation could be improved by using skilful forecasts in a real-time control scheme. Given the current levels of accuracy of the available forecasts and generated EF, the daily stochastic TB-MPC appeared to be the best on-line control scheme for Lake Como, allowing to use forecasts while considering their uncertainty and consistency.

Keywords: hydrology, ensemble forecasts, optimization, predictive control

Sommario

Il Lago di Como, situato nel nord Italia, é una delle riserve idriche piú importanti del paese. Supporta una grande varietá di attivitá antropiche, e viene regolato per soddisfare due obiettivi conflittuali molto importanti: (i) soddisfare il fabbisogno idrico a valle dell'estuario, e allo stesso tempo (ii) prevenire allagamenti lungo le banchine del lago, specialmente nella cittá di Como. Questo lavoro é diretto alla valutazione delle performance di differenti approcci di controllo per la gestione di questa importante risorsa, utilizzando previsioni idrologiche sia deterministiche che probabilistiche ("ensemble"), le quali sono sempre più usate ed accurate. I metodi di controllo oggetto di studio vanno dal "Model Predictive Control" (MPC) deterministico, alla sua relativamente nuova modifica chiamata "Tree Based Model Predictive Control" (TB-MPC), un approccio stocastico che sta ricevendo crescente interesse dalla comunitá scientifica grazie alla sua maggiore robustezza e adattibilitá di fronte all'incertezza delle previsioni. Questi controllori predittivi verranno confrontati alla gestione storica e a due approcci off-line di riferimento, chiamati "Stochastic Dynamic Programming" (SDP) e "Deterministic Dynamic Programming" (DDP). MPC e TB-MPC usano previsioni a breve termine dell'afflusso, in particolare MPC usa previsioni deterministiche e TB-MPC probabilistiche sotto forma di "Ensemble forecast" (EF), un insieme di traiettorie deterministiche che rappresenta l'incertezza delle stesse. TB-MPC potrebbe essere un metodo chiave verso il miglioramento della gestione di risorse idriche affrontando meglio le incertezze grazie alla sua natura adattativa. Tuttavia, questi EF non sono sempre disponibili per diversi motivi (per esempio il loro costo, il grande bias, etc.). Questo studio esegue quindi uno dei primi tentativi nella letteratura di sperimentare l'implementazione di TB-MPC con EF sintetici generati con un metodo basato su "machine learning" implementabile operativamente. I risultati delle simulazioni suggeriscono che la regolazione del lago può essere migliorata usando previsioni di qualità in uno schema di controllo in tempo reale. Data l'accuratezza delle previsioni disponibili e di quelle d'ensemble generate, il TB-MPC con frequenza di controllo gionarliera risulta essere il migliore schema per la gestione del Lago di Como tra (TB-)MPC orario/giornaliero, permettendo di considerare l'incertezza delle previsioni nell'ottimizzazione.

Parole chiave: Idrologia, previsioni d'ensemble, ottimizzazione, controllo predittivo

Contents

1 | Introduction

Surface water resources like rivers, canals and lakes or reservoirs, both natural and man-made, play a key role in the supply of water for the population in the surrounding regions, as well as for agricultural, industrial and energy production uses. Water supply is essential for many different human interconnected sectors in the Water-Energy-Food nexus, so water scarcity that is expected to increase in the future with climate change can lead to significant food and energy security problems due to intensified water supply stress (Zhang et al. 2018). Moreover, these surface water resources support directly and indirectly countless local human activities, as for example fishing, livestock sustainment, recreation, tourism, manufacturing, and navigation as well as the ecosystems and wildlife inhabiting it. These resources are particularly invaluable not only for their usefulness in a wide array of sectors but also because they are sparsely distributed around the world and generally cannot be created artificially wherever they may be needed, given the constraint of surface water availability. Another key characteristic is that this surface water availability is also highly variable in time with water levels that change according to the seasons and climate. Therefore, it is critical to manage and use the available water resources to their full potential, also facing and anticipating possible future changes.

1.1. Context: Forecasts value for regulated lakes

Water reservoir management is a particularly challenging task, considering for how many human processes lakes and reservoirs are used, many of these characterized by conflicting interests and very different intrinsic requirements. In this context, Integrated Water Resources Management (IWRM) is the unanimously recognized key reference paradigm used to meet society's long-term needs for water resources while ensuring essential ecological services and economic benefits, balancing human water needs and ecological requirements. This systemic approach has frequently been subject of debate and has no unambiguous definition, hence why this is a set of practices that must be autonomously developed by regional authorities and national institutions depending on the particular case at hand. Nevertheless, it is increasingly being acknowledged as the standard instrument to re-

place the traditional, sector-specific approach to water resource management that has led to poor services, fragmented and unsustainable resource use (Castelletti et al. 2008a). An essential tool for effective water resources management is hydrological forecasting, which can be used with different forecast-based optimal control techniques to increase the efficiency of water system operations, also for multi-objective problem formulations, maximizing performances with respect to all involved stakeholders. This integration of forecasts in the control strategy makes the management *proactive*, as before the realization of the disturbance (e.g. rainfall or inflows) the controller sets the system to a state which is optimal to accommodate said expected disturbance. An example would be lowering the water level of a regulated lake before an expected storm event, in order to keep a buffer of storage available to attenuate flooding along the lake shores.

Giuliani et al. (2021) highlighted the extensive research efforts on more and more advanced methods for optimal reservoir operations of the last decades. Stochastic Dynamic Programming (SDP) has traditionally been regarded as the best paradigm for reservoir control since the '60. SDP consists in an off-line management paradigm formulating a Markov Decision Process (MDP). A MDP formulation has been found to be particularly suitable for modeling and operating reservoirs, as it can capture both non-linear system dynamics and objective functions, requiring only discrete domains of state, decision, and disturbance variables, along with the time separability of objective functions and constraints. However, this approach has several limitations and problems that prevent its real world application (see Section 4.1). One of the most promising options for overcoming these problems is Model Predictive Control (MPC), a member of the class of optimal real-time control methods, that can use exogenous information including predicted inflows to the reservoir. Despite the extensive research efforts and the increasingly available open-source tools, recent surveys (Pianosi et al. 2020) found a low adoption of these methods in practice (Dobson et al. 2019). Reservoir operators often refuse to use available optimization models and forecasts to inform their actual real-time operations and prefer simpler tools, such as rule curves (Loucks & Beek 2017).

1.2. Literature review

1.2.1. The value of real-time information

In the last two decades, there have been several studies in the literature proving the effectiveness and the value of hydrological forecasts for water management and reservoirs control, especially in on-line control schemes. On-line control approaches solve the pol-

icy design problem with a computational effort distributed in time, with a new control problem optimization formulated and solved over a finite horizon, potentially just before each time step of a new control implementation (Castelletti et al. 2008b). This allows exploiting newly available information and possibly deterministic or stochastic forecasts of disturbances, providing overall more flexible and adaptive control actions. On the other hand, off-line control approaches compute the sequence of optimal control laws over a given planning horizon with a unique execution prior to the actual management. These off-line methods are still very frequently used for reservoir management but they generally lack flexibility and suffer of several computational limitations, as they might compute control actions for states that never occur in reality.

Castelletti et al. (2008a) highlighted this potential short-term improvement that can be brought by on-line control optimization techniques with respect to off-line control methods by exploiting newly available information in real-time, such as inflow predictions, information about a power plant temporarily out of service, etc. For this reason, on-line methods are expected to lead to an effective technical implementation of IWRM through the better adaptation of control actions to the actual conditions the system is currently facing. In particular, they focused on the case study of Lake Como in northern Italy, a multi-purpose regulated lake, and investigated the possible advantages of refining the off-line management policy computed in a previous planning process through an on-line extension based on a receding horizon control scheme combined with a stochastic inflow predictor. The objective of this on-line extension is to better cope with extreme events, particularly those occurring in unusual periods of the year. Castelletti et al. (2008a) considered the Partial Open Loop Feedback Control (POLFC) framework for on-line control, and SDP as off-line method and they concluded that through the use of real-time information the former POLFC scheme can indeed improve SDP based policies, designed with only a-priori available information instead. Their research also revealed that longer prediction horizons can lead to an increase in performance, but this increase is less than proportional to the increase in length of the prediction horizon, and can be achieved only when the predictor is accurate. It is important to note that predictors have an inherent cost increasing with the increase of prediction horizon, so the trade-off between these costs and related improvement has to be thoroughly and carefully evaluated in each context.

Other authors followed a similar narrative of a comparative analysis between off-line and on-line control, such as Galelli et al. (2014) and Pianosi & Soncini-Sessa (2009). They also brought similar findings and conclusions, discussing some additional facets depending on their particular case studies.

Galelli et al. (2014) focused on the Marina reservoir in Singapore, a recently constructed reservoir with an highly urbanized catchment. The very fast and extreme storm events pose a significant control challenge given also the additional peculiarity of the interplay of this multi-purpose system with the tidal cycle. This reservoir has been built primarily to mitigate those frequent high peak inflows while also turning this storm water runoff in drinking water for the inhabitants, which previously had to be outsourced. This is complicated by the fact that the Marina of Singapore is effectively a tidal barrier on the external sea, so the discharge of excessive stored water has to ideally be done with low tides, minimizing its discharge during high tides as this operation would need a series of pumps that consume a great amount of power. Conclusions highlighted how, even in such a complicated setting, the MPC on-line management not only proved to be employable for drinking water supply but also outperformed the off-line control (SDP). Moreover, MPC provided a better compromise between flood control, pumps usage, and drinking water supply objectives. Longer control horizons seemingly increased that behaviour, but only up until the time constant of the reservoir (in the order of about 10 h). This improvement is based on the fact that the off-line scheme can only react to inflow events, while the on-line scheme can *anticipate* them thanks to the real-time predictions. Additionally, any event occurring further than the time constant of the reservoir does not have to be anticipated but can be resolved during the next low tide. Following work proved MPC to be perfectly capable of managing the same reservoir Marina with the additional burden of minimizing water salinity levels along the other previously considered objectives (Galelli et al. 2015).

Pianosi & Soncini-Sessa (2009) focused instead on another multi-purpose reservoir, Lake Verbano in northern Italy, on which they analysed in more detail the well known limitations of the off-line approaches. Firstly, SDP requires a discretization of all system variables, thus providing an approximate solution, but most notably suffers from mainly two bigger limitations known as "curse of modeling" and "curse of dimensionality". The first one represents the necessity for every component of the system that is being treated to be explicitly modeled to be taken into account, this forced description of every dynamic involved directly worsen the second one, which is already the most problematic between them. The curse of dimensionality implies that the computing time increases *exponentially* with the dimension of the state, decision and system disturbance. This may even cause some problems to become intractable with SDP, or only with strong simplifications to the model of the system leading to on-line methods, like POLFC and MPC, becoming the only viable approach, such as in this particular case study. Similarly to previous work, it was found that the on-line policy brings noticeable improvements thanks to the antici-

patory action brought by the run-time inflow prediction. An additional counter intuitive finding was that greater prediction ability does not necessarily imply greater effectiveness in management, as results with an heteroscedastic (LOG-ARMAX) forecast model outperformed those with the perfect forecast model.

Finally, both Galelli et al. (2014) and Pianosi & Soncini-Sessa (2009) emphasized how in addition to inflow forecasts a reliable rainfall prediction could definitely help push performances of on-line management even further, to the point of being almost comparable to the one obtained with *perfect* inflow predictions.

Table 1.1: State of the art review table, summarizing the main studies found in the literature that analyse some advanced applications of on-line control approaches.

Carachteristics	Castelletti et al 2008	Pianosi et al 2009	Galelli et al 2014	Galelli et al 2015	Ficchi' et al 2016	Raso et al 2014	Tian et al 2017	Tian et al 2019
Approach type:								
Denomination	POLFC	POLFC	POLFC	POLFC	TBMPC	TBMPC MMPC	MSMPC ACR	GAMPC
Benchmark	SDP	MPC-PF	SDP		MPC-PF	MPC-PF	MPC	MSMPC
Objectives	2	2	3	4	4	2	2	1,2
Lin./Non Lin.	NL	NL	NL	NL	NL	NL	NL	NL
Forecasts:								
Variable	Inflow	Inflow	Inflow,Tide	Inflow,Tide	Inflow,Rain	Inflow	Inflow	Inflow
Type	Stoch.	Stoch.	Stoch.	Det.	EF (6)	EF (15)	EF (20)	EF (20)
Operator		$E[\cdot]$	$E[\cdot]$		Tree	Tree	W. sum	Score m.
Others:								
Prediction h	1,2,4d	1h	1h	3h	9d	15d	24,72,144h	6h
Control h	1,2,4d	1h	1h	3h	9d	15d	24,72,144h	6h
Simulation h	50d	4y	1y	3y	15y	15M	1y	60d
Decision's f	Daily	Daily	Hourly	Daily	Daily	Every 6h	Hourly	Hourly
Info. update f	Daily	Hourly	Hourly	Hourly	Daily	Every 6h	Every 6,24h	Hourly
Reservoirs #	1	1	1	1	4	1	1	1

NOTE: h stands for horizon, f for frequency, all the other acronyms can be found in the acronyms section

1.2.2. Ensemble forecasts and stochastic approaches

In more recent years, there has been some work pushing further innovation in the context of forecast use in on-line control schemes with relatively new and more advanced implementations. These make use of the so-called Ensemble Forecasts (EF), which have been increasing in availability and skill over the last few decades, and are the cause of a gradual shift from deterministic to stochastic operations (Thielen et al. (2008), Chang & Shenglian (2020) and Buizza & Leutbecher (2015)).

Over the last 25-30 years, weather forecasting has undergone a radical change, from a deterministic forecasting procedure where at each initial time of forecast issue (or update), a single prediction is made "from a best-guess set of initial conditions using a best-guess

deterministic computational representation" of the underlying physical equations, to a probabilistic one where an ensemble (i.e. a set) of forecasts is made from a sample of initial conditions, possibly using stochastic computational representations of the underlying physical equations Palmer (2019). EF are essentially a collection of trajectories, called "members", obtained by running a deterministic Numerical Weather Prediction (NWP) model multiple times with slightly different initialization or numerical atmosphere representation schemes. This is based on the chaotic nature of the meteorological system: in a chaotic system, like the atmosphere and climate, the known "butterfly effect" leads to exponential divergence of trajectories in time for small perturbations of the initial state of the system. In order to create reliable forecast systems for decision making, methods are needed to determine when the butterfly effect will compromise the accuracy of the forecast or when the uncertainty is too large to use the forecast information to make decisions. The process for defining and constructing the previously mentioned initial perturbations is not uniquely defined, and is a critical ingredient in generating ensembles that do not suffer from sampling errors (Palmer et al. 1997), as those are the components that contribute the most to the uncertainty of deterministic forecasts (Gneiting & Raftery 2005). The most commonly used ensemble predictions generally come from the Ensemble Prediction System (EPS) of the European Center for Medium-Range Eather Forecasts (ECMWF), and range from 5 to 100 members.

Ensemble Forecasts traditionally refer to meteorological predictions, but are also increasingly used for derived predictions, such as hydrological ones (e.g. streamflow). A visual example of a hydrological ensemble forecast is provided in **Figure 1.1**. In this study, with ensemble forecasts we will refer to hydrological forecasts only, which are sometimes called Ensemble Streamflow Predictions (ESP) (Raso et al. 2014). In the context of control problems, EF allow to inherently take into account into the management problem the uncertainty related to the predictions themselves, something that cannot be done with single deterministic forecasts. The possibility of taking into account uncertainty via ensembles is a powerful benefit in the search of an optimal control policy, but to create these EFs are required powerful computers able to run computationally demanding NWP at high resolution and a large number of times (for each ensemble member), and a high computational time nevertheless. Additionally, to maximise their potential, EFs require a considerable amount of statistical post processing to address model bias (e.g. Arnal et al. (2018); Wetterhall & Giuseppe (2018)), insufficient uncertainty representation and different spatial scales of model grid-boxes (Gneiting & Raftery 2005). All these additional requirements are also computationally intensive, and only few large international centres in the world, like ECMWF, can run EPS for medium-range or longer lead times with a

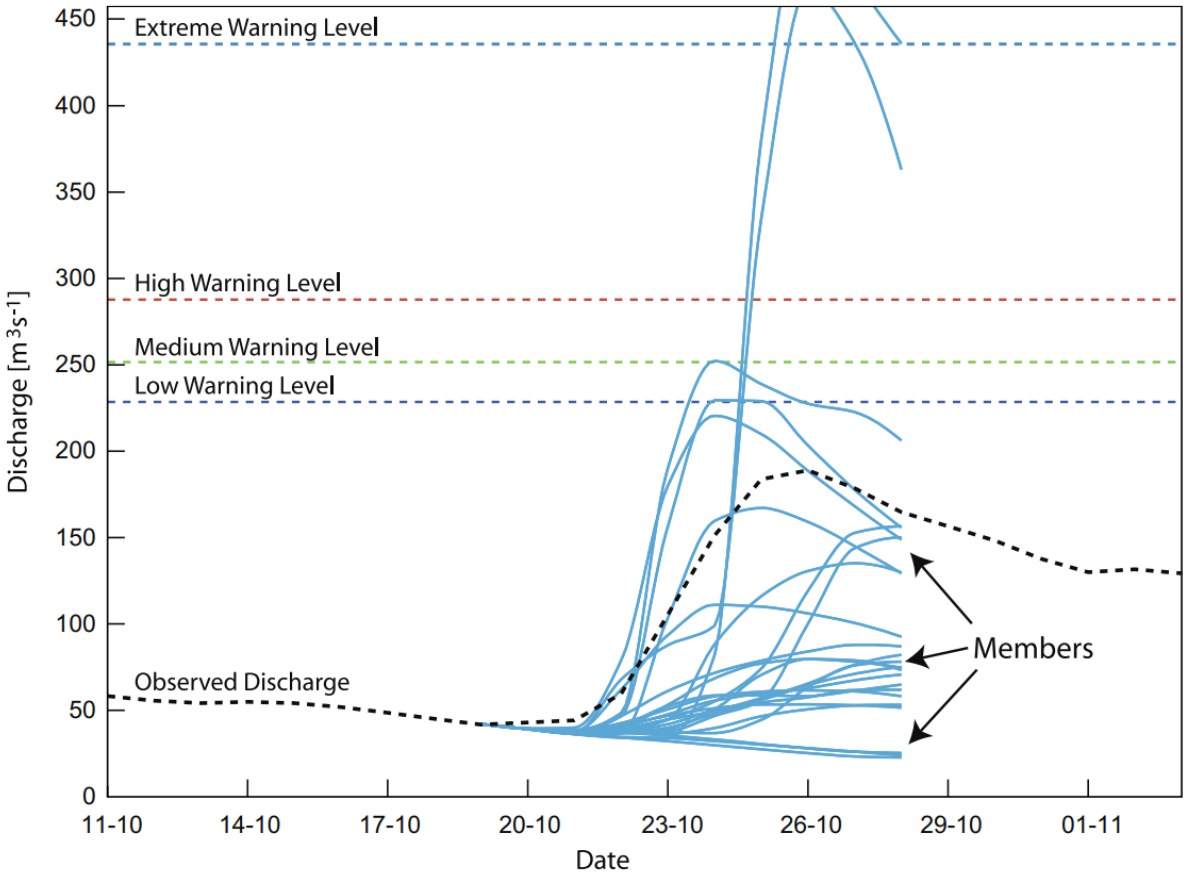

Figure 1.1: An example of ensemble forecast hydrograph for a hindcasted flooding event (Pappenberger et al. 2015). The black dotted line refers to the real observed inflow, while the blue solid lines represent the different forecast inflow trajectories, and the horizontal dashed lines represent four flood discharge warning levels. Example from the October 2007 Flood in Romania on the river Jiu as modelled by the European Flood Awareness System (EFAS), Pappenberger et al. (2015)

physical model at high resolution.

For the last several decades, weather ensemble forecasting has been dominated by NWP models, but because of the computational requirements and other reasons listed above, in recent years there has been a growing interest in generating ensembles through other means. For example, data driven methods based on Artificial Neural Networks (ANN), which were previously implemented only for post processing activities on EFs (Rasp & Lerch 2018) or for assigning a confidence measure to forecast skills (Scher & Messori 2018), have been tested more recently to generate ensemble forecasts from deterministic ones. This is an "output-oriented" method that neglects all the real-world assumptions

and physical laws dictating how meteorological variables or stream flow evolve in time and aim at just reproducing the data looking at their historical behaviour (Scher & Messori 2021). These prediction tools are effective for synthetic short-term ensemble forecasts only though, while in the medium range (between 3 to 14 days) they are currently still outperformed by NWP models. Despite that, they provide a wide array of benefits, this is a very flexible approach that yielded satisfactory results in terms of skills of the produced ensembles, and does not require high performance machines nor long run times, effectively making it available to the average end user.

Only a few studies in literature have used real ensemble forecasts in real-world applications, relying on different stochastic on-line control methods all derived from the original MPC. The claim that EFs contain very valuable information for optimal control of water systems is further reinforced by Raso et al. (2014), a study that compared MPC against TB-MPC and another stochastic modification called Multiple MPC (MMPC). While MMPC deploys the entire original ensemble in a pure stochastic optimization of a series of control actions, TB-MPC differentiates itself by transforming the ensemble into a tree structure and optimizing a tree of controls instead. The key difference between the two is that TB-MPC does contemplate the possibility of the devised strategy to be adjusted in the next decision instants thanks to the information on which member of the EF is actually occurring or, while MMPC does not. Thus MMPC is a *non-adaptive* approach often providing an over-conservative policy, with the high risk of producing an infeasible solution in presence of state constraints as well (uncertainty is overestimated). Expectedly, it was found that TB-MPC outperforms once again MPC and also its counterpart MMPC. These conclusions suggest that, while still requiring more computational effort, by working on a tree structure TB-MPC effectively makes better use of the information contained in the EF. However, authors have acknowledged that the involved performance statistics were based off a few events, due to the shortness of available data (see **Table 1.1** for reference), these findings should be further tested on longer simulations and with possibly with longer control horizons.

Another example of application of this innovative stochastic approach was brought by Ficchí et al. (2016b), who discussed the advantages of TB-MPC over the standard deterministic MPC in a real-world application in the Seine River Basin (France) over a longer horizon of multiple years, see again **Table 1.1**. Their conclusions show that TB-MPC provides results almost as good as those obtained with standard deterministic MPC using perfect forecasts (MPC-PF), although computational time grows about 7 times greater, remaining however feasible for the needs of real-world decision makers even for complex

multi-purpose multi-reservoir systems. Authors suggest that this approach could become more and more valuable with the projected increase of climate change's impact and could also be extended to account for the uncertainties characterizing the hydrological-hydraulic model. Most notably, both studies (Ficchí et al. (2016b) & Raso et al. (2014)) underline the importance of having an EF correctly representing the relevant uncertainties affecting the hydrological system, as the powerful adaptivity of TB-MPC is a direct consequence of that, and the approach could quickly lose ground if that requirement was not to be met.

A later interesting study tackled the issue arose in previous research of low computational efficiency of these advanced stochastic MPC approaches, making use of EF such as TB-MPC MMPC and Multi Scenario MPC (MSMPC). Tian et al. (2017) proposed the Adaptive Control Resolution (ACR) scheme to reduce the number of control variables involved in these management approaches, which in brief consists of dividing the control horizon into 3 distinct phases:

- **Phase I:** Near future, forecast expected to be very accurate, control resolution is the finest possible
- **Phase II:** Moderate future, forecast accuracy is presumably lower, so the control resolution is decreased
- **Phase III:** Distant future, characterized by the lowest forecast accuracy, control resolution greatly decreased

This idea that the control action is significantly more important in the nearest future thanks to the maximum forecasting accuracy is simple and intuitive, yet proved to be very effective in increasing computational efficiency, alleviating previous concerns regarding future practical implementation of these powerful approaches for complicated large scale systems. This could definitely come in handy with the Genetic Algorithm (GA) modifications of MPC derived approaches recently proposed in Tian et al. (2019), which still use EF to represent uncertainty in their frameworks. GAs (MOEAs) are algorithms architecturally built to solve multiple conflicting objectives problems with their greedy and "brute-force" nature leading to high computational demands. The study highlights how promising they could be in bettering management objectives when deployed to solve multi-scenario optimization problems.

1.3. Objective of the thesis

From this extensive literature review, it can be argued that it is well recognized that hydro-meteorological forecasts are an invaluable tool in the water management context. Forecasts can be directly used in combination with on-line control schemes, but also to provide short-term improvements to policies designed off-line and lead to more robust and adaptive policies. However, even if a few studies showed the value of real forecasts in real-world applications of reservoir control, there have been too few studies addressing the value of ensemble forecasts which are increasingly recognised as the standard technique in hydrometeorological forecasting. Moreover, as suggested by the review of Giuliani et al. (2021) an unresolved challenge for all optimal reservoir management techniques is their limited uptake by practitioners. Thus, there is an urgent need of showing the value of available *operational* forecasts in real-world case studies, to show the advantages of investing in the re-operation of existing infrastructure to optimise their benefits and bridge the gap between science and policy-making. An exemplary very well-studied case study in this context is Lake Como, in Italy. Local forecasts are available for the lake inflows, and open-source optimisation tools have been developed in previous studies, but the actual management of the system is not yet using these forecasts and tools (in a systematic and documented way, at least).

Using forecasts to better adapt to climate variability is more and more urgent nowadays as global warming is causing more and more climate extremes, with increased variability. Several areas, including the Lake Como region, have been experiencing more frequent extreme droughts or torrential rain causing large damages to human activities and assets, and making water resources management increasingly challenging. On-line control approaches, being able to exploit real-time information and forecasts, are a powerful tool that could be capable of mitigating the increased weather variability and extreme events expected with climate change.

This thesis will address these gaps, by assessing the skill of available short-term deterministic hydrological predictions for the Lake Como basin, and also by investigating in what measure they can bring benefits with respect to flood control and downstream water supply in the optimal regulation of this important regulated lake. To date, this case-study has never been investigated with such an extensive forecast analysis, and with an operative application study of a real-time controller based on real deterministic and ensemble forecasts. The authority in charge of the lake regulation has had the deterministic hydro-meteorological predictions readily accessible for some years, but has not

yet incorporated this data in an Integrated Decision Support System (IDSS) to improve the historical management of the reservoir. Major improvements could be expected by integrating these forecasts, especially for coping with the projected more frequent extreme events caused by climate change. Another gap that this thesis will try to cover is the enhancement from deterministic to stochastic on-line control, deploying ensemble forecasts generated with an innovative data driven method based on neural networks, as a computationally-effective alternative to traditional ensembles from NWP. While some studies have brought attention to this promising procedure, to our best knowledge this will be the first assessment of their usefulness in a real-world water management problem.

2 | Case Study

2.1. Case study region: the Lake Como basin

The Lake Como Basin is located in Northern Italy (Lombardia Region) and is part of the Adda River Basin, a very important basin with a key strategic role in supporting irrigated agriculture in the region. The Adda River (the fourth longest river of Italy) is the main tributary of Lake Como and its only emissary, which comes out of the lake close to the city of Lecco and flows southwards until it reaches the Po River, see **Figure 2.1**. Lake Como is the third largest lake of Italy after Lake Garda and Lake Maggiore and has a total active water storage capacity of $247 Mm^3$. The lake regulation is committed to "Consorzio dell'Adda" (www.addaconsorzio.it), an Italian regional authority controlling the lake regulation since 1946 by operating a dam located in Olginate, a province of Lecco, and providing for its maintenance.

The Lake Como basin is characterized by a topology common to many other Alpine watersheds. The upstream part is characterized by the mixed snow-rain dominated regime typical of southern Alps, with a catchment area of 4.552 Km^2. There are 16 sub-alpine reservoirs upstream of the lake, with a total storage capacity of 545 Mm^3, approximately twice as much as the main lake storage, distributed along several small to medium artificial hydropower reservoirs contributing together to roughly 12% of the national electricity demand (Denaro et al. 2017). Contrarily to the main lake (Lake Como), which is governed by a public authority, the aforementioned reservoirs are managed independently by four private companies with the sole scope of maximizing energy production. This affects considerably the seasonal inflow pattern of the lake, possibly causing increased inflows in periods with high flooding incidence, but most notably it may lead to less water volumes available during periods with high irrigation demands, a ground for potential disputes with downstream farmers. One instance of this issue was during the summer droughts of 2003 and 2005 when the situation escalated to the point where the Consorzio forced power companies to release extra water to cope with said droughts and support agricultural needs. This however caused *significant economic losses* to these companies with a little benefit to the farmers, leading eventually to a lawsuit (Anghileri et al. 2013).

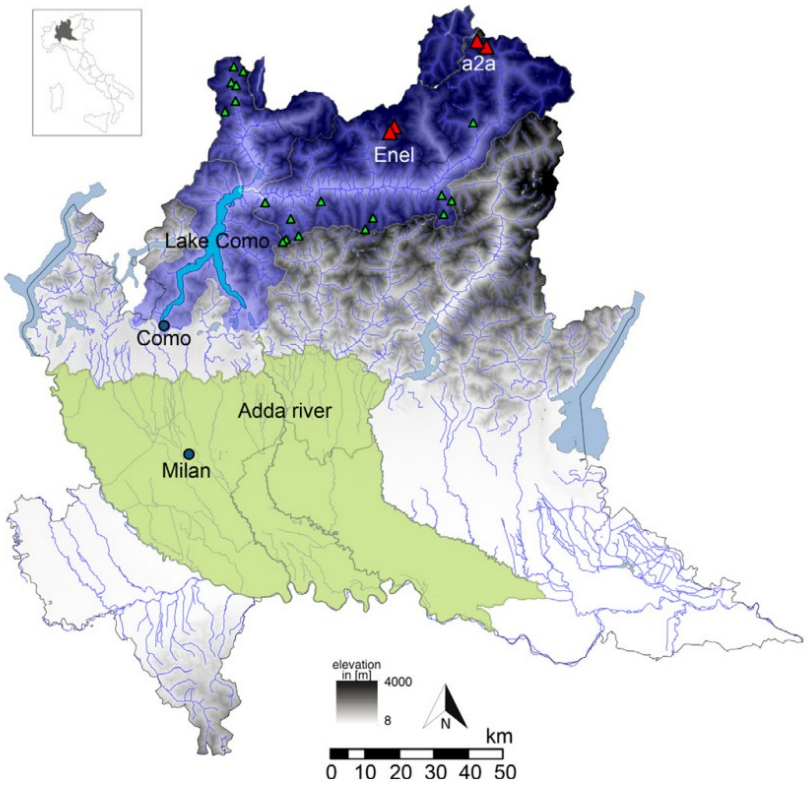

Figure 2.1: Map of the Lake Como basin, from Denaro et al. (2017), highlighting the catchment area (blue) and the downstream agricultural districts (green). The triangles denote hydropower reservoirs with the red ones being the four main ones considered in Denaro et al. (2017).

The regulation of the lake is essential to aim to satisfy two primary competing objectives: water supply, mainly for irrigation, and flood control along the lake shores. Downstream agricultural districts prefer to store water from snowmelt in early spring in the lake to satisfy the peak summer water demands, when the natural inflow is insufficient to meet the irrigation demand, but storing such water increases the lake level and flood risk (Giuliani et al. 2016).
Additional interests are related to hydropower, navigation, tourism, and ecosystems services. These multiple objectives further challenge the current water management strategy and motivate the research on this case study to look for more efficient solutions also relying on forecasts. Hydrometeorological forecasts can contribute to improving the reliability of the irrigation supply as well as to mitigating existing conflicts between competing sectors (Giuliani et al. 2016).

In conclusion, the Lake Como Basin is a very complex system with long lasting issues of conflicts between stakeholders, mainly for flood protection of lake shores, hydropower production, agriculture and food production objectives, as well as other economical, environmental and recreational issues such as navigation, tourism, ecosystems and fish conservation. For these reasons, the regulation of Lake Como can be seen as a prominent example of *multi-objective optimization problem* which is expected to grow more and more difficult to treat, given the relatively small regulation capacity of the system itself and due to a steady growth of global temperatures and a consequent expected reduction of water availability (García-Herrera et al. 2010). Climate change is expected to cause an increase in the frequency and intensity of water crises over the next years (Denaro et al. 2017).

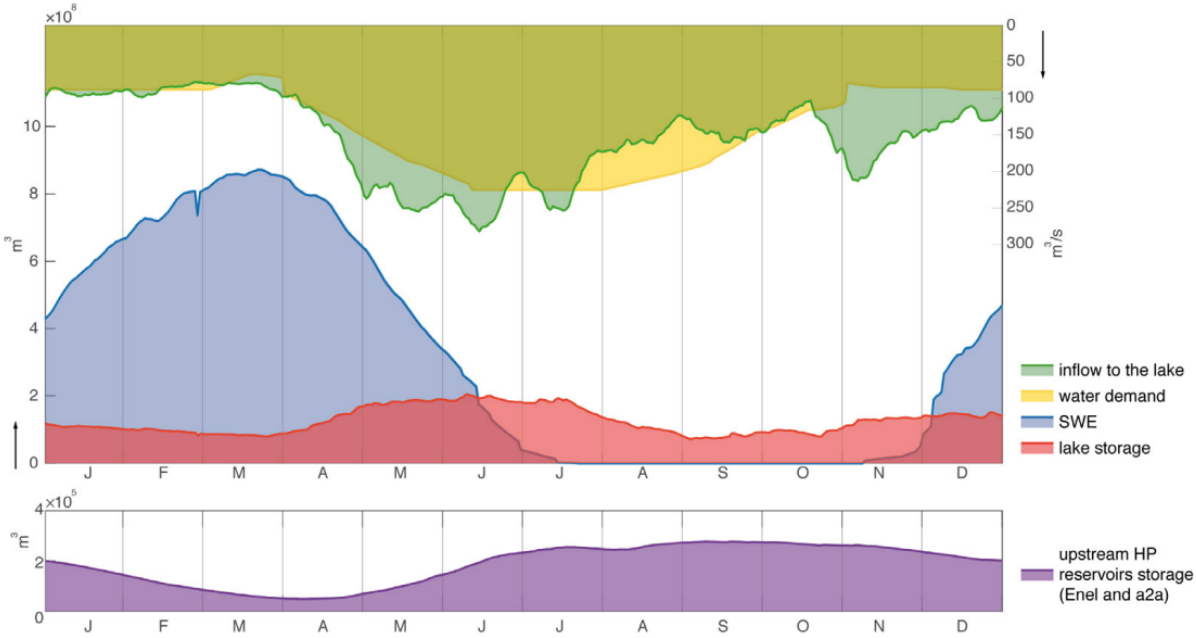

Figure 2.2: Main hydrological components of the inflow in the study area, the patterns represent moving averages computed from observed data over the period 2006 - 2013 (Denaro et al. 2017).

2.2. Climatology of the hydrological variables

The Lake Como (and Adda River) Basin is a catchment of alpine nature, meaning that the inflows to Lake Como are characterized by lower values in summer and winter, while higher values are observed in autumn and late spring. Low temperatures during winter result in snow accumulation at high altitudes; this volume of stored water is called Snow

Water Equivalent (SWE) and has a delayed contribution to the lake inflow, observed in spring when the temperature gets warmer and the snow melts.

Figure 2.2 shows that this SWE represents the main contribution to the lake's seasonal storage supporting the downstream water demand (Denaro et al. 2017). This great volume of water is thus critical for satisfying the water demand throughout the year (see yellow curve), especially during summer time, when it increases as the surge in temperatures and more sporadic precipitations challenges crop growth. However, this slow dynamic and the necessity of keeping the water stored in spring is in conflict with the faster dynamics of inflows (due to precipitations) affecting flood risk, which is more pronounced in spring and autumn. The snowmelt accumulation for long-term water scheduling directly interferes with the short-term flooding mitigation objective that requires keeping an appropriate storage pool to buffer inflow peaks caused by heavy precipitation. This is a key issue that the multi-objective management of the lake should deal with.

2.3. Available hydrological observations

Measured data of lake levels and releases is available at daily resolution over a long period (1946-present), while the same data is also available at hourly resolution for some recent years (2007-present). Hourly data of lake levels is affected by more uncertainty than daily data because of the problem of sudden oscillations in the lake levels, due to the known "seiches" phenomena, which are standing waves in an enclosed body of water such as lakes and reservoirs. In the presence of seiches, gauge readings can sometimes pick up negative or locally anomalous values that do not represent the real hourly average level. However, upon further inspection and data pre-processing, this problem turned out to have rather low incidence (1,4% of data points) with a very low number of registered negative values. Consequently, it was easily fixed with a simple linear interpolation.

Measuring directly the total distributed inflow to the lake is not feasible, as there are several small tributary rivers that drain to the lake, in addition to distributed lateral runoff and the main tributaries, i.e. Mera and Adda Rivers. Many river gauging stations would be needed as well as the modeling of complicated processes like evaporation and seepage bank loss to get an accurate direct estimate of the total inflows hence the indirect estimation of the total inflow. This is done by an inversion of the mass balance equation, based on the time series of the storage and release, leading to an efficient estimation of the *net* inflow of the lake, corresponding to the sum of all distributed inflows to the lake minus the evaporation from the water stored in the lake.

The upstream hydro-power basins are not considered in this study as there's no centralised control authority, and a lack of data of all their release time series, because they are managed independently to maximize hydropower energy production.

Figure 2.3 shows the daily cyclostationary inflows for the selected period of study, ranging from the 1st Jan, 2014 to the 1st Jan, 2022, this was chosen since it coincides with the period for which the hydrological forecasts are available.

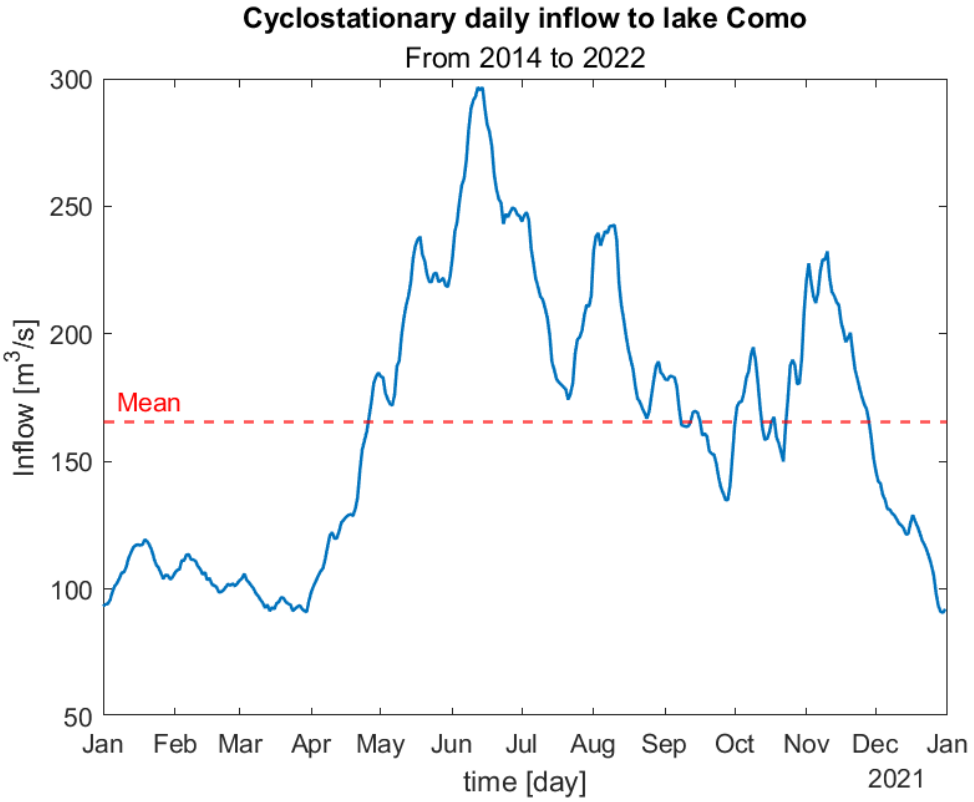

Figure 2.3: Daily cyclostationary inflows to Lake Como, computed with a moving average with a 5 days window over the study period from January 2014 to January 2022, from aggregated hourly data provided by Consorzio dell'Adda.

2.4. Available hydrological forecasts

Inflow forecasts used in this study include the real operational forecasts produced by PROGEA s.r.l. (www.progea.net), an italian company specialised in developing hydrological forecasting models and services. Forecasts for Lake Como come from their registered software called EFFORTS (European Flood Forecasting Operational Real-Time System), which is a decision support system with a GIS interface for real-time flood forecasting and hydrological monitoring. Hourly forecasts are produced through EFFORTS combining a

hydro-meteorological database containing real-time information and forecasts for the area where it is deployed (in this case the Adda River Basin) and past hydro-meteorological data, with an hydrological model called TOPKAPI. The original temporal resolution of the forecasts is 1 hour, the maximum lead time is 60 hours and the forecast update frequency is every hour. Daily forecasts were obtained from these hourly forecasts through aggregation from 8 a.m. of each day (as this is the convention for the start of the day considered in the observed hydrological data).

TOPKAPI is a physically based, fully distributed rainfall-runoff model based on the lumping of a kinematic wave assumption in the soil, on the surface and in the drainage network. It is constructed around ten main modules: interception, evotranspiration, snowmelt, infiltration, interflow and percolation, vertical recharge to groundwater table, groundwater flow, surface flow, channel flow and finally lake-routing producing three non-linear differential equations describing the catchment (Liu et al. 2018).

The meteorological inputs of the TOPKAPI model for the Lake Como catchment come from COSMO's high resolution meteorological forecasts. This is a group of meteorological services from several European Countriesd (Germany, Greece, Italy, Poland, Romania, Russia, and Switzerland) that pool their research and development resources in the field of regional NWP (Baldauf et al. 2011).
The COSMO meteorological forecasts used in EFFORTS have a spatial resolution of up to 7Km until 2018 and 5Km from late 2018 to present time. These resolutions are comparable to other studies that obtained satisfactory results, and also showed that there is not necessarily benefit from finer resolutions (Alessandrini et al. 2013).

The forecasts of hourly inflows from EFFORTS were available over the period 2013-2022 (present) with several gaps, in total they miss around 14,5% of the data, but these gaps are severe only until the year 2014 where they miss 99.9% of the data while from the year 2014 onwards just 10% is sporadically missing. The study period was thus selected from 1st Jan, 2014 to 1st Jan, 2022 to allow for a more valid reconstruction of the missing data, these gaps were filled with the value initialized at lead time 0 of the same forecast model (TOPKAPI, EFFORTS), this for all time steps when the 60-hour forecasts were missing. This procedure corresponds to using the persistence benchmark commonly used as benchmark in hydrological forecasting that is known to be more skilful than climatology as a forecast benchmark for short lead times (Pappenberger et al. 2015).

2.5. Lake regulation

The management of the lake has started with the installation of a dam in Olginate during World War II to transform the lake into a reservoir (Guariso et al. 1986), it has then been committed to the Consorzio of Adda since 21st Nov, 1938 which has historically regulated this reservoir to balance water needs between downstream utilities, farmers and run-of-the river hydroelectric plants, while preventing floods along the lake shores and preserving the ecological processes related to the lake.

2.5.1. Downstream water supply

The cultivated area fed by the estuary river Adda counts six agricultural districts, with the main crops being corn wheat and forage on an irrigated surface totaling 144'000 hectares. The downstream canals also provide water to seven hydroelectric power plants with a total capacity of 92 MW, with an annual production of around 473 GWh (Guariso et al. 1986).

The most important water derivation from the Adda river is the Muzza canal, located south-east of Milan, managed by the Consorzio di bonifica Muzza-bassa Lodigiana. This canal has a length of 60 Km and a maximum flow of 112 $\frac{m^3}{s}$, supplying 36 secondary canals in an area of 55000 hectares (Consorzio dell'Adda, 2013). The second most important is canal Vacchelli, situated in the province of Cremona, managed by Consorzio Irrigazioni Cremonesi, and has a total length of 34 km and a maximum flow of $\frac{m^3}{s}$, supplying an area of 58864 hectars. The other canals fed by Adda river are: canal Retorto, canal Vailata, canal Martesana, canal Bergamasco and lastly canal Rivoltana.

The nominal irrigation water demand to be satisfied is defined as the aggregation of the historical water rights of all the downstream water users. These water rights, which were originally established in 1942 and only marginally modified through the years, represent the irrigation water requirements under *normal* conditions and do not account for the type of crops cultivated or the meteorological conditions actually occurring in a specific year(Giuliani et al. 2016). Favorable hydrological conditions throughout history allowed for agricultural sustainment without any particular effort even with such a simplistic definition , but due to the projected increase in water crisis (Lehner et al. 2006) call for a more dynamic and flexible management, exploiting hydrological predictions.

2.5.2. Environmental flow

An additional component to the water demand for irrigation is what is known as Minimum Enviromental Flow (MEF), defined as the minimum downstream water flow needed to preserve the ecological functions of the river and associated aquifer.

In the Lake Como Basin, the MEF was defined in 2006 by a regional plan and is updated regularly. In the last update, the regional authority decreed that a daily water release of 22 $\frac{m^3}{s}$ is required to maintain an healthy river ecosystem for the Adda. Although important, this requirement has been the subject of heavy debate during the years. As per March 2022 (at present), the water crisis in the basin is at an all time high with water inflows down 60% with respect to the historical mean. The Consorzio dell'Adda is currently requesting the regional authority to half the MEF to cope with these unprecedented droughts, as reported in official documents (Official bulletin of Regione Lombardia, 22nd Apr, 2022).

2.5.3. Flood control

Flood management is one of the most important aspects in the regulation of Lake Como because of the great infrastructural damages caused by the floods on the lake shores, particularly in the city of Como, the lowest point on the lake shoreline (Giuliani et al. 2019). In recent years, the problem of the operation of the dam is becoming increasingly critical, as the risk of flood during periods of high inflows increased. This risk increase has been driven by the subsidence of the main square of Como, Piazza Cavour, and the surrounding area that have been progressively sinking since the beginning of the 1960's. This is likely because of a mix between the stratigraphic characteristic and the geological setting of the basin as well as changes in the urban configuration of the city (Nappo et al. 2020) and other anthropic activities such as land reclamation and overpumping from the underground aquifer (Guariso et al. 1986). For all these reasons, the flooding threshold on the lake level has been progressively reduced from 1.5 meters to 1.1 meters, and is expected to be reduced even further with the years.

2.5.4. Integrated model

The overall system modelled includes the lake and the release from the regulation dam situated in Olginate. As it was previously mentioned, the contribution of upstream alpine reservoirs is considered only indirectly in this study within the total estimated inflows, but they are not explicitly modeled.

The lake is modeled as a discrete time daily mass balance with an hourly integration, where the contribution of evaporation is not considered directly, while the use of estimates of net inflows compensate for this. The following mass balance equation is used to model the dynamics of the lake:

$$S_{t+1} = S_t + n_{t+1} - r_{t+1} \tag{2.1}$$

Where S_t is the storage of the lake a the current time step, n_{t+1} the incoming net inflow to the lake, also considering indirectly evaporation losses, and r_{t+1} the outflow of water defined as as non linear function $f(u_t, S_t, n_{t+1})$ of the requested control action u_t and the state of the system itself. Together they give the storage at the next time step S_{t+1}. Note that the function $r_{t+1} = f(\cdot)$, while representing the dam, is not modeled with respect to detailed hydraulics but is simple saturation of the input control decision u_t depending on the state of the system. These maximum and minimum saturations allow for the maneuvering of the system within the range [-0.4:1.1] meters, outside of these boundaries the release is either forced to 0 or the maximum natural release of the estuary.

Depending on the chosen control frequency for the lake operation, the model is slightly different, with daily operations the control u is computed every day at 8 a.m. and then kept constant for the next 24 hours, while in the case of hourly operation, u is computed and applied every hour of the day, see Section 4.2.1.

2.6. Objective formulation

The mathematical formulation of the stakeholders' interests is necessary to be able to compare the behaviour of all the control alternatives under study. These interests are formulated as a set of indicators that are computed via the simulation of the system management. The indicators used here follow a standard definition of such objectives in water management problems and have also been developed and used for this case study by previous authors. For example, Anghileri et al. (2011) used these for the quantitative assessment of the climate change impacts on the key water related activities in the basin, while Denaro et al. (2017) to derive the off-line policies (DDP and SDP) that are also considered as benchmarks in this study.
The indicators used are defined in the equations below:

$$I_{flood} = \frac{1}{N_{sim}} \sum_{t=0}^{N_{sim}} g_t^{flood} \quad with \quad g_t^{flood} = \begin{cases} 1, & \text{if } x_t \geq x_{flood} \\ 0, & \text{if } x_t < x_{flood} \end{cases}$$

$$I_{wdef} = \frac{1}{N_{sim}} \sum_{t=0}^{N_{sim}} g_t^{wdef} \quad with \quad g_t^{wdef} = \begin{cases} (MEF + w - u_\tau)^{2-r_w}, & \text{if } MEF + w \geq u_\tau \\ 0, & \text{if } MEF + w < u_\tau \end{cases}$$

$$I_{lowlvl} = \frac{1}{N_{sim}} \sum_{t=0}^{N_{sim}} g_t^{lowlvl} \quad with \quad g_t^{lowlvl} = \begin{cases} 0, & \text{if } x_t > x_{lowlvl} \\ 1, & \text{if } x_t \leq x_{lowlvl} \end{cases}$$

The flood and low-level indicators (I_{flood} and I_{lowlvl}) are computed as a percentage of respectively flooding and low level days along the whole simulation span (N_{sim} days) to account for the incidence of such events in time. Flood occurrence is defined when the lake level (x_t) exceeds the flood level threshold (x_{flood}) equal to 1.1 m. A low-level event is defined when the lake level falls below the low-level threshold (x_{lowlvl}) equal to 0.2 m. An alternative formulation could take into account the magnitude of these events. The optimal value of both the formulations is 0, but the first formulation (temporal incidence) can be interpreted more easily than the second.

The water deficit indicator (I_{wdef}) is defined as an average squared water deficit, but the quadratic exponent is corrected with the addition of a rain weight term r_w representing the contribution of rainfall. This rain weight assumes the value of 0 from June to July and 1 all the other months, in order to accentuate the deficit of water during periods of more aridity, while being more permissive in those periods of the year characterized by more precipitation which can help compensate for the lower water supply from the lake. In the equation defining the water deficit, the water demand (w) is defined as the daily target release values (u_t) over the year, and corresponds to the ideal amount of downstream water flow to support crops growth (yellow area in **Figure 2.2**). The squared power is particularly important here as it penalizes more concentrated higher shortages instead of distributed small shortages, as the former one is much more damaging to the crops than the latter. In principle the water demand should change with the type of crop that is grown, as different plants require different watering schedules across the seasons, but in this study this detail is neglected. The MEF is the minimum environmental flow equal to 22 m3/s, which is an additional component of the downstream water demand, as previously introduced.

3 | Methods: Forecast performance and Ensemble generation

This chapter describes in detail the methods used for the evaluation of the performances of the available deterministic forecasts, as well as the procedures used to generate and evaluate synthetic Ensemble Forecasts (EFs). EFs were generated based on the historical observed data available, with two different techniques for the daily and hourly ensembles. The specific performance indicators and skill scores used to verify the synthetic ensembles are presented too. These analysis steps are then performed prior to the implementation of the real-time control schemes to assess whether or not the prediction systems available are able to produce skillful forecasts of the lake inflows, which are the core asset required to develop a Model Predictive Control framework able to generate meaningful on-line policies.

3.1. Introductory notes: Study period and time conventions

The hydrological observed data available and used in this study are time series of the hourly and daily mean inflows to the lake ($[m^3/s]$), that are available from 2007 for the hourly resolution and from 1946 for the daily one (see Section 2.3). The lake operator collects these data at both a hourly and daily sampling frequency, with the daily data starting at 8 a.m. of each day. These data can be used to inform the control decision-making for the following hours or days, respectively. Hourly deterministic forecasts from PROGEA are available over the period 2014-present. Thus, the shorter period of forecast availability that is in common with the observations available is used to define our study period (from 1st Jan, 2014 to 1st Jan, 2022) in the following simulation tests and analysis. Daily forecasts are obtained through a standard temporal aggregation in time of the hourly forecasts, i.e. the mean of the first 24 hourly forecast values issued each day at 8

a.m. gives the daily forecast value with lead time 1 day. Similarly, the average for the next 24 hours gives the daily predicted value for the lead time of 2-days ahead, and so on.

As already mentioned in Section 2.1, observed data is produced by a gauging station on the Adda River, while the forecasts are produced by a *deterministic hydrological model*, called TOPKAPI, that is fed with observed and forecast precipitation data (for initialisation and forecasting, respectively). This model is initialized on the starting conditions of each hour and the run is reiterated to generate inflow predictions up to h time steps ahead with an hourly forecast update frequency. For this particular basin, the forecasts are available with a maximum lead time of 60 hours. So the aggregated daily forecasts are generated at h=1,2,3 days ahead (where the average daily forecast value for the third day is obtained by using the average of the first 12 hours of that day).

3.2. Data processing and subsets

The skill analysis has been carried out not only over the whole forecast data set (a), but also on particular conditions of high interest or data subsets, these being the *medium-high flow condition* (b) and the *flooding condition* (c). The additional subsets can be particularly informative as it is crucial to have a skillful forecast of these important conditions, especially for the flood peaks timing and magnitude, which if missed or severely underestimated could lead to a wrong control action and heavy infrastructural damages.

Their identification procedure is done by drawing out of the whole data set those data points which are above a certain threshold defined by a percentile drawn from the whole daily time series, and are then grouped together as two new sets, on which the same performance analysis is carried out. The two thresholds are respectively:

- 80-th percentile for the medium high flow conditions (b);
- 99.5-th percentile for the flood peaks (c).

3.3. Performance metrics and skill scores for deterministic forecasts

In this study, standard performance metrics and skill scores for deterministic (and ensemble) forecasts of inflows are used to tell how accurate the forecasts are, also with respect to other simple predictors like the *(static) mean* and *cyclostationary mean*, used

as benchmarks. In this section, the performance metrics used for deterministic forecasts (or ensemble mean) are presented, while Section 3.5 will present the metrics used to evaluate EF.

Forecasts have an inherent cost in being realized or acquired, so if they wouldn't prove to be much more informative than those basic, easily and cheaply computable metrics they are confronted with, the forecasting systems needed to generate them would prove to be an inherent waste of economical resources to build or upkeep.

Note that cyclo-stationarity of the data must be assumed before attempting to compute the corresponding cyclo-stationary mean, but this is the case in hydrological systems as the Lake Como Basin where all the variables are influenced by the seasonal regime of snow-melt and rainfall (see Chapter 2).

3.3.1. Mean Square Error, Root Mean Square Error, and Mean Absolute Error

The Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are all standard and well known error metrics, and are defined as:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(Q_{sim}(t) - Q_{obs}(t))^2 \tag{3.1}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Q_{sim}(t) - Q_{obs}(t))^2} \tag{3.2}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|Q_{sim}(t) - Q_{obs}(t))| \tag{3.3}$$

where: $Q_{obs}(t)$ is the observed value at time step t, $Q_{sim}(t)$ is the corresponding forecast (or simulated) value for the same time step and T is the length of the data set (i.e. the total number of time steps in the paired observation-forecast dataset).
Due to their squared formulation, MSE and RMSE weight bigger errors more than smaller ones, and this is useful as in hydrology such errors tend to happen during the most critical conditions of high-flow or flooding. MSE measures the variance of the residuals, RMSE measures the standard deviation of the residuals and MAE measures the average of the

absolute values of the residuals.

3.3.2. Nash-Sutcliffe Efficiency

The Nash-Sutcliffe Efficiency (NSE) score is a traditional metric used in hydrology, defined as in the following equation:

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_{sim}(t) - Q_{obs}(t))^2}{\sum_{t=1}^{T}(Q_{obs}(t) - \overline{Q_{obs}})^2} \tag{3.4}$$

where: $Q_{obs}(t)$ is the observation at time t, $Q_{sim}(t)$ is the forecast at time t and $\overline{Q_{obs}}$ is the mean of the observations, i.e. the inherent benchmark, and T is the total number of time steps in the paired observation-forecast dataset.

For an unbiased model, a $NSE = 0$ indicates that the simulation has the same explanatory power of the observation's mean, respectively, a $NSE < 0$ indicates that the model is a worse predictor than the observation's mean.

The NSE score can also be modified by substituting the standard benchmark, $\overline{Q_{obs}}$, with the more refined cyclostationary mean, $\overline{Q_{cyclo}}$, without any additional computation (Knoben et al. 2019). This alternative NSE score will be referred to as NSE_{cm} and, for sake of clarity, the traditional NSE score with the static mean will be written as NSE_{sm}.

3.3.3. Kling-Gupta Efficiency

The Kling-Gupta Efficiency (KGE) score is based on a decomposition of the NSE score into its constitutive components, being Pearsons's correlation r, bias ratio β and variability ratio γ. Defined as in the following equation:

$$KGE' = 1 - \sqrt{(r-1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{3.5}$$

with

$$r = \frac{\sum_{t=1}^{T}(Q_{obs}(t) - \overline{Q_{obs}})(Q_{sim}(t) - \overline{Q_{sim}})}{\sqrt{\sum_{t=1}^{T}(Q_{obs}(t) - \overline{Q_{obs}})^2}\sqrt{\sum_{t=1}^{T}(Q_{sim}(t) - \overline{Q_{sim}})^2}} \tag{3.6}$$

$$\beta = \frac{\mu_{sim}}{\mu_{obs}} \tag{3.7}$$

$$\gamma = \frac{CV_{sim}}{CV_{obs}} = \frac{\sigma_{sim}\mu_{sim}}{\sigma_{obs}\mu_{obs}} \tag{3.8}$$

As it can be seen from Equation 3.8, in the computation of the variabilty ratio the adimensional Coefficient of Variation (CV) was used instead of the standard deviation σ. This is coherent with the second ('modified') proposed version of the KGE definition, ensuring that bias and variability ratios *are not cross-correlated* (Kling et al. 2012).

Considering as a benchmark the mean flow, a model performance in the range $-0.41 < KGE' \leq 1$ means that the model outperforms said benchmark (Knoben et al. 2019), with an ideal value of 1 for a perfect simulation/forecast model.

3.3.4. KGE as a skill score

The standard definition of KGE implicitly refers to the standard mean as a benchmark, although its interpretation as a skill score is less straightforward than the NSE, which is a skill score per se. Here we use also a modification of the KGE as a skill score, KGE Skill Score (KGEss), that refers to the cyclostationary mean as a benchmark, defined as by Knoben et al. (2019) and in the following equation:

$$KGEss = \frac{KGE_{sim} - KGE_{bench}}{1 - KGE_{bench}} \tag{3.9}$$

where: KGE_{sim} is the previously computed KGE score of the forecasts and KGE_{bench} is a new KGE score computed using the cyclo-stationary mean (at each time step t) instead of the simulated/forecast dataset.

This KGE_{ss} skill score is scaled exactly like the NSE score, meaning that positive values indicate a model better than the benchmark while negative values indicate that the opposite is true, which is a more intuitive threshold. However, because of its nature it should not be used by itself as a metric. With a scaled metric, the "potential for model improvement over a benchmark" always has a range of [0,1], but the information about how large this potential was in the first place is lost and must be reported separately for proper context (Knoben et al. 2019). If the benchmark is already very close to a perfect simulation/forecast, a KGE_{ss} skill score of 0.5 might indicate no real improvement in practical terms. In cases where the benchmark constitutes a poor simulation, a KGEss score of 0.5 might indicate a large improvement by using the simulation/forecast model. This issue applies to any metric that is converted to a skill score.

3.4. Advanced event-based metrics

Since one of the main focuses of this work is to use forecasts to anticipate flooding events, an additional set of more advanced skill metrics aimed at assessing the accuracy of the forecasts in terms of *flood timings* and the *magnitude of peaks* have been computed.

3.4.1. Flood peak efficiency

As a score of flood peak efficiency, we use a simple relative difference in magnitude between the forecast peak and the actual observed flood peak, defined from Ficchí et al. (2016a), as in the following equation:

$$\Delta Q_p = \frac{Q^p_{sim} - Q^p_{obs}}{Q^p_{obs}} \tag{3.10}$$

where: Q^p_{sim} is the predicted magnitude of a particular flood event peak P and $Q_{obs}p$ is the actual observed peak. The ideal value for this indicator is 0, if positive it denotes an overestimation of the event magnitude and if negative it indicates an underestimation.

3.4.2. Time to peak error

The Time to peak error represents the time delay between the forecast and the actual observed flood peak, defined from Ficchí et al. (2016a), as in the following equation:

$$\Delta t_p = t(Q^p_{sim}) - t(Q^p_{obs}) \tag{3.11}$$

where: $t(Q^p_{sim})$ is the time at which the corresponding simulated (or forecast) peak is registered and $t(Q^p_{obs}$ is the time at which the real observed peak is observed. The ideal value for this indicator is 0 again, while positive values denote a delay in the modelled event and negative values denote their anticipation from the model. Control-wise, while anticipating peaks from a few hours to 1-2 days might bring benefits to the derived policy, the real issue is predicting them too late (or, of course, even missing them).

3.4.3. Volumetric Efficiency

Originally proposed by Criss & Winston (2008), the Volumetric Efficiency (VE) represents the fraction of water properly delivered at the right time, defined as in the following equation:

$$VE = 1 - \frac{\sum_{i=1}^{N_p} |(Q_{sim}^p - Q_{obs}^p|}{\sum_{i=1}^{N_p} Q_{obs}^p} \tag{3.12}$$

In this study, the VE has been used as a unique indicator for all the N_p flood peaks events. VE is a measure of the overall goodness of the simulated streamflow timing (Ficchí et al. 2016a). VE ranges from 1, being the ideal perfect score, to 0 for the worst model behaviour, as all other skill scores.

In this study flood peaks are identified in a simpler way than the procedure reported in Ficchí et al. (2016a). Here, flood peaks are searched within an user-defined time window around the real observed peaks, and all found peak values are considered valid if they are above the 99^{th} percentile. If no peak was to be found it is considered *missed*. The size of the window was chosen to be 5 days. This seemed a reasonable margin of time to consider a certain peak as neither an excessively anticipated nor excessively delayed forecast, and to be actually useful in the forecast-based control simulation.

3.5. Ensemble Generation

This section will present the two approaches used to obtain stochastic (probabilistic) predictions. Reliable and bias-corrected hydrological Ensemble Forecasts (EF) were not immediately available prior to this study. The well-known EF from the European Flood Awareness System (EFAS), provided by ECMWF within the Copernicus EMS service, are known to have a large bias in this basin to be useful in practice without bias-correction procedure (Wetterhall & Giuseppe 2018). This happens because no data from river gauges in the basin were available to calibrate EFAS in the basin, to date (EFAS v. 4.0). Thus, the hydrological model used in EFAS (LISFLOOD) could not be calibrated in this region yet and EFAS is expected to have large biases. Moreover, even if EFAS was calibrated in the region as it might be soon, the access to real-time EF products is restricted to some partners (while older data is open access) which may limit its uptake. Also, building alternative local hydrological EFs with similar physically-based models (as PROGEA's deterministic model) may have associated monetary costs, requiring large computational

resources. For these reasons, in this study EFs will be generated, as anticipated, from the historical observations and deterministic forecasts, with a relatively new method that has been recently proposed in the literature. This method consists of a data driven synthetic EF generation through a series of Feed Forward Neural Networks (FFNN). This approach circumvents the complexity, the prohibitive computational burden and run times related to the NWP models used by large data centers, allowing operational users to have access to alternative operational synthetic ensembles in real-time, potentially effective more immediately and free-of-charge.

Both daily and hourly ensembles that will be generated will have 25 members, which is in line with the size of EFs produced by international centres like the ECMWF, for long-term reforecasts (e.g. EFAS v.4.0 reforecasts range from 11 to 25 ensemble members at extended and seasonal range respectively). Additionally, while the computational requirements decrease with a decreasing number of branches, the skill tends to decrease as well (Fan et al. 2016).

3.5.1. Daily Ensemble Forecast Generation via a Neural Network

Scher & Messori (2021) proposed four distinct methods with increasing complexity to generate EFs in a data-driven fashion:

- **Random initial perturbations:** Conceptually the simplest method, even though not necessarily the best, it consists in applying a random noise to the training data set of the neural network.
- **Singular Value Decomposition:** A technique from linear algebra with a wide set of applicability, used here to find and apply the minimum input perturbation leading to the maximum output perturbation through the network's Jacobian.
- **Network Retraining:** Training a neural network is inherently a "random" procedure, it consists in carrying out a non linear optimization with a random initialization of weights and biases as well as training and validation subsets of the input data. This leads to slight differences in the end network.
- **Random Dropout:** A known regularization technique, it consists in a random deactivation of one or more neurons and all their connections during training. This helps reducing overfitting, but it can be applied during the forecasting stage instead to diversify predictions.

This study will adopt the third proposed method (Network Retraining) with additional sources of randomness. This choice was made since results by Scher & Messori (2021) concluded that the method that retrains the network achieves the best results, while keeping a relative simplicity.

The network of choice is a classic FFNN, the most basic and reliable type of neural networks in the literature, which will be used for all the ensemble members without any modification to its architecture (**Figure 3.1**).

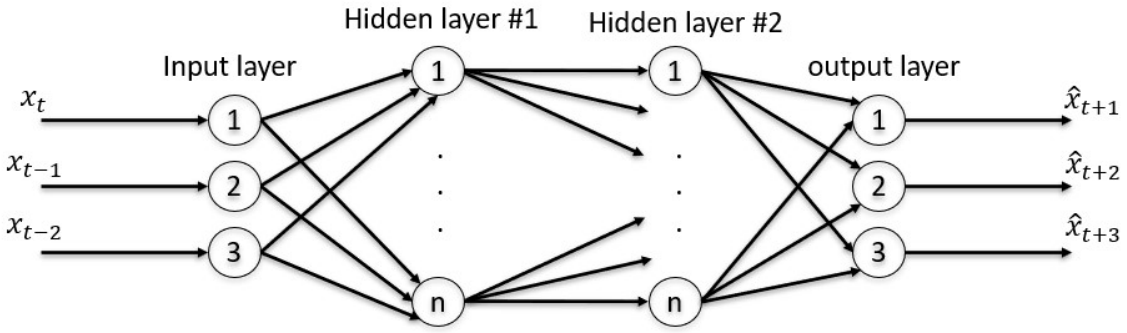

Figure 3.1: FFNN architecture used for synthetic daily ensemble forecast generation, applying the Network Retraining method by Scher & Messori (2021).

The FFNN has its input and output layers of size three. The input layer takes the current inflow value x_t and those of the two days before x_{t-1} and x_{t-2}, to estimate the future three day ahead inflow values $\tilde{x}_{t+1}$, $\tilde{x}_{t+2}$ and $\tilde{x}_{t+3}$.
The numbers of hidden layers and their respective size have been chosen with some trial and error tests and empirical rules. A choice of two hidden layers of $n = 10$ neurons yielded good prediction capabilities. Moreover, the overall training data set is the historical daily inflow data from 1946 to 2022, which has a sample size well over ten times the total number of parameters $N_{tot} = 183$ of the FFNN network ($N_{tot} = N_w + N_b$, with $N_w = 3*10+10*10+10*3 = 160$, number of weights, and $N_b = 0+10+3 = 23$, number of biases), ensuring that overfitting is avoided.

As FFNN are a very powerful fitting and predicting tool, the first results showed that a simple re-training of the network did not produce a good spread of (diversified enough) output trajectories, so at *every training cycle* one or more of these additional randomness factors were added:

1. change the seed of random number generator before every calibration phase;

2. use different partitions of the whole data set for training and validation;
3. switch the training and validation data sets;
4. add a Gaussian perturbation to the input data set;
5. randomly choose a training algorithm from those available;
6. use the available deterministic predictions (PROGEA) as the input data set, instead of the observed historical values at past time steps.

In total four EFs with varying degree of skill were generated, this throught the progressive implementation of these randomness factors, in order to asses the effect of EF performances on those of the management policy produced by TB-MPC. See later in Section 5.1.2.

3.5.2. Hourly Ensemble Forecast Generation

While it is technically possible to generate an hourly EF with the proposed data-driven method based on neural networks, our results of first tests were relatively poor (ensemble performance) and the computational time was too high to perform more tests. Theoretically, the underlying FFNN could be extended from the daily prediction (i.e. 3 daily time steps ahead) to the hourly setting (i.e. 72 or 60 hourly time steps ahead), by taking as inputs observations (or deterministic forecasts) over the previous 72 hours, i.e. $x_{t-72}, ..., x_t$. However, the hourly resolution case would lead to a neural network to be trained to reproduce $\tilde{x}_{t+1}, ..., \tilde{x}_{t+72}$ hourly data up to 72h in the future, with a new suitable number of neurons and hidden layers, which is a much more complex problem that would require more training data than what available too. Neural networks are a powerful tool but cannot be expected to skillfully predict over such a long horizon, as also noted by Scher & Messori (2021), let alone with the simple architecture adopted here. Due to the nature of the neural networks training requiring a large quantity of parallel computations, bigger networks require exponentially large memories and time to train. In addition to the poor performance, the problem gets so computationally taxing that it loses its prospected advantage over the NWP models, to the point where it loses its "user-friendliness". The method by Scher & Messori (2021) should be further adapted to work with longer hourly records (see Section 6.1).

Consequently, an alternative simpler method was used to generate an hourly synthetic EF, as proposed firstly by Nayak et al. (2018). This alternative method, originally proposed to extend the period of available re-forecasts, was used here to get short-term ensemble

forecasts starting from the historical record of deterministic forecasts paired with observations. This hourly ensemble forecast generation method is not directly operationally applicable though, being based on 'dressing' past observations with errors with given statistical properties. So, this alternative EF generation is used just for a proof-of-concept analysis, to investigate the value of moving from daily to hourly EF resolution in TB-MPC, but could not be used operationally, unlike the previous data-driven method used for daily ensembles. More precisely, this alternative stochastic generation approach is used to generate a short-term 60-hour synthetic ensemble forecast of inflows with statistical properties matching the original observed and deterministic forecast data. This ensemble re-forecast generation procedure is done through a K-Nearest Neighbor (KNN) algorithm based on a resampling of the existing forecast residuals. The approach is a novel inverted application of the common Model Output Statistics (MOS), which employs statistical postprocessing to correct forecast output based on the error distribution of past forecasts (Wilks & Hamill 2007). Some statistical properties of our original deterministic forecasts for the Lake Como basin are extracted and used to develop a synthetic hourly ensemble.

The key idea is to resample the logarithmic residuals of the existing deterministic forecasts based on the euclidean distance (d) of each hourly observed inflow at the time step of forecast issue. Within the KNN algorithm, in addition to the actual value of the observed inflows (at the current time step t), forecast residuals are resampled based also on the observed values and the residuals of the previous 3 hours (t-1, t-2, t-3), to ensure that the auto-correlation of the synthetic forecasts matches or gets close to the one of the original forecasts (Nayak et al. 2018). Then a residual is randomly selected from those resampled and a new forecast is generated at each time step by adding the re-sampled residual to the observed inflow. The procedure is repeated a number N of times to generate N ensemble members. The number of members for this ensemble forecast of hourly inflows has been fixed to 25, as in the previous method for the daily ensemble, which is in the range of typical numbers of ensemble members for operational reforecasts like those coming from the EPS of ECMWF.
Among the vector of K nearest residuals in the reforecast period, the selection of the residual to use was based on a kernel density function:

$$f(d_j) = \frac{1/j}{\sum_{j=1}^{K} 1/j} \tag{3.13}$$

$f(d_j)$ provides ordered resampled weights increasing with increasing distance d_j, K was chosen as $\sqrt{\frac{n}{n_s}} = 47$ with n the number of hours in the reforecasted period and n_s the number of 'seasons' or sub-periods, here set equal to 12 (by month).

Then, the forecast error $\hat{\epsilon}$ is resampled for each lead time (up to 60 hour), $t^*+1 : t^*+60$, from the deterministic reforecast period in this way:

$$\hat{\epsilon}_{Q,t^*+1:t^*+60} = \log \hat{Q}_{t^*+1:t^*+60} - \log Q_{t^*+1:t^*+60} \tag{3.14}$$

where: $\hat{\epsilon}_{Q,t^*+1:t^*+60}$ is the observed error between each observed value $Q_{t^*+1:t^*+60}$ and the corresponding reforecast $\hat{Q}_{t^*+1:t^*+60}$. Once $\hat{\epsilon}$ is resampled, the synthesized forecast over the 60 hours is:

$$\hat{Q}_{t+1:t+60} = Q_{t+1:t+60} e^{\hat{\epsilon}_{Q,t^*+1:t^*+60}} \tag{3.15}$$

The choice of the logarithmic error model ensures the non negativity of the newly synthesized inflow forecast (Nayak et al. 2018).

This method, although not easily employable in an operative setting as it relies on observations (over the forecast horizon), leads to ensembles with similar properties and accuracy to the ones that would be generated with operationally feasible methods, and are useful for comparisons with their daily EF counterparts.

3.6. Ensemble Performance Metrics

To asses performances of the generated ensembles a set of extensively known indicators will be computed, all of them are computed with respect to the particular h lead time.

3.6.1. Root Mean Square Error

The first introduced metric is the RMSE, easily interpretable and traceable back to its computation for deterministic forecasts showed in Section 3.3.1. It is defined as in the following equation:

$$\begin{aligned} RMSE_h &= \sqrt{\frac{1}{N}\sum_{\tau=1}^{N}(\overline{X}_{\tau,h} - y_\tau)^2}, \quad for \quad h = 1, 2, 3d \\ with \quad \overline{X}_{\tau,h} &= \frac{1}{K}\sum_{k=1}^{K} X_{\tau,h,k} \end{aligned} \tag{3.16}$$

where: N is the total number of data entries generated, K is the number of ensemble members, $\overline{X}_{\tau,h}$ the ensemble mean at time tau and at lead time h and y_τ the corresponding observation.

3.6.2. Spread

The second metric is the ensemble Spread, it is essentially the square root of the average ensemble variance. This is a more correct and informative way of estimating it in terms of exchangeability and ensemble size with respect to the square root of the standard deviation of the ensemble (underestimation) (Fortin et al. 2012). It is defined as in the following equation:

$$Spread_h \cong \frac{1}{N}\sum_{\tau=1}^{N}\left(\sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(\overline{X}_{\tau,h}-X_{\tau,h})^2}\right), \quad for \quad h=1,2,3d \tag{3.17}$$

It is widely known that a "perfect" ensemble has the mean forecast spread coinciding to the mean forecast error, respectively estimated by the Spread and RMSE formulated in this way (Scher & Messori 2021).

3.6.3. Continuous Ranked Probability Score

The CRPS is another well known performance metric for EFs, that compares the cumulative probability distribution of the ensemble forecast to an observation. It is sensitive to the mean forecast bias as well as the spread of the ensemble (Wetterhall & Giuseppe 2018). For each lead time (h), the CRPS is defined as in the following equation:

$$CRPS_h = \int (F(X_h) - H(X_h - y))^2 \, dX, \quad for \quad h=1,2,3d \tag{3.18}$$

where F(X) is the Cumulative Distribution Function (CDF) of the forecast distribution of the probabilistic forecast X (at lead time h), y is the correspondent observed variable, and H(X-y) is the Heaviside step function (unit step function, with a value of 1 for positive arguments, and 0 for negative ones). The CRPS is a generalisation of the MAE for the case of probabilistic forecasts, so is expressed in the same units as the forecast/observed variable of $[m^3/s]$ and its ideal value is zero.

4 | Methods: Optimal Control Benchmarks and Model Predictive Control

This chapter is devoted to a brief description of the control benchmarks (off-line control methods) and to a more detailed description and formal definition of the general framework of the on-line control problem. This is firstly considered as deterministic, thus implementing Model Predictive Control (MPC) using deterministic forecasts from PROGEA. Then, it will be extended to a stochastic framework that makes use of ensemble probabilistic forecasts within the Tree-based MPC (TB-MPC). The extension to a stochastic framework makes the control less vulnerable to the forecast uncertainty, especially in the presence of non linearity in the cost function or in the system to be controlled (Raso et al. 2012). The stochastic TB-MPC approach is also expected to allow for a partial compensation of performance loss caused by high-variance disturbances and poor prediction capabilities (i.e. non-skilful forecasts), while the performance of a deterministic MPC can quickly deteriorate with increasing inaccuracy of the forecasts.

Water resources have historically been managed with off-line approaches. The algorithms implemented to do this (like SDP) are mathematically guaranteed to produce an optimal solution, but they are sometimes not feasible or have limitations and high prices in a real world setting, for their computational and practical requirements (see Section 4.1). Also, the literature has shown how additional information gathered in real-time could yield a performance improvement in the short term (Castelletti et al. 2008a).

4.1. Benchmarks: DDP and SDP

To properly assess the performance of the implemented on-line control approaches under study, one or more benchmarks are required.

The benchmarks selected here come from two off-line approaches that have commonly been used in IWRM: Deterministic Dynamic Programming (DDP) and its stochastic counterpart, Stochastic Dynamic Programming (SDP). Both of them solve the operating policy design problem treating it as a sequential decision making process based on the key idea that, when choosing the *optimal* solution at time t, all the future decisions chosen from *t+1* and on-wards *will be optimal*. This is formalized by the Bellman's equation (Bellman 1957), providing the optimal *cost-to-go* function at time t (H_t^*), given the one at time t+1, through a cost-to-go g_τ and an end-state cost g_h in this way:

$$H_t^*(x_t) = \underset{p_{[t,h)}}{minimize} \underset{\tau=t+1,\dots,h-1}{\mathbb{E}} \left[\sum_{\tau=t+1}^{h-1} g_\tau(x_\tau, u_\tau, \epsilon_{\tau+1}) + g_h(x_h) \right] \tag{4.1}$$

where: ϵ_τ is a disturbance with known probability distribution in SDP (or a known deterministic trajectory in DDP), x_τ the state of the system, u_τ the decision variable at time τ, defined by the policy p adopted. The solving procedure then is performed backwards from the final stage (time instant) to the initial one. In SDP, the function H is computed for every possible one step ahead permutation of the state. SDP minimizes the total expected cost H over the considered time horizon h granting optimality of the solution (Bellman 1957).

DDP represents the ideal (but practically infeasible) best possible scheduling of release, under the condition of perfect knowledge of the disturbances (inflows) over an infinite horizon, so this benchmark is also called Perfect Operating Policies (POP). On the other hand, SDP is based on the search of an optimal sequence of control laws over the entire horizon but with degraded input knowledge, caused by the stochastic description of the disturbances. However, this degradation is to be expected in real applications, as the knowledge about future inflows is not perfect. Thus, the comparison between the two benchmarks, DDP and SDP (that will be shown in Chapter 5), is useful to prove that the knowledge about future inflows brought by skilful forecasts can indeed bring an useful contribution in water management.

4.1.1. SDP limitations

SDP has little to no requirements, being the discretization of state, decision and disturbance variables, as well as time-separability of objective functions and uncorrelated disturbances (Giuliani et al. 2016). It is theoretically the second best possible approach behind DDP, particularly suited for reservoir modeling and consequential operation, because applicable under general hypotheses that are usually satisfied (Giuliani et al. 2021). Yet, its practical implementation is limited and not feasible in most cases, because of a series of well known issues, as already mentioned in Chapter 1.
These issues are:

- **The "Curse of Dimensionality":** the computational cost grows exponentially with the dimensions of the state, control and disturbance vectors; this is because every combination of these discretized variables has to be checked at every time step during the optimization (Bellman 1957). Thus the computational time grows with the dimension of the problem as in the following equation:

$$time = N_x * N_u * N_\epsilon * h \tag{4.2}$$

  where: N_x, N_u and N_ϵ are respectively the dimension (number of grids) of the discretized state, control and disturbance, and h is the total simulation horizon.
  This makes SDP generally unsuitable for systems with more than two or three reservoirs (Castelletti et al. 2008a).

- **The "Curse of Modeling":** SDP requires any information used in the optimization to be explicitly modeled by a dynamic model or to be considered as a disturbance (Powell 2007). This directly exacerbates the Curse of Dimensionality as well.

- **The "Curse of Multiple objectives":** SDP is at all effects a single objective optimization algorithm, so if it was to be used to solve a multiple objective optimization problem it would need to repeat its whole optimization procedure for every single objective separately (Bertsekas & Tsitsiklis 1995). With every additional objective computational time grows very quickly to the point of making SDP intractable for more than two or few objectives.

4.1.2. Multi-Objective Problem: weighting method

Although an intuitive ideal choice would be a management policy that could somehow maximise all the indicators, this solution generally does not exist, in multi-objective optimization problems with conflicting objectives like our case-study. There is often a trade-off between stakeholders and objectives, which can be highlighted with the so-called Pareto Frontier that maps all the Pareto efficient solutions, where a solution is Pareto efficient if there is no alternative solution where improvements can be made to at least one objective without worsening any other objective (see, for example, Denaro et al. (2017) and Section 5.2.2 for our case study).

As done by Denaro et al. (2017), in this study we use the *weighting method* to convert the multi-objective problem (two in our case, floods and water deficit) into a single-objective one via convex combinations: the objective of the single-objective problem is obtained as a weighted sum, of the individual objectives, with the sum of weights equal to one. The exploration of the trade-off between the objectives is performed by varying the weights used in the objectives' aggregation. In our case study, the DDP and SDP were solved using multiple pairs of weights for the flood incidence and water deficit objectives as described in Section 5.2.2, while the weight for the low-level indicator was kept at zero.

4.2. Deterministic on-line framework: MPC

MPC is an advanced on-line control method which makes use of a dynamical model of the system under study to predict its future behavior and response to the control policy. Based on the optimisation of an objective function, MPC is able to provide an optimal sequence of controls over a finite horizon, by solving a potentially constrained optimization problem. At every control time step, the MPC controller takes as inputs the input w and the current state of the system y (see Figure 4.1). The prediction and optimization is run by feeding these to the system model, to provide an optimal control sequence over the finite control horizon, from which only the first action u is applied to the real system (receding horizon strategy). Then, at the next time step, the system evolves to a new state and the procedure is repeated by using that same real-time information, now updated.

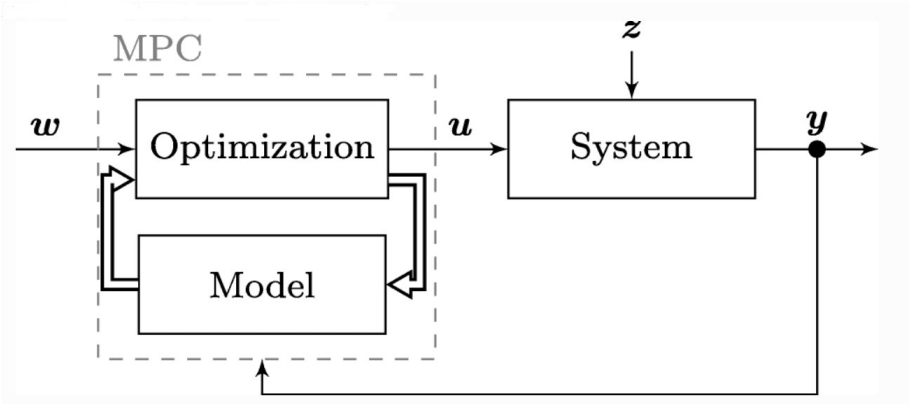

Figure 4.1: A simplified scheme of deterministic MPC, from Schwenzer et al. (2021).

4.2.1. The control problem

The MPC control problem is formalized as:

$$\underset{u_t,\dots,u_{t+H}}{minimize} \quad f_{tot} = \sum_{\tau=t}^{t+H} g_\tau(x_\tau, u_\tau) + g_\tau^{end}(x_{\tau+H}) \tag{4.3}$$

subject to

$$\begin{aligned} x_{\tau+1} &= x_\tau + \Delta_T(q_\tau - u_\tau) \\ x_{\tau=t} &= x_0 \quad given \end{aligned} \tag{4.4}$$

$$u_{min} \leq u_\tau \leq u_{max} \tag{4.5}$$

where: the system's state is represented by x, while u represents the control decision (in our case, the release of the dam); H is the control horizon of the MPC optimization (in this study, up to 3 days ahead, due to forecast availability), that coincides with the horizon of the short term inflow predictions available.

Note that the temporal resolution of the approach is not constrained a-priori in the equations, as the same scheme can be used both for **daily or hourly operations**, with a suitable choice of all the parameters (e.g. horizon H, model time step Δ_T, etc.). For example, the horizon H could be either 3 days or 72 hours and the state variables (arrays) can be scaled accordingly. The model and control time step, Δ_T can be 24*3600 or 3600 seconds. An hourly approach is expected to take more time to solve but to be more precise, as it follows better the sub-daily evolution of the real system.

4.2.2. System Model and Constraints

The system model is represented by Equation 4.4 , which in our case is defined as a simple mass balance of the first order, where $x_{\tau+1}$, the storage at time step $\tau + 1$, is equal to x_τ, the storage at the previous time step, plus the total inflow minus the released water amount over the time step. This simple mass-balance equation could take into account explicitly also the contribution of rainfall over the lake, as well as evaporation from the lake, but in this case study these were neglected for simplicity and clarity sake (however, the estimated net inflow data were used in the simulation, as described in Section 2.3).

Equation 4.12 represents the control saturation constraint, which is implemented in such a way to mimic the same saturation acting on the real world system, so that MPC searches and provides only feasible control actions along the optimization. It is implemented as a non linear constraint since the maximum and minimum releases related to the dam are non linear and a function of the state of the system.

4.2.3. The optimization problem: step-cost and penalty cost

The optimization subroutine is implemented in Matlab, specifically, with the non linear optimization toolbox's function "fmincon.m". As Equation 4.3 is a non linear problem with linear and non linear constraints, it it set to be solved with Sequential Quadratic Programming (SQP), a small-medium scale algorithm suitable to solve these problems. The optimal control sequence $u_1 ... u_H$ corresponds to the one minimizing the total cost in Equation 4.3, composed by two elements, the step-cost g_τ and the end state penalty cost g_τ^{end}.

$$g_\tau = g_\tau^{flooding} * W_1 + g_\tau^{wdeficit} * W_2 + g_\tau^{lowlevel} * W_3 \tag{4.6}$$

The step cost is a weighted aggregation of different g_τ^i components that formalize the objectives and requirements of the case study, along the particular time window t...t+H. These are chosen specifically to mimic in the best way possible the real requirements of the stakeholders, defined in Section 2.6, while being suitable for the optimization routine to handle (i.e. the flood indicator is adapted to be managed by MPC). The weights W_i are a user-defined set of numbers $\in [0, 1]$ expressing the relative importance given to the components of g_τ.

The step-cost components are defined as follows:

$$g_\tau^{flooding} = (\max_{t...t+H}(x_\tau - h_{flood}, 0))^2 \tag{4.7}$$

$g_\tau^{flooding}$ is **related to flooding condition**, measuring the squared difference between the simulated lake level (storage) x_τ and a flooding threshold h_{flood}, penalizing the solution when the trajectory of the lake level is predicted to exceed the threshold value within the time horizon H. This threshold is defined as 1.1m (see Section 2.6), as crossing this level would lead to flooding and damages in Como.

$$g_\tau^{wdeficit} = (\max_{t...t+H}(MEF + w - u_\tau, 0))^2 \tag{4.8}$$

$g_\tau^{wdeficit}$ is **related to water deficit** downstream of river Adda, and penalizes the solutions giving a release decision u_τ that does not satisfy the MEF and the water demand w along the time horizon H (see Section 2.6).

$$g_\tau^{lowlevel} = (\max_{t...t+H}(h_{low} - x_\tau, 0))^2 \tag{4.9}$$

$g_\tau^{lowlevel}$ is **related to low-level condition** and is based on a threshold, similarly to (4.7); it measures the squared difference between the simulated lake level x_τ and a threshold, now h_{low}, defined as -0.2m. This is a secondary objective related to navigation and preservation of the shores ecosystem which will be generally neglected in this study (apart from some secondary tests), by putting its respective weight W_3 to zero, because this objective

is less critical than the other two.

The penalty cost, g_τ^{end}, is a critical component of the total cost (Equations 4.3 and 4.10) for the performances of the closed loop system in the long-term, but its choice is far from immediate. It is a cost associated to possible less desirable states of the system, particularly critical for natural systems, as when the optimization ends the system must be left into a non detrimental condition. This study makes use of a penalty function derived off-line, as in (Castelletti et al. 2008a), by solving an off-line Dynamic Programming horizon problem. This problem has to be traceable back to the original one, then the optimal cost-to-go functions H_k expressing the future cost associated to each state value, at each time instant, can be used as penalty functions over the final state of the finite horizon on-line problem. In this way, the penalty cost g_τ^{end} contains crucial information from the end of the foreseeable future H to the end of the simulation. In this way, the MPC optimization has at hand the information on which state would be preferable at every time t to properly mitigate future critical events that cannot be seen in the short term horizon (3 days), based on the optimality of the devised dual off-line problem.

4.3. Stochastic on-line framework: Tree Based MPC

The stochastic version of MPC used in this study, called Tree-based Model Predictive Control (TB-MPC), was originally proposed by (Raso et al. 2014). In TB-MPC, ensemble forecasts can be used to set up a Multistage Stochastic Programming, which finds a different optimal strategy for each likely future trajectory of the predicted disturbance (represented as a branch of a tree) and enhances the adaptivity to forecast uncertainty. The underlying key idea of TB-MPC, that differentiates it from deterministic MPC, is to search for an optimal *decision tree* instead of an optimal decision sequence. In this way, TB-MPC provides a more adaptive control implementation that makes the best use of ensemble forecasts, while other stochastic approaches using them might yield less efficient or infeasible solutions due to the burden of multiple disturbance in combination with hard constraints (Uysal et al. 2018).

The approach is characterized by the possibility of changing future control actions with respect to the actual trajectory that the system takes among those predicted; this causes a prominent increase in the optimization problem dimensions influencing heavily computational complexity and increasing the required optimisation time.

The TB-MPC problem is formalized as:

$$\begin{aligned} &\underset{u_t^{M_\tau},\dots,u_{t+H}^{M_\tau}}{minimize} \quad f_{tot} = \sum_{z=1}^{Z} p(z) \left[\sum_{\tau=t}^{t+H} g_\tau(x_{\tau,z}, u_{\tau,z}^{M_\tau}) + g_\tau^{end}(x_{\tau+H,z}) \right] \\ &\textit{subject to} \end{aligned} \tag{4.10}$$

$$\begin{aligned} x_{\tau+1,z} &= x_{\tau,z} + \Delta_T(q_{\tau,z} - u_{\tau,z}^{M_\tau}) \quad \forall z \\ x_{\tau=t,z} &= x_0 \quad given \quad \forall z \end{aligned} \tag{4.11}$$

$$u_{min} \leq u_{\tau,z}^{M_\tau} \leq u_{max} \quad \forall z \tag{4.12}$$

The notation is the same as the deterministic version, with the addition of the subscript z indicating that every variable depends also on the particular ensemble member, $z \in Z$, where Z is the total number of ensemble members. The model equation and the optimization constraints representing the system model and control saturation are the same as for the deterministic MPC framework.

Every ensemble member has a total cost f_z coinciding with the one for the single deterministic optimization of MPC showed in Section 5.2.3. They are combined together into the actual total cost f_{tot} by a weighted sum with their probability of outcome $p(z)$. The simulation aims to minimize this comprehensive expected cost for every ensemble forecast by devising a set of controls $u_t^{M_\tau}, ..., u_{t+H}^{M_\tau}$.

The so-called *"scenario tree nodal partition matrix"*, M_τ, is a way to represent the forecast tree structure from the ensemble members: M_τ is a matrix responsible for structuring the set of controls into a tree. In other words, M_τ provides *"the necessary labeling scheme to set up the Multistage Stochastic Programming problem"* (Raso et al. 2014). It is built from the ensemble forecast at every time step τ, respecting the following reported **Non-Anticipatory Condition**:

$$u_\tau^i = u_\tau^j \quad when \quad \begin{cases} P(j) = i \\ t < B(j) \end{cases} \quad \forall\, i, j \in Z \tag{4.13}$$

where: P(j) represents the *Parent* of member j, and B(j) is the *Branching point* of member j, that identifies when the member j is distinguishable from its parent P(j).
This key condition states that controls should not depend on the outcome of stochastic variables that have not been extracted yet. At every time τ it is unknown which member

of the ensemble will actually happen, so the controller has to calculate control actions that are valid for all the scenarios in the branch. Once the bifurcation point has been reached, the uncertainty is solved and the controller can calculate specific control actions for the scenarios in each of the new branches (Grosso et al. 2014).

It is important to underline that TB-MPC does not manipulate directly the EF, before the start of the optimization the ensemble undergoes two important procedures:

- **Scenario Reduction:** The ensemble is reduced from k members to k_{red} members by grouping together similar enough trajectories that compose it, and then their probability of realization p gets updated with Bayes's rule.
- **Tree Generation:** The reduced ensemble is transformed into a tree structure, without any loss in carried information. This procedure generates a labelling scheme, called Tree Nodal Partition Matrix M_τ, that is used to represent the tree structure (of forecast inflows) and used in the optimization to search for the optimal control tree.

These procedures will be explained in detail in the following respective subsections.

4.3.1. Scenario reduction

Ensembles generally have from 10 to 50 members composing them. Although this size is good for catching most of the variability, if not all the possible scenarios of the system evolution in time, they would definitely slow down a lot the optimization if all of them were to be considered. However, reducing the size of the ensemble in a smart way does not necessarily lead to any loss in performances: a previous study showed that even after a 50% reduction of the number of ensemble scenarios, about 90% of relative accuracy is still reached (Dupačová et al. 2000).

The scenario-reduction procedure itself is carried out through an heuristic algorithm called Backward Reduction, this is a simple and fast recursive algorithm that groups togheter the most similar members of the original ensemble to create a reduced one with an an a priori defined number of members N_{red}, chosen in this study as 6.

The actual way in which the algorithm discerns between similar enough members (or not) is by comparing the actual raw difference between their trajectories and their probability p, and when two members are selected to be aggregated their total probability is updated as their sum.

4.3.2. Tree generation

Ensembles tend to have small differences for small lead times and then diverge, as this is one of their key peculiarities. Thus, a natural and efficient way to contain their information is to transform the ensemble itself into a tree structure, with branches diverging as the forecast horizon increases.

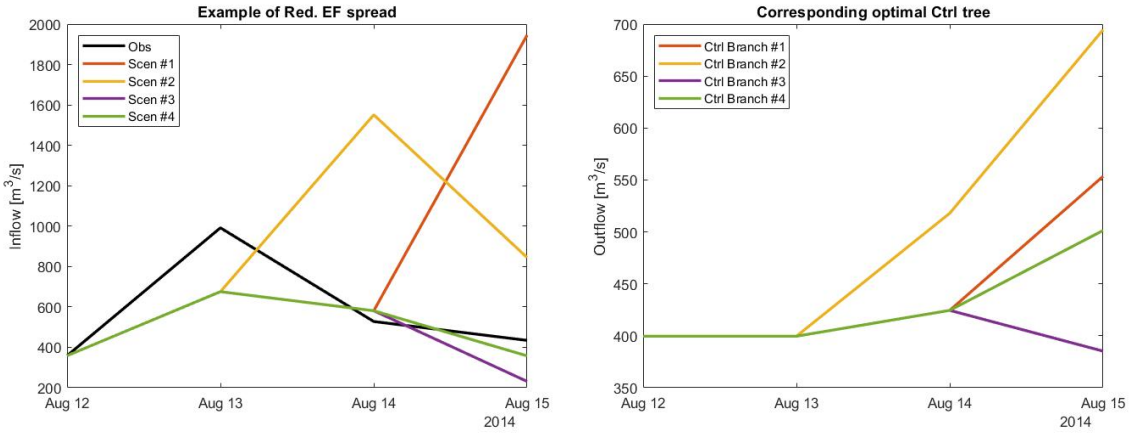

Figure 4.2: An example of reduced ensemble forecast (left panel) and the correspondent computed control tree (right).

The most compact way to mathematically represent a tree structure is with the paired set of variables $[P(\cdot), B(\cdot)]$ for every ensemble member $z \in Z$, where P(z) represents the *parent* of member z, i.e. the branch of the tree from which z diverges, and B(z) is the *branching point* of member z, that identifies when z is distinguishable from its parent P(z) (Ficchí et al. 2016b).

This formalization at the code level is done with the introduction of the matrix M, the previously mentioned tree nodal partition matrix; M is a matrix with a number of columns equal to N_{red} and a number of rows equal to h, that contains labels referring to the time step when a 'branch' divergence (branching point) occurs, indicating in a way the moment when uncertainty on the occurring sub-branch is going to be resolved (Raso et al. 2014).

5 | Results and discussion

This chapter will first present the results of the skill evaluation of the deterministic forecasts from PROGEA and the ensemble forecasts from the two data-driven methods, through the performance metrics and skill scores introduced in Chapter 3.

Then, the second Section 5.2 will report and discuss the results of the different control algorithms tested comparing their performances between them and with the historical management, following these three main groups of control schemes: (i) the two off-line control benchmarks, i.e. DDP and SDP; (ii) the deterministic MPC with perfect and real forecasts, at both daily and hourly time steps; (iii) the stochastic TB-MPC with synthetic ensemble forecasts, at both daily and hourly time steps.

5.1. Forecast analysis

This section will show the results of the forecast skill analysis, using the different performance metrics introduced in Chapter 3. As anticipated, all metrics are computed for the whole time series and for different subsets of the complete original data set, in order to investigate in further details the capability of the forecasts to carry information on the most critical conditions of above seasonal mean flow and flooding periods.

This section will first report the analysis of the accuracy of the deterministic forecasts from PROGEA, and then also report the skill of the daily and hourly ensemble forecasts generated with the data-driven methods (FFNN and KNN algorithms; see Chapter 3).

5.1.1. Deterministic forecasts from PROGEA

Table 5.1 reports the results of the forecast performance metrics, computed for the three data subsets (see Section 3.2): the whole data set (a), and the subsets over high flow (b) and flooding conditions (c). The three mean (quadratic) error-based scores (MSE, RMSE and MAE) have very good values for the whole dataset (a) with an expected worsening from 1- to 3-day lead time, with the RMSE and MAE always below 100 m3/s. RMSE

and MAE are well below the static mean of $146 m^3/s$ also at 3-day lead time, and are less than half of it ($<$ 65 m3/s) at 1-day lead time. The fact that they increase for high flow and flooding condition is something to be expected as the most difficult part of hydrological forecasting is to skillfully predict the flood peaks, both in timing and in magnitude. The quite high errors in (c) suggest that peaks are either under/over estimated or even predicted in advance/with a delay. Predicting peaks in advance can be still valuable in terms of flood control, while the opposite is potentially very detrimental.

From both NSE_{sm} and NSE_{cm} of (a) it can be inferred that the overall prediction capabilities of PROGEA's forecasts are better than both the proposed benchmark statistics (static and cyclo-stationary mean), although with an expected decrease with growing lead times. This seems to be valid even for the other high-flow and flood conditions, (b) and (c), when considering the cyclo-stationary mean, while they decrease substantially when compared to the static mean. However, this is certainly caused by the reduced size and variability of these smaller sets over 8 years (e.g. only about 15 days over the 99.5-th percentile), that makes the mean benchmarks very difficult to beat. Looking at the bias component of KGE, β, for the whole dataset (a) and high-flow condition (b), it can be seen that predictions at all lead times are almost perfect ($\beta = 1$), as they are almost completely unbiased. On the other hand, when focusing only on flood conditions, (c), we found a clear tendency for PROGEA's forecasts to underestimate the highest peak flows.

The same conclusions can be derived after analysing the KGE' and KGEss results. Thanks to the KGE peculiarity of being decomposed into its specific components, it is clear that the performance loss in high flow conditions (b) is mainly caused by a higher variability of forecasts with respect to observations (γ>1), while for the highest flood peaks (c) there is an average underestimation (β<1). For the latter, it was previously explained that the smaller size of these data sets is the issue with an apparent degraded explanatory performance with respect to the mean. Also, the smaller sample size and higher mean (in (b) and (c)) lead to a smaller coefficient of variation of the observations CV_obs in turn increasing γ, as per **Equation 3.8**. This however is not something to be alarmed of, as overall KGE' scores are still within the range $[-0.41, 1]$ indicating "reasonably good" performances with respect to an average-flow benchmark also in high flow conditions.

Table 5.1: Performance metrics for PROGEA's deterministic forecasts over the whole dataset (a), high flows (b), and flood conditions (c) across lead times.

	Dataset (a)			Dataset (b)			Dataset (c)		
	1d	2d	3d	1d	2d	3d	1d	2d	3d
MSE	$4.1*10^3$	$6.4*10^3$	$8.9*10^3$	$1.5*10^4$	$2.2*10^4$	$3.1*10^4$	$1.3*10^5$	$1.4*10^5$	$2.1*10^5$
RMSE	64	80	94	$1.2*10^2$	$1.5*10^2$	$1.8*10^2$	$3.5*10^2$	$3.8*10^2$	$4.5*10^2$
MAE	37	49	56	78	99	$1.2*10^2$	$3*10^2$	$3.1*10^2$	$4*10^2$
NSE_{sm}	0.7	0.53	0.35	0.41	0.13	-0.24	-2.2	-2.6	-4.2
NSE_{cm}	0.63	0.41	0.18	0.65	0.48	0.26	0.75	0.71	0.58
r	0.85	0.8	0.72	0.69	0.6	0.43	0.51	0.46	0.38
β	1.1	1.1	1.1	0.99	1	1	0.67	0.72	0.55
γ	0.91	0.95	0.94	0.97	1	1	2.2	2.6	2.7
KGE'	0.82	0.75	0.68	0.69	0.6	0.43	-0.34	-0.72	-0.9
KGE_{ss}	0.76	0.69	0.59	0.74	0.67	0.53	0.19	-0.047	-0.16

Note: For each dataset (whole dataset (a), high flow condition subset (b), and the flooding condition subset (c), skill metrics are evaluated at each lead time h (1-, 2- and 3- day).
Note 2: Units of measurements for RMSE, MAE are $[m^3/s]$, for MSE $[m^3/s]^2$, while all other scores are adimensional.

Table 5.2 reports the more advanced flood-event peak and timing metrics, to show how skillfully the PROGEA's forecasts predict the flood event occurrence.

In terms of flood peak magnitude errors, ΔQ_p shows that for every lead time the forecasts tend to slightly underestimate the observed magnitude of flood peaks, with however moderately low differences ($< 6\%$) even at the longest lead time (3 days).

In terms of flood timing errors, Δt_p values show that the daily forecasts of PROGEA struggle in correctly reproducing the flood hydrograph's timing. It can be seen that almost all ($> 90\%$) flood peaks are predicted (within the margin of magnitude/timing error limits defined), but roughly half of them are predicted to occur later or in advance, on average 2-3 days late or 1d in advance, at 1-2 days lead times. Yet all the other events are correctly predicted in timing too (Δt_p=0). At 3-day lead time, the proportion of forecasts anticipating the flood peak timing (of 1.5 days on average) increases significantly, while the one with delayed timing is the same as at shorter lead times. The errors in flood timing are a well known problem in daily hydrological models as highlighted by Asadzadeh et al. (2016), especially arising when the watershed area has a fast rainfall-runoff response, e.g. a daily or sub-daily concentration time constant.
Despite some errors in flood peak timings for about half of the events, the overall timing of predicted flows over the whole flood event is satisfactory at all lead times, as reflected

by the scored VE values.

In conclusion PROGEA's forecasts proved to be overall quite skillfull, and thus may contain useful knowledge for the generation of an optimal control policy in a MPC implementation.

Table 5.2: Advanced flood timing metrics for PROGEA's deterministic forecasts

	Lead time h		
	1d	2d	3d
Median ΔQ_p	-1.69%	-5.23%	-5.78%
# Events with $\Delta t_p = 0$	15/31	14/31	8/31
# Events missed	2/31	3/31	2/31
# Events with $\Delta t_p > 0$	6/31	8/31	8/31
Mean delay	3	3.5d	2.5d
#Events with $\Delta t_p < 0$	8/31	6/31	13/31
Mean delay	-1d	-1.5d	-1.5d
VE	0.83	0.79	0.77

5.1.2. Synthetic ensemble forecasts

To generate daily synthetic Ensemble Forecasts, the training of the k=30 parallel FFNN took in total around 30 minutes, employing the whole available historical dataset of daily inflows from year 1946 to early 2022. The cited sources of randomness (see Section 3.5) were inserted progressively to generate ensembles with different levels of skill; this was done in order to investigate how the simulations and performance of TB-MPC are overall heavily influenced by the ensemble quality. Less randomness factors yield a more precise,

	Horizon			
	1d	2d	3d	Average
EF#1				
RMSE	84.9	109.2	110.8	101,6
Spread	173.5	233.9	266.7	224,7
CRPS	32.5	41.1	43.9	39.2
EF#2				
RMSE	177.5	169	166.9	171,1
Spread	59.1	55.1	46.1	53,43
CRPS	209.4	196.1	189.1	198
EF#3				
RMSE	112.3	213.2	170.8	165,4
Spread	225.9	383.7	489	366
CRPS	228.3	223.5	236.7	229.3
EF#4				
RMSE	210.7	207	209.2	208.8
Spread	53.8	46.5	53.6	51,3
CRPS	230.5	246.1	321.9	266.2
Forecast				
RMSE	64	80	94	79.3

Table 5.3: Skill metrics for the synthetic short therm ensemble forecasts

but less diverse prediction (i.e. with smaller ensemble spread), while on the other hand increasing these disturbances leads to more diverse ensemble members, diverging with lead time. **Table 5.3** reports the skill metrics of the generated ensembles. EF#1 and EF#2 used historical data only at previous time steps (hence its operational feasibility), while EF#3 and EF#4 used real predictions. EF#1 and EF#3 had switched calibration and validation datasets, while EF#2 and EF#4 did not. All of them had a Gaussian noise added on the input, different splitting for every member and randomized training algorithm and random number seed (see Section 3.5.1).

For the hourly implementation of TB-MPC, only one ensemble was generated as its different method required a lot more computational time (around 10h). The skill metrics of the hourly EF are reported in **Figure 5.1** for conciseness. This ensemble is as skilled as the best daily one, EF#1, and in particular, it is an "under-dispersive" ensemble (Fortin et al. 2012), with a smaller spread than the RMSE.

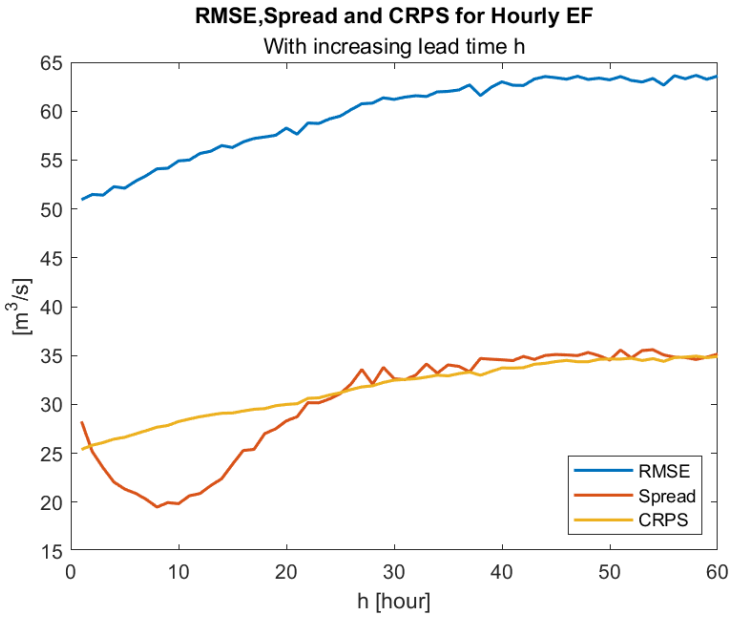

Figure 5.1: Skill metrics of the synthetic hourly ensemble forecast, varying with lead time h (1h-60h). It denotes and "under-dispersive" behaviour.

5.2. Performance analysis of control schemes

5.2.1. Benchmark: Stochastic vs Deterministic DP

The best possible option in terms of hydro-meteorological prediction would irrefutably be a perfect knowledge of inflow with a infinite horizon in the future. This is naturally impossible in a realistic setting but it would lead to the upper-bound of control performances, or the so called Perfect Operating Policy (POP), in a fictitious simulation setting. This POP is embodied by the DDP algorithm (see Section 4.1) and was used as a benchmark for all the other proposed sub-optimal control methods.

DDP can effectively schedule control action for the system in the best possible way, while its stochastic counterpart SDP is closely behind, as expected given the less accurate knowledge available. They both use the same mathematical principle to optimize a control policy based on solving the Bellman's equation (Section 4.1). The usefulness of forecasts as a concept is firstly evaluated by comparing them.

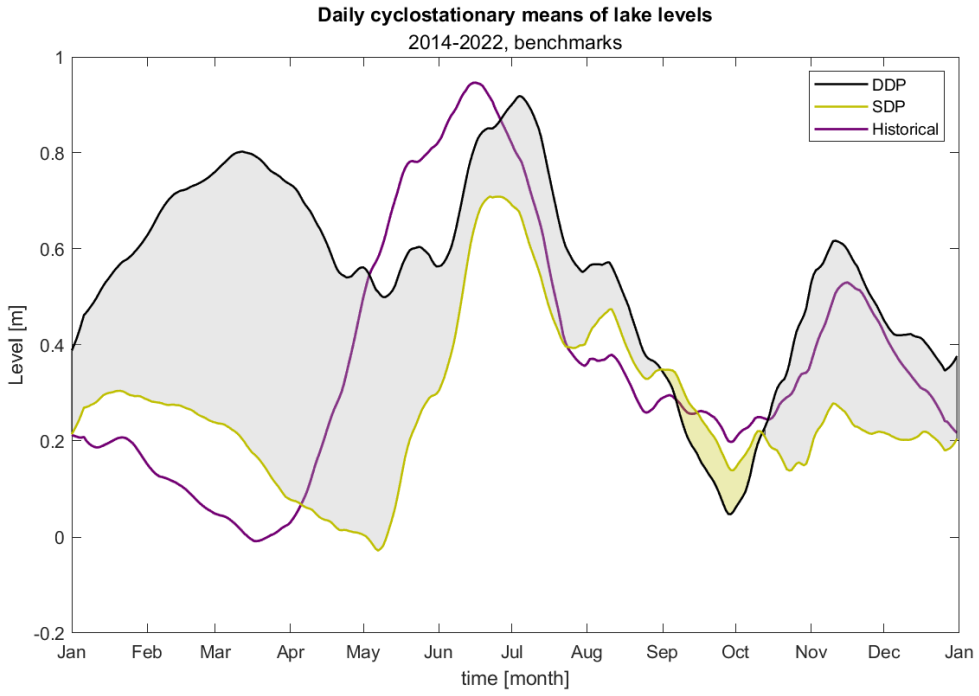

Figure 5.2: Cyclo stationary behaviour of system state with DDP, SDP and historical management

From **Figure 5.2** it can be inferred that over the course of the years SDP tends to be more conservative like the historical management, while DDP thanks to its perfect foreseeing can store higher water levels for drought seasons while still being able to decrease storage to attenuate inflow peaks when necessary. This tendency can be seen better in a snapshot on period May 2020 to Nov 2020 reported in **Figure 5.3** where SDP is seen releasing less water during inflow peaks, causing the all-high level for the year in October to surpass the one related to DDP.

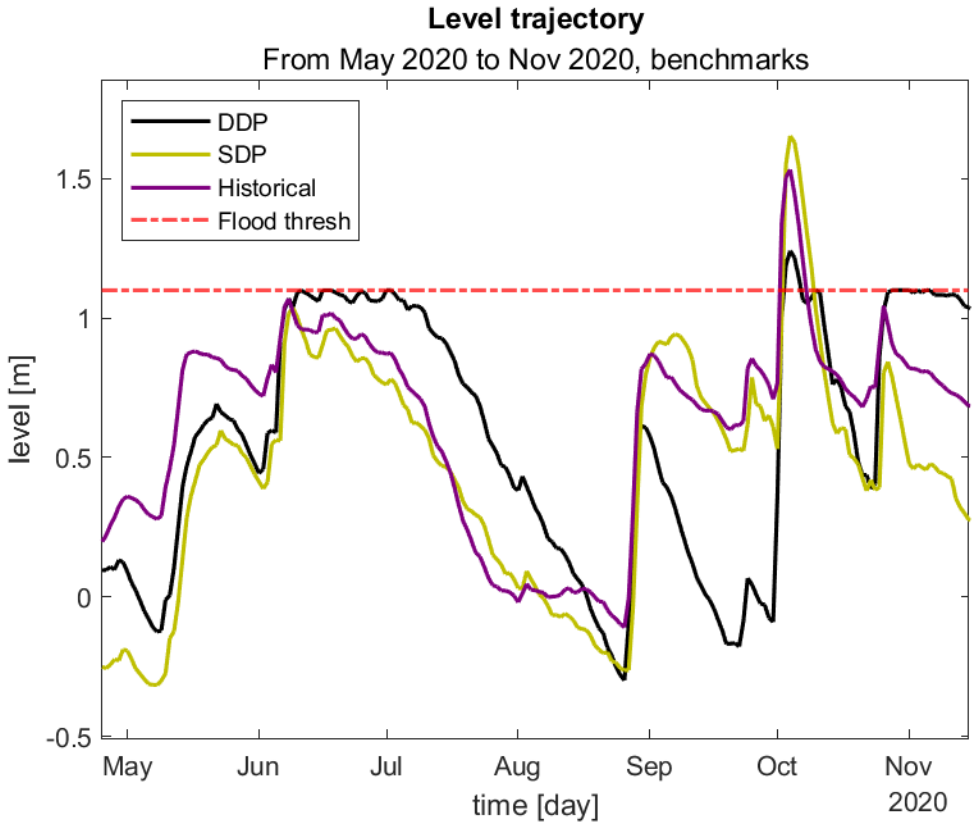

Figure 5.3: System state with DDP, SDP and historical in flooding condition, year 2020

5.2.2. Optimal weight choice

The earlier simulations used coinciding weights for objective aggregation of $W = [0.667\ 0.333\ 0]$, this particular choice comes from an a-priori examination of performance sensitivity with respect to weight choice.

A set of consequent simulations with progressively varied weight choice was carried out for both SDP and Daily MPC with perfect predictions and varying horizons to create a Pareto frontier on the two most important objectives, flooding incidence and water deficit. Through this graph, a unique best set of weights was identified and chosen for all other simulations (for MPC and TB-MPC) to maintain consistency.

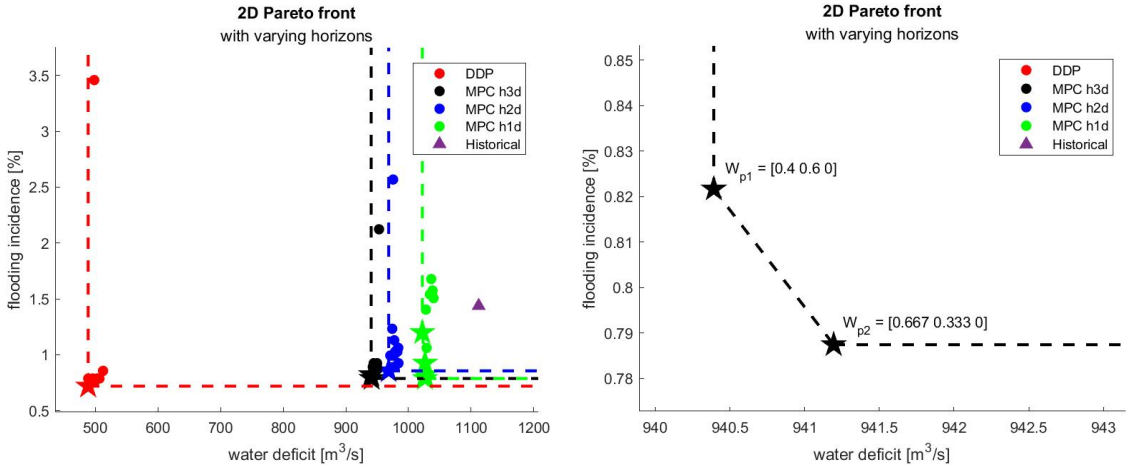

Figure 5.4: 2D Pareto frontier of Daily deterministic MPC with perfect predictions and varying horizons with respect to DDP, with also a snapshot zoomed onto the the two Pareto-dominating solutions for the 3 days ahead horizon case

From the degenerated characteristic of the obtained DDP in **Figure 5.4** it can be deduced that the flooding and water deficit are not really conflicting objectives, as a unique best alternative weighting choice exists. This does not directly translate to MPC though, which appear to have a set of two dominating Pareto solutions W_{p1} and W_{p2}, the former dominating in terms of water deficit and the second dominating in terms of percentage of flooding days. In light of the facets of the particular case study already presented in Chapter 2, it was chosen the weight [0.667 0.333 0] of solution W_{p2} as the best weight choice for all simulations from this point forward.

5.2.3. Deterministic MPC: Perfect predictions

The next step towards a real world implementation of forecasts is to deploy them in a real-time control scheme such as deterministic MPC. This however is a management paradigm very different from Dynamic Programming approaches, so the very first results will focus on assessing if, and to what extent, MPC can use effectively weather predictions with respect to this benchmark.

This comparative analysis has been carried out on a fair common ground, with same system requirements, performance indicators, weight factors on costs to minimize and with *perfect predictions*.
Additionally, these differences are investigated considering different possible control frequencies. Both of them with 3 days ahead of control and prediction horizon.

As defined in Chapter 4, MPC can be set to operate with either a D*Daily* or *Hourly* decision frequency by properly scaling the inner system model and adopting respectively daily or hourly forecasts. Daily forecasts are obtained simply by aggregating the original hourly forecasts over 24h, from 8 a.m. of a day to 8 a.m. of the next.

Daily MPC

- Control action u computed at 8 a.m. and applied all day
- Cannot fully satisfy the saturation of the control of the system
- Rougher control action
- Very low computational demand

Hourly MPC

- Control action u computed every hour
- Completely satisfies the saturation of the control of the system
- Finer control action
- High computational demand

By the premises, Hourly MPC is expected to achieve better performances than the Daily one, especially with respect to the water deficit objective, since an hourly control resolution can better follow and satisfy the real control saturation of the system.

The following are the results of a simulation from January 2014 to January 2022 of Daily and Hourly management of the lake compared to the historical management and both DDP and SDP benchmarks.

Figure 5.6 reports the score indicators for the whole simulation, proving that MPC can in fact make and effective use of weather predictions almost as good as the benchmark when it has perfect knowledge of inflow as well, while also confirming how the Hourly adaptation is superior to the Daily alternative and even beats the "perfect" benchmark DDP with respect to flooding.

Figure 5.5 provides a more detailed snapshot of the MPC behaviour during a particular period characterized by a flooding event, all the alternatives besides SDP behave similarly to the best one (black) especially during October when they start discharging water to absorb the predicted peak of November. It is also worth noting how MPC releases more water than SDP and the historical management effectively reducing the reached level of around 0.4m.

Figure 5.7 displays the cyclostationary mean of the system state over the whole period of the simulation, it can be intended as the average amount of water that is kept in storage throughout the year. DDP is unsurprisingly the less conservative of the alternatives with respect to flood risk, having perfect knowledge of the magnitude and timing of the flooding peaks it can keep more water in storage for irrigation while simultaneously being able to release it right before the flooding events to attenuate them. SDP is more conservative with respect to flood risk, as it operates based on a statistical description of the inflows it tends to keep more storage available at all times to avoid the risk of having the reservoir too full. A similar behaviour is found for both MPC schemes. Although Daily and Hourly MPC score different indicators, it can be seen from this figure that their long-term (yearly average) behaviour is mostly the same, while differences arise from the capability of the latter to provide a finer release on a hourly basis that can be useful in rapid flood events.

Finally, **Figure 5.8** shows the cyclostationary mean for the releases alongside with the yearly demand. All of the alternatives show a difficulty in satisfying the demand during the last weeks of July and August, even DDP. This can be explained by the cyclostationary inflows shown in **Figure 2.3**, this particular catchment is characterized by sudden inflows much higher than the seasonal mean in the 3 months preceding July and early August, with then very sudden and sharp decreases. While this great amount of water could in principle satisfy the high irrigation demand characterizing Summer if stored, it arrives in such a short span of time that most of it has to be released immediately to avoid flooding.

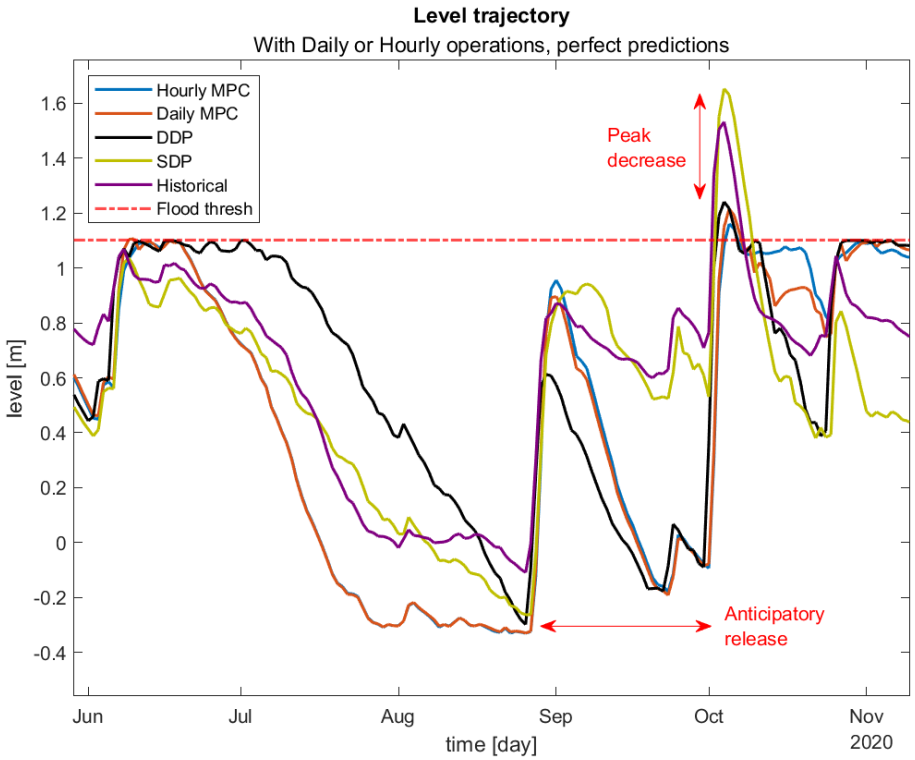

Figure 5.5: System state evolution in a 6-month period with flooding from Jun 2020 to Nov 2020, with daily and hourly MPC using perfect predictions alongside the benchmarks (DDP, SDP) and the historical management.

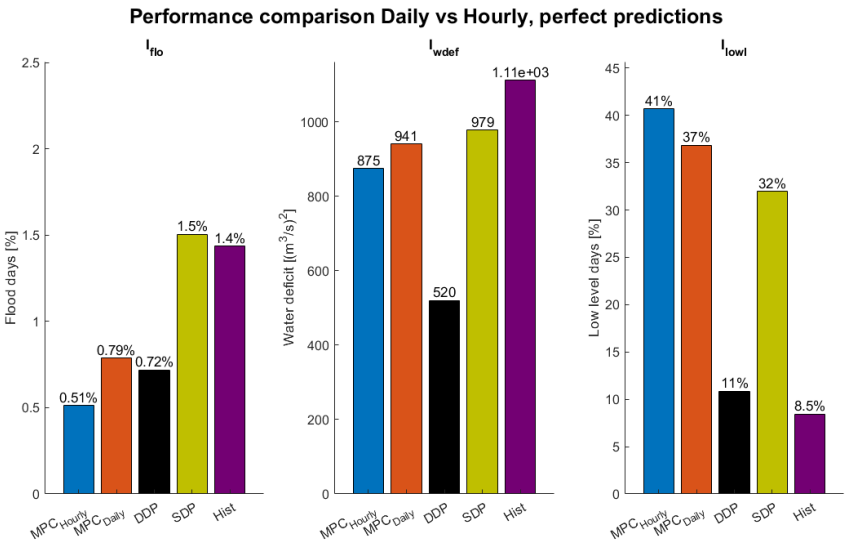

Figure 5.6: Score indicators throughout the entire simulation period of 8 years, with daily and hourly MPC using perfect predictions alongside the benchmarks (DDP, SDP) and the historical management.

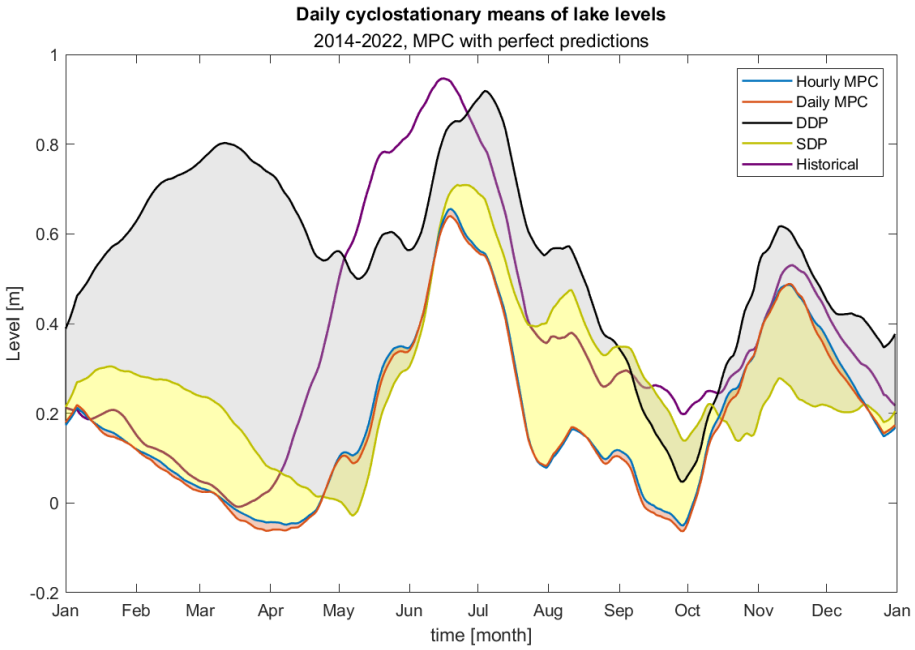

Figure 5.7: Daily cyclostationary levels over the whole simulation period of 8 years, daily/hourly MPC with perfect predictions alongside the benchmarks (DDP, SDP) and the historical management.

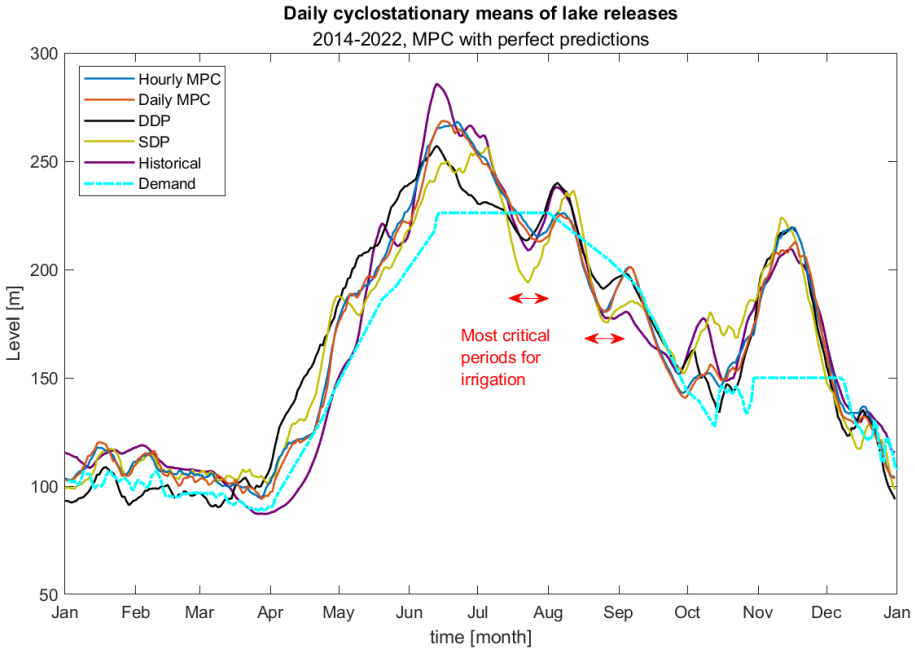

Figure 5.8: Daily cyclostationary releases for the whole simulation period of 8 years on top of the the yearly water demand characteristic, daily/hourly MPC with perfect predictions alongside the benchmarks (DDP, SDP) and the historical management.

5.2.4. Deterministic MPC: Real predictions

Results brought in the previous section show how the MPC paradigm provided with perfect knowledge of the future can provide a satisfactory management of the system.

Now the same simulation is repeated but using real deterministic predictions from PROGEA in both Daily and Hourly MPC, keeping the rest of the setting as before (demand, requirements, horizon etc...). This deterioration of knowledge about the future inputs to the system (from the perfect forecast case) is expected to cause a deterioration in the produced control policy. However, performances of this simulation are now truly indicative of real world objective accomplishment while the previous one has to be intended more as a first proof of validity showing the upper bound of performance for MPC, prior to its implementation.

Figure 5.10 Shows the updated score indicators for the whole simulation, the expected performance loss is more prominent on the water deficit objective than the flooding one, which remained quite satisfactory and just under DDP. This increase in deficit is still acceptable as it remains in the neighborhood of the historical and SDP alternatives.

Figure 5.9 shows the same snapshot shown before for the perfect predictions, the anticipatory discharge of water to mitigate the October peak is now significantly different. Both MPC implementations decreases the level of the lake a lot less due to an under-estimation of the incumbent peak's magnitude and/or time of arrival. Peak level is still decreased with respect to SDP and the historical management, but only of around 0.2m this time.

Figure 5.11 reflects the tendency of MPC to discharge more water in prevention of floods with a slightly higher level around October, while there is big difference in storage when possible to accumulate water from November to December.

The same previous conclusions can be drawn from **Figure 5.12**, now with MPC releases reaching moderately higher deficits in the critical final weeks of July and August.

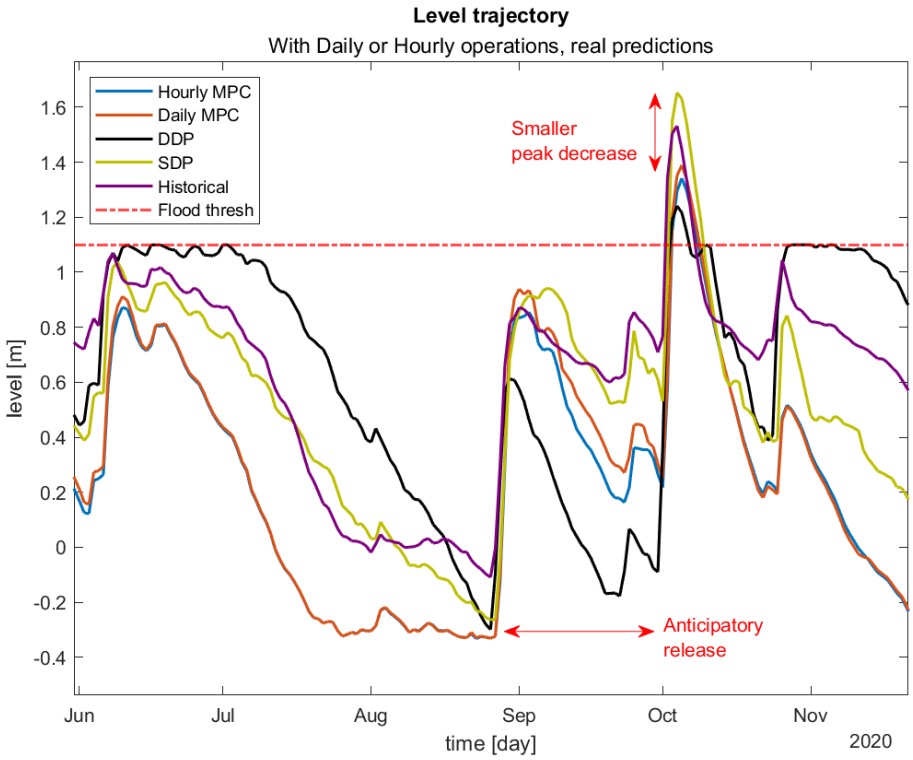

Figure 5.9: System state evolution in a 6-month period with flooding from Jun 2020 to Nov 2020, with daily and hourly MPC using real predictions.

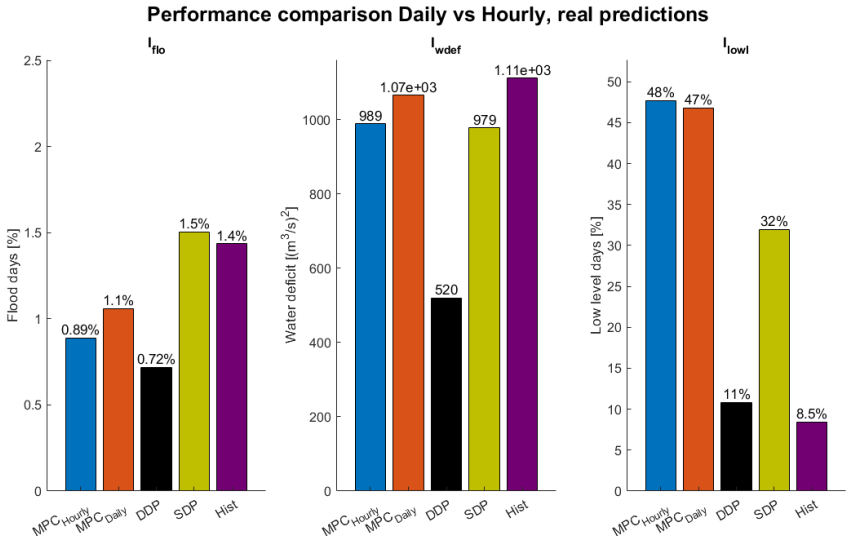

Figure 5.10: Score indicators throughout the entire simulation period of 8 years, with daily and hourly MPC using real predictions.

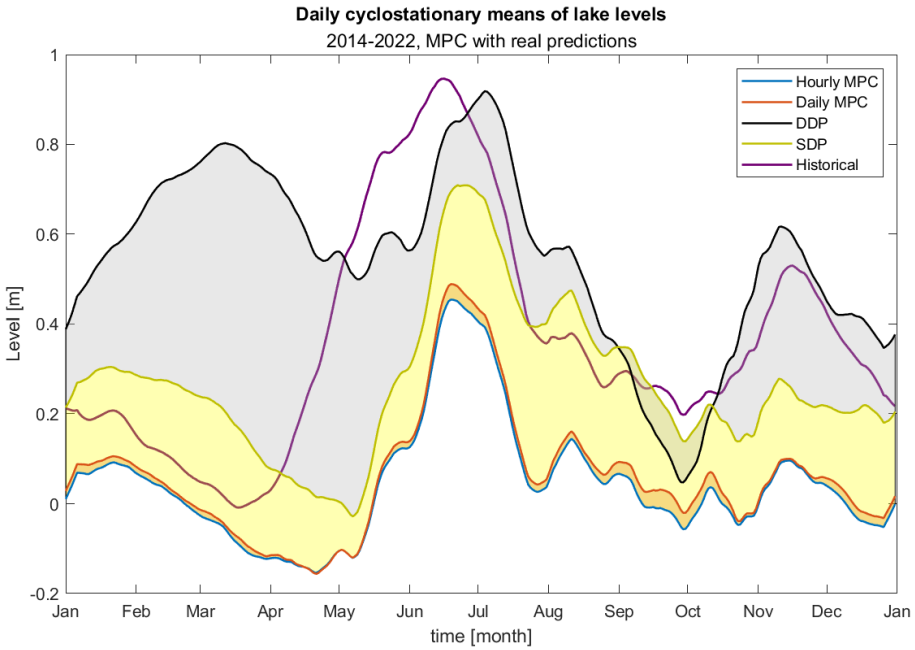

Figure 5.11: Daily cyclostationary levels over the whole simulation period of 8 years, daily/hourly MPC with real predictions alongside the benchmarks (DDP, SDP) and the historical management.

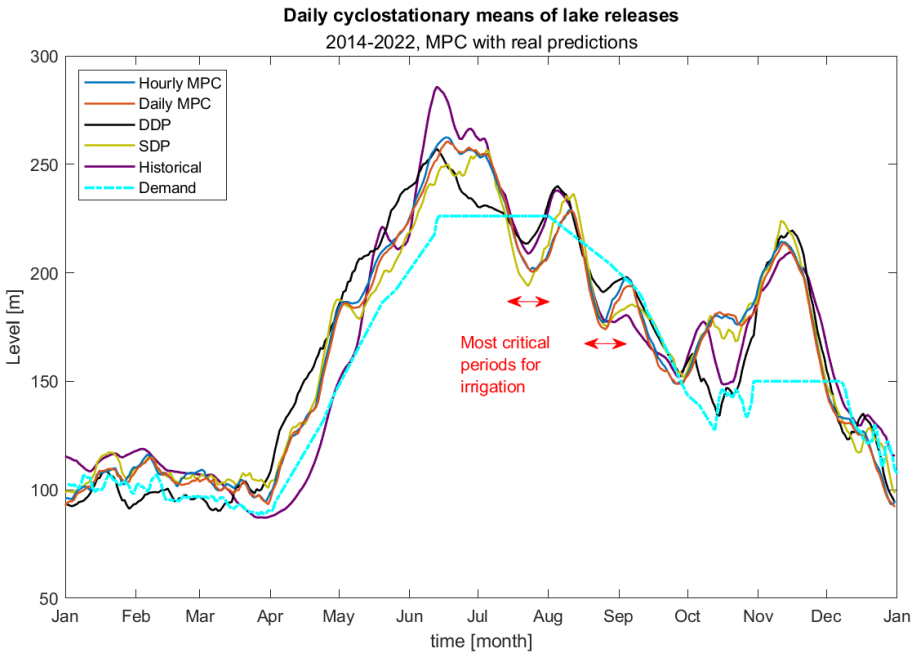

Figure 5.12: Daily cyclostationary releases for the whole simulation period of 8 years on top of the the yearly water demand characteristic, daily/hourly MPC with real predictions alongside the benchmarks (DDP, SDP) and the historical management.

5.2.5. Daily TB-MPC vs Daily deterministic MPC

This section will describe the results related to the daily stochastic implementation of TB-MPC, adopting the synthetic ensemble forecasts generated based on the real deterministic forecasts from PROGEA and the algorithms described in the Section 3.5.
Daily TB-MPC will be compared to the corresponding Daily deterministic MPC with real predictions in order to bring a comparison between approaches truly applicable in the real world and to highlight the changes in performance moving from a deterministic to a stochastic approach with real forecasts. The comparison with MPC with perfect predictions is not reported in this section, but the reader could refer to the previous Section 5.2.3 to see the upper bound of performance of the deterministic MPC approach. The benchmarks will still be SDP, DDP and historical management.

Figure 5.13 and **Figure 5.14** show score indicators across the same 8 year simulation as before, here produced with the Daily TB-MPC implementation using each of the four selected generated short-term ensembles. The TB-MPC performances fluctuate more on the flooding objective rather than the water deficit one, showing that different ensemble forecasts can lead to different control performances (also with respect to deterministic forecasts) particularly in flood conditions. Additionally **Figure 5.13** shows that EF#1 (the best ensemble with respect to all of forecast skill metrics) is the one associated to the best TB-MPC performance, as it could be expected. The other ensembles are closely behind with the exception of EF#4, the one with worst skill metrics, having 2 times the error of the deterministic forecasts with a quite lower spread than other EFs. As foretold the spread is a very important feature of an ensemble, and this can be seen with the behaviour of TB-MPC with EF#3 beating the TB-MPC with EF#4 on both objectives, while having similar high RMSE and CRPS but with a much larger spread. These results suggest that there is a link between TB-MPC performance and ensemble forecast skill and spread, that further work could aim to validate using a larger set of ensemble forecasts with a wider range of skill and spread metrics.

Note that low level indicators have been omitted now as there are more options to show and they are not optimized.

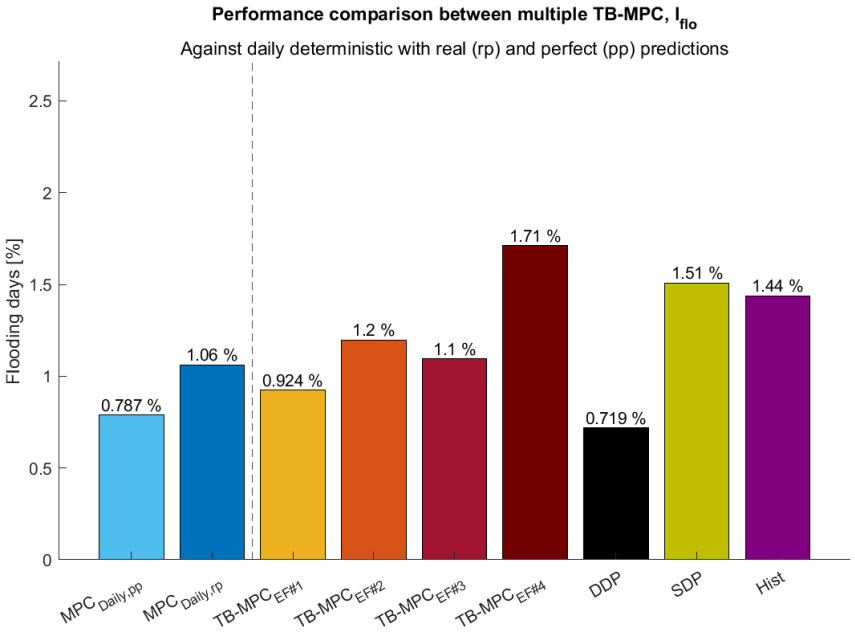

Figure 5.13: Flooding indicators for all the four TB-MPC options (with different ensemble forecasts) over the 8-year study period, alongside the daily deterministic MPC with perfect and real forecasts (this for comparison between operationally-feasible paradigms), and against the benchmarks (DDP, SDP and the historical management).

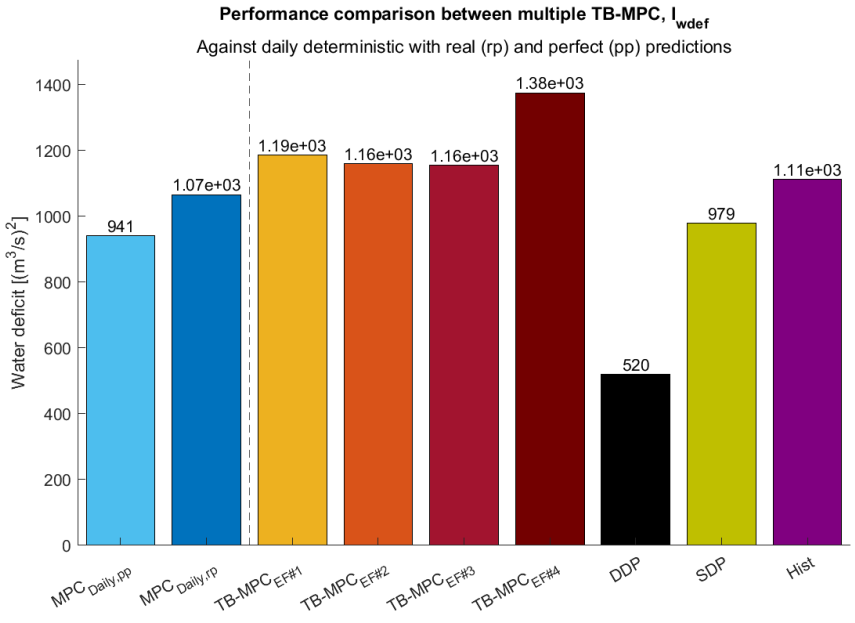

Figure 5.14: Water deficit indicators for all the four TB-MPC options (with different ensemble forecasts) over the 8-year study period, alongside the daily deterministic MPC with perfect and real forecasts (this for comparison between operationally-feasible paradigms), and against the benchmarks (DDP, SDP and the historical management).

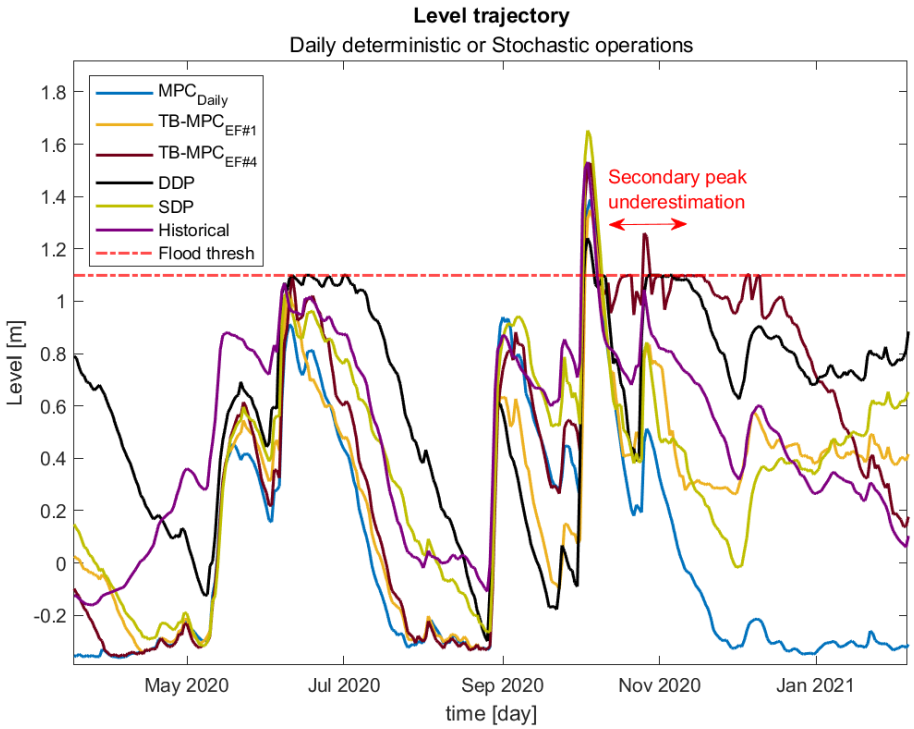

Figure 5.15: System state evolution in a 9-month period with flooding events (from May 2020 to Jan 2021) with different control strategies: TB-MPC (with the best and worst ensembles), daily MPC using real predictions alongside the benchmarks (DDP, SDP) and the historical management.

Figure 5.15 reports a snapshot of the system's state evolution with daily TB-MPC using the best and worst ensembles, EF#1 and EF#4, alongside daily MPC with real forecasts together with all the other benchmarks. Using EF#1, the TB-MPC controller produces a good trajectory, similar to all the others, while the information contained in the EF#4 likely has a widespread underestimation in high flow conditions, judging by the incapability of the controller to mitigate the secondary burst in inflows after the main peak.

Figure 5.16, and especially **Figure 5.17**, suggest that there is no consistent difference in normal conditions between TB-MPC and MPC operations,apart from the fact that having access to ensemble predictions allows TB-MPC to behave less conservatively around those months with high flooding incidence, while having an overall lesser storage during high water demand months. This is consistent with the indicators in the previous Images. Performances are highly influenced by the capacity of the ensemble to capture variegated possible scenarios and which one actually happens.

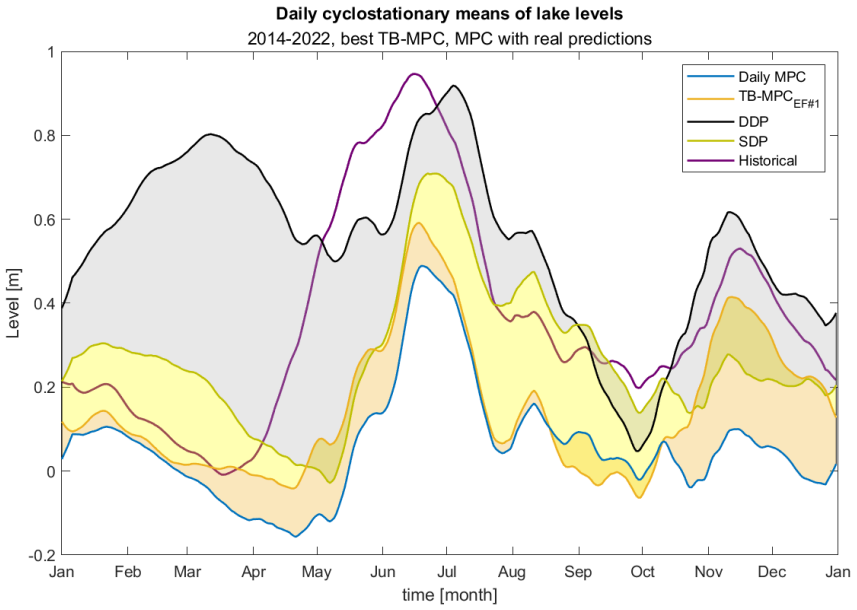

Figure 5.16: Daily cyclostationary mean lake levels computed from the whole simulation period of 8 years, with different control strategies: TB-MPC (with only the best ensemble, to maintain clarity), daily MPC using real predictions alongside the benchmarks (DDP, SDP) and the historical management.

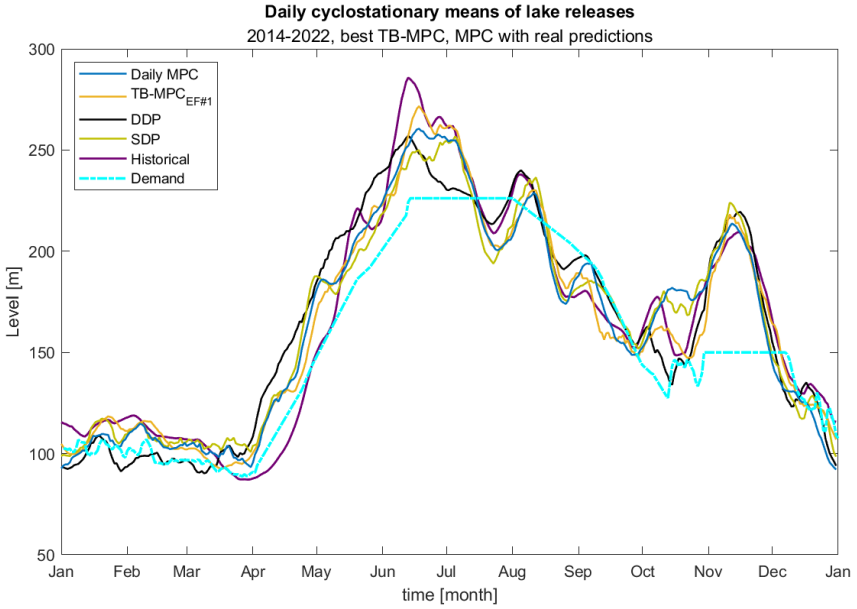

Figure 5.17: Daily cyclostationary releases over the whole simulation period of 8 years, with different control strategies: TB-MPC (with only the best ensemble, to maintain clarity), daily MPC using real predictions alongside the benchmarks (DDP, SDP) and the historical management.

5.2.6. Hourly TB-MPC vs Daily TB-MPC

The comparison between hourly and daily TB-MPC is presented as a potential "proof-of-concept", since the hourly ensemble forecast generation method used is not directly operationally applicable (being based on 'dressing' past observations), differently from the operative method used for the daily EF generation used in the previous Section 5.2.5. It is reminded here that the reason for this is that the data-driven method used for daily ensemble is usable operationally, but needs to be further adapted to work with longer hourly records (see Section 6.1). Anyway, the skill of the hourly EF is comparable with the deterministic operational forecasts, so it is a good basis to clarify if the hourly TB-MPC brings more benefits than the daily counterpart. This analysis aims to prove if it is worth to develop further an operative hourly EF data-driven generation method and implement a more computationally-expensive hourly TB-MPC as well. However, from the previous results comparing daily and hourly deterministic MPC operations, we do not actually expect any significant improvement moving from daily to hourly resolutions, except for a possible slight improvement in flood conditions, as in low or average flow conditions the hourly forecasts did not provide any additional information.

Before proceeding with the discussion of the results, it's important to remind that, for the sake of consistency, the daily TB-MPC uses a daily ensemble derived from the aggregation of the hourly synthetic ensemble. Moreover, both daily and hourly TB-MPC simulations start with the same initial state over each period, obtained from the daily TB-MPC simulation over the whole 8-year period.

While the simulation of the optimal daily operation with TB-MPC take around six times the computational time of the daily deterministic MPC, the hourly TB-MPC takes 8-10 times the one required by daily TB-MPC (i.e. the hourly TB-MPC requires about 30-min per day of simulation). Thus, the complete 8-year simulation with the hourly TB-MPC would require a prohibitive amount of time to be carried out, up to two months with the current undistributed and unclustered computational setting. Future research should aim at reducing the computational time of the TB-MPC optimisation (see Section 6.1). Consequently, the hourly TB-MPC was tested only for specific periods characterized by flooding events (i.e. the two highest and most recent registered historical peaks, October 2020 and September 2021), where we may expect some differences between daily and hourly optimal operations. For the sake of a more comprehensive analysis, also one drought event (summer 2020) was used for running the tests with the hourly version of TB-MPC, to corroborate our expectation on the null benefits from the higher control

frequency in average and below-average inflow conditions.

Figure 5.18 and **Figure 5.19** show the lake level (state of the system) and the release trajectories during the two selected flooding periods (October 2020, left panel, and September 2021, right panel). They both show an expected similarity between the behaviour of the two different TB-MPC control frequencies, as suggested by previous results with hourly/daily deterministic MPC too. Both daily and hourly TB-MPC outperform SDP and the historical management, with a behaviour very similar to the perfect benchmark DDP (see **Table 5.4**).

Figure 5.20 shows that the hourly synthetic forecasts with 60h ahead lead time underestimated both flood peaks on average, but some ensemble members correctly picked up the flood peaks and timing (especially for the first flood event of 2020). It is reminded here that the level of accuracy of the daily ensemble used in this Section is the same as the hourly EF, given its construction by simple temporal aggregation of the hourly one.

These first hourly TB-MPC results suggest that there does not seem to be any advantage from the development of a more sophisticated data-driven procedure to generate hourly EF, as they are likely to not bring any significant improvement during critical conditions with respect to daily EF. The only notable difference between hourly and daily TB-MPC operations is that the hourly one tends to be less conservative than the latter, closely behaving like DDP at times. For example, after the peak of September 2021, both DDP and hourly TB-MPC keep the level right below the threshold instead of lowering it more quickly as the daily TB-MPC and the historical management.

	Event #1		Event #2		Event #3	
	I_{flood}	I_{wdef}	I_{flood}	I_{wdef}	I_{flood}	I_{wdef}
DDP	13.33	1	19.35	358	0	905
SDP	23.33	22	29.03	1768	0	1355
Historical Man.	20	268	16.13	578	0	1783
Hourly TB-MPC	13.33	245	29.03	2025	1.54	2218
Daily TB-MPC	20	5	22.58	1545	0	1808

Table 5.4: Performance indicators for for hourly/daily TB-MPC over the three events under analysis (I_{flood} is the percentage of flooding days [%], I_{wdef} is the mean (squared) water deficit [m^3/s]). Event #1 corresponds to flooding event of October 2020, Event #2 the flooding event of September 2021 and Event #3 drought of Summer 2020, with the addition of daily MPC with perfect/real forecasts and benchmarks

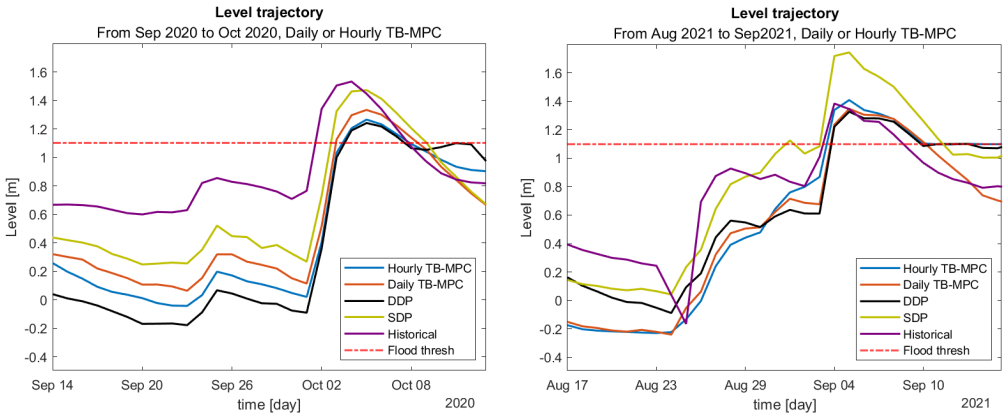

Figure 5.18: System state evolution for the hourly TB-MPC compared to the daily TB-MPC and benchmarks in flooding conditions: October 2020, left panel; September 2021, right.

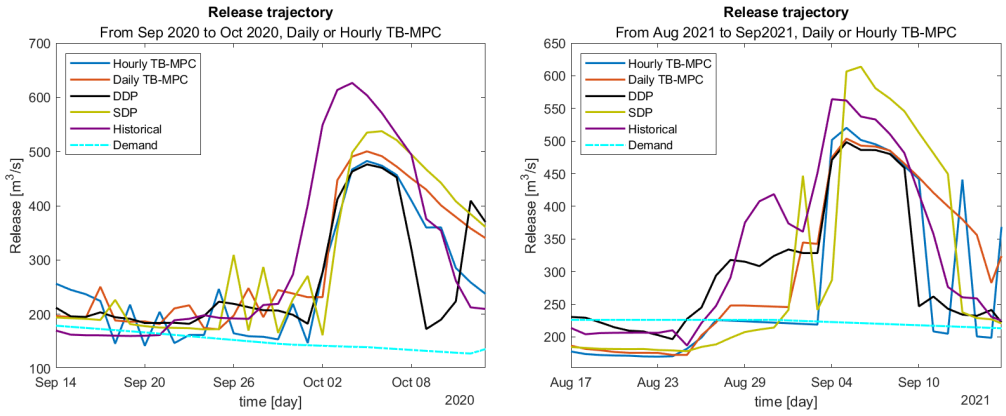

Figure 5.19: Release trajectory for the hourly TB-MPC compared to the daily TB-MPC and benchmarks in flooding conditions: October 2020, left panel; September 2021, right.

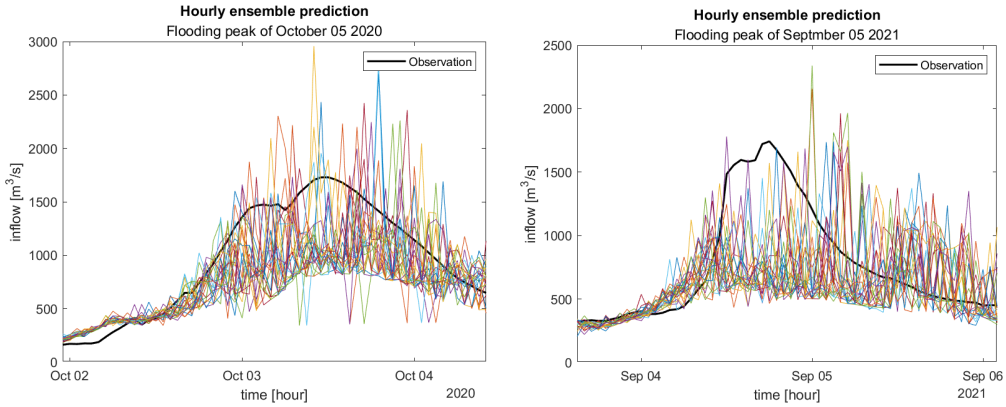

Figure 5.20: Hourly synthetic ensemble forecasts at 60-hour lead-time, for the two selected flooding periods: October 2020, left panel; September 2021, right.

Figure 5.22 and **Figure 5.23** show the simulated state of the system and the corresponding release on a longer period (9 months) covering the entire Summer of year 2020, using the hourly and daily TB-MPC against the different benchmarks. As seen in **Figure 5.21**, this period is characterized by below-average inflows in late winter and early spring, and one high inflow peak in late Spring (June), typical of those caused by snowmelt in this alpine basin.

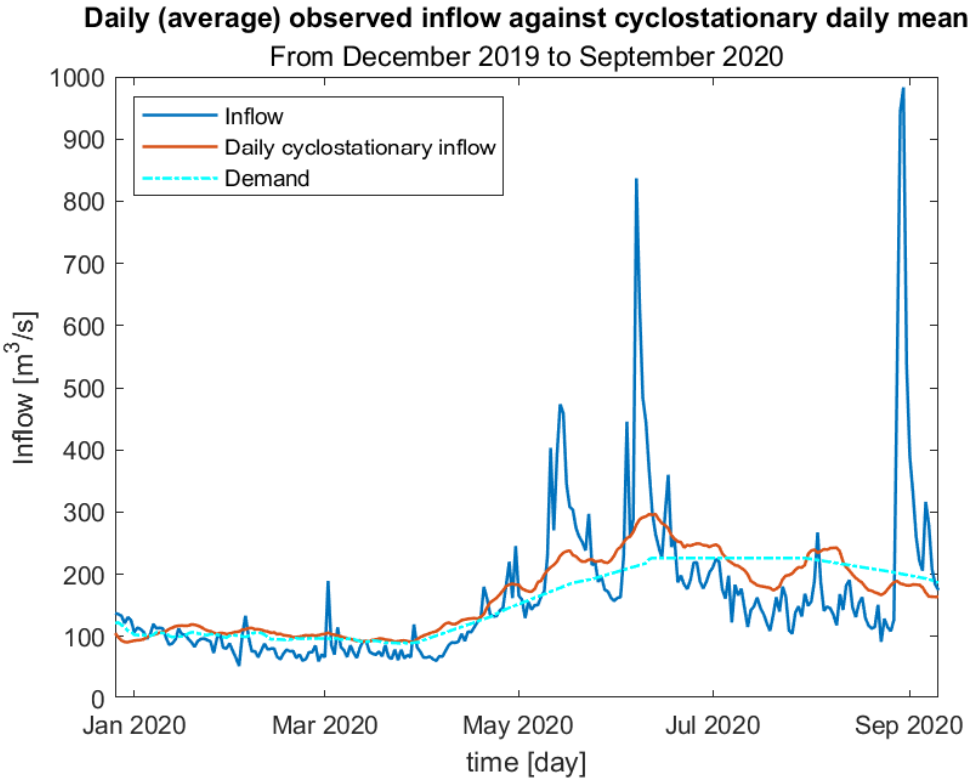

Figure 5.21: Observed inflows over January-September 2020 against daily cyclostationary inflows and water demand. The observed inflows over most of this period (26$^{\text{th}}$ Dec, 2019 to 10$^{\text{th}}$ Sept, 2020) are lower than the seasonal average and than the water demand, apart from a few flood events in June and September 2020.

It can be seen that the hourly TB-MPC leads it to only partially attenuate the sudden inflow peak in June. Interestingly, the daily TB-MPC making use of the same ensemble predictions (aggregated) completely attenuates this peak instead. **Figure 5.23** shows in fact how in June the hourly TB-MPC has a considerably lower release than the other approaches right before this critical event, causing the level of the lake to not be properly lowered in time (**Figure 5.22**) before the inflow peak. This did not happen with the flood events of October 2020 and September 2021 previously analysed. For the October 2020 event, for example, as it can be seen from **Figure 5.20** and **Figure 5.18**, the incoming peaks are predicted better than in June 2020 and this leads to a better control: most

of the ensemble members pick the actual observed flood event timing, while the sudden peak in **Figure 5.24** is under-estimated by about 25 members of the ensemble (out of 30).

This difference in behaviour between daily and hourly TB-MPC has a counter-intuitive explanation: in such cases of underestimation of the flooding peak by the forecasts, the controller might schedules a lower release than the ideal for that period. Due to the finer hourly frequency, the optimal decision of hourly TB-MPC can (wrongly) follow the forecasts more closely and control the system towards an unfavorable state of higher water level. Finally, when the scenario that actually happens is one with (moderately) higher than expected inflow, a (small) flooding event occurs. In this instance, the level only slightly exceeds the flooding threshold, but it definitely brings to light how an hourly ensemble needs to have more accuracy than a daily one.

The daily TB-MPC has instead a rougher control capability which makes it much more 'balanced' (i.e. less reactive), but also more resilient in case of wrong predictions. Occasionally, when the flood peak forecasts are accurate, the daily TB-MPC might struggle to lower the lake levels and achieves slightly lower performance than the hourly TB-MPC, as during the flood of October 2020. On the other hand, it also does risk less to run into following too closely forecast errors, like false alarms (releasing more water) or underestimation of peaks (with lower release than the ideal).

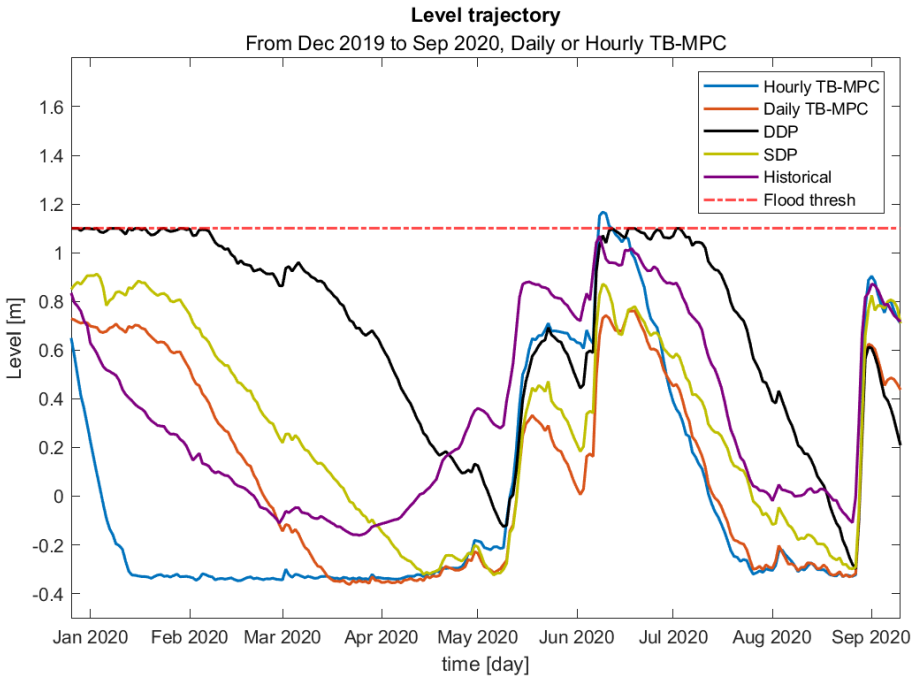

Figure 5.22: System state evolution for the hourly and daily TB-MPC, against benchmarks, over a 9-month simulation period from 26$^{\text{th}}$ Dec, 2019 to 10$^{\text{th}}$ Sept, 2020, with drought conditions in Summer 2020.

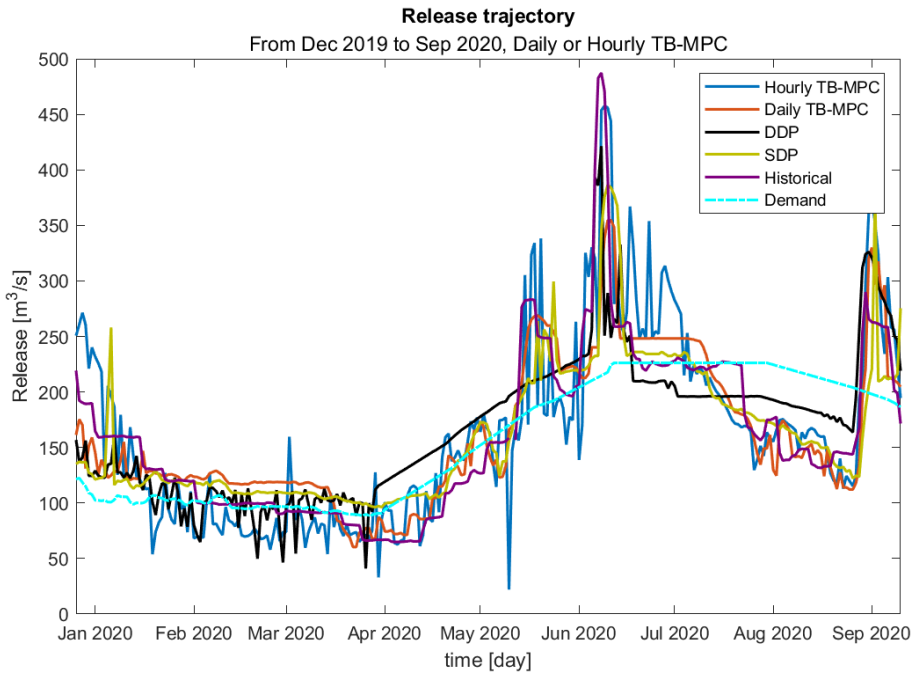

Figure 5.23: Release trajectory for the hourly and daily TB-MPC, against benchmarks, over a 9-month simulation period from 26$^{\text{th}}$ Dec, 2019 to 10$^{\text{th}}$ Sept, 2020, with drought conditions in Summer 2020.

Another possible contribution to the too reactive behaviour of the hourly TB-MPC is the forecast discontinuity, also known as 'jumpiness' (Zsoter et al. 2009). The stochastic ensemble generation used creates some very 'jumpy' trajectories especially in flooding conditions, which is a feature that can be found also in EFs from NWP in some cases, but is not desirable. These results suggest that the ensemble forecast generation should be further tuned to liming extracting too variable data in following time steps for flooding conditions and reach too high jumpiness, while keeping a significant ensemble spread. This is particularly important to avoid providing a detrimental contribution to the optimization, which is already stressed by the great amount of variables involved. This could cause many instances of no convergence to the optimal value or potentially even a fail (i.e. no feasible point found), although this did not occur in our tests.

Figure 5.24 and **Figure 5.25** show an exemplary case of the higher jumpiness of hourly predictions with respect to daily aggregated predictions, from the same hourly ensemble, for the same days reported in **Figure 5.24**, empirically showing that aggregating data on a daily scale smooths out members trajectories, giving a more consistent signal to the controller. This further reinforces the explanation suggested above on why the daily TB-MPC may perform better than the hourly one in the event analysed here. Further work should analyse the link between forecast jumpiness and the performance of TB-MPC over a larger set of flood events and more case studies (Chapter 6).

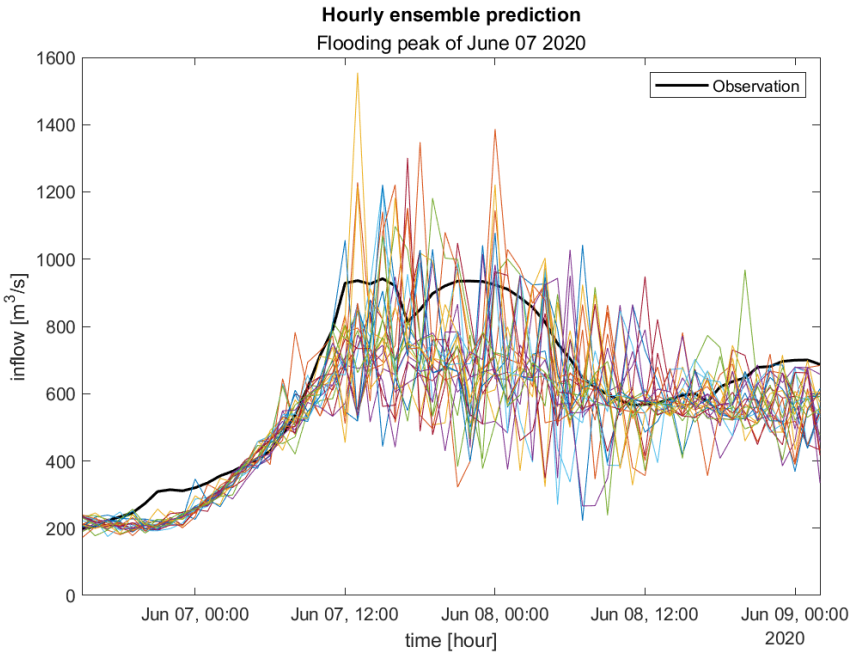

Figure 5.24: Hourly ensemble prediction of the flooding event with peak dated 7th Jun, 2020. Most ensemble members moderately underestimated the flood peak and the trajectories of the hourly EF members are mostly jumpy.

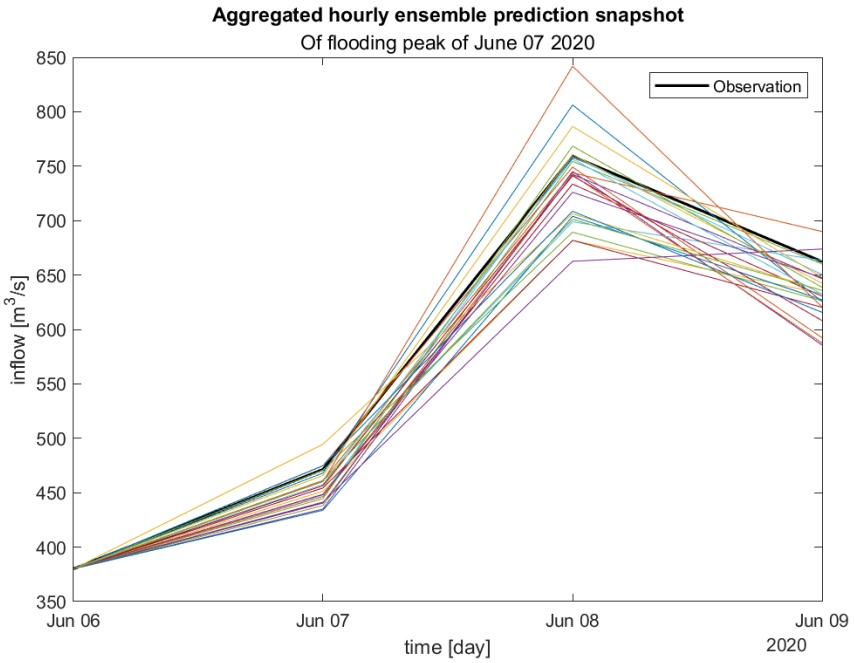

Figure 5.25: Aggregated daily ensemble prediction of the flooding event with peak dated 7th Jun, 2020 (from the original hourly EF above). Most ensemble members moderately underestimated the flood peak. The daily aggregation filters (smooths out) the jumpy trajectories of the hourly EF members.

6 | Conclusions and future work

The objective of this thesis was to assess whether and to what extent short-term hydrological forecasts can help improve the management of multi- purpose water reservoirs and regulated lakes with short- and long-term objectives. Lake Como, in northern Italy, is used as a case study. Located in a snow-dominated alpine basin, Lake Como is regulated for flood control and water supply. This study was one of the first to implement a deterministic and stochastic Model Predictive Control (MPC) exploiting real-time exogenous information for the management of this important regulated lake, after that only a few existing studies tested these approaches for other water reservoirs in the literature. For Lake Como, this was done in a simulation setting to investigate the possible improvements of these on-line control methods with respect to the historical management and off-line approaches considered as benchmarks, and to investigate the benefits of using available operational short-term (3-day) forecasts. Operational deterministic forecasts are already available in real-time with a 60-hour lead-time, but are not used in any optimal control scheme by the lake operator yet. Here, MPC was tested with both daily and hourly control frequencies and its performance was compared against two off-line control approaches that have already been extensively studied in literature, i.e. Deterministic (DDP) and Stochastic (SDP) Dynamic Programming. The results of deterministic MPC are also compared to the stochastic formulation of Tree-Based MPC (TB-MPC) using data-driven synthetic ensemble forecasts produced with a novel approach tailored to this specific case study, to investigate the possibility of unlocking even more benefits by considering the uncertainty associated to the inflow forecasts. TB-MPC was tested for both daily and hourly control frequencies (as well as MPC). The daily and hourly ensemble forecasts were produced with two synthetic stochastic forecast generation algorithms, driven by real deterministic forecasts and past observations. To our best knowledge this is the first study (or at least one of the first) where such data-driven probabilistic forecasts are applied to a real-world reservoir management problem, even if only through simulations. Ensemble forecasts of different quality were generated to investigate the impact of forecast skill characteristics on the simulations and performance of TB-MPC.

The first comparative analysis between the benchmarks showed that the Perfect Operating Policy (POP) embodied by the DDP significantly outperforms both SDP and the historical management for both objectives, suggesting the existence of room for improvement of the lake regulation by using perfect or skilful forecasts of inflows. Deterministic MPC fed with perfect forecasts and both daily and hourly control frequencies proved that indeed an on-line control scheme has the potential to reach the performance of the DDP, and surpass all benchmarks (in the hourly case). This is especially true in terms of flood control, as water scheduling capabilities are mostly tied by the length of the short control horizon of MPC and are all obtained by the use of a penalty cost. The refinement from daily to hourly control frequencies seems to bring only minor improvements for some flood events, as the long-term behaviour in both storage and release is mostly the same. The next comparative analysis followed the same procedure with real deterministic predictions instead of perfect predictions for MPC. The real short-term forecasts are found to be overall quite skillful in terms of error metrics and flooding prediction capabilities. The results of MPC fed by these forecasts reflect this, as apart from a minor degradation in performances this operative MPC proves to be quite competitive. It still outperforms SDP and the historical management, while is now slightly worse than DDP, which is however to be expected.
The multiple randomized calibration of neural networks method implemented is one of the simplest that have been proposed in the literature, yet it was remarkably able to produce satisfactory ensembles with performance metrics similar to the local operational (PROGEA) deterministic forecasts. This led to similar performances of deterministic MPC and stochastic TB-MPC at daily resolutions, with the latter scoring slightly better for the flooding objective with only a slightly higher water deficit. The ability of the neural networks to produce ensemble forecasts as skilled as deterministic forecasts of operational data centers specialized in producing them is an undeniable success. This is also because, even if in this particular scenario (i.e. our simulation of 8 years on historical data) there was no clear cut improvement, the key strength of TB-MPC optimizing over inflow trees is to, on average, beat deterministic MPC operations across more variable or unexpected scenarios. A stochastic control of this type is a "no regret" implementation that increases robustness and is able to cope with the drastic environmental changes and extreme events increasing frequency expected with climate change.

Finally, the comparison between daily and hourly deterministic MPC operations showed that although performances could slightly improve on the short term with the finer hourly control (on some flood events) at the cost of higher computational time, the long-term behaviour and water deficit performance is mostly the same, as expected. The results

have shown a lower return in benefits with the hourly implementation of the stochastic MPC. TB-MPC heavily depends on the quality of the EF utilized in the optimization and the finer control capabilities require finer representation of the inflow of the system. The hourly TB-MPC could theoretically achieve supplemental benefits if forecasts were more accurate, but these improvements seem to be minor with respect to the higher risk of over-confidently following wrong or jumpy forecasts. So it does not seem to be worth the effort to work towards extending the re-calibration of ANN method for hourly ensemble forecast generation and run a hourly TB-MPC given its extremely high computational burden and low expected return.
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, allowing to use forecasts with their uncertainty, which is more consistent at the daily scale given the current accuracy and jumpiness of forecasts.

6.1. Future work

The successful development of a local, computationally cheap and user friendly alternative to generate ensemble forecasts would definitely be of great value for water management around the globe, especially for those areas where available global or continental hydrological ensemble forecasts are not accurate or have too much bias to be used. Usually, local physically-based hydrological models with EF are not available yet, in most areas of the world (like for Lake Como), given the high computational burden and large amount of meteorological data needed. So future research should focus on refining the data-driven approach used in this study, for example testing other ANN with more refined structures than FFNN or testing more advanced techniques for randomization like singular value decomposition (SVD) or network dropout during reproduction (Scher & Messori 2021). This could improve the skill of the synthetic ensemble forecasts and allow for an extension to longer lead times. It would also be beneficial to investigate in a more systematic way the relationship between EF skill metrics and overall TB-MPC performances, including the ensemble spread-error relationship and its link with the control performance.

Future work could also investigate the use of alternative operational forecasts at longer lead times for the Lake Como case study, using the continental forecasting systems available, like the European Flood Awarness System (EFAS) at sub-seasonal and seasonal lead times. This would be possible only after applying post-processing and bias correction techniques to deal with the known low levels of forecast skill and large biases that such continental systems have in areas where they are not calibrated yet, as the Lake Como

basin (Arnal et al. (2018) and Wetterhall & Giuseppe (2018)). The control performances following this approach could be compared to the application of synthetic ensemble forecasts developed in our study.

Further work should assess in more quantitative terms the link between forecast accuracy and jumpiness at different temporal resolutions (e.g. daily/hourly) and the performance of daily/hourly TB-MPC over a larger set of extreme events (floods and droughts) and other case studies.

Future research should also tackle other practical aspects for the implementation of TB-MPC that would help its real-world deployment for water reservoir management, aiming particularly at reducing its computational time. As highlighted in this and other studies, TB-MPC has a significantly higher computational time than a standard MPC procedure, e.g. around seven times more than MPC with the setting used here and by Ficchí et al. (2016b). This is due to a higher number of optimization variables, and this requirement would grow even more larger ensemble sizes and longer lead times, that would probably require more refined optimization routines. It would be very beneficial to develop ways to distribute computational time of the optimization problem between different agents, like with a dual decomposition method showed by Maestre et al. (2012). In case of multiple reservoirs, a smarter way to alleviate computational burden could be developing filtering methods to pass local controllers only information of use for their task extracted from the ensemble, a very promising procedure initially proposed by Velarde et al. (2018).

Bibliography

Alessandrini, Sperati & Pinson (2013), 'A comparison between the ecmwf and cosmo ensemble prediction systems applied to short-term wind power forecasting on real data', *Water Resources management* **107**, 271–280.

Anghileri, Castelletti, M.ASCE, Pianosi, Soncini-Sessa & Weber (2013), 'Optimizing watershed management by coordinated operation of storing facilities', *Journal of Advances in Modeling Earth Systems* **139**(5), 492–500.

Anghileri, D., Pianosi, F. & Soncini-Sessa, R. (2011), 'A framework for the quantitative assessment of climate change impacts on water-related activities at the basin scale', *Hydrology and Earth System Sciences* **15**(6), 2025–2038.

Arnal, Cloke, Stephens, Wetterhall, Prudhomme, Neumann, Krzeminski & Pappenberger (2018), 'Skilful seasonal forecasts of streamflow over europe?', *Hydrology and Earth System Sciences* **22**(4), 2057–2072.

Asadzadeh, Leon, Yang & Bosch (2016), 'One-day offset in daily hydrologic modeling: An exploration of the issue in automatic model calibration', *Journal of Hydrology* **534**, 164–177.

Baldauf, Seifert, Frstner, Majewski & Raschendorfer (2011), 'Operational convective-scale numerical weather prediction with the cosmo model: Description and sensitivities', *American Meteorological Society* **139**, 3887–3905.

Bellman (1957), 'Dynamic programming', *Science* **30**, 7–34.

Bertsekas, D. P. & Tsitsiklis, J. N. (1995), 'Neuro-dynamic programming: an overview', *Proceedings of 1995 34th IEEE Conference on Decision and Control* **1**, 560–564 vol.1.

Buizza & Leutbecher (2015), 'The forecast skill horizon', *Quarterly Journal of the Royal Meteorological Society* **141**(693), 3366–3382.

Castelletti, de Rigo, Soncini-Sessa, Tepsich & Weber (2008b), 'On-line design of water reservoir policies based on inflow prediction', *IFAC Proceedings Volumes* **41**(2), 14540–14545. 17th IFAC World Congress.

Castelletti, Pianosi & Soncini-Sessa (2008a), 'Receding horizon control for water resources management', *Journal of Hydrology* **204**, 621–631.

Chang & Shenglian (2020), 'Advances in hydrologic forecasts and water resources management', *Water* **12**(6).

Criss & Winston (2008), 'Do nash values have value? discussion and alternate proposals', *Hydrological Processes* **22**(14), 2723–2725.

Denaro, Anghileri, Giuliani & Castelletti (2017), 'Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data', *Journal of Hydrology* **103**, 51–63.

Dobson, Wagener & Pianosi (2019), 'An argument-driven classification and comparison of reservoir operation optimization methods', *Advances in Water Resources* **128**, 74–86.

Dupačová, Grwe-Kuska & Rmisch (2000), 'Scenario reduction in stochastic programming an approach using probability metrics', *Mathematics Subject Classification* **95**, 493–511.

Fan, Schwanenberg, Alvarado, dos Reis, Collischonn & Naumman (2016), 'Performance of deterministic and probabilistic hydrological forecasts for the short-term optimization of a tropical hydropower reservoir', *Water Resources Management* **30**, 3609–3625.

Ficchí, Perrin & Andréassian (2016a), 'Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events', *Journal of Hydrology* **538**, 454–470.

Ficchí, Raso, Dorchies, Pianosi, Malaterre, van Overloop & Jay-Allemande (2016b), '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.

Fortin, Abaza, Anctil & Turcotte (2012), 'Why should ensemble spread match the rmse of the ensemble mean?', *Journal of Hydrometeorology* **15**(2014), 1708–1713.

Galelli, Castelletti & Goedbloed (2015), 'High-performance integrated control of water quality and quantity in urban water reservoirs', *Water resource research* **146**, 9053–9072.

Galelli, Goedbloed, Schwanenberg & van Overloop (2014), 'Optimal real-time operation of multipurpose urban reservoirs: Case study in singapore', *Journal of Water Resources Planning and Management* **140**(4), 511–523.

García-Herrera, Díaz, Trigo, Luterbacher & Fischer (2010), 'A review of the european

summer heat wave of 2003', *Critical Reviews in Environmental Science and Technology* **40**(4), 267–306.

Giuliani, Lamontagne, Reed & Castelletti (2021), 'A state-of-the-art review of optimal reservoir control for managing conflicting demands in a changing world', *Water Resources Research* **57**(12).

Giuliani, Li, Castelletti & Gandolfi (2016), 'A coupled human-natural systems analysis of irrigated agriculture under changing climate', *Water Resource Research* **52**, 6928–6947.

Giuliani, Zaniolo, Castelletti, Davoli & Block (2019), 'Detecting the state of the climate system via artificial intelligence to improve seasonal forecasts and inform reservoir operations', *Water resource research* **55**(11), 9133–9147.

Gneiting & Raftery (2005), 'Weather forecasting with ensemble methods', *Science* **310**(18), 248–249.

Grosso, Maestre, Ocampo-Martinez & Puig (2014), 'On the assessment of tree-based and chance-constrained predictive control approaches applied to drinking water networks', *Journal of Hydrology* **47**(3), 6240–6245.

Guariso, Rinaldi & Soncini-Sessa (1986), 'The management of lake como' a multiobjective analysis', *Water resource research* **22**(2), 109–120.

Kling, Fuchs & Paulin (2012), 'Runoff conditions in the upper danube basin under an ensemble of climate change scenarios', *Journal of Hydrology* **425**, 264–277.

Knoben, Freer & Woods (2019), 'Technical note: Inherent benchmark or not? comparing nash–sutcliffe and kling–gupta efficiency scores', *Hydrology and Earth Systems Sciences* **23**(10), 1–9.

Lehner, Bernhard, Dll, Petra, Alcamo, Joseph, Henrichs, Thomas, Kaspar & Frank (2006), 'Estimating the impact of global change on flood and drought risks in europe: A continental, integrated analysis', *Climatic change* **75**(3), 273–299.

Liu, Martina & Todini (2018), 'Flood forecasting using a fully distributed model: application of the topkapi model to the upper xixian catchment', *Hydrology and Earth System Sciences* **9**(4), 347–364.

Loucks & Beek, V. (2017), 'Water resource systems planning and management: An introduction to methods, models, and applications.'.

Maestre, Raso, van Overloop & de Schutter (2012), 'Distributed tree-based model predictive control on an open water system', **117**, 1985–1990.

Nappo, N., Ferrario, M. F., Livio, F. & Michetti, A. M. (2020), 'Regression analysis of subsidence in the como basin (northern italy): New insights on natural and anthropic drivers from insar data', *Remote Sensing* **12**(18), 2072–4292.

Nayak, Herman, Yang & Steinschneider (2018), 'Balancing flood risk and water supply in california: Policy search integrating short-term forecast ensembles with conjunctive use', *Water Resources Research* **54**, 7557–7576.

Palmer, Barkmeijer, Buizza & Petroliagis (1997), 'The ecmwf ensemble prediction system', *Meteorological applications* **4**, 301–304.

Palmer, T. (2019), 'The ecmwf ensemble prediction system: Looking back (more than) 25years and projecting forward 25years', *Quarterly Journal of the Royal Meteorological Society* **145**(S1), 12–24.

Pappenberger, F., Ramos, M., Cloke, H., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A. & Salamon, P. (2015), 'How do i know if my forecasts are better? using benchmarks in hydrological ensemble prediction', *Journal of Hydrology* **522**, 697–713.

Pianosi, F., Dobson, B. & Wagener, T. (2020), 'Use of reservoir operation optimization methods in practice: Insights from a survey of water resource managers', *Journal of Water Resources Planning and Management* **146**(12), 02520005.

Pianosi & Soncini-Sessa (2009), 'Real-time management of a multipurpose water reservoir with a heteroscedastic inflow model', *Journal of Hydrology* **45**, 621–631.

Powell (2007), 'Approximate dynamic programming: Solving the curses of dimensionality'.

Raso, Schwanenberg, van de Giesen & van Overloop (2014), 'Short-term optimal operation of water systems using ensemble forecasts', *Advances in water resources* **71**, 200–208.

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

Rasp & Lerch (2018), 'Neural networks for postprocessing ensemble weather forecasts', *Journal of Hydrometeorology* **10**, 3885–3899.

Scher & Messori (2018), 'Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning', *Geophysical Research Letters* **45**(22), 12,616–12,622.

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

Schwenzer, Ay & Abel (2021), 'Review on model predictive control: an engineering perspective', *The International Journal of Advanced Manufacturing Technology* **117**, 1327–1349.

Thielen, Schaake, Hartman & Buizza (2008), 'Aims, challenges and progress of the hydrological ensemble prediction experiment (hepex) following the third hepex workshop held in stresa 27 to 29 june 2007', *Atmospheric Science Letters* **9**(2), 29–35.

Tian, Goedbloed, Schwanenberg & van Overloop (2019), 'Multi-scenario model predictive control based on genetic algorithms for level regulation of open water systems under ensemble forecasts', *Journal of Water Resources Planning and Management* **140**(4).

Tian, Negenborn, Peter-Jules, van Overloop, Maestre, Sadowska & van de Giesen (2017), 'Efficient multi-scenario model predictive control for water resources management with ensemble streamflow forecasts', *Advances in water resources* **109**, 58–68.

Uysal, Alvarado-Montero, Schwanenberg & Sensoy (2018), 'Real-time flood control by tree-based model predictive control including forecast uncertainty: A case study reservoir in turkey', *Water* **10**(340), 1–22.

Velarde, Tian, Sadowska & Maestre (2018), 'Scenario-based hierarchical and distributed mpc for water resources management with dynamical uncertainty', *Water Resources management* **33**, 677–696.

Wetterhall & Giuseppe, D. (2018), 'The benefit of seamless forecasts for hydrological predictions over europe', *Hydrology and Earth System Sciences* **22**(6), 3409–3420.

Wilks & Hamill (2007), 'Comparison of ensemble-mos methods using gfs reforecasts', *Monthly Weather Review* **135**(6), 2379–2390.

Zhang, Chen, Li, Ding & Fu (2018), 'Water-energy-food nexus: Concepts, questions and methodologies', *Journal of Cleaner Production* **195**, 625–639.

Zsoter, Buizza & Richardson" (2009), '"jumpiness" of the ecmwf and met office eps control and ensemble-mean forecasts', *Monthly Weather Review* **137**(11), 3823–3836.

A | Low level days as a constraint

This appendix presents a small note on the tests carried out to include the low level days indicator (see Section 2.6) as a constraint. This has been ignored in the cost function used in the (TB-)MPC optimisation in the rest of the study, as the low-level days indicator is the less important of the three. However, it could still be taken into account by making use of a hard constraint, as one of the most distinguishing features of MPC is the possibility of including hard constraints within the optimization. This is quite straightforward from an implementation standpoint, the framework in Section 4.2 remains the same but the contribution to the total cost g_τ^3 reported in Equation 4.9 is cancelled and it is added an hard constraint of the form $h_\tau > h_{low}$, where $h_{low} = -0.2m$ is the threshold value defined by the regional authority.

The simulations with this alternative setting are carried out on the same period of 8 years ranging from 1st Jan, 2014 to 1st Jan, 2022, with the same weight choice for the flood and decifit objectives, same objective formulations, same requirements, and using perfect predictions.

The performance indicators reported in **Figure A.1** show that the low level days can be completely nullified, throughout the entire 8-year simulation period, as it can also be seen by looking at the (green) trajectory in **Figure A.2**. As expected, prohibiting MPC to lower the level of the lake under a certain threshold limits the quantity of water releasable to satisfy the water demand. The increase in the water deficit indicator is still very small, with an even lower influence on flooding incidence.

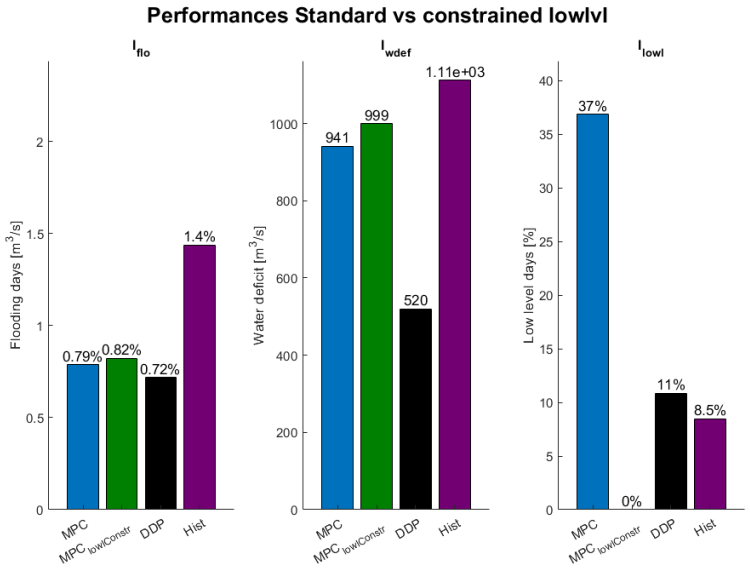

Figure A.1: Performance indicators throughout the entire simulation period of 8 years, for the deterministic MPC with the low-level hard constraint using perfect forecasts, standard MPC with no constraint, DDP and the historical management.

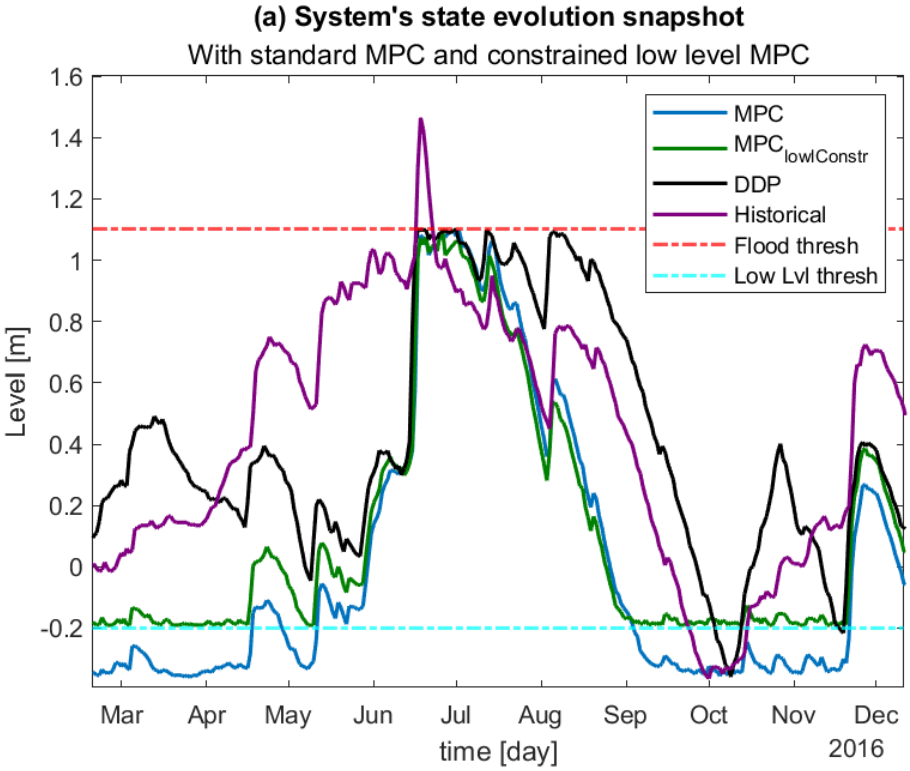

Figure A.2: System state evolution in a 9-month period with flooding from Mar 2016 to Dec 2016, for MPC with the low-level hard constraint using perfect forecasts, standard MPC with no constraint, DDP and the historical management.

List of Figures

List of Tables

Acronyms

ACR Adaptive Control Resolution. 9

ANN Artificial Neural Networks. 7, 79

CV Coefficient of Variation. 27

DDP Deterministic Dynamic Programming. i, 38, 55, 56, 58, 77

ECMWF European Center for Medium-Range Eather Forecasts. 6, 29, 30

EF Ensemble Forecast. i, 8, 9, 32, 35, 46

EPS Ensemble Prediction System. 6, 33

FFNN Feed Forward Neural Networks. 30, 31, 32, 53, 79

GA Genetic Algorithm. 9

IDSS Integrated Decision Support System. 11

IWRM Integrated Water Resources Management. 1, 3, 38

KGE Kling-Gupta Efficiency. 26, 27

KGEss KGE Skill Score. 27, 50

KNN K-Nearest Neighbor. 33

MAE Mean Absolute Error. 25, 49

MEF Minimum Enviromental Flow. 20, 22, 43

MMPC Multiple MPC. 8, 9

MOS Model Output Statistics. 33

MPC Model Predictive Control. i, 2, 4, 8, 9, 41, 42, 44, 52, 57, 58, 59, 62

MSE Mean Square Error. 25, 49

MSMPC Multi Scenario MPC. 9

NSE Nash-Sutcliffe Efficiency. 26, 27

NWP Numerical Weather Prediction. 6, 7, 8, 18, 30, 32

POLFC Partial Open Loop Feedback Control. 3, 4

POP Perfect Operating Policies. 38, 55

RMSE Root Mean Square Error. 25, 49

SDP Stochastic Dynamic Programming. i, 2, 3, 4, 38, 39, 55, 56, 57, 58, 77

SWE Snow Water Equivalent. 15, 16

TB-MPC Tree Based MPC. i, 8, 9, 44, 46

VE Volumetric Efficiency. 29, 52

Acknowledgements

I would like to thank Professor Andrea Castelletti for giving me the opportunity to work on such an interesting project regarding an often taken for granted natural resource. I would also like to thank Dr. Andrea Ficchí for guiding me and helping me fill up my lacks of knowledge in this field of study and for his invaluable support throughout the entire work, without whom I could have never been able to complete. I would like to thank Raffaele Cestari, for all of his clever inputs on encountered issues and smart takes on code troubleshooting. Last but not least, thanks also to Professor Matteo Giuliani, for the useful discussion and advice on the case study details and facets.

Special thanks go also to my family, and to my friends Eric Davide and Marco for being there when I needed it.