A centralized real-time controller for reservoirs management on the Seine river using ensemble weather forecasting

A. Ficchì

POLITECNICO DI MILANO
Facoltà di Ingegneria
Scuola di Ingegneria Civile, Ambientale e Territoriale
Master of Science in Environmental and Land Planning Engineering

A centralized Real-Time Controller for Reservoirs Management on the Seine River using Ensemble Weather Forecasting

Supervisor:
Andrea Castelletti

Assistant supervisors:
Francesca Pianosi
David Dorchies

Master of Science Thesis by:
Andrea Ficchì
Student Id. Number: **766183**

Academic Year: 2012-2013

POLITECNICO DI MILANO
Facoltà di Ingegneria
Scuola di Ingegneria Civile, Ambientale e Territoriale
Corso di Laurea Magistrale in Ingegneria per l'Ambiente e il Territorio

A centralized Real-Time Controller for Reservoirs Management on the Seine River using Ensemble Weather Forecasting

Relatore:

Prof. Andrea Castelletti

Correlatori:

Prof. Francesca Pianosi

Ing. David Dorchies

Tesi di Laurea di:

Andrea Ficchì

Matricola: **766183**

Anno Accademico: 2012-2013

Alla mia famiglia che mi ha sempre sostenuto e incoraggiato lungo il mio percorso.

Acknowledgements

This report describes the research I did for my Master thesis at UMR G-Eau of Irstea - Montpellier, France, in collaboration with the Water Resources Management group at Delft University of Technology, the Netherlands. It has been realized under the supervision of Prof. Francesca Pianosi as Master thesis supervisor of Politecnico di Milano, Italy and David Dorchies as internship tutor at Irstea - Montpellier. I am grateful to them for the opportunity they gave me to work on this interesting project and for their continuous teaching and support.

I owe a big thank to all the team that collaborated with me and my supervisors in the development of this work for the kind and constant support and the inspiring conversations. In particular, thanks to: Luciano Raso and Peter-Jules Van Overloop of TU Delft, for their teaching on control theory and their hospitality at Delft; Maxime Jay-Allemand and Pierre-Oliver Malaterre for their close collaboration at Irstea and their useful suggestions.

I'm grateful also to all the other persons that collaborated with us in this project; in particular, thanks to: Guillaume Thirel and Maria-Helena Ramos of Irstea - Antony, that provided us essential data and useful support for this work.

I would like to thank also all the researchers and friends of G-Eau for the pleasant time spent together at Montpellier during the period of development of this thesis.

Last but not least, I can't forget my family that always supported me throughout my studies and my girlfriend for all the love and attention in the last months.

This work has been for me a stimulating experience from both a scientific and human point of view. It has been a great opportunity for learning, not only about water management, that is the subject of this thesis, but also about the world of scientific research, increasing my interest in this field.

Contents

Abstract

The reservoirs on the Seine River basin, upstream of Paris, are regulated with the objective of reducing floods and augmenting low flows. Their current management is empirical, reactive and decentralized, mainly based on filling curves, constructed from an analysis of historical floods and low flows. Therefore, the efficiency of this management strategy could be enhanced when inflows are significantly different from their seasonal average. Adaptation to climate change is also a challenge, for the possible modification of future hydrologic conditions.

To improve such management strategy, the use of a centralized real-time controller is investigated. The control method used is Tree-Based Model Predictive Control (TB-MPC), a proactive method that uses all the information available in real-time, including ensemble weather forecasts. This information is used in the model predictive control framework, to optimize an objective function over a finite receding horizon, using a model to predict the evolution of the system in response to the forecasted inputs.

The TB-MPC controller is implemented in combination with the model of the Seine river basin, including a semi-distributed hydrologic model of the watershed, a simplified hydraulic model of the river network and the four reservoirs models. The controller optimizes a global cost function that takes into account the costs associated to high and low flows, based on critical thresholds at some key downstream stations, and a penalty based on the final storages of the reservoirs, to guarantee a sustainable management in the long term.

The reservoirs management is tested using different weather forecasting models: (i) a hypothetical perfect predictor that takes observations as forecasts; (ii) a real deterministic forecasting model; (iii) a stochastic model producing ensemble forecasts. The performance of MPC using perfect predictions is compared with the actual management, to assess the expected improvement by moving to a centralized and proactive controller. The performance of TB-MPC is compared with that of deterministic MPC, to evaluate the possible improvement due to the integration of the uncertainty of predictions.

Riassunto

I serbatoi nel bacino idrografico della Senna, a monte di Parigi, sono gestiti con l'obiettivo di ridurre le piene e aumentare le basse portate. La loro gestione attuale è empirica, reattiva e decentralizzata, basata principalmente su curve di riempimento, costruite a partire da un'analisi statistica di piene e magre storiche. Pertanto, l'efficienza di questa strategia di gestione può essere migliorata quando gli afflussi sono significativamente diversi dalla loro media stagionale.

Per migliorare tale strategia di gestione, si indaga l'uso del *Tree-Based Model Predictive Control* (TB-MPC), un metodo di controllo centralizzato e proattivo, che utilizza tutte le informazioni disponibili in tempo reale, tra cui le previsioni meteorologiche di *ensemble*. Questa informazione viene utilizzata nel quadro del controllo predittivo, per ottimizzare una funzione obiettivo su un orizzonte finito, usando un modello matematico per prevedere l'evoluzione del sistema in risposta agli ingressi previsti.

Il controllore TB-MPC è implementato in combinazione con il modello del bacino della Senna, che integra un modello idrologico semi-distribuito, un modello idraulico semplificato della rete fluviale ed i modelli dei quattro serbatoi. Il controllore ottimizza una funzione di costo globale che tiene conto dei costi associati ad alte e basse portate, calcolati sulla base di soglie critiche definite per alcune stazioni a valle, e una penale basata sui volumi finali dei serbatoi, per garantire una gestione sostenibile nel lungo termine.

La gestione dei serbatoi è simulata testando diversi modelli di previsione meteorologica: (i) un ipotetico previsore perfetto che usa osservazioni come previsioni; (ii) un vero modello di previsione deterministico; (iii) un modello stocastico che produce previsioni di ensemble. La performance del MPC deterministico usando previsioni perfette è confrontata con quella della gestione attuale, per valutare il miglioramento atteso dal passaggio a un controllore centralizzato e proattivo. La performance del TB-MPC è confrontata con quella del MPC deterministico, per valutare il possibile miglioramento dato dall'integrazione dell'incertezza delle previsioni.

List of figures

List of tables

Acronyms and symbols

Acronyms

fr.: French acronym or words
BV : watershed *(fr. Bassin Versant)*
BVI : intermediate watershed *(fr. Bassin Versant Intermédiaire)*
EPST: Public institution for Scientific and Technological research *(fr. Etablissement Public à caractère Scientifique et Technologique)*
ETP : Potential Evapotranspiration
FC *(or OFC)* : Objective Filling Curve
GCM : Global Circulation Model
G-EAU : IRSTEA research group in water management *(fr. Gestion de l'Eau, Acteurs et Usages)*
GR4J : hydrological model with 4 daily parameters *(fr. Génie Rural à 4 paramètres Journaliers)*
Irstea : French public research institute; *fr. Institut de Recherche en Science et Technologies pour l'Environnement et l'Agriculture*
LR *(or LLR)* : hydraulic model *(fr. Linear Lag and Route)* with delay and propagation
MPC : Model Predictive Control
PP : Perfect Predictor
RF : Real Forecasts
SGL : *fr. Seine Grands Lacs*
TB-MPC : Tree-Based Model Predictive Control
TGR : rainfall-runoff model; *fr. Transfert Génie Rural*
UMR : *fr. Unité Mixte de Recherche*

Symbols

t : time
E : evapotranspiration
P : rainfall
$\mathbf{P^f}$: forecasted rainfall
I : inflows
$\mathbf{Q_{river}}$: river flow
V : volume
$\mathbf{Q_{sim}}$: simulated flow
$\mathbf{Q_{obs}}$: observed flow
$\mathbf{Q_{ref}}$: reference flow
$\mathbf{Q_{res}}$: reserved flow
u : control decisions
$\mathbf{V_{curve}}$ **or** $\mathbf{V_{FC}}$: theoric volume of the lake given by the management filling curve
$\mathbf{V_{lake}}$: lake volume

1. Introduction

The present work is a case study application of *Climaware*, an European project on the study of climate change impacts on water resources management. The Climaware project is funded by the Integrated Water Resources Management Network (IWRM-Net). The duration of the project is three years from September 2010 to December 2013.

The project has five partners: (i) the Department of Hydraulic Engineering and Water Resources Management, University of Kassel, Germany; (ii) the Center for Environmental Systems Research (CESR), University of Kassel, Germany; (iii) Irstea (ex Cemagref) at Antony and Montpellier in France; (iv) The Public Establishment of Territorial Basin (EPTB) and Seine Grand Lake (SGL), based in Paris, France; (v) Istituto Mediterraneo Agronomico di Bari, Bari, Italy.

The Climaware project aims to study the effects of climate change on the hydro-morphological conditions as well as on the changes in river flows and their consequences (frequency of floods and droughts). Another of its goals is to examine the uncertainties of models and scenarios. Finally, it should help to define adaptation strategies for the programs within the framework of the IWRM-NET Funding Initiative, the management of dams and irrigation practices.

Climaware includes three case studies used to investigate the impacts of climate change on water resources at a regional scale and to focus on different aspects:

- The first case study focuses on the hydro-morphology; it is located in Germany on the Eder River watershed.
- The second regards the dam management of the Seine lakes in France; it is the focus of this work.
- The latter is about the agricultural use of water in Apulia region in southern Italy.

Once these three studies are completed, the results can then be compared and it will be possible to see whether regional measures could be transferred on a larger scale, in an European perspective. This organizational perspective is represented in Figure 1.

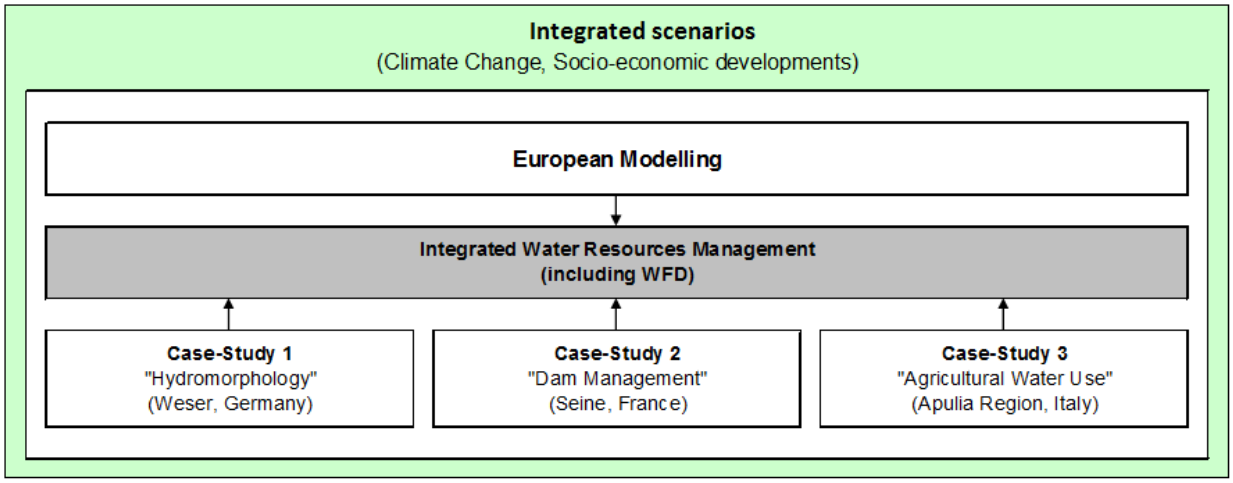


Figure 1. The organizational framework of the Climaware Project and its 3 case-studies. [ClimAware website - www.uni-kassel.de/fb14/wasserbau/CLIMAWARE/home/home.html]

The work presented in this thesis deals with case study 2, that is the multi-objective reservoirs management on the Seine River in France. This work has been conducted in partnership between Irstea - Montpellier (UMR G-Eau), TU - Delft (Water Resources Management group) and Politecnico di Milano (DEIB).

In the Seine river basin (total size of 78,650 km²), located in northern France, four large dams with a total storage capacity of 800 hm^3 are managed for low-flow augmentation and flood alleviation by the "Seine Grands Lacs" institution [www.seinegrandslacs.fr]. They were built after dramatic floods (in 1910 and 1924) and droughts (in the 1920s), which had major consequences on the Paris area [Ambroise-Rendu, 1997]. Currently, the 30% of the French population live on the Seine basin, including the capital city of Paris and its suburbs. The basin has a major economic role in France because it provides drinkable water to almost 20 million persons, includes a lot of factories, and is a major agricultural and touristic region.
For all these reasons, better knowing and anticipating the impacts of climate change on the Seine River basin becomes critical. Indeed, more frequent droughts or flood events would have major consequences given the already high pressure that local population and activities put on the available water resources, and the dense urbanization of the river banks in the Paris area. To further protect Paris, some experts recommend building a new reservoir lake, the Bassée to absorb a flood on the Yonne; the cost of this project is estimated to 500 million of euros, but the debate opened in 2001 is still open [Landrin, 2013].

The objective of the case study is to provide the managers of the reservoirs in the Seine River basin an analysis framework to evaluate the possible consequences of climate change on the basin hydrological behavior and to assess possible adaptation strategies they could consider in the future, as more efficient management strategies. These adaptation strategies could be necessary to continue to fulfill the objectives of floods prevention and water supply in the future, and may represent a viable alternative to building new dams.

The key phases of the work on this case-study are to:

1. develop an integrated model of the basin (hydrologic and hydraulic), including the artificial influences of the dams;
2. define climatic as well as socio-economic changes over the basin by the mid-century;
3. evaluate the sustainability of current management rules of the reservoirs (filling curves and river flow thresholds);
4. define adaptation strategies for a long term (adaptation of storage capacities), medium term (adaptation of management rules) and short term (management of the reservoirs in real-time) perspective.

In particular the research presented here is related to the last phase of the work proposing the use of a Real-Time Controller (RTC) as approach to solve the issue of reservoirs management adaptation to changes in hydrological conditions due to climate change. Our goal is to overcome the limitations of the current management approach by implementing on the four reservoirs a centralized RTC using the Model Predictive Control (MPC) method and all data available, including *ensemble weather forecasts*. We will be able to simulate this kind of management on the past climate scenario with the observations available and on some future hypothetic scenarios derived by climate models. The results could help to investigate the effectiveness of a centralized RTC as mean to increase water resources efficiency and better tackle water stress due to climate change.

The present thesis is organized as follows:

- Chapter 2 briefly describes the Seine River basin case study, the current management rules of the reservoirs and the data used in this work;

- Chapter 3 presents the theory behind classic Model Predictive Control (MPC) and Tree-Based MPC (TB-MPC), the new method that will be used to face with forecasts uncertainties;
- Chapter 4 is dedicated to the formulation of the control problem for the case study, for implementing the MPC and TB-MPC management methods;
- Chapter 5 shows some results of the MPC and TB-MPC simulation experiments done and provides performance comparisons between MPC and current management and between TB-MPC and MPC;
- Chapter 6 highlights some conclusions, and presents the on-going research and the possible future developments of the work.

2. Case study: reservoir's management on the Seine River

In this chapter the case study will be presented and contextualized. It regards the reservoirs management in the Seine river watershed upstream of Paris. Following the Participatory and Integrated Planning procedure (PIP, [Soncini-Sessa et al., 2007]), the first phase of the problem definition for water resources planning and management is the *Reconnaissance* phase. This is concentrated on defining the spatial and temporal boundaries of the system being considered, the data available and to be collected, the normative and planning context and the goal of the project. The reservoirs and their actual management rules will be presented, highlighting the limitations of the current management strategy faced in the past and the possible limitations that could be encountered in the future due to climate changes. The objective of this work will be introduced as to develop a possible adaptation strategy in order to respond to these limitations.

2.1 The Seine river basin

The system being considered in this study is the Seine river basin upstream of Paris, with outflow in Paris-Austerlitz, including the sub-basins of the tributaries: Aube, Seine, Yonne, Marne and Blaise. This study area has a surface of 43 824 km^2. The position of the case-study area in France is highlighted in the following Figure 2.

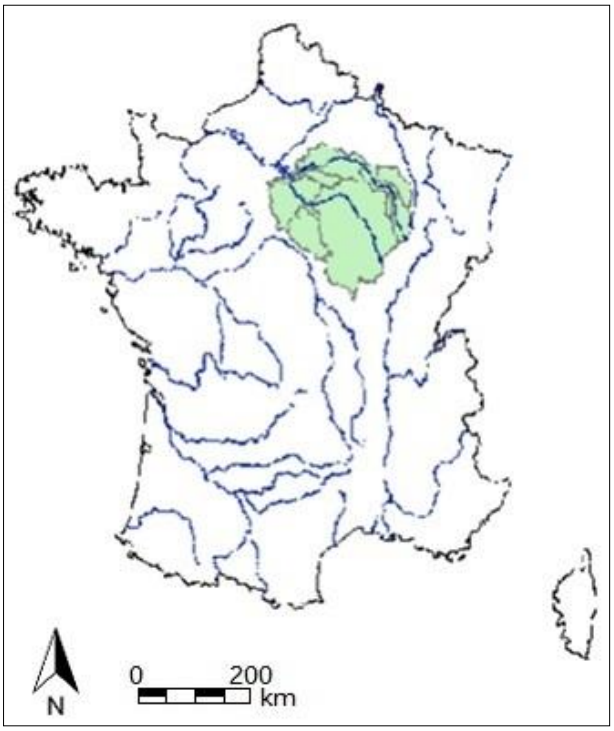


Figure 2. Position of the case-study area of the Seine river basin in France. [Dorchies et al., 2013]

The Seine river basin represents important socio-economic stakes in France, especially because of the Paris urban area along the Seine River. The presence of cities and industries is the cause of high water demand and vulnerability to floods.

In order to limit flooding in Paris and supporting low flows, four lakes-reservoirs were constructed on the main tributaries: the *Aube* and *Seine* lakes (on the homonymous rivers), the *Pannecière* on the Yonne, and the *Marne* between the Marne and the Blaise. These four reservoirs have a total capacity of 810 hm^3 and are managed by *Seine Grand Lacs* (SGL). Only one of these reservoirs, Pannecière, is directly on the river; the other three lakes have inlet derivations (one derivation for Aube and Seine, two derivations for Marne) and one outlet derivation. The reservoirs were built in different years between 1950 and 1990: Pannecière in 1958, Seine in 1965, Marne in 1973 and Aube in 1987.
Figure 3 shows a map of the case-study catchment locating the four reservoirs and the 25 gauging stations spread over the main stream and its tributaries. A technical scheme for each lake is provided in Appendix A, reporting the capacity of each reservoir and its inlet/outlet channels. A list of the 25 gauging stations with the respective positions and areas of influence is reported in Appendix D.

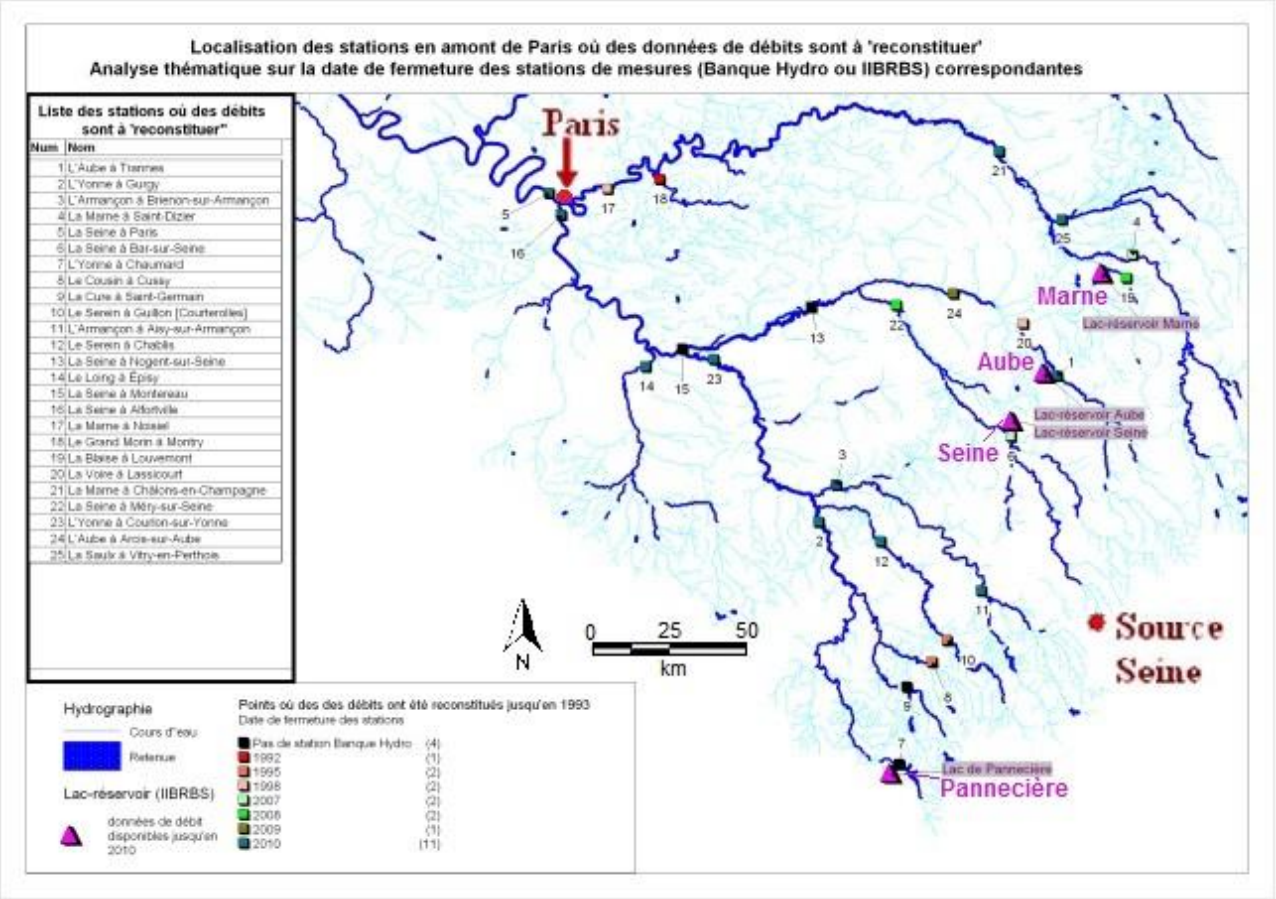


Figure 3. Case study area of the Seine river basin with the localization of the four reservoirs (triangles) and the 25 runoff gauging stations (square dots).

The climate of the basin is the typical western European oceanic climate affected by the North Atlantic Current (Köppen climate classification: *Cfb*). The temperatures ranges are moderate, with cool summers (between 15 and 25 °C) and mild/cold winters (around 7° C). The precipitations occur all year with an annual average of 700 mm/y. The natural flow regime is characterized by low flows in summer and high flows in winter (see Figure 4). For an exhaustive description of the physical, meteorological and hydrological characteristics of the Seine river basin, the reader may refer to Ducharne et al. [2007].

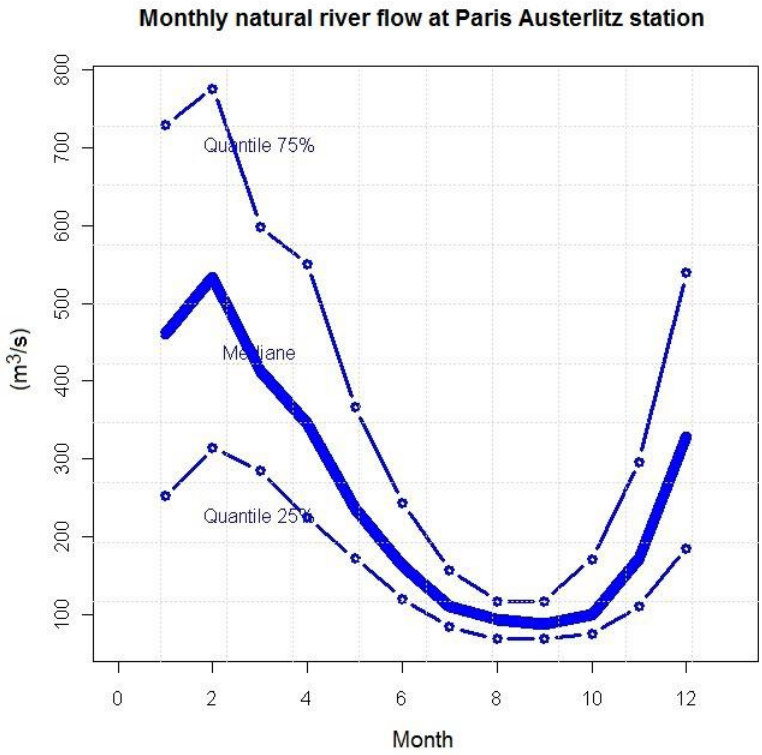


Figure 4. Monthly natural runoff on the Seine river at Paris Austerlitz station, for the period 1958-2010 (the 25%, 50% and 75% percentiles are drawn).

The water demand is high due to both city and industrial demands, and can be met in summer period thanks to the oceanic climate, the positive influence of aquifers and the management of reservoirs that sustain low flows.

2.2 Data used and temporal boundaries of the study

The *temporal boundaries* considered in this study depend on the time horizon considered in the previous part of the work for this case-study of Climaware and on data availability for the present work. In the Climaware project, for the simulation of the actual management rules over past and present scenarios, the horizon from 1961 to 2009 has been used, since climate and inflows measures were available over this period. Time-series of ensemble and deterministic weather forecasts were recovered for the period 2005 – 2009.

2.2.1 Data survey

Hereafter we present the data used for the present research.

- Data-base of naturalized flows series at a daily time-step (calculated removing the influence of the dams from the observed flows) provided by Seine Grands Lacs for the period 1900-2009.

- Meteorological observations at a daily time step on an 8x8 km grid over the past period 1958 - 2009 from the SAFRAN database from Meteo-France. They include: total precipitation (P), temperature (T) and potential evapotranspiration (ETP) calculated with the Penman-Monteith formula.

- Hydrologic, topographical and morphological data related to the Seine basin (distances, areas, etc.) used before this work to calibrate and build the hydrologic and hydraulic models of the basin.

- Seven different climatic scenarios at a daily time step, over the periods 1961-1991 and 2046-2065, produced by seven Global Circulation Models (GCM), forced by the moderate A1B green-house gases emissions scenario. These simulations were statistically downscaled at an 8 km x 8 km scale, using the DSCLIM downscaling method [Boè et al. , 2007].

- Data and information on the four reservoirs management:
    - current management rules, expressed as objective filling curves (see section 2.4.1);
    - thresholds for minimum authorized flows and reference flows (see section 2.4.2);
    - time-series of inlet/outlet flows and storages of the 4 reservoirs;
    - maximum and minimum volumes;
    - capacity of the inlet/outlet channels.

- Thresholds values for the flows of the river system at different control stations used by SGL for the performance assessment of the management.

2.2.1.1 *Weather forecasts*

The weather forecasts of rainfall are extracted from the ECMWF data-base [www.ecmwf.int] of ensemble weather forecasts for the case study area.

The ensemble forecasts data available are composed by 52 members (predictions):
- The 1^{st} - 50^{th} members are the actual ensemble forecasts members; they come from the same model as the 51^{st} member but the initial conditions of the model are perturbed.
- The 51^{st} member is a deterministic forecast called “control” forecast; it's used for generating the ensemble members, i.e. forecast realized at the same resolution of the ensemble members, but not perturbed as the members 1^{st} - 50^{th}.
- The 52^{nd} member is an accurate deterministic forecast, i.e. forecast realized at a more detailed resolution.

The temporal features of the EF are:
- Forecast time step: 1 day;
- Forecast horizon: 9 days;
- Ensemble and deterministic forecasts available from 11/03/2005 to 01/10/2008.

The spatial features of the EF are:
- Spatial resolution: 0.5° x 0.5°;
- One forecast for all the watershed.

For the spatial resolution, the rainfall data in the semi-distributed hydrological model (GR4J) are aggregated at the BVI scale. The extension of these BVIs is between 200 km^2 and a few thousand km^2. So, the spatial resolution of the ensembles is not too far from our needs and in the future development of the research we could be able to downscale the data at the BVI scale.

2.2.2 Scenarios

Based on the data available, it was therefore possible to run simulations over different periods and scenarios. The following ones were defined:

- “NAT” (Natural): observed climate over the period 1958-2009;

- "NTP" (Natural Present Time): observed climate over the period 1961-1991, with all the lakes virtually in operation from the beginning of this period;
- "TP" (Present Time) : GCMs scenarios over the period 1961-1991;
- "TF" (Future Time) : GCMs scenarios over the period 2046-2065.

2.3 Project goal definition

Since the project goal depends on the interests being considered and on the expectations that one wants to fulfill, it is good to start the definition of the objectives for the case-study from the identification of the *stakeholders* and *Decision Makers (DM)* involved and their *interests*.

The *stakeholders* are:

- the *urban areas*, i.e. *population of Paris* and other cities in the basin and *industries*;
- users and providers of *navigation* service on the Seine;
- the hydroelectric production plants manager.

The *Decision* Maker is:

- EPTB «Seine Grands Lacs» (SGL, [www.seinegrandslacs.fr]), also named «Institution Interdépartementale des Barrages-Réservoirs du Bassin de la Seine» (IIBRBS). SGL is a 40-year old French public establishment gathering the city of Paris and three administrative departments (Hauts de Seine, Seine-Saint Denis and Val de Marne). SGL is the public authority responsible for the coordination of the upper Seine basin management and the management of the four artificial reservoirs.

The objective of the dams is to satisfy the interests of the stakeholders that can be summarized as:

- protecting houses and industry against floods;
- sustaining water demand for drinking water, industry, navigation and energy production.

The purposes of sustaining the water demand and sustaining the navigation are not conflicting, so we can take into account them together. The hydroelectric production is a minor goal and hereafter it's not considered.

2.4 Actual management of the lakes

2.4.1 Objective Filling Curves (FC) and current management rules

The actual management of the reservoirs is done *empirically* and *decentralized* following some *Objective Filling Curves* (FCs) constructed off-line for each reservoir and respecting some constraints downstream the connections of the lakes. The FCs are curves that plot the target volume of the lake in function of the day of the year, based on historical flood and low-flows statistics. They are constructed in order to store water for the low-flows season (from beginning of July to end of October) filling the reservoirs from autumn (November) to spring-summer (June). The target volume to reach at the end of the filling season is calculated in order to have enough water to satisfy a minimal flow threshold in the river for all the releasing season. The slopes of the curve in the filling period are calculated on the basis of the average inflows to the reservoirs. The releasing period can be prolonged if the river flow at the control stations immediately downstream the dams is under the second low-flow threshold (see section 2.5.1) until the 1st of January at most. In this condition of prolonged or delayed droughts the releasing period ends when the flows begin to

be higher than this threshold. For example, the filling curve of the lake Marne is shown in Figure 5. The filling curves of all the lakes and the maximum and minimum volumes are provided in Management Filling Curves of the lakes

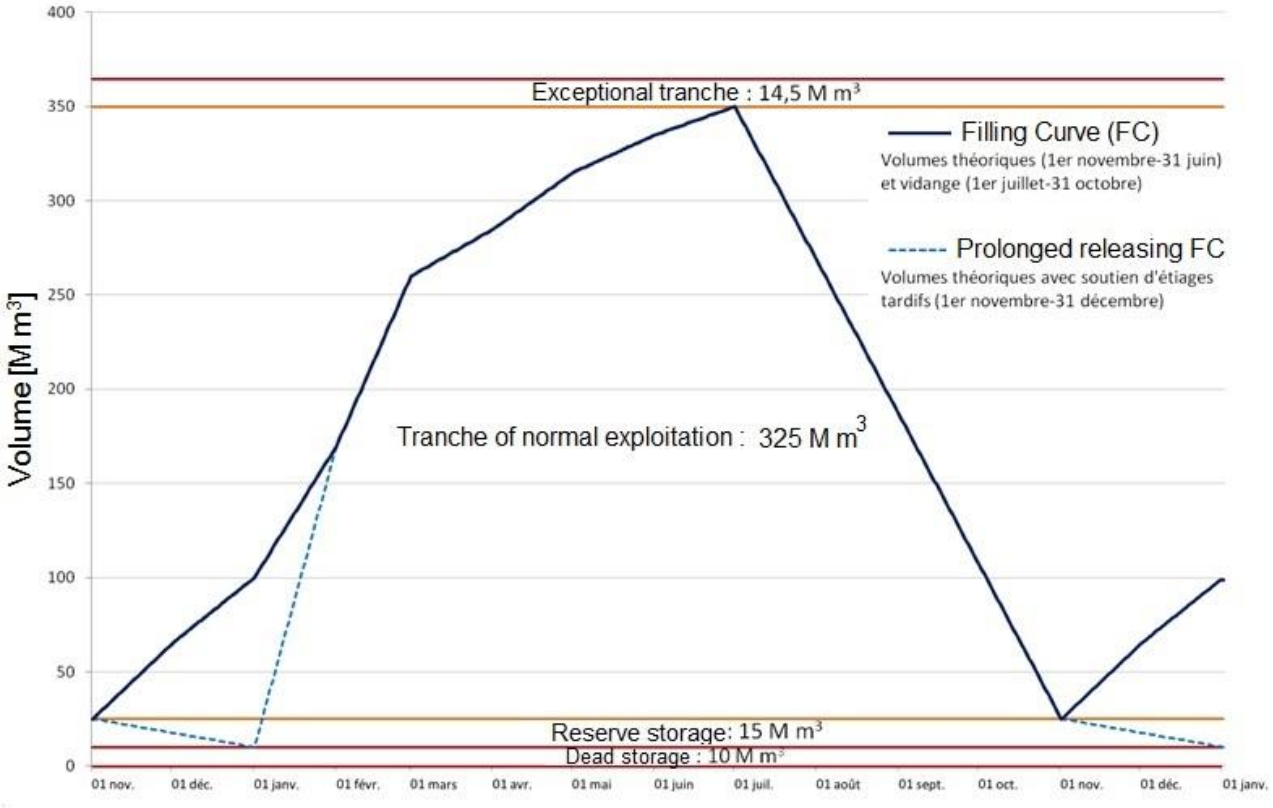


Figure 5. Filling Curve (FC) of the lake Marne (continuous blue line) and curve for prolonged releasing (dashed blue line).

It should be noted that in the releasing season the actual behavior of the managers of SGL is emptying the lakes from the actual point of maximum filling and if this point does not coincide with the maximum point of the filling curves, the rule is trying to empty the lakes to get to the end-point of the filling curves at the end of the releasing season. In other words the manager replaces the descending branch of the FC with a straight line joining these two points.

The filling trajectory is calculated taking into account the historical floods of the 20th century so that there is enough space (storing capacity) available for these floods control. The storing capacity available is given by the difference between the maximum capacity and the current volume. For example, Figure 6 reports the filling curve of lake Marne and the storing capacities necessary for controlling the historical floods of 20th century. The storing capacity required for these floods is represented by the distance of the red points from the line of the maximum capacity of the lake. So this capacity is not available when a red point is below the FC, considering the volume of the lake equal to the target one. There were two flood events of the 20th century that with the actual management rules couldn't be controlled: those of June 1983 and August 1910. As shown in Figure 6, following the filling curve, in June and August the reservoir is full and there was not enough storing capacity to control these two floods. Besides the 1983 and 1910 events, there are some other critical events (1970, 1988, 1999 and 2001) in which the capacity available in the reservoir Marne is just enough to control the floods.

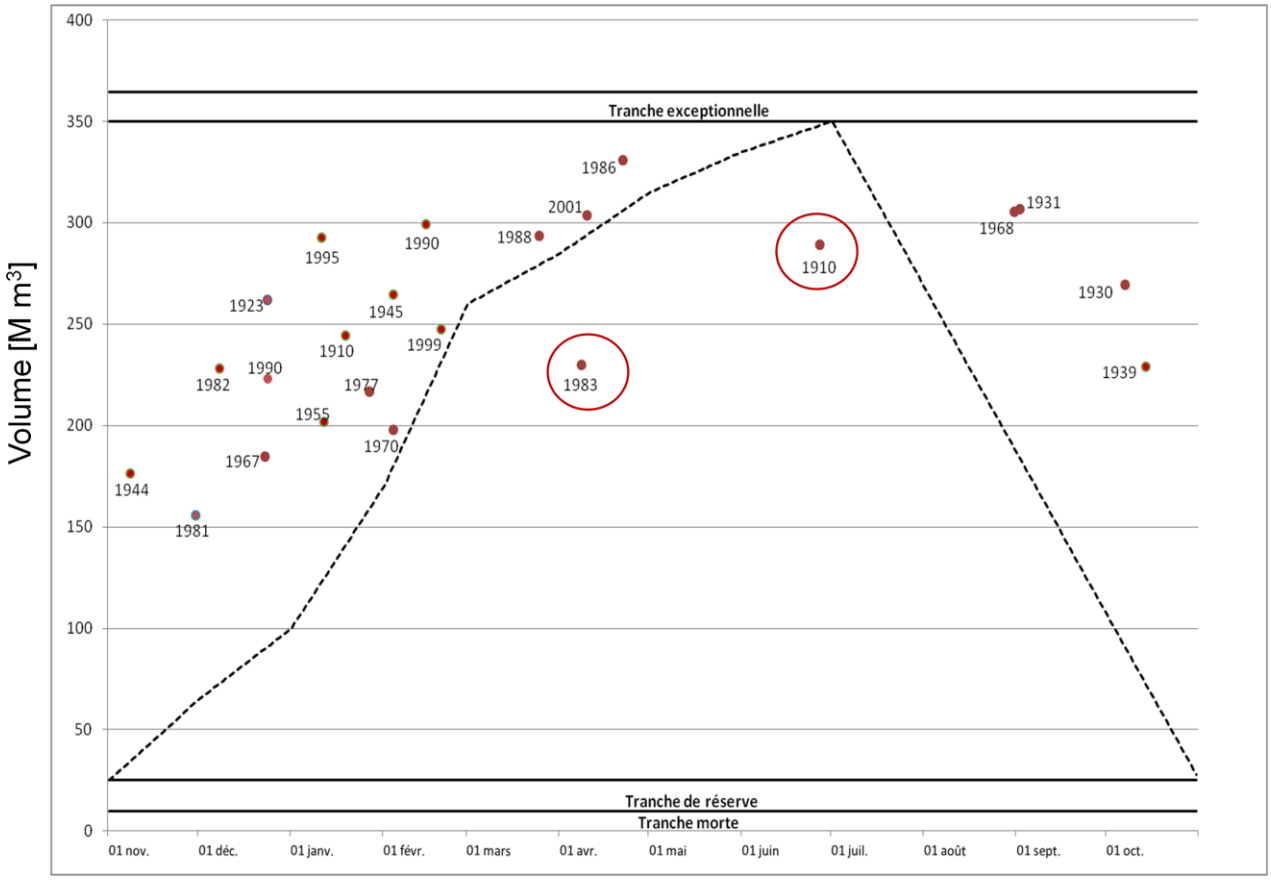


Figure 6. Filling Curve (FC) of the lake Marne and storing capacity necessary for controlling the historical floods of 20th century. The red dots represent the historical floods: the distance of these points from the line of maximum capacity of the lake is the required storing capacity for containing these floods.

2.4.2 Constraints on river flows: reserved and reference flows

In addition to following the filling curves, the actual management rules has to respect some legal constraints on the river flows, if possible. There are the following constraints:

- ***Reserved flow***: is a minimum flow to be left in the river in order to assure life, movement and reproduction of all species in the river. This discharge is a *legal obligation* (Article L432-5 of the French Environmental code); the reserved flow value changes from a season to another. In case of reservoirs with derivations this threshold defines the minimum flow to let in the river downstream the inlet channel, while for reservoirs directly on the river, as Pannecière, it indicates the minimum flow to be discharged.

- ***Reference flow (or Retention flood level)***: is the threshold which indicates the *occurrence of a flood* on the sub-basin *downstream the dam*. In the current reservoirs management, if downstream the inlet and outlet channels the flow overcomes this value, *the excess flow is stored in the reservoir*, over-filling it respect to the Filling Curve volume. The value of the retention flood level depends on the season, because the flood areas depend on agricultural use of the adjacent territory.

The values of reserved and reference flows taken into account in our work are reported in Appendix A.

2.5 Efficiency criteria for reservoirs management

2.5.1 Low and high flows thresholds

For assessing the management of the reservoirs, the manager SGL monitors the flow in the river at several strategic stations downstream the four lakes. At each of these gauging stations, some flow thresholds have been defined for critical low and high flows.

For low-flows, the thresholds are defined at 9 downstream gauging stations, from regulatory thresholds corresponding to restrictions on the water withdrawals:

- *Vigilance threshold*: at this first threshold any restriction of uses is defined but the river is extremely sensitive to pollutions;
- *Alert threshold*: at which 30% restriction of uses;
- *Reinforced alert threshold*: 50 % restriction of uses;
- *Crisis threshold*: all uses are prohibited except a minimum use for drinking water.

The first *vigilance threshold* can be used to derive the filling curves, calculating the volume to reach at the end of the filling season in order to be able to satisfy this flow in the river throughout the releasing season, discharging a constant flow from the reservoirs.

For high flows, the same control stations as the ones used for low flows are used for convenience. The thresholds are close to the ones used by forecast prediction services for alerting population in case of flood. The thresholds correspond to three critical levels, respectively*: limit of flooding, flooding in regular area, exceptional flooding*.

The monitoring stations and their thresholds are presented in the following table.

Monitoring stations				Low flows (m^3/s)				High flows (m^3/s)		
#	Gauging station	River	Influenced by Reservoirs...	Vigilance	Alert	Reinforced alert	Crisis	Vigilance	Regular	Exceptional
24	Arcis-sur-Aube	Aube	Aube (A)	6,3	5,0	4,0	3,5	110	260	400
22	Méry-sur-Seine	Seine	Seine (S)	7,3	5,0	4,0	3,5	140	170	400
13	Nogent-sur-Seine	Seine	A+S	25,0	20,0	17,0	16,0	180	280	420
02	Gurgy	Yonne	Pannecière (P)	14,0	12,5	11,0	9,2	220	340	400
23	Courlon-sur-Yonne	Yonne	Pannecière (P)	23,0	16,0	13,0	11,0	550	700	900
16	Alfortville	Seine	A+S+P	64,0	48,0	41,0	36,0	850	1 200	1 400
21	Châlons-sur-Marne	Marne	Marne (M)	12,0	11,0	9,0	8,0	330	520	700
17	Noisiel	Marne	Marne (M)	32,0	23,0	20,0	17,0	350	500	650
05	Paris	Seine	A+S+P+M	81,0	60,0	51,0	45,0	950	1 600	2 000

Table 1. Monitoring stations downstream the reservoirs and corresponding flow thresholds for low-flows and high-flows.

2.5.2 Historical critical low and high flows events

For high-flows, the overcoming of the first threshold is a frequent event that does not imply damages; in fact the first high flows threshold is only a vigilance threshold that represents a warning for possible future critical events. Instead the second threshold is a critical threshold, that was rarely overcome in the past. This can be seen in Figure 7 that shows the river flow at the downstream station of Paris with the current management simulation from 1961 to 1991 over the "NTP" scenario (considering all the dams operational from 1961). It can be noticed that from 1961 to 1991 there's only one flood event over the second threshold (in 1982, with a flow peak of about 100 m^3/s above the threshold), while the first vigilance threshold is largely overcome many times.

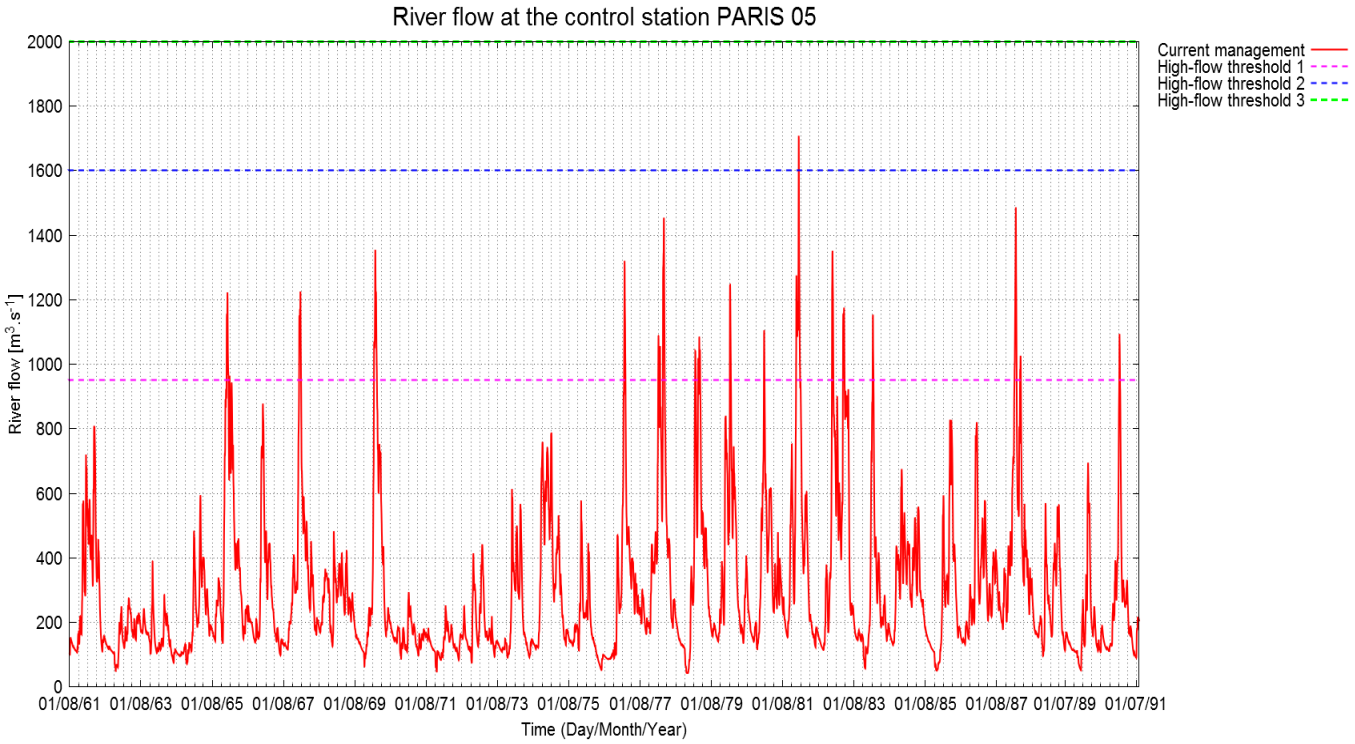


Figure 7. Simulated river flow at Paris from 1961 to 1991 with current management over NTP scenario. The thresholds for floods are the dashed horizontal lines.

As for the drought events, there were some critical events. For example, one of the most critical events is in 1976 with flows below all the critical thresholds, for a long drought during all the spring and summer.

The RTC that we implement in our work will be tested on these critical events of the past, flood of 1982 and drought of 1976, using the observations as forecasts ("perfect predictions").

2.6 Expectations of improvement by using a centralized Real-Time Controller

It can be noticed that the current management of the lakes proves inadequate when inflows are out of seasonal averages, as highlighted in Figure 6. This is due to a management strategy based on filling curves calculated off-line and based on the average meteo-hydrological conditions over the year, without using real-time measurements and forecasts of meteo-hydrological variables, which may be very different from their seasonal averages. Moreover the current management is designed under historical climate conditions

and socio-economic environment, so it might prove inadequate in the future, under the evolution of hydrologic conditions due to climate change and the environment evolution.

The possible benefits of a centralized real-time controller (or Model Predictive Controller, MPC) derive from the use of the information available in real-time. This information includes:

- The predicted inflows generated from weather forecasts:
  - the inflows upstream the reservoirs;
  - the intermediate inflows between the reservoir and the downstream control points.

- The actual storing capacity of each reservoir for reaching an objective at a downstream point concerned by several reservoirs.

- The current river flows downstream the reservoirs.

The possible improvements that derive from the use of this information can be summarized in the following points:

- The use of forecasts, including ensemble weather forecasts, makes the controller *proactive,* acting in advance to prevent a forecasted flood event.

- A *centralized controller* may coordinate the contributions of all the lakes to the achievement of the objectives, using the information of the actual storages and available storing capacities of all the lakes.

- The direct use of the *current river flows* at the *monitoring downstream stations* introduces a feedback in the control, because the decisions are calculated taking directly into account the effects at downstream.

3. Model Predictive Control (MPC) and Tree-Based MPC (TB-MPC)

In this chapter we will present the theory about the two Real-Time Control (RTC) methods used in this study: the deterministic Model Predictive Control (MPC) and the stochastic Tree-based Model Predictive Control (TB-MPC). In the first part of the chapter, the MPC method is introduced and in the second part the MPC framework is enlarged to exploit the information in the ensemble forecasts in the TB-MPC method. In section 3.1, we will explain the receding horizon strategy which is the basis of the MPC and TB-MPC methods and we will briefly describe the essential components of the predictive control; in section 3.1.2 the formalization of the deterministic MPC problem is introduced; then we highlight some remarks about the control and prediction horizon (section 3.1.3) and the advantages of MPC respect to the off-line approach (section 3.1.4). In section 3.2, the TB-MPC method is presented, introducing the *ensembles* and trees as uncertainty models and briefly explaining the tree-generation methods; in section 3.2.3 the TB-MPC control problem is formally defined.

3.1 MPC and Receding Horizon Strategy

Model Predictive Control is a form of control in which the current control decisions are obtained by solving *on-line,* at each sampling instant, a *finite horizon open-loop* optimal control problem. MPC uses a model of the system being controlled to predict its behavior in response to the control actions over a finite future horizon, called *prediction horizon*. The model takes as inputs the current measured state of the system, as the initial state, and the deterministic forecasts of the disturbances that act on the system. The idea is to select the control trajectory that promises the best predicted behavior in the future *prediction horizon* [Maciejowski, 1999]. Thanks to the use of forecasts the management becomes *proactive,* acting in advance to deal with expected problems caused by the disturbances. For example if an high inflow is predicted within the prediction horizon, before its expected realization, the controller will set the system to a state optimal to accommodate it, for example by lowering the water level in the reservoirs. Since MPC takes into account the current state of the system, it provides implicitly a *feedback control law*. For this reason MPC is also called *Naïve Feedback Control* (NFC). This implicit formulation represents an advantage since it allows obtaining a closed-loop control law at the price of an open-loop control problem, that has a lower computational cost.

Once the future control trajectory has been chosen at time step t, only the *first element* of that trajectory is applied as input to the system: the control at time t. Then the whole cycle of outputs (states) measurement, prediction, and control trajectory determination is repeated one sampling interval later, t+1. Since the prediction horizon remains of the same length than before, but slides along by one sampling interval at each step, this way of controlling a system is called *receding horizon strategy* [Maciejowski, 1999]**.** This strategy is represented in the following two figures.

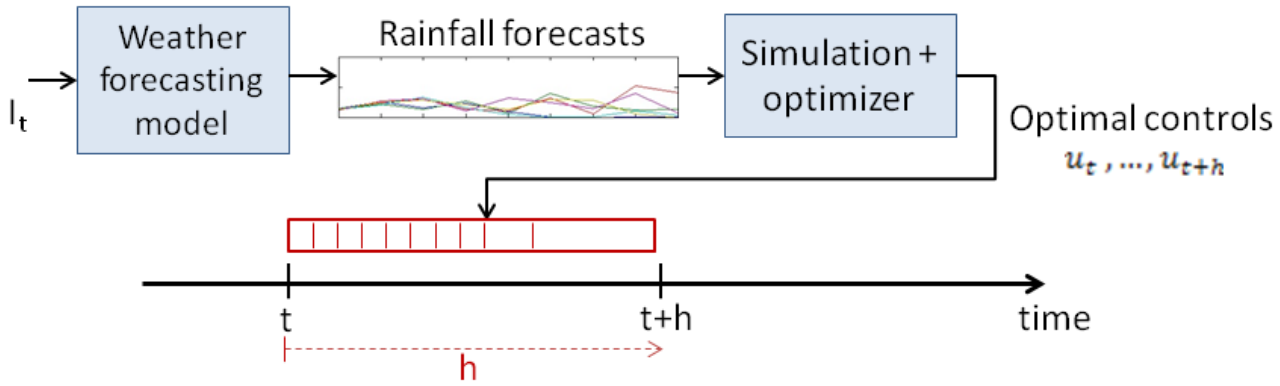


Figure 8. Receding Horizon Strategy (1): RTC operation scheme at time step t.

In the figure above there's the scheme of Real-Time Control operation at time step t for control of water systems, as our case. A control sequence over the prediction horizon [t, t+h) is calculated at each time step t, using all the exogenous information I_t available at time t. The information vector I_t contains uncontrolled exogenous variables, like, for example, measures of temperature, that are used as input to a weather forecasting model to produce the rainfall forecasts. The forecasts are used by a simulation model and an optimizer to find the optimal controls u_t, …, u_{t+h} . Only the first control, u_t, is applied to the system, and the system evolves from time-step t to t+1. Then the same problem is solved at time step t+1, as represented in the next figure.

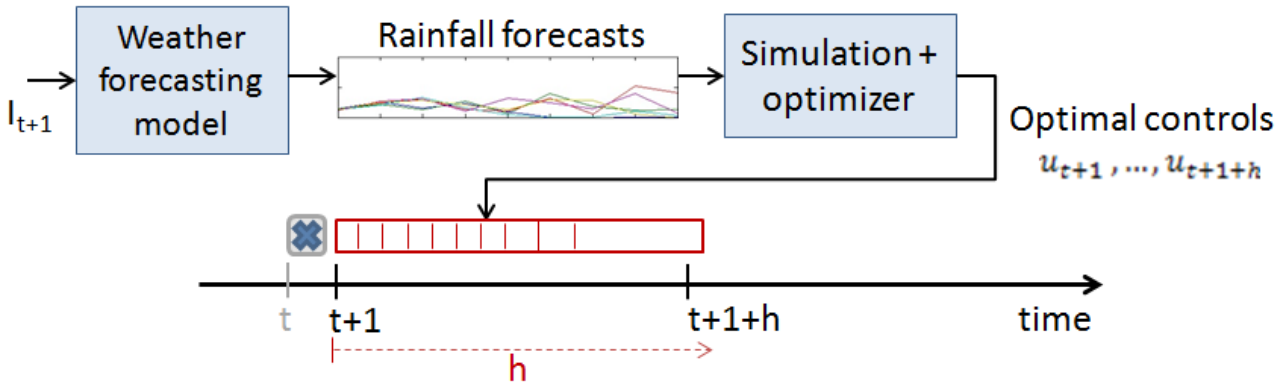


Figure 9. Receding Horizon Strategy (2): RTC operation scheme at time step t+1.

MPC is a deterministic algorithm, because the disturbance description is deterministic. At each time step of application of MPC the optimization yields an optimal control sequence, which is found by means of optimization (mathematical programming or heuristic methods). The drawback of a deterministic formulation is that it does not take into account the uncertainties of forecasts and models. However, this problem is alleviated by the on line approach based on the receding horizon principle. In fact the use of on-line information available in real-time (new states measures and disturbances forecasts) solves the uncertainties of forecasts and model, realigning the states of the system with the reality at each time-step. So the result is that there's no increase of the deviation between forecasted and real behaviors of the system in the long period, due to these uncertainties.

3.1.1 Components of MPC

The essential components of MPC are:

1. The ***internal model*** that is used to predict the future trajectories of the controlled variables (states and outputs of the model) over the prediction horizon as a result of the forecasted disturbances, the initial conditions of the system (i.e.: inputs of MPC) and the controls (i.e.: outputs of MPC). The model is then

used in the optimization process. For this reason it is important to use a *reduced model* of the system, simplified to a considerable extent.

2. The ***objective function*** that formalizes the goals that the controller has to try to achieve. It can be calculated after simulating the future trajectories of the system, using the internal model.
3. The ***constraints*** that must be respected by the controlled system. The constraints can be regarded in two different ways: as "hard" boundaries that can never be crossed or as "soft" boundaries that can be crossed occasionally, but only if really necessary. There's an important distinction between input and output constraints. In general input constraints can't really be exceeded and so they are "hard" constraints and there is no way in which they can be softened. The reason is that they represent the most of the time physical limitations for the control actions. Instead output constraints are usually "soft" constraints [Maciejowski, 1999].
4. The ***optimizer*** that is used by the controller to find the optimal solution that minimizes (or maximizes) the objective function, respecting the constraints. Since the MPC problem must be solved in real-time at each time step, it's very important to limit the calculation time. For this reason efficient algorithms must be used.

3.1.2 Formalization of deterministic MPC problem

At each decision time step t, the optimal controls can be determined by solving *on line* the following *open-loop* problem:

$$\min_{\boldsymbol{u}_t,\dots,\boldsymbol{u}_{t+h-1}} \sum_{\tau=t}^{t+h-1} g_\tau\left(\boldsymbol{x}_\tau,\boldsymbol{u}_\tau,\boldsymbol{d}_{\tau+1}\right) + g_{t+h}(\boldsymbol{x}_{t+h})$$

subject to:

$$\begin{array}{ll}
\boldsymbol{x}_{\tau+1} = \boldsymbol{f}_\tau(\boldsymbol{x}_\tau,\boldsymbol{u}_\tau,\boldsymbol{d}_{\tau+1}) & \tau = t,\dots,\ t+h-1 \\
\boldsymbol{u}_\tau \in U_\tau(\boldsymbol{x}_\tau) & \tau = t,\dots,\ t+h-1 \\
\boldsymbol{d}_{t+1},\dots,\boldsymbol{d}_{t+h} & \text{given} \\
\boldsymbol{x}_t & \text{given} \\
\text{any other constraint} & \tau = t,\dots,\ t+h-1
\end{array}$$

where:

- All the variables written in bold are vectors.
- h is the length of the prediction horizon that recedes at every time step.
- $\boldsymbol{x}_\tau$ is the *state vector* of the system at time τ; the initial state $\boldsymbol{x}_t$ can be measured or calculated since we solve the problem on line and so the solution implicitly depends on the actual state (*implicit control-law*).
- $\boldsymbol{u}_\tau$ is the *controls vector,* that contains all the decisions to be applied at time τ.
- $\boldsymbol{d}_{\tau+1}$ is the *disturbances vector*, that contains all the stochastic disturbances affecting the system during the time interval $[\tau;\ \tau+1)$.
- $\boldsymbol{f}_\tau$ is the *state transition function.*
- $\boldsymbol{u}_\tau \in U_\tau(\boldsymbol{x}_\tau)$ is a constraint that defines the *feasible controls* as a function of the state vector.
- $\boldsymbol{d}_{t+1},\dots,\boldsymbol{d}_{t+h}$ is the *disturbances trajectory* that is assumed to be deterministic and is derived from a forecasting model; the time subscripts denote the time instant when the disturbance value is deterministically known.
- g_τ is the *step indicator* or *step cost function* at time τ that expresses the aggregate cost associated to the transition from time τ to time τ+1 (from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$); the arguments of the step-cost function are all variables related to the time interval [t, t+1) and in general it depends on the state, the control, and the disturbance at time τ. The word "*cost*" must be considered in a wide sense of measure of performance and not necessarily as an expense.
- $g_{t+h}(\boldsymbol{x}_{t+h})$ is the *penalty cost* on the final state, that expresses the cost paid for the fact of being at the end of the horizon in the state $\boldsymbol{x}_h$. This penalty tries to compensate for the limited horizon, in

order to get to the end of the horizon in a desirable state to allow to achieve good performances later.

In a *centralized MPC*, as the one we consider in our study, the vector of controls is expressible as a function *m* of the vector of the states. This function *m* is not determined mathematically but is contained implicitly in the MPC procedure to determine the optimal controls. For example assuming that the controls to be determined are four, in a centralized controller, they will eventually be a function of all the *n* states that affect the performance of the controlled system, as:

$$\begin{vmatrix} u_t^1 \\ u_t^2 \\ u_t^3 \\ u_t^4 \end{vmatrix} = m\left(\begin{vmatrix} x_t^1 \\ x_t^2 \\ \cdots \\ \cdots \\ \cdots \\ x_t^n \end{vmatrix}\right)$$

3.1.3 Prediction and control horizon

The prediction and control horizon are two essential parameters to be determined for the formulation of the management problem. The prediction horizon is limited by the prediction ability, i.e. the horizon of available forecasts. Inside the prediction horizon the control decisions are taken at discrete time-steps. The control horizon is the time lapse between two consequent decisions. The real control horizon depends on physical constraints of the system and management characteristics (operational costs). In the formulation of the problem the control horizon can be assumed of constant or variable length. The control horizons can coincide with the real ones or be enlarged in order to have less decisions to calculate and lower calculation time.

As for the prediction horizon, one might wonder why the controller bothers to optimize over h = P future sampling periods and calculate M future controls when it discards all but the first control in each cycle. Indeed, the horizon lengths have an important impact. Some illustrative examples (taken from [Bemporad et al., 2012]) follow:

- ***Constraints management***
  Given sufficiently long horizons, the controller can "see" a potential constraint and avoid it, or at least minimize its adverse effects. For example, it's useful to consider the situation depicted below in Figure 10, in which one controller objective is to keep the output y of the system below an upper bound y_{max}. The current sampling instant is k, and the model predicts the upward trend y_{k+i}. If the controller were looking P_1 steps ahead, it wouldn't be concerned by the constraint until more time had elapsed. If the prediction horizon were P_2, it would begin to take corrective action immediately. This example is very appropriate in our case-study, in which water level thresholds are to be respected.

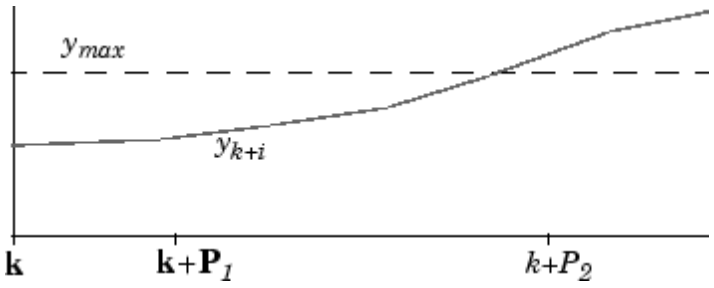


Figure 10. Example of the impact of the prediction horizon on constraints management. [Bemporad et al., 2012].

- ***System delays***
  Suppose that the controlled system includes a pure time delay equivalent to D sampling instants. In other words, the controller's current move, u_k, has no effect until y_{k+D+1}. In this situation it is essential that h > D and M< h −D, as this forces the controller to consider the full effect of each decision. This situation is typical for water systems problems as our case study.

3.1.4 Advantages of MPC respect to the off-line approach

Where MPC differs from the other feedback controllers is that it solves the optimal control problem on-line, for the current state of the system, rather than calculating off-line a feedback policy (that provides the optimal control for all states). Determining the feedback solution off-line would require the solution of Stochastic Dynamic Programming (SDP), a vastly more difficult task from a mathematical point of view. This approach requires to solve the recursive Bellman equation [Bertsekas, 1995] and in theory provides the exact, theoretical solution of the optimal control problem, but in practice its resolution is limited by a dual curse [Castelletti et al., 2010]:

i) The computing time increases exponentially with the dimensions of the state, decision and disturbance vectors: *curse of dimensionality* [Bellman, 1957]; for this reason, with the computational capacity available today, the maximum number of states that SDP can take into account with reasonable calculation time is about 3-4.

ii) A model of each component of the water system is required to anticipate the effects of the system transition: *curse of modeling* [Bertsekas and Tsitsiklis, 1996]; so using exogenous information (as precipitation or evaporation measurements) requires a dynamical model for each included information, thus adding to the curse of dimensionality.

For this reason, it can be recognized that the *raison d'être* for model predictive control is its ability to handle control problems where off-line computation of a control law is difficult or impossible [Mayne et al., 2000]. The difference between using dynamic programming and MPC is of implementation. In MPC *a receding horizon control* law is employed rather than an infinite horizon control law. This difference makes MPC more suitable to be used in detailed problems, where the large number of states makes the problem unsolvable by using SDP or it would require an over-simplification of the model.

3.2 Tree-Based Model Predictive Control

Since MPC includes the information of the disturbances forecast (weather forecast) in the control strategy the management becomes proactive. However, weather is difficult to forecast and meteorological models are affected by uncertainty. For this reason the control is vulnerable to forecasts uncertainty, especially when using only one deterministic trajectory of forecasts. If there's a predicted event that will not occur, the controller runs the risk of taking an action to counteract it and to reach a not desirable state. For example if an high inflow is predicted, before its expected realization the controller could lower the water level in the reservoirs in vain losing water stored for the drought season. This problem can be faced using Ensemble Forecasting that recognizes this uncertainty producing a large number of possible future trajectories, that are members of an ensemble. Ensemble Forecasts (EF) can be used for control by setting up stochastic programming. This makes the control less vulnerable to the forecast uncertainty.

Ensembles have been already used in stochastic optimal control in recent applications as Multiple Model Predictive Control (M-MPC) for open water systems [Van Overloop et al., 2008]. This method searches the control series optimal on average for all the trajectories of the ensemble. Even if this method seems to be more robust than a deterministic optimization, it overestimates the uncertainty because it is considered the same along the entire forecasting horizon. M-MPC takes conservative decisions considering all the entire ensemble and all the predicted events as possible scenarios to deal with.

In reality, uncertainty differs in time and its possible future reduction can be taken into account in the optimal control problems. In fact, as time passes the new measures could provide useful information to know which member of the ensemble is the most likely to occur and which ones can be excluded. For this reason the proper problem formulation using EFs is multistage stochastic programming (MSP): a framework for modeling optimization problems that involve uncertainty in recursive decision making. The level of uncertainty can be estimated by some information (e.g. probability distributions) that can be updated at each stage in which the optimization problem is solved and so uncertainty can be reduced. In this perspective we propose the use of a new control method based on the framework of Model Predictive Control, called Tree-Based MPC (TB-MPC). TB-MPC reduces the sensitivity to wrong forecasts, enhancing the control performance [Raso et al., 2013]. In TB-MPC the ensemble is transformed into a tree, that serves the scope of embedding ensemble data in MSP, specifying the moments when some uncertainties are resolved. At these moments control strategies can be changed to consider the feedback mechanism of the receding horizon approach.
The tree structure generation and the optimal control problem are two separate and consequent problems, as it is schematized in the following figure.

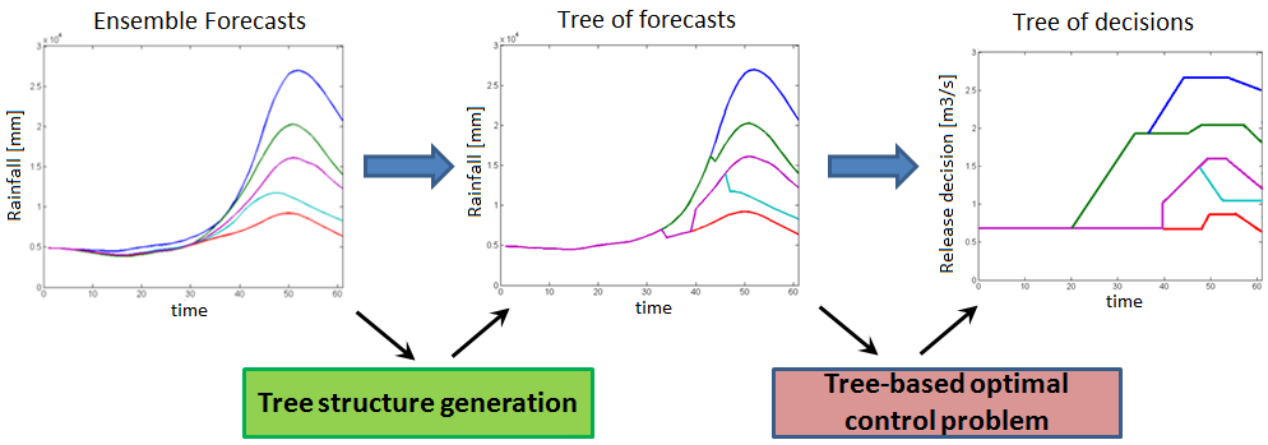


Figure 11. Flow scheme of the two stages of TB-MPC application: tree structure generation and optimal control problem.

The tree generation is the first step for setting the TB-MPC problem and it is both difficult and of a critical importance, as discussed in Stive [2011]. The tree generation can be done by bundling ensemble members at their initial stages, until the point where they are similar to each other, using the methodology proposed in Raso et al. [2012] that will be presented in section 3.2.2. The tree so defined is made up of different branches. This structure is used in the TB-MPC algorithm to find a different optimal control strategy for each branch, as it will be explained in section 3.2.3.

3.2.1 Ensemble forecasts and trees as uncertainty models

3.2.1.1 *Ensemble Forecasts*

Ensemble Forecasts (EFs) are a set of representative future trajectories of outputs of a dynamical system, as meteorological or hydrological models. EFs typically refer to weather predictions as in our case. Today ensemble predictions are commonly made at most of the major weather prediction centers worldwide, including the European Centre for Medium-Range Weather Forecasts (ECMWF), the National Centers for Environmental Prediction (US) etc.

EFs are uncertainty models that present both a deterministic and a stochastic component assuming uncertainty among the members, but determinism within each member. Each trajectory is produced by a deterministic physically based model. It can include different variables, as temperature, precipitation, etc.

In most cases, EFs are produced running the same meteorological model many times by slightly perturbing the initial conditions. The members of the EFs so obtained are usually close to each other at the early stages of the forecasting horizon and then spread out in time. This is due to the fact that the meteorological models are chaotic systems, owing to the chaotic nature of the fluid dynamics equations involved.
The ensemble trajectories are $d_t|x$, where: t is the time instant, with $t \in T : \{1, 2, \ldots, H\}$, H being the length of the forecast horizon; x is an index representing the ensemble member, i.e. outcome of a random variable; its sample space Ω is composed of N values, where N is the number of ensemble members, generally a number between 5 and 100. So the ensemble member is:

$$x \in \Omega = \{1, 2, \ldots, N\}$$

Each ensemble member x has a probability *p(x)>0,* such that $\sum_{x=1}^{N} p(x) = 1$. Generally all the members are equiprobable, in this case the distribution of x is discrete uniform.

An example of weather Ensemble Forecast is represented in the following Figure 12.

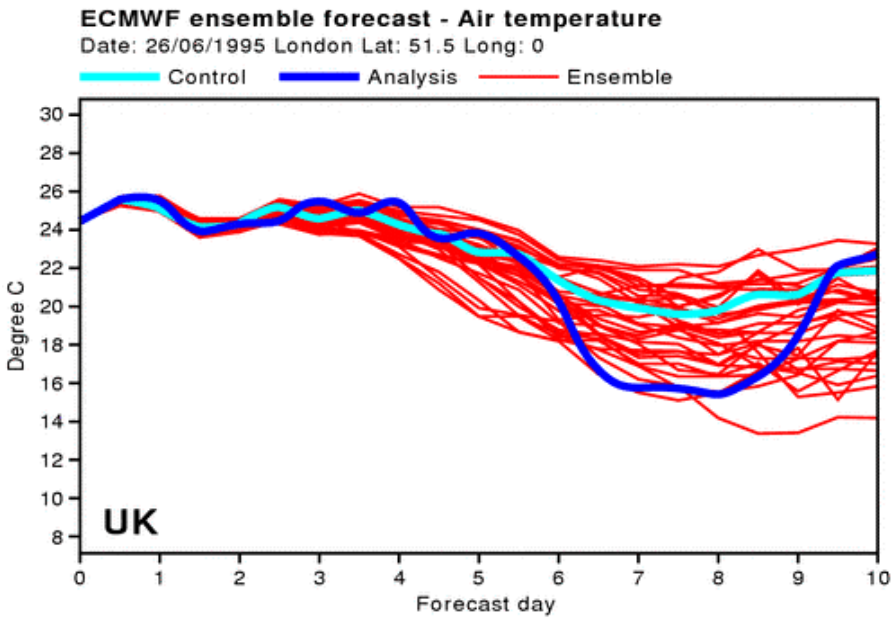


Figure 12. Ensemble Forecast of the air temperature in London (UK) produced by ECMWF for 10 days ahead the 26th of June of 1995. (Roberto Buizza, ECMWF)

The ensemble members in the figure above (red lines) are given by 33 different forecasts started from very similar initial conditions for the 26th of June of 1995. It can be noticed that all the ensemble members stay close together at the initial time steps and then gradually diverge. The control trajectory (light blue) is the unperturbed forecast using the same model than for the ensemble members. The analysis trajectory (blue) represents the observations.

3.2.1.2 *Trees*

A tree is a model of uncertainty derived from an ensemble, locating uncertainty resolution at specific time-steps along the forecasting horizon. A tree respect to an ensemble contains more information about when the members branch out from each other, specifying the so called *branching points*. At each branching point the sample space of the ensemble members splits in two subsets that are mutually exclusive, only one of them will occur. So the branching point is the instant when the uncertainty about which subset will occur is resolved.

A tree can be defined in different ways. Raso et al. [2012] define it in the most compact way by assigning to each ensemble member x a *parent* $P(x)$ and a *branching point* $B(x)$:

$$\forall\, x \;\exists\, P(x), B(x)$$

The parent $P(x)$ indicates the member from which the member x branches out and so it is a function defined from Ω to Ω; the *branching point* $B(x)$ represents the time instant when the out-branching of x happens and so it is a function defined from Ω to T. One ensemble member is assumed to be the one from which all the others branch out, the so called '*root'.* For convention the root's parent is the root and its branching point is the first time-step.
In alternative to the definition of parental relations and branching points, another way to represent a tree structure is by use of the *scenario tree nodal partition matrix M* [Dupacova et al., 2003]. M can be derived from the latter formulation and viceversa. Parental relations and branching points can be easily identified in the matrix. For different members z and P(z), before a branching point M gives the same label number; when the member z branches out, its label number becomes different from the one of its parent.
The matrix M has dimension (h, n): each element in position *(i,j)* in this matrix contains the label number of the tree branch for time-step i and the member j. So the maximum value of the matrix M corresponds to the number of distinct branches of the tree. M provides the necessary labeling scheme to set up the MSP problem pointing at the 'right position' in the matrix of *controls tree,* as it will be presented more in detail in section 3.2.3.

3.2.2 Tree generation from ensemble

There are different methods to build up a tree starting from an ensemble. In the present work we use the "information based" method by Raso et al. [2012] that takes into account the available information along the forecasting horizon. This method implies the explicit definition of the observations available in the future for the controller and their degree of uncertainty.
The first step before the proper tree generation process is the scenario reduction. With a large number of scenarios the calculation time needed for the optimization process will be large as well. To decrease the calculation time, the number of scenarios needs to be reduced. The optimal control action at the first time step based on this reduced ensemble should be close to the one obtained with the complete (non-reduced) ensemble. This problem is solved by heuristic algorithms [Growe-Kuska et al., 2003]: the backward reduction algorithm and the scenario tree construction algorithm, which together form the scenario reduction algorithm. The backward reduction algorithm reduces the original ensemble to a reduced ensemble with a predefined number of scenarios (N_{red}), by aggregating the most similar members of the original ensemble. Which scenarios are aggregated is based upon the difference between scenarios and their probabilities as described in Stive [2011]. The method is based on the minimal mass transportation problem, minimizing the probability mass of the scenarios deleted and the distance from these scenarios and the remaining ones. It's important to remark that at the end of the scenario reduction the new members of the reduced ensemble have different probabilities resulting from the sum of the probabilities of the aggregated scenarios.
The second step is to create a tree of scenarios from this reduced ensemble. In this step only parts of the scenarios (and not complete trajectories) are aggregated until the moment they diverge. The aggregation only occurs if the difference is smaller than a predefined threshold. The key idea is to identify the moment τ when a specific ensemble member x can be considered certain to happen with a confidence level p^*, chosen sufficiently large (say equal to 0.95). At this moment a branching point $B(x)$ is placed, i.e. the ensemble member x branches out from its parent $P(x)$:

$$B(x) = \tau : \; p_\tau(x) > p^*$$

This information is important for the control problem, because from the moment τ a decision strategy optimal for the member x can be applied.

The ensemble member's probability $p_t(x)$ changes in time, when new information is available. Information at time t is composed of observations available at that instant I_t and their likelihood of coming from an ensemble member $f_{I_t}(I_t|x)$. This function is the observational uncertainty, i.e. the probability distribution of the observations that depends on σ, i.e. the variance of the observations. The probability $p_t(x)$ at time t increases if the collected observations I_t are more likely to have been produced by this member, decreases otherwise. Bayes' theorem describes how new information changes the conditional probability of an event.
The knowledge of the probability dynamics of each ensemble member x can be used to determine the instant in which the member x is expected to be distinguishable from another. Information on these instants for all the members pairs is contained in the *distinguishability matrix D*. The procedure to build up the matrix D is described in Raso et al. [2012]. Here it's important to remark that the matrix D(p^*, σ) is a function of the confidence level p^* and the variance of the observational uncertainty σ:

- A larger value of p^* requires more evidence from the observations, shifting ahead in time the instant when two members are distinguishable, delaying the branching point.
- The better we can observe (little value of σ) the sooner we can distinguish a member from another and change the control strategy, anticipating the branching point.

A tree can be then generated from the *distinguishability matrix* D as described in Raso et al. [2012].

3.2.3 Formalization of TB-MPC problem

The key idea of TB-MPC is to consider a different control strategy for each ensemble member (or set of them) to be applied from the moment in which this ensemble member (or set) branches out [Raso et al., 2013]. This is expressed in the following equation in which for different ensemble members, *i* and *j*, up to their branching points, the controls must be the same:

$$u_t^i = u_t^j \ when \ \begin{cases} P(i) = j \\ t < B(i) \end{cases} \ i,j = 1, \dots, n$$

The equation above translates the *non-anticipativity condition*, saying that controls should not depend on the outcome of stochastic variables that have not been extracted yet [Birge and Louveaux, 1997].
Having introduced M (see section 3.2.1), the problem of optimal control for TB-MPC at each decision time step t can now be defined as:

$$\min_{u_{M(t)}} \sum_{z=1}^{n} p(z) \left[\sum_{\tau=t}^{t+h-1} g_\tau\left(\boldsymbol{x}_{\tau,z}, \boldsymbol{u}_{M(\tau,z)}, \boldsymbol{d}_{\tau+1,z}\right) + g_{t+h}(\boldsymbol{x}_{t+h,z}) \right]$$

subject to:

$\boldsymbol{x}_{\tau+1,z} = \boldsymbol{f}_\tau(\boldsymbol{x}_\tau, \boldsymbol{u}_{M(\tau,z)}, \boldsymbol{d}_{\tau+1,z})$	$\tau = t, \dots, t+h-1$
$\boldsymbol{u}_\tau \in U_\tau(\boldsymbol{x}_\tau)$	$\tau = t, \dots, t+h-1$
$\forall z$: $\boldsymbol{d}_{t+1,z}, \dots, \boldsymbol{d}_{t+h,z}$	given
$\boldsymbol{x}_t$	given
any other constraints	$\tau = t, \dots, t+h-1$

where:
- u_M is the control space, a matrix of dimension max(M) x N_u where N_u is the control dimension, i.e. for each separate branch of the tree of controls we calculate all the N_u control variables;
- $p(z)$ is the probability of the member z;
- the trajectory $\boldsymbol{d}_{t+1,z}, \dots, \boldsymbol{d}_{t+h,z}$ is the forecasts trajectory given by the member z of the ensemble;
- $\boldsymbol{x}_{\tau,z}$ is the *state vector* of the system at time τ derived applying to the system the disturbances trajectory of the member z.

Matrix M provides the "position" in u_M (i.e. the row in the vector) given by $M(\tau, z)$. Before a branching point, for different members z, M gives the same address, thus the same control value, respecting the non-anticipativity condition. Considering for example a member z and its parent $P(z)$, for t< $B(z)$, when it is not known which member will happen, the same control must find an optimal compromise between the effects of z and $P(z)$, weighted by their probability, $p(z)$ and $p\big(P(z)\big)$. After the branching point, matrix M returns different addresses for different members, thus different control values.

4. Formulation of MPC problem for the case study

In this chapter we will formalize the problem to be solved, based on the case-study system's information and data available, already presented in Chapter 2. The first element that will be presented is a scheme of the overall information flow of the work, focusing on the inputs and outputs for the optimizer. Following this scheme we will present all the components necessary to build our real-time controller, in the order: the *model* that is used to predict future trajectories of the controlled variables, then the *objective function* and the *constraints* of the problem, and at last the *optimizer*. A key element for the formalization of the MPC problem is the *definition of the cost function* to be used in the optimization. Since the lakes have different goals, already identified in the *Reconnaissance* phase (in Chapter 2), it's possible to define different criteria and objective-functions (*sub-objectives)* to express them formally. For this reason, the problem is a *multi-objective* one. A spontaneous way to solve this kind of problem is to use a global cost function that is the sum of the different components, using the so-called *weighting method*. In the exposition we will follow the chronological order of decisions taken in the work, starting defining the sub-objectives in the most spontaneous way and then analyzing how to *normalize* and *weight* them in order to be used within the weighting method. After defining the cost function, we will discuss about the horizons used in our MPC problem: optimization, forecast and control horizons. Then we will enter into the heart of the optimization problem, presenting the algorithm used, *Nelder-Mead* [Nelder and Mead, 1965], and some modifications operated to this algorithm to make it more efficient for the case study, as the normalization of the variables to be optimized and the expert-based initialization of the decisions. In section 4.8 we will present the extensions in the formulation of the MPC problem to be made in order to use the rainfall ensemble forecasts instead of a deterministic trajectory. In the end we will briefly describe the code architecture for the model predictive control simulation.

4.1 Information flow of the problem: inputs and outputs

The information flow of the present work can be summarized by the following scheme, in Figure 13, in which there's the representation of the main conceptual blocks to be built and used and their interactions. The perspective of this flow chart is the open loop simulation performed over the optimization horizon.

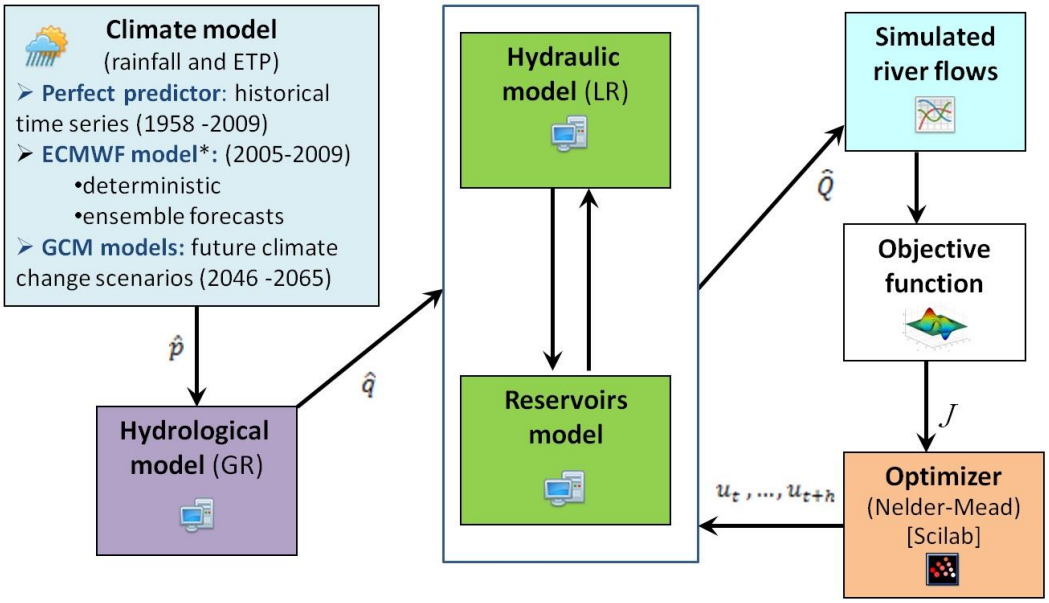


Figure 13. Flowchart of the information flow for the MPC problem.

The first element of the scheme is the climate model that provides the meteorological forecasts, producing time-series of rainfall and evapotranspiration as outputs, that are used as forecasts over the prediction horizon at each time-step of MPC application. We will test different climate models in our work, belonging to the two following classes:

- The so-called “perfect predictor", that assumes that the forecasted value of rainfall at each time-step of the prediction horizon is equal to the observed value (assuming a variance of the forecasts error equal to zero). As for the evapotranspiration we take the values calculated by the Penman equation using the observations for the weather inputs (temperature). The perfect predictor is a fake model that in theory can be applied only for the past, using historical time series of weather observations. In practice it can be used also for the future, assuming to know a hypothetical reality (for example assuming the simulations of GCM models as the future realizations). For our case-study we can use the past period 1958-2009 of data availability for the observations. The goal of using this perfect predictor is to evaluate the upper bound of the prediction ability and the relative performance of the real-time controller.
- A real predictor, represented for the rainfall forecasting model by the ECMWF (European Centre for Medium-Range Weather Forecasts) meteorological models [www.ecmwf.int] with two cases:
  - deterministic model;
  - stochastic model, that produces Ensemble Forecasts;

  The output trajectories of this models are available over the period 2005-2009.
  For the evapotranspiration we take as forecasts the *pluri-annual* average of the 'observed' values (calculated by Penman equation using weather observations).

The second element is the hydrological model that takes in input the rainfall and ETP information and gives in output the inflows to the river network. The hydrological model used (GR4J) is presented in section 4.2.1. It’s important to note that the hydrological model is not affected by the decisions operated on the four dams and so it can be put out of the model directly used by the controller. For this reason and since the rainfall forecasts are already available for all the period of application of MPC, we can generate all the inflow forecast scenarios off-line (before starting the optimization process) for all the time-steps of the period of application of MPC (and TB-MPC). So we give in input to the GR model the rainfall forecasts (one trajectory for MPC / ensemble of trajectories for TB-MPC) to obtain the correspondent inflows forecasts.

These inflows forecasts are inputs of the internal model of MPC that is composed by:

- hydraulic model "LR", presented in section 4.2.2;
- reservoir model, presented in section 4.2.3.

These two models make up the internal model of MPC that interacts with the optimizer. The optimizer seeks an optimal sequence of controls over the forecast horizon, minimizing an objective function, expressed in terms of costs, not to be interpreted as economic costs but in a wider sense. These costs are function of the simulated river flows, but also of the reservoirs volumes, as discussed in detail in section 4.4.

4.1.1 Classification of inputs/outputs of the models and MPC

For convenience of the exposition, here below we define all the variables involved in the formulation of the MPC problem and their symbols.

4.1.1.1 *Notation of the variables*

- h = length of the optimization horizon (days)*;*
- Ti = first time-step of application of MPC;
- Te = last time-step of application of MPC;
- u(t) = vector of decisions for all the lakes, i.e. inlet/outlet flows over the period [t, t+1) (m^3/s); there are 8 decisions to be calculated at each time step: two decisions for Aube and Seine lakes; three for Marne lake; one for Pannecière lake;
- P(t+1) = rainfall observations over the time period [t, t+1) (mm);
- $P^f(t+1)$ = rainfall forecasts over the period [t, t+1) (mm);
- E(t+1) = evapotranspiration over the period [t, t+1) (mm);
- s(t) = state of the production reservoir (soil moisture) in the hydrological model (GR) at time t (mm);
- r(t) = state of the routing reservoir in the hydrological model (GR) at time t (mm);
- $I^{(f)}(t+1)$ = inflows (forecast) over the period [t, t+1), calculated by the hydrologic model, using observations P_{t+1} (or forecast P^f_{t+1}) (m^3/s); note: I is also called $Q_{lateral}$.
- $Q_{river\ (up/downstream)}(t+1)$ = average flows of the river (upstream/downstream the BVIs) at the gauging stations over the period [t, t+1) (m^3/s);
- $V_{lakes}(t)$=vector of lakes volumes at time t (Mm^3);
- $g(t)=g(u_t,Q_t,I_{t+1})$=step-cost at time t depending on: control u_t, disturbance I_{t+1}, state of the system Q_t.

4.1.1.2 *Proprieties of the variables*

- Frequency: time step of 1 day for all the climatic, hydrological and hydraulic variables;
- Spatial distribution: climatic and hydrological data for each intermediary basin (BVI); uniform conditions on each BVI;
- Position:
  - upstream stations for the inflows i to each BVI;
  - connection stations for the derived flows $Q_{connect.}$.

It's important to do a classification of the variables involved as inputs/outputs of the model and of the MPC optimizer. Here below this classification is reported, considering all the variables necessary for the simulation of the model and the application of MPC at time t.
Note: The notation (t_i; t_e) indicates all the time-steps between t_i and t_e. E.g.: $P(t_i; t_e)$ is the serie: $P(t_i)$; $P(t_i+1)$; … ; $P(t_e)$.

4.1.1.3 *Inputs of the models*

- Initial conditions for:

- the hydrological model GR: two main initial states soil moisture s(t) and routing reservoir r(t); some hidden states are also present and for this reason the GR model is called for a period of warming up before t, to realign all the states;
- the hydraulic model LR: delayed flow for each upstream sub-catchment $Q_{river}^{up.i}(t-\tau_0^i : t)$, for a period back in time long as the maximum delay of each BVI;
- the reservoirs models: initial volumes of the reservoirs $V_{lakes}(t)$;

- Deterministic forecast/ensemble forecasts of precipitation, $P^f(t+1; t+h)$, over the optimization horizon h, where $P^f(t+1)$ is the rainfall occurring over the time interval [t, t+1);
- Potential evapotranspiration *ETP* (average ETP on each intermediary basin);
- Decisions $u(t; t+h-1)$ for all the reservoir connections and time steps of the forecast horizon h.

4.1.1.4 *Inputs of the MPC optimizer*

- Arguments of the cost-functions:
  - flows Q_{river}, i.e. states of the LR model, at the control stations for each simulation step of the optimization horizon (t; t+h);
  - final volumes of the reservoirs at the end of the optimization horizon $V_{lakes}(t+h)$, i.e. output of the reservoirs models.
- Arguments of the constraints:
  - flows in the reservoirs connections $Q_{connect.}$ for the horizon (t; t+h-1);
  - volumes of the reservoirs during the optimization horizon $V_{lakes}(t; t+h)$;
  - flows $Q_{river}(t; t+h)$ downstream the reservoirs connections.

4.1.1.5 *Outputs of the MPC optimizer*

- Optimal decisions u* for the four dams (inlet and outlet flows) for the optimization horizon (t; t+h-1);
- Objective value J*, that is function of the optimal decisions.

4.1.2 Flow chart of the variables in operational perspective

In order to describe the cause-effect relationships among the variables of interest for the MPC problem, we can analyze the flow process of inputs and outputs with a flow chart more detailed than the one in Figure 13. Moreover here we want to represent the closed-loop simulation, emulating the presence of the real controlled system by using a model more accurate than the one used in the MPC optimization loop. The implementation of this accurate system is planned to be done within the Climaware project, but it's not presented in this thesis.

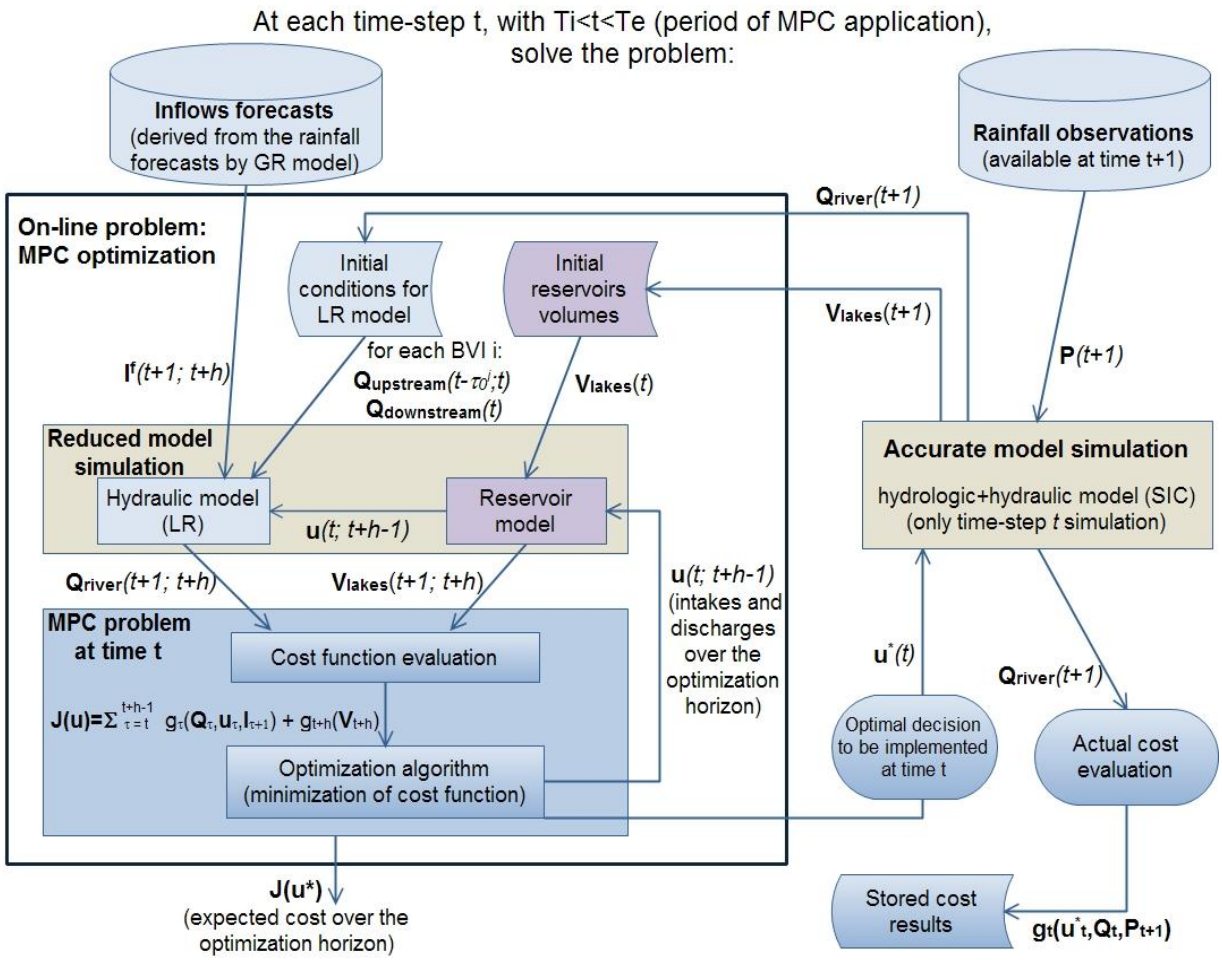


Figure 14. Flowchart of the MPC problem at each decision time step t in a closed loop simulation from Ti to Te. The standpoint of this scheme is the operational perspective, emulating the application of MPC to the real system by using an accurate model.

The flowchart in Figure 14 represents the schematization of the simulation of an operational use of the MPC controller at time-step t. At each instant of decision t, the MPC problem involves the information flow regarding the countered box on the left side of the diagram.

The on-line optimization problem uses a reduced model of the controlled system, with a small number of states of the model that allows to keep low the calculation time. In the real world application of MPC, at each time step the real physical systems evolves under the optimal controls calculated for the current step. Then observations are then used to realign the states of the reduced model. For simulating a operational perspective, instead of playing with the real dams, we can use a more accurate model for trying to reproduce the complexity of the real world. Since in the receding horizon strategy only the first optimal decision must be applied to the real system, in the operational perspective the first control of the sequence calculated over the optimization horizon is used as input by the accurate model for only one time-step simulation. This accurate model allows to realign the states of the reduced model at the next time-step of simulation of MPC, using rainfall observations in input and by means of a more accurate description of the real system. The accurate model simulation shown on the right side of the diagram in Figure 14 is going to be integrated after the end of this work in SIC [J.P. Baume et al., 2005]. For the moment, in the implementation used for obtaining the results of this thesis, the accurate model is substituted by the same reduced model (TGR) used as internal model in the MPC optimization.

4.1.3 Uncertainty sources

The MPC optimization is based on the use of the *meteorological forecasts* and a *reduced model*, so its solution is affected by the uncertainty of these two sources. Since the same reduced model is used so far for both on-line optimization and closed-loop simulation, as explained in the previous section (4.1.2), the realignment of the states (of the hydraulic and reservoirs models) is due only to the elimination of the uncertainty of the forecasts and not that of the model. Using the internal model of MPC as reality emulates a perfect knowledge of the rainfall-runoff process, thus considering to have a perfect model and underestimating uncertainty.
In this study, our goal is to evaluate the possible performance improvement by using ensemble instead of deterministic forecasts. So, we will focus only on the impact of the *uncertainty of the meteorological forecasts*, thanks to the comparisons between MPC with perfect predictions (MPC-PP), with real deterministic forecasts (MPC-RF) and with ensemble forecasts (TB-MPC).
For all the three methods (MPC-PP, MPC-RF and TB-MPC), at each time-step of the closed-loop simulation, the MPC solution is used for simulating the system using the *rainfall observations* instead of the forecasts. Here it must be noticed that even the rainfall observations are affected by the uncertainty of the measurement process, that we will not take into account in this study.

4.2 The model of the system (TGR)

In order to model the Seine watershed we used the TGR model (*fr.: modèle de Transfert couplé au modèle du Génie Rural*), a semi-distributed conceptual model developed in Munier [2009]. The TGR model couples a hydrological lumped rainfall-runoff model with a simplified hydraulic model. Since TGR is a conceptual model, its states do not correspond to physical properties of the system.
In TGR calculations are made at the scale of the intermediate river basin (*BVI = fr.: "Bassin Versant Intermédiaire"*) that is the sub-basin obtained removing the upstream sub-basins [Dehay, 2012]. Using a semi-distributed approach, it's possible to take into account the heterogeneity of the basin. For each sub-basin a module of TGR is identified, with its own parameters. This structure is schematized in Figure 15. Each module calculates the downstream flow of the BVI starting from climate data (E and P). The rainfall is transformed by the hydrological component into run-off. For all the sub-basins, except the very upstream ones, this runoff is then taken by the hydraulic component with the flows of the upstream BVIs.

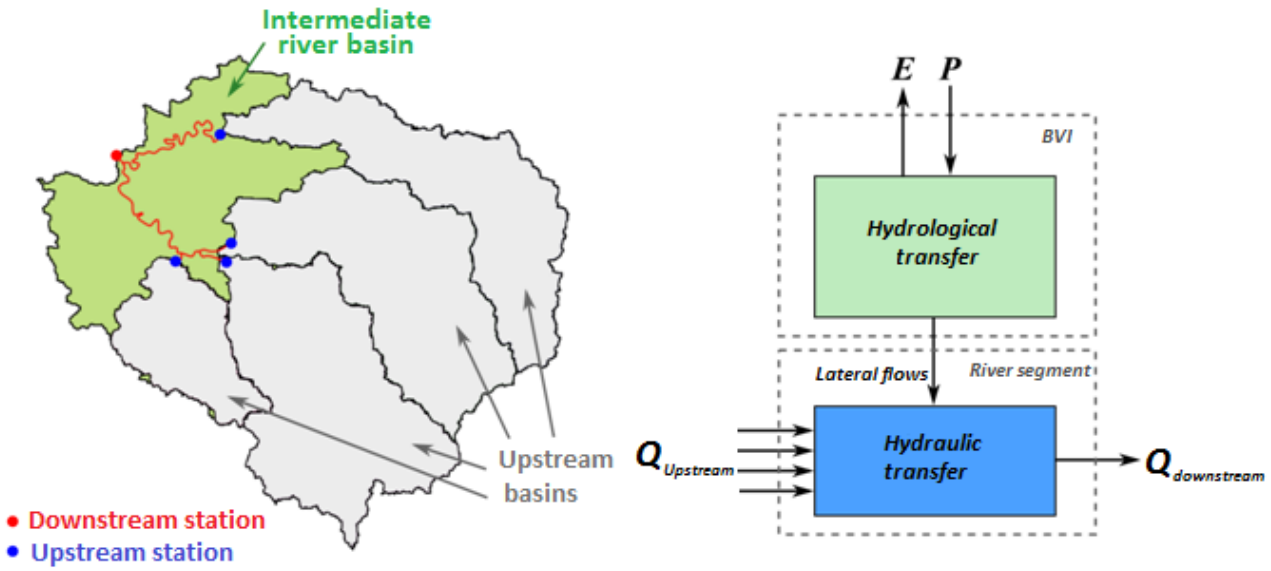


Figure 15. TGR model structure: an intermediate river basin and its relative TGR module.

The two sub-models used in TGR are:

- the GR (*fr.: Génie Rural)* model for the hydrological part, described in section 4.2.1;
- the LR (*fr.: Lag and Route*) model for the hydraulic part, described in section 4.2.2.

The watershed of the Seine is divided into 25 intermediate river basins (*BVIs)*, having as outlets the 25 gauging stations for river flows upstream of Paris. A map of these BVIs and stations is provided in Appendix C. The TGR module schematized in Figure 15 is applied to the 25 BVIs in a proper order. For performing this operation a network of the intermediate river basins configuration is identified (see Appendix E). TGR is first applied to the upstream BVIs, before being applied to the downstream BVIs. Flows from upstream BVIs are considered as punctual upstream inflows. For example, a networking scheme of the BVIs and TGR modules is reported in the figure below.

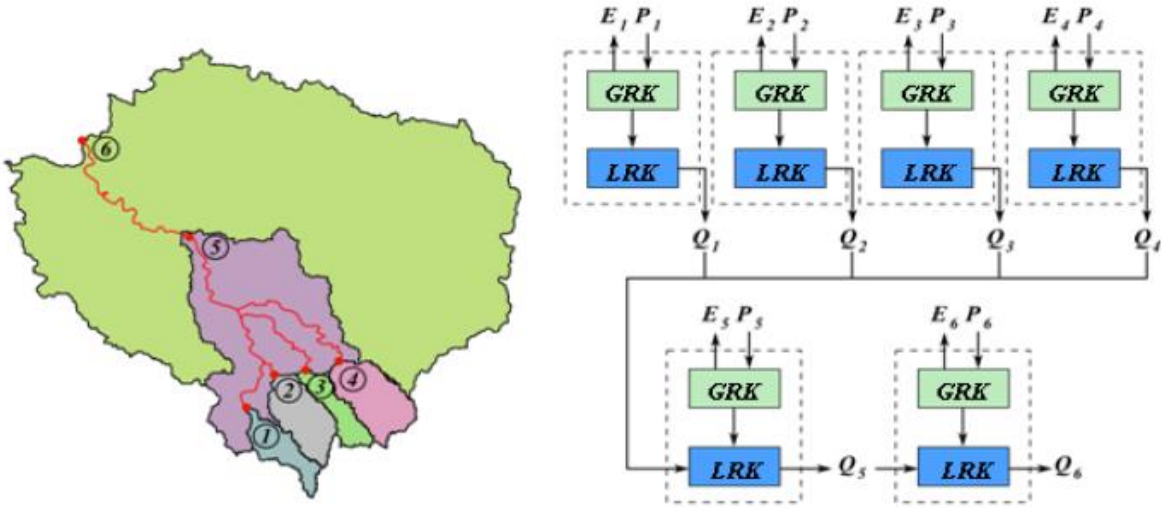


Figure 16. Networking scheme of BVIs and TGR modules.

The TGR model was calibrated for the case study using daily average observed flows, called “naturalized” because the influence of the lakes is removed. The calibration was performed minimizing the sum of the squares of the deviations between simulated and observed flows, using the optimizer Nelder-Mead (see section 4.7.2.1).

The observed data used for calibration are the historical time-series of daily river flows over the period 1958-2009. It can be remarked that in this period the big floods of 1910 and 1924 are not included, so the model it’s supposed to reproduce better normal conditions or high-flows and floods of moderate level, as the flood of 1982, than the exceptional conditions of a 100 year return-period flood as the case of 1910. It’s important to remember this issue about calibration representativeness when using the same model for future climate scenarios. The daily time-scale of the calibration data is good for our use of the model that will be operated at the same scale.

The calibration was done sequentially from upstream to downstream. For the sub-basins located at the upstream ends (i.e. no other sub-basin is located upstream), the calibration of the sub-basin parameters only consists in calibrating the four GR4J parameters. For the other sub-basins (i.e. the sub-basins that route flows from upstream sub-basins), there are two hydraulic parameters plus one extra parameter for the routing of each upstream flow. This extra parameter is used for defining the relative routing distances between the sub-basin outlet and each connected upstream sub-basin outlet.

The TGR model was first used to simulate natural flows. Then the four reservoirs were explicitly included in the TGR model structure to account for their influence in order to respond to the objectives of Climaware. This modification is made taking into account the reservoirs to define new sub-catchments, starting from the 25 original ones, as described in section 4.2.2.1, and then adding a model for the reservoirs. This modified model uses the same calibration as for natural flows.

4.2.1 The hydrologic model GR4J

The GR4J (*Fr.: "Génie Rural à 4 paramètres Journaliers*") model is a non linear rainfall-runoff model used to calculate the net rainfall production and the exchanges with the aquifer [Perrin et al., 2003]. It takes in input the climate data, rainfall and ETP, and gives in output the run-off for each BVI. The model depends on four parameters: the capacity of the production reservoir (S), the gains or losses from exchanges with the aquifer (IGF), the capacity of the routing reservoir (R), and a time constant of the unit hydrograph (T). In this project, we assume that all the sub-basins have the same hydrological behavior, so the same model is used for all of them, with different parameters, obtained by the calibration at the BVI scale. In addition, rainfall and ETP are uniform throughout the BVI, as with any global hydrological model. The output of the GR4J is then considered as distributed lateral inflow to the hydraulic river system of the BVI and it is added at downstream.

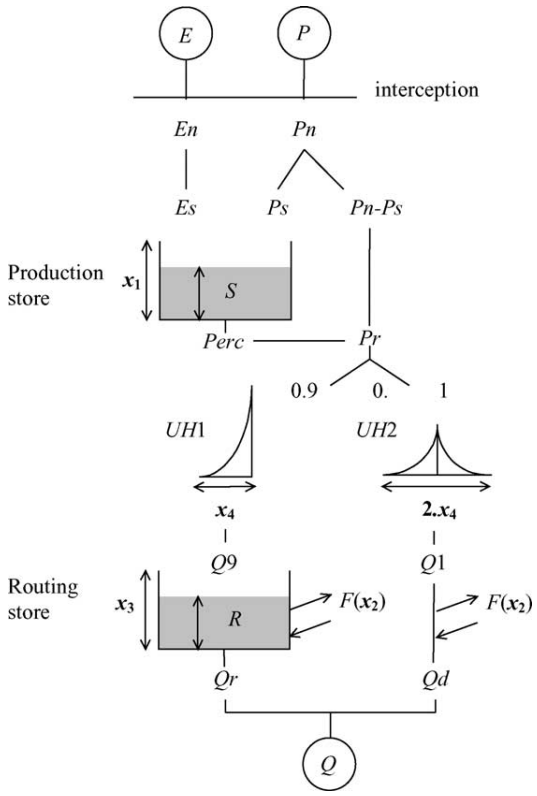


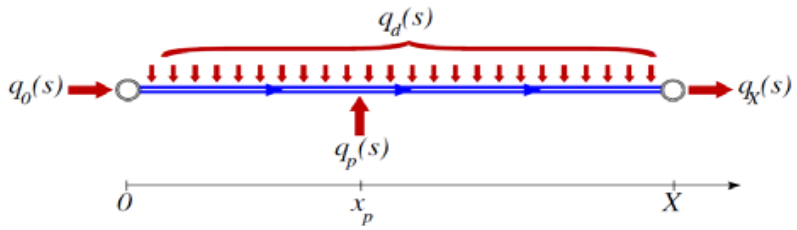
Figure 17. Scheme of the GR4J rainfall-runoff model. [Perrin et al., 2003]

States of the hydrologic model GR and warming up

The main states of the system for each BVI are the production reservoir S and the routing reservoir R. These two reservoirs represent the "long" memory of the system GR. However, there are other internal states in this model, corresponding to the rainfall waiting in the unitary hydrographs. Since in the implementation of the model available (at the moment) only the two external states are saved as outputs, it's necessary to pay attention to have the good initial conditions. If we take as initial conditions for our problem only the reservoirs states (S and R) we make the implicit assumption that there is no rain waiting in the unit hydrographs. But if this is not the case, it will cause errors. As the unit hydrographs have a short memory if we run the model for a period of warming up long enough (some weeks) before the target period of simulation, using the observations as inputs, the impact of not good initial conditions in the hydrographs is annulled. In this way we need only the initial conditions of S and R. So we decided to run the simulation of the hydrologic model GR for a period of warming up of 2 months before the real period of interest, because we found that in this way the errors are negligible (1E-5 m3/s).

4.2.2 The hydraulic model LR

It allows the calculation of the downstream flow q_X depending on the upstream flow q_0, the punctual lateral inflow q_p and the diffuse lateral inflow q_d. Each of these three contributions to downstream flow is calculated independently by a *transfer function* of first order, with a delay representing the *attenuation* phenomena and a delay typical of streams *propagation*. These transfer functions are calculated using two different parameters for the different inputs: a time constant τ and a delay K. Since these inputs regard the same river branch, it's natural to imagine that there's a relationship between these two parameters and the physic features of the BVI. These relationships were studied in Munier [2009]. Using the surface of each BVI and the distance between upstream/downstream stations, it's possible to calculate the two parameters τ_0 and K_0.



$$q_X(s) = TF_0(s)q_0(s) + TF_p(x_p, s)q_p(s) + TF_d(s)q_d(s)$$

$$TF_0(s) = \frac{e^{-\tau_0 s}}{1+K_0 s}$$
$$TF_p(x_p, s) = \frac{e^{-\tau_p(X-x_p)s}}{1+K_p(X-x_p)s}$$
$$TF_d(s) = \frac{1}{1+K_d s}$$

$$\begin{cases} \tau_p(x) &= \frac{x}{X}(\tau_0 + K_0) - \sqrt{\frac{x}{X}}K_0 \\ K_p(x) &= \sqrt{\frac{x}{X}}K_0 \\ K_d &= \frac{\tau_0+K_0}{2} \end{cases}$$

Figure 18. LR hydraulic model scheme and general equations.

The scheme and equations in Figure 18 show that the propagation of the flow rate q0 from upstream to downstream occurs through rigid modules (plug flow). Moreover in the LR model the BVI have also a detention capacity (not shown in the figure). For this reason the states of the hydraulic system are all the flows in the river network that are going to be transferred towards the downstream outflow of the watershed. For our case study a simplification on the LR model was made by considering all the lateral inflows as punctual inflows added at the outflow of each river branch.

4.2.2.1 *Introduction of the reservoirs in the model*

The work of introduction of the reservoirs in the TGR model is done in Dehay [2012] where all the data of the lakes that must be taken into account in our model are summarized. These data include: the maximum and minimum volumes of the lakes and the capacities of the channels, the distances between the upstream and downstream stations and the connections, the reference and reserved flows in the river downstream the inlet/outlet channels, the eventual presence of withdrawal channels, etc. These data are reported in the schemes in Appendix A.

Adding the dams, the intermediate river basins are split at the connection points of the channels (intersections of inlet and outlet of the lakes with the river). Each connection of the lake cuts an existing BVI in two new BVIs with outflow in the connection point for the upstream BVI (see next figure). For this reason, after adding the dams in the model *the number of BVIs increases from 25 to 32*, because there are 7 connections (3 for Marne, 2 for Aube and 2 for Seine).

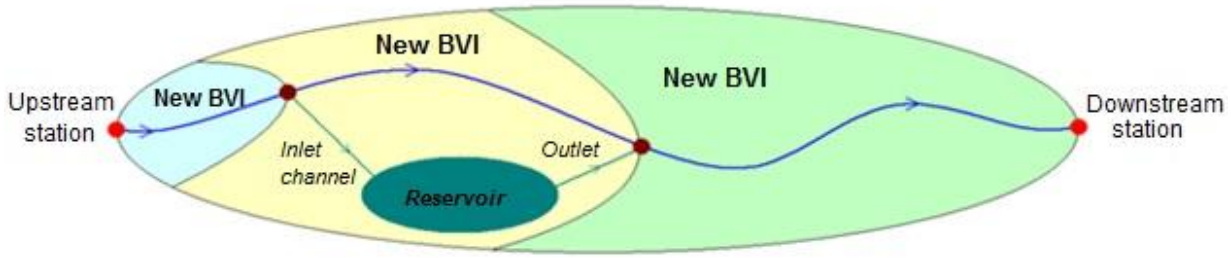


Figure 19. General scheme for lake's connections with one inlet and outlet and subdivision into new BVIs . [Dehay, 2012]

The scheme in the figure above represents the situation of a lake with one inlet and one outlet. In particular each lake has a different configuration, as summarized in Figure 20.

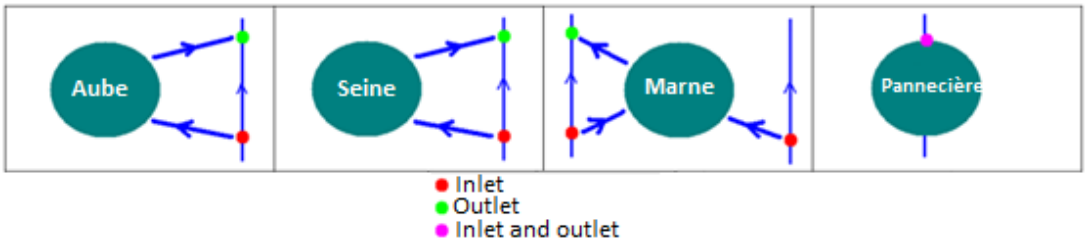


Figure 20. Summary scheme of lakes connections configurations. Scheme taken from Dehay [2012] and corrected for Marne lake.

Only the lake Aube has exactly the configuration with two connection channels (one inlet and one outlet) on the Aube river. For the lake Seine, there's an inlet connection, while there are two outlet connections with two branches of the Seine river. However for one of these the flow in unknown since there's no upstream station for one of the two river branches. Moreover, this restitution is very distant (41.8 km) from the downstream station of the model (Mery-Sur-Seine). So it was considered that all the outlet flow is restituted to the Seine in only one connection. The lake Marne configuration is the more complex, with two connections for the inlet and two for the outlet, one on the Blaise and one on the Marne. The restitution on the Blaise, being very inferior to that on Marne, can be considered as a complement. Moreover, the Blaise and the restitution connection with the Marne rejoin 450 m before arrive in the Marne. For this reason the two connections can be reduced to one. The Pannecière is directly on the Yonne, without derivations. Moreover, the station of Chaumard, used in the model, is located in the lake and so has not physical meaning. It can be considered as the intake and restitution connection at the same time.

4.2.2.2 *Equations of the hydraulic model*

The implementation of the hydraulic model has been modified in the framework of this work in order to manage to work with the receding horizon strategy of MPC. In fact, the implementation already available for the case study carried out the calculation of the model over the whole simulation horizon (many years) starting from a random initialization. Since the influence of the initialization is lost after a few days, a good initialization was not important compared to the length of the simulation horizon of many years used to simulate the current management of the lakes. On the contrary, in the case of our MPC the influence of initialization is important and so it's necessary to make the initialization as accurate as possible at each time step of application of MPC. In order to understand how to initialize the model, we must analyze in detail the equations. From the equations we can identify the states of the model and then the initialization will be done at each time step saving the initial states of the next time-step.
The hydraulic model propagates the flows in the river network BVI by BVI. The calculation order of the BVIs is reported in Appendix E. The equations for each BVI can be seen as *linear combinations of the different*

components involved: propagation of upstream BVI flows, attenuation of the previous delayed flow of the same BVI, withdrawal by the reservoirs, addition of lateral inflows and restitutions from the reservoirs.

The explanation of the LR model provided above shows us which are the states of the system. Calling $\hat{\tau}_{0,j}$ the *floor* value of the real parameter $\tau_{0,j}$, delay for the BVI *j*, at each time-step it's necessary to have the previous $(\hat{\tau}_{0,j}+1)$ flows for each BVI *j* so that the LR model can take them in input to forecast the future dynamic of the whole system. In fact for calculating the downstream flow $Q_{out,j}(t+1)$ of BVI j we need the upstream flows $Q_{up,i}(t-\hat{\tau}_{0,i})$ and $Q_{up,i}\left(t-\hat{\tau}_{0,i}+1\right)$ coming from each upstream BVI *i* and the previous downstream flow $Q_{out,j}(t)$. So in total, considering all the 32 BVI and the corresponding parameters τ_0, we have 51 states.

We can better visualize these states and their role in the equations provided here below. For clearness in the exposition, the equations will be summarized in different classes, considering all the possible positions for a BVI j in the river network. Referring to the BVI numeration reported in the scheme in Appendix E, we have the following cases:

- BVI *j* upstream the rest of the system and without channels of derivation (case of BVI number 1, 3, 6, 12, 13, 15, 17, 21, 23, 29 and 31):

$$Q_{out,j}(t+1)=\left(1-\frac{T_s}{K_0^j}\right)\cdot Q_{out,j}(t)+Q_{lateral,j}(t+1)$$

- BVI *j* with n_{up} upstream BVIs and an inlet channel to a reservoir (case of BVI number 2, 7, 24 and 25):

$$Q_{out,j}(t+1)=\frac{T_s}{K_0^j}\left[\sum_{i=1}^{n_{up}}\left[r_i\cdot Q_{up,i}(t-\tau_{0,int.}^{i})+\ (1-\ r_i)\cdot Q_{up,i}(t-\tau_{0,int.}^{i}+1)\right]\right]+\left(1-\frac{T_s}{K_0^j}\right)\cdot Q_{out,j}(t)$$
$$+\ Q_{lateral,j}(t+1)-u_j(t)$$

Note: the component of propagation of upstream flows, i.e. part between square brackets, is absent for BVI 25 where n_{up} is zero.

- BVI *j* with n_{up} upstream BVIs and an outlet channel (case of BVI number 4, 8, 11, 27 and 28):

$$Q_{out,j}(t+1)=\frac{T_s}{K_0^j}\left[\sum_{i=1}^{n_{up}}\left[r_i\cdot Q_{up,i}(t-\tau_{0,int.}^{i})+\ (1-\ r_i)\cdot Q_{up,i}(t-\tau_{0,int.}^{i}+1)\right]\right]+\left(1-\frac{T_s}{K_0^j}\right)\cdot Q_{out,j}(t)$$
$$+\ Q_{lateral,j}(t+1)+u_j(t)$$

Note: for BVI 11 the part between square brackets is absent, since n_{up} is zero; moreover, all the inflow from t to t+1, $Q_{lateral,j}(t+1)$, feeds the reservoir, whose outlet should be considered downstream the same BVI. So the downstream flow of BVI 11 is equal to the outlet from Pannecière lake.

- BVI *j* with n_{up} upstream BVIs and without channels (case of BVI number 5, 9, 10, 14, 16, 18, 19, 20, 22, 26, 30, 32 and 33):

$$Q_{out,j}(t+1)=\frac{T_s}{K_0^j}\left[\sum_{i=1}^{n_{up}}\left[r_i\cdot Q_{up,i}(t-\tau_{0,int.}^{i})+\ (1-\ r_i)\cdot Q_{up,i}(t-\tau_{0,int.}^{i}+1)\right]\right]+\left(1-\frac{T_s}{K_0^j}\right)\cdot Q_{out,j}(t)$$
$$+\ Q_{lateral,j}(t+1)$$

Where, in the equations:

- $Q_{out,j}(t+1)$ is the average river flow downstream the BVI j [m^3/s] over the horizon [t, t+1);
- $Q_{up,i}(t)$ is the river flow of the upstream BVI i [m^3/s] over the horizon [t-1, t);

- $Q_{lateral,j}(t+1)$ is the lateral inflow of the BVI j [m^3/s] over the horizon [t, t+1), that is the output of the hydrological model depending on rainfall over the horizon [t, t+1) and initial conditions at time t;
- $u_j(t)$is the decision at time t, that is the flow of the channel located in BVI j over the horizon [t, t+1) [m^3/s] ; the connection of the channel is considered to be downstream the BVI j;
- $\tau^i_{0,real}$ is the real delay parameter (real number) of the upstream BVI i [hours];
- $\tau^i_{0,int.}$is the integer part of the delay parameter$\tau^i_{0,real}$ [days] ($\tau^i_{0,int.}$=floor($\tau^i_{0,real}/24$))
- $r_i = r(\tau^i_{0,real})$ is the decimal part of the delay parameter $\tau^i_{0,real}$ [-];
- K_0^j is the attenuation coefficient for the BVI j [hours] that is zero for almost all the BVIs;
- T_s is the simulation time-step [hours] (assumed equal to 24 hours).

The initial states of the hydraulic model for the first time-step of the simulation horizon with MPC management are set equal to the values obtained by simulation of the current management on the same scenario (NTP/NAT/TP/FUT) at the first day of the simulation horizon. After that, at each time-step of MPC simulation, we save the simulated river flows that appear as inputs of the equations above for the next time-step simulation, to be used as initial conditions.

4.2.3 Reservoirs models

The four reservoirs can be described by a simple linear equation, derived from the mass conservation equation:

$$\mathrm{V_k(t+1) = V_k(t) + \left[u_{inlet,k}(t) - u_{outlet,k}(t)\right] \cdot F_C}$$

where:

- $\mathrm{V_k(t)}$ = volume of the reservoir k [Mm^3] at the end of the time-step t (=[t, t+1]); if Δt=1day the convention assumed is to consider $\mathrm{V_k(t)}$ the volume at the end of the day t; $\mathrm{V_k(t)}$ is the state of the model for the lake k;
- $\mathrm{u_{inlet,k}(t)}$ = flow(s) of the inlet channel(s) of the lake k, over the horizon [t, t+1) [m^3/s];
- $\mathrm{u_{outlet,k}(t)}$ = flow of the outlet channel over the horizon [t, t+1) [m^3/s];
- $\mathrm{F_C}$ = conversion factor from m^3/s to Mm^3 ($3600*24*10^{-6}$).

The inlets and outlets channels are configured differently for the four lakes as already shown in Figure 20 and summarized below:

1. Aube: 1 inlet and 1 outlet channel;
2. Seine: 1 inlet and 1 outlet channel;
3. Pannèciere: no inlet channel (dam directly on the river) and 1 outlet channel;
4. Marne: 2 inlet channels and 1 outlet channel.

The reservoirs equation for each lake k takes in input a different number of inputs $\mathrm{u_{inlet,k}}$ and $\mathrm{u_{outlet,k}}$. In total there are 8 decisions to be applied at each time step for the four dams: two decisions for Aube and Seine; three for Marne; one for Pannecière.

For a scheme of the position of the dams and the derivation channels in the river network the reader may refer to Appendix E.
The initial state of the reservoirs model at the beginning of the simulation period is set equal to the storage value obtained by simulation of the actual management rules on the same scenario at the beginning of the first day of the simulation horizon.

4.3 Definition of criteria and indicators

An evaluation criterion is an attribute, or a factor, with which the Decision Maker (DM) or a Stakeholder judges the performance of an alternative from the viewpoint of one of his/her interest [Soncini-Sessa et al., 2007].
A quantitative indicator *i* is a function of the trajectories x_0^h, u_0^{h-1} and d_0^h of the states, of the controls and of the disturbances that act on the system in a time horizon H=[0, ... , h], termed design horizon.
When it is not possible to associate an indicator to a criterion a proxy indicator should be used i.e. a variable in a logical relationship with the criterion and related to the effects of the alternatives.

Following the Participatory and Integrated Planning procedure (PIP, [Soncini-Sessa et al., 2007]), in order to formalize the objective, first, it's good to define the ideal criteria and the indicators, expressed in words, corresponding to the *sector of interests* identified in section 2.3. Following these criteria we will be able to build an objective function in the next section. The definition of these criteria and cost-functions presents some *arbitrary aspects* that will be highlighted for clearness. All the sub-objectives will be defined as costs to be minimized.

1. *Floods*: the criterion can be the total damage produced by floods; a possible proxy indicator is the difference between the actual river flow/level and a warning maximum threshold flow/level.
2. *Water demand*: the criterion can be the satisfaction of the demand; a possible indicator is the supply deficit or the difference between the actual river flow and a minimum threshold flow;

As we have already discussed in section 2.5.1, warning thresholds for the river flows are defined at some downstream control stations and are actually used by the manager to assess the performance of his management strategy. The current information about the thresholds was already reported in Table 1. They were defined in two groups, maximum thresholds for floods and minimum thresholds for low-flows. So it's natural to think to use these thresholds for expressing our indicators.
Moreover, we know that reference flow thresholds, defined at other stations immediately downstream the reservoirs, are used by the manager as a constraint to change his decisions respect to the filling curves. In fact, overcoming these reference flows corresponds to floods in the sub-basins just downstream the dams. For this reason we must take into account also these reference thresholds for the flood indicator.
At this point it becomes possible to define in a formal way the indicators for high and low flows.

4.4 Step-costs on system transitions and penalty-costs on final states

4.4.1 High flows step-costs

As for floods, we can express the effects at time t with a proxy indicator defined as a linear combination of the differences between the current flows at time t and the thresholds flows of the river at the control stations, when this difference is positive. Since there are three thresholds that we want to be respected in order of priority (1st of *vigilance*, 2nd of *regular flooded areas* and 3rd of *exceptional flooded areas*) it's natural to think to use a *piecewise linear cost function for the cost at each station, with increasing slopes for increasing flows*. The first threshold is the less important since it doesn't correspond already to floods occurrence, but anyway it's dangerous to overpass it. So, at the beginning of this work we supposed the cost being zero until the first threshold flow and linearly increasing after. However, after some simulation experiments we've found that having a cost already increasing from the first threshold led to the result that the reservoirs were filled too early in a flood event, consuming their capacity too soon.

For this reason, we have then redefined the *first threshold where to start to increase the step-costs, as the first vigilance threshold plus a percentage p of the interval between the first and the second thresholds (e.g. p=0.9)*. Then the slope increases more and more in correspondence to the second and the third thresholds.

The thresholds where the slopes of the step-costs change can be so defined:

$$\begin{cases} \overline{q^J_{hf\,1}} = \overline{q^J_{vig.,HF}} + p \cdot \left(\overline{q^J_{al.,HF}} - \overline{q^J_{vig.,HF}}\right) \\ \overline{q^J_{hf\,2}} = \overline{q^J_{al.,HF}} \\ \overline{q^J_{hf\,3}} = \overline{q^J_{cr.,HF}} \end{cases}$$

where:

- $\overline{q^J_{vig.,HF}}$, $\overline{q^J_{al.,HF}}$ and $\overline{q^J_{cr.,HF}}$ are the thresholds for high flows at the control station j, whose values are reported in Table 1;
- p is a parameter between 0 and 1 (e.g. $p = 0{,}9$) defining the level where the step-cost begins to increase, in the interval $\left(\overline{q^J_{al.,HF}}; \overline{q^J_{vig.,HF}}\right)$.

The control stations, that are to be taken into account for evaluating the performance of the controlled system, are placed at strategic points in the river network. For example, among the nine control stations, there are:

- *Paris* : extreme downstream of the model;
- *Arcis-sur-Aube* : downstream of Aube Lake before the confluence with Seine river;
- *Courlon-sur-Yonne* : downstream of Pannecière Lake just before the confluence with Seine river;
- *Chalons-sur-Marne* : downstream of Marne Lake;
- *Méry-sur-Seine* : downstream of Seine Lake before the confluence with Aube river.

For the detailed information about the thresholds at all the monitoring stations the reader should refer to section 2.5.1.

We can define the floods indicator for the whole system as a weighted sum of the costs associated to all the control stations. For each station j the total cost for high-flows on the optimization horizon h can be expressed as the sum of h functions g_t^j, called *step costs*. We assume that each step cost is a function $g_t^j(\cdot)$ that depends linearly on the state of the system in correspondence with the station *j*, that is the current flow q_t^j at that point. Thus, the step-cost is expressed so far in units of measurement of flow rate [m^3/s]; however, this cost will be then normalized and expressed in dimensionless units (see section 4.6.1).
So, the step-cost function can be expressed in the way described by the following piecewise function:

$$g_t^{\,j,HF} = \begin{cases} 0 & if\ q_t^j < \overline{q^J_{hf\,1}} \\ \alpha \cdot \left(q_t^j - \overline{q^J_{hf\,1}}\right) & if\ \overline{q^J_{hf\,1}} \le q_t^j < \overline{q^J_{hf\,2}} \\ \beta \cdot (q_t^j - \overline{q^J_{hf\,2}}) + \alpha \cdot (\overline{q^J_{hf\,2}} - \overline{q^J_{hf\,1}}) & if\ \overline{q^J_{hf\,2}} \le q_t^j < \overline{q^J_{hf\,3}} \\ \gamma \cdot (q_t^j - \overline{q^J_{hf\,3}}) + \beta \cdot (\overline{q^J_{hf\,3}} - \overline{q^J_{hf\,2}}) + \alpha \cdot (\overline{q^J_{hf\,2}} - \overline{q^J_{hf\,1}}) & if\ q_t^j \ge \overline{q^J_{hf\,3}} \end{cases}$$

where:

- q_t^j is the river flow at time t at the reference point j (state of the model);
- $\overline{q^J_{hf\,1}}, \overline{q^J_{hf\,2}}$ and $\overline{q^J_{hf\,3}}$ are the high-flows thresholds defined above;
- $\alpha < \beta < \gamma$ are the slopes of the piecewise linear cost, steeper as flows q_t^j increase.

It should be noted that these choices are arbitrary. We know that the cost must increase with the river flow, but we could have chosen another class of function, such as a quadratic one. However the linear form is the most simple and implies a steep growth of the cost already in proximity of the thresholds. Also the choice of the slopes values is arbitrary and depends on the preferences of the decision maker. We will explain how the slopes are defined in section 4.6.2.2, in order to balance the high-flows step-costs with the penalty cost on the final storages.

4.4.2 Low flows step-costs

For low flows, in order to evaluate the satisfaction of the demand for drinking water and the other water uses of the river, the thresholds of low flows can be used as we have done for high flows. The thresholds taken into account for low flows are three: *vigilance, alert* and *crisis*; their values are reported in Table 1. We adopt the following notation for the thresholds values:

$$\begin{cases} \overline{q^J_{LF\,1}} = \overline{q^J_{vig.,LF}} \\ \overline{q^J_{LF\,2}} = \overline{q^J_{al.,LF}} \\ \overline{q^J_{LF\,3}} = \overline{q^J_{al.r.,LF}} \\ \overline{q^J_{LF\,4}} = \overline{q^J_{cr.,LF}} \end{cases}$$

where $\overline{q^J_{vig.,lf}}$, $\overline{q^J_{al.,lf}}$, $\overline{q^J_{al.r.,lf}}$ and $\overline{q^J_{cr.,lf}}$ are the thresholds for low flows at the reference station j.

So, at each reference station j, we define the low-flows step-cost as a piecewise linear cost function with a changing slope in correspondence of the different thresholds. We assume to start with a cost of zero for flow at $\overline{q^J_{LF\,2}}$. So, the step cost $g^j_t(\cdot)$, at the station j and time-step t, linearly depends on the flow at the station j, in the way described by the following function:

$$g_t^{j,LF} = \begin{cases} 0 & if\ q_t^j > \overline{q^J_{LF\,2}} \\ \alpha \cdot (\overline{q^J_{LF\,2}} - q_t^j) & if\ \overline{q^J_{LF\,3}} < q_t^j \le \overline{q^J_{LF\,2}} \\ \beta \cdot \left(\overline{q^J_{LF\,3}} - q_t^j\right) + \alpha \cdot (\overline{q^J_{LF\,2}} - \overline{q^J_{LF\,3}}) & if\ \overline{q^J_{LF\,4}} < q_t^j \le \overline{q^J_{LF\,3}} \\ \gamma \cdot \left(\overline{q^J_{LF\,4}} - q_t^j\right) + \beta \cdot (\overline{q^J_{LF\,3}} - \overline{q^J_{LF\,4}}) + \alpha \cdot (\overline{q^J_{LF\,2}} - \overline{q^J_{LF\,3}}) & if\ q_t^j \le \overline{q^J_{LF\,4}} \end{cases}$$

where:

- q_t^j is the river flow at time t at the reference point j (state of the model);
- $\alpha < \beta < \gamma$ are the slopes of the piecewise linear cost, steeper as flows q_t^j decrease.

The same observations done in the previous section, about the arbitrary choices made for the high-flows step-costs definition, are valid for the low-flows step-costs. As for the slopes, we will explain how they are defined in section 4.6.2.1, in order to balance the low-flows step-costs with the penalty cost on the final storages.

4.4.3 Reference flows step-costs

Overcoming these reference flows corresponds to occurring floods in the sub-basins just downstream the dams. For this reason we must take into account also the reference flows at each time step of the simulation. We can define the step-cost $g^j_t(\cdot)$ for reference flow, at the station j and time-step t, that linearly depends on the flow at the station j, in the way described by the following piecewise function:

$$g_t^{j,Qref} = \begin{cases} 0 & if\ q_t^j < \overline{q_{ref.}^J} \cdot fs \\ \alpha \cdot (q_t^j - \overline{q_{ref.}^J}) & if\ \overline{q_{ref.}^J} \cdot fs \le q_t^j < \overline{q_{ref.}^J} \\ \beta \cdot (q_t^j - \overline{q_{ref.}^J}) & if\ q_t^j \ge \overline{q_{ref.}^J} \end{cases}$$

where:

- q_t^j is the river flow at time t at the reference point j (state of the model);
- $\overline{q_{vig.,lf}^J}$, $\overline{q_{al.,lf}^J}$ and $\overline{q_{cr.,lf}^J}$ are the thresholds for low flows at the reference station j;
- $\alpha < \beta$ are the slopes of the piecewise linear cost, steeper as flows q_t^j increase.

4.4.4 Total step-costs

The *total step cost* at time t g_t^{tot}, for both floods and low flows, is the weighted sum of the step costs at all the reference stations:

$$g_t^{tot} = \sum_{j=1}^{N} (w_j^{floods} g_t^{j,floods} + w_j^{lowflows} g_t^{j,lowflows} + w_j^{Qref}\ g_t^{j,Qref})$$

where:

- N is the number of control stations for high/low flows (see Table 1) plus the number of control stations for reference flow (see Appendix A);
- w_j^{floods}, $w_j^{lowflows}$, w_j^{Qref} are the weights of the reference station j for floods, low flows, and reference flows used to balance these components of the total step-cost.

4.4.5 Penalty-cost on the final storages of the lakes

Since the optimization horizon is finite, in the MPC approach it's usually necessary to take into account a penalty-cost on the final state of the system that must assure that the choice of the controller is not influenced only by the short time horizon and that instead the management is sustainable in the long term. In our case study, this sustainability requirement can be formalized as trying to stay as close as possible to the Filling Curves of the lakes. This is an arbitrary but natural choice derived from the meaning of the FC. Therefore the desire of a sustainable policy can be implemented adding a cost on the final state of the reservoirs that penalizes the decisions that cause large deviations of the volumes of the lakes from the filling curves at the end of the optimization horizon. So we can define the penalty on the final storage $g_h^{tot}(V_h)$ as the sum of the squares of the differences between the volumes of the lakes at time t=h and the objective filling curves at the same time-step. In this way minimizing this penalty-cost we will assure that the state $\boldsymbol{V}_h$ of the reservoirs, which would result from the application of the sequence of optimal decisions, will be as nearest as possible to the state defined by the objective filling curves.

$$g_h^{tot}(V_h) = w_{penalty} \cdot \sum_{i=1}^{4} (V_h^i - V_{curve,h}^i)^2$$

where:

- V_h^i is the volume of the i-th reservoir at time h simulated applying the controls proposed by MPC;
- $V_{curve,h}^i$ is the volume of the filling objective curve of the i-th reservoir at time h.

It should be noted that the quadratic form of the penalty-cost $g_h^{tot}(\boldsymbol{x}_h)$ has been chosen to penalize in the same way positive and negative values of the difference $V_h^i - V_{curve,h}^i$. Moreover the quadratic form penalizes more large deviations than small ones.

In order to counterbalance correctly the step-costs and the penalty on the final storages, we found the necessity to develop the penalty function definition in a more refined way, using:

- two different functions for the cases of being over or under the FC (as it will be explained in section 4.6);
- a piece-wise quadratic function for each of the two cases (over/under the FC).

4.4.6 Penalty-cost on by-passing the river

Since for the controlled reservoirs, except for Pannecière, there are both inlet and outlet channels to derive water into/from the lakes, there are countless equivalent solutions taking and discharging water at the same time. This management behavior implies by-passing the river without changing effectively the effects downstream the lakes. The solutions opening inlet and outlet derivations at the same time are not good for the manager that would have to take care of two flows (opening two gates) instead of one. For this reason, to help the optimizer to discern between these equivalent decisions, we defined a penalty cost to avoid to take and discharge at the same time g_t^{bypass}.
One might wonder why the controller bothers to calculate two decisions, inlet and outlet, for each reservoir when the effects depend on the difference between these two values. The reason is that there are some constraints downstream the inlet derivations (legal constraints as reserved flows) that we must take into account as we will see in the next section.

4.5 Constraints

After having defined our objective function we need also to define the constraints that the solution has to respect. There are some *general constraints* regarding the whole system (all the lakes/channels/river stations) that are essential and some *special constraints* that concern only specific points of the system. The latter ones in a first implementation could be neglected without jeopardizing too much the correctness of the implementation. So we will implement only the general constraints.

General constraints:

- **Maximum capacity of intake and discharge channels**: it is the *physical capacity* of the channels; this constraint can be mathematically formulated as *boundaries on the decisions domain (hard constraint).*
- **Maximum and minimum volumes of the lakes:** there's no sense to overcome this range of admissibility, over the maximum we can't fill more/below the minimum we can't empty more the lakes; this constraint can be mathematically formulated in two ways:
  - *Hard constraint* on the volumes; since volumes are linearly dependent from decisions as seen in section 4.2.3; this formulation is equivalent to *linear constraints on decisions;*
  - *Soft constraint* on the volumes.
- **Minimal reserved flow in the river**: it is the minimum flow for assuring life to the aquatic species, that is a *legal environmental obligation* (Article L432-5 of the French Environmental code); this constraint can be mathematically formulated in two ways:
  - *Soft constraints* on the river flows (controlled system states) or on *inlet decisions* (to limit the inlet flow rate depending on the upstream flow).

Additional special constraints:

- **Maximum allowable flow for the Seine inlet:** since the Seine inlet connection has no gates at the derivation, differently from other lakes, the derivable flow rate depends on the upstream level in the river; however at the moment the exact formula it's not available, so this constraint will not be implemented. In the future it could be added as:
  - *upper bound on inlet decision*, to limit the inlet flow rate depending on the upstream flow; it's a *physical* relation *(hard constraint).*
- **Maximum daily variation of flow rate for the discharge from Pannacière lake**: it can't be more than 2 m^3/s for 24 hours. This constraint can be mathematically formulated as:

- as *soft constraint* on the sequence of decisions for Pannacière.

4.5.1 Soft-constraints

4.5.1.1 *Reserved flow*

We have seen in section 2.4.2 that the reserved flow is a legal constraint for the manager of the lakes that defines the minimum flow to be left in the river. The flow in the river downstream an inlet derivation or the dam (for Pannecière that is directly on the river) must be greater than the *reserved flow,* when it's possible. In other words, the inlet flows to the reservoirs must let always this minimum reserved flow or more, if it's possible; moreover the inlet flow must be equal to zero if the river flow upstream the inlet channel is already less than the reserved flow. In our problem formulation we chose to implement this constraint as a soft constraint, since it is not a physical constraint and it's better to use soft constraints in this case. The soft constraint formulation can be translated in the sentence: "*when it's possible, the reserved flow must be ensured in priority respect to the other objectives*". So we can define a penalty $g_{Qres}^{sc}(\boldsymbol{Q}_t)$ as the sum of the squares of the differences between the current flow of the river and the reserved flow, multiplied by a very high weight w_{sc_Qres}. Moreover it's natural to think to use a security factor fs, as for prudence we want the controller to respect the limits with a security margin.

- For each river station i downstream an inlet derivation, when $Q_{inlet}^{i} > 0$:

$$\begin{cases} if\ Q_t^i > Q_{reserved}^i \cdot fs & g_t^{Q_{res}}(Q_t^i) = 0 \\ if\ Q_t^i \leq Q_{reserved}^i \cdot fs & g_t^{Q_{res}}(Q_t^i) = w_{sc\_Qres} \cdot (Q_t^i - Q_{reserved}^i \cdot fs)^2 \end{cases}$$

- For the whole system (for the 5 stations downstream an inlet derivation):

$$g_t^{Q_{res}\,TOT}(Q_t^i) = \sum_{i=1}^{5} g_t^{Q_{res}}(Q_t^i)$$

where:

- $fs = 1{,}05$ is a security factor, to respect the limit with a margin;
- $w_{sc_Qres}$is the weight, that must higher than the other cost-components weights;

The reserved flows values for each lake are reported in the schemes in Appendix A.

4.5.1.2 *Limits for the volumes of the lakes*

The reservoirs have a fixed capacity and so a maximum volume is defined. Moreover, a minimum volume is also defined for a correct reservoir's operation. These constraints are a priori hard constraints, but for easiness of implementation and to reduce hard-constraints for MPC, they can be implemented as soft-constraints, associating a very high penalty cost to the eventual overcoming of these volume thresholds. We can define the soft-constraint cost $g_V^{sc}(\boldsymbol{V}_t)$ as the sum of the squares of the differences between the current volume of the lakes and the minimal or maximal volume, multiplied by a very high weight w_{sc_V}. Moreover it's natural to think to use a security factor fs , since, for prudence, we want the controller to respect the limits with a security margin.

- For each lake *i*:

$$\begin{cases} if\ V_{min}^i < {V_t}^i < V_{max}^i & {g_t}^{V\,sc}(\boldsymbol{V}_t) = 0 \\ if\ {V_t}^i \leq (V_{min}^i + fs) & {g_t}^{V\,sc}(V_t) = w_{sc\_V} \cdot ({V_t}^i - V_{min}^i - fs)^2 \\ if\ {V_t}^i \geq (V_{max}^i - fs) & {g_t}^{V\,sc}(V_t) = w_{sc\_V} \cdot (V_{max}^i + fs - {V_t}^i)^2 \end{cases}$$

- For the whole system:

$$g_t^{V\,TOT}(\boldsymbol{V}_t) = \sum_{i=1}^{4} g_t^{V\,sc}(\boldsymbol{V}_t)$$

where:

- $V_t^{\,i}$, $V_{min}^{\,i}$ and $V_{max}^{\,i}$ are respectively the current, minimum and maximum volumes of the lake i;
- $fs =$ security factor (e.g. it can be defined as one thousandth of the range $[\mathrm{V_{min}}; \mathrm{V_{max}}]$);
- w_{sc_V} is the weight, that must higher than the other cost-components weights and also higher than w_{sc_Qres}.

The maximum and minimum volumes of all the lakes are reported in the schemes in Appendix A.

4.5.2 Hard-constraints

Limits for the inlet/outlet flows

The capacities of the channels are implemented as hard constraints on the decisions, since they represent physical constraints that the system is not allowed to overpass. So, the equation that describes these limitations can be written as:

$$0 < \boldsymbol{u}_\tau < \boldsymbol{u}_{max} \qquad\qquad \tau = t, \dots\ ,\ t + h - 1$$

where $\boldsymbol{u}_\tau$ is the control decisions vector at time τ and $\boldsymbol{u}_{max}$ is the vector containing the capacities of the channels.
The capacities for all the lakes connections are reported in the schemes in Appendix A.

4.6 Sub-objectives normalization and weighting

4.6.1 Normalization of the sub-objectives

Since there are many sub-objectives and each one contains different components that are added up, a good approach is to normalize each cost-component using the maximum value that it may assume (making the ratio between the original cost and maximum value). In this way each component has *dimensionless units of measurement* and can be summed and compared with the others.
A sensible procedure of normalization is performed using the Maximum Allowed Value Estimate (MAVE) that is an estimate of how much the state we want to control may vary [Van Overloop, 2006].
Each sub-objective is normalized on its maximum range of variation. Since the objective functions are defined in a piecewise form, we have ranges within which the objective function is the same and therefore its value is normalized in each range on its amplitude.
As for the step-costs for floods and low-flows, the normalization is carried out on the basis of the differences between the thresholds, for each band of the piece-wise linear function.
As for the penalty on the final storages of the reservoirs, the normalization is done on the differences between different filling curves, defined to provide a support for counter-balancing the different sub-objectives weights as it will be explained in the next section. These new curves are calculated as reference storages, under the FC, for supplying water for each low-flow threshold and, above the FC, for letting a fixed storage capacity to avoid exceeding the higher flood thresholds (see next section 4.6.2).

4.6.2 Weighting the sub-objectives

Since we have different sub-objectives in the cost function and some of them are conflicting, one might formalize the problem in a multi-objective framework and be interested in finding a representative set of

the Pareto optimal solutions. However, this is not our goal, since we need to find a single decision to be applied for the first day of the horizon. Our target is not to quantify the trade-offs in satisfying the different objectives. Moreover, in our case-study the conflict among the objectives is of a particular kind: it is a conflict between the same objectives but seen in different time perspectives (in the finite horizon and in the long term). The step-costs considered in the finite horizon for floods and droughts are not in contrast between them, since floods and droughts are not simultaneous events in space and time. Instead the step-costs are in contrast with the penalty on the final storage of the reservoirs. In other words the prevention of floods and droughts in the short term is in contrast with trying to stay as close as possible to the objective filling curves of the lakes. In fact the step-costs for low-flows lead the controller to release water from the reservoirs in case of necessity and this can cause the volumes of the reservoirs to go under the filling curves. Viceversa, the step-costs for high-flows lead the controller to store water into the reservoirs in case of necessity and this can cause the volumes of the reservoirs to be over the filling curves. The solution to this conflict can be achieved defining a *compromise* to counterbalance:

- what one should do *in the immediate horizon,* filling/emptying the reservoirs, to face floods/droughts;
- what one should save for later leaving the capacity to fill/empty the reservoirs to be perpetuated in time *after the immediate horizon*.

This compromise involves a *political choice* but can be framed in a precise and explicit formalization that could help the decision-maker to express his preferences. We developed a new framework to solve this conflict given the specificities of the case-study. The core idea is defining correctly the preferences between what to do now and what to save for later, using the concept of filling curves. In fact we know that the FCs are constructed in order to have enough water to ensure a minimal flow threshold in the river for all the releasing season taking care that there is always enough space available for floods control. We assumed that the minimal flow threshold that the FCs can ensure is the *vigilance low-flow threshold* $\overline{q^J_{vig.,LF}}$ at the extreme downstream station of Paris.
So starting from the FCs we can define some reference levels above or below the original curves that are less desirable than the FCs level. These levels are defined by new filling curves whose scale of priority is directly deducible by the neighborhood to the original curve. After that it's necessary to understand in which cases the manager of the lakes can allow to get to these levels, defining the magnitude of the flood/drought events whose immediate avoidance in the optimization horizon is more important than remaining at those levels. Once this trade-off is established the weights to counterbalance step-costs for floods/droughts and penalty on final storage are consequently determined. This approach will be followed in the next two sections for both flood/drought events calculating a set of weights for balancing the step-costs and the penalty on the final state. It should be noted that defining this scale of priorities is a political choice that involves the preferences of the DM. However it's easy to see the importance of applying this procedure even defining an arbitrary scale of priorities, because it's natural to accept some costs in the present to be sure that in the future we won't pay an enormous cost. These considerations will be better explained in the following paragraphs.

4.6.2.1 *Balancing low-flows step-costs and penalty to be under the FC*

The importance of weighting low-flows step-costs and penalty on the final storages can be seen analyzing the results of the MPC optimization using a formulation with a very little weight on the penalty-cost respect to the low-flows step-costs. These results show that the water resource available in the reservoirs is used as soon as necessary in the short optimization horizon (some days). In a long run of MPC (at least some months) the effect of this policy is that the lakes are emptied too much time before the end of the releasing season or even at the beginning. For example, this behavior can be seen in the results reported in the following plot showing the volume of lake Marne simulated with MPC over a simulation horizon of one year.

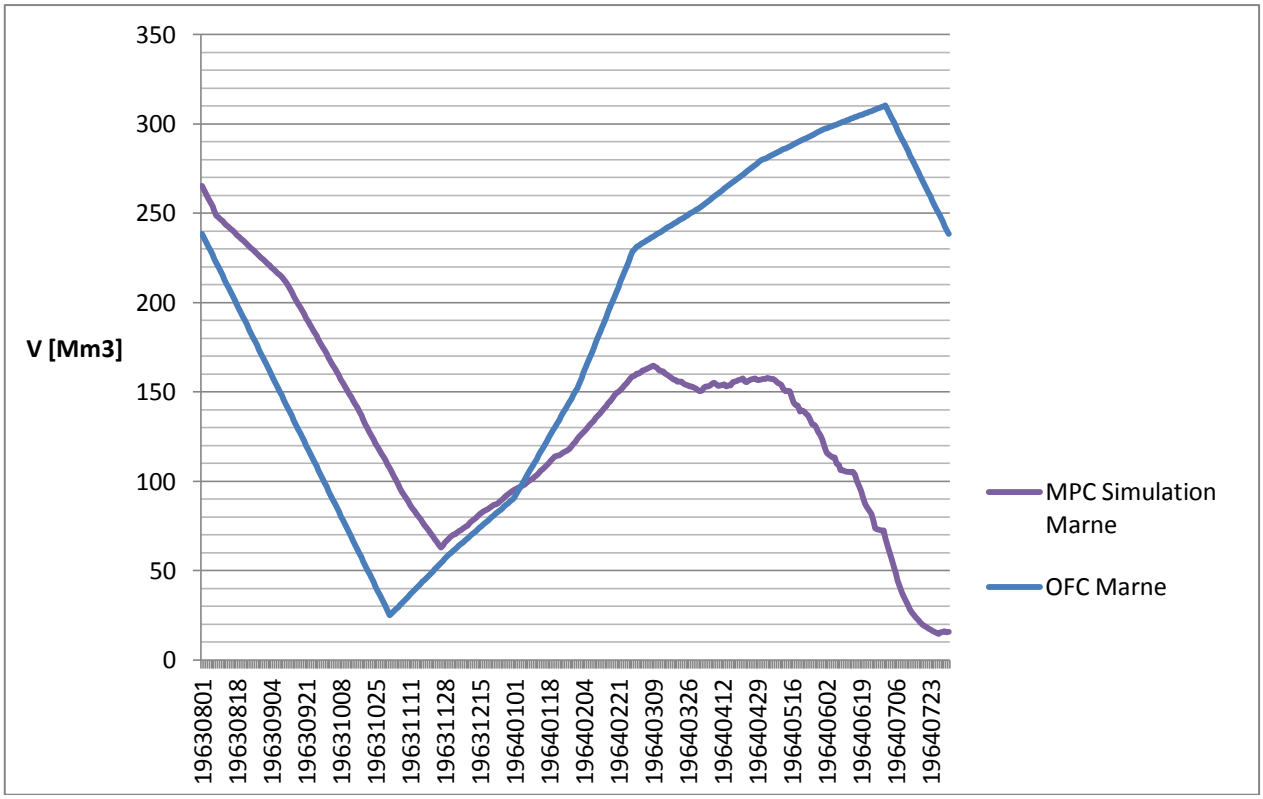


Figure 21. Volume of Marne lake [Mm3] simulated with MPC management from 01/08/1963 to 01/08/1964. Example of the behavior of MPC management with a smaller weight on the penalty-cost respect to the low-flows step-costs.

In Figure 21 the volume of lake Marne simulated with the MPC management drops down in the filling period because of the small weight on the penalty cost for the filling curve respect to low-flows step-costs. The period of the simulation (1963/1964) is not one of the most characterized by droughts, presenting only a few events under the first or second thresholds at some stations. However, the lake is completely emptied already at the beginning of the releasing season, in July, after having faced some low-flow events, that not being critical, would not deserve this priority. This kind of solution is unacceptable, because it means that we do not care of the costs in the future in order to pay a lower cost for low-flows in the short term. By acting like this, the costs for droughts in the long term are enormous. To avoid this behavior of the MPC management we have to balance appropriately the low-flows step-costs and the penalty-cost to be under the FC. This problem is complicated since it consists in a comparison between a penalty on the storages calculated after an horizon of a few days and the costs of drought events that may last a few months. Here we propose an original solution for this problem trying to design an MPC controller that adapts its hedging policy to the states of the reservoirs, taking into account the remaining capacity to satisfy the different thresholds. In other words we want to try to reduce the costs in the short term as much as possible without compromising the future adaptive capacity of the system.

Our solution is based on the presence of different levels for the objective of sustaining low-flows, i.e. different low-flows thresholds. So, it's important to remind here the meaning of the low-flows thresholds that from the 1st ($\overline{q^J_{vig.,LF}}$) to the 4th ($\overline{q^J_{cr.,LF}}$) correspond respectively to less or more important restrictions of water uses, as explained in section 2.5.1. Therefore, given these different levels of low-flows severity, to design our policy we must answer to the following questions:
"*Do we allow the flows in the river to go sometimes under the 2nd or 3rd low-flows threshold because it's more important to respect the 4th (most critical) threshold (and so we want to save more water for later)*? *Or do we prefer to stay above the 2nd or 3rd threshold the most of the time and to risk to pass under the 4th one later, since we have not kept enough water in the reservoirs?*".

Evidently these questions concern policy decisions. However, we can make the two following logical assumptions:

A. we associate the biggest costs to the lowest thresholds (4th), in reason of the most important impacts on water uses;
B. we assume the temporal equivalence of the costs over a season.

Therefore, from assumptions A and B we derive the following response to the questions above:
"*Yes, sometimes we may allow the flows in the river to go under the 1st, 2nd or 3rd low-flows threshold because it's more important to be able to satisfy the 4th (most critical) threshold and so to keep water in the reservoirs for this objective for later.*"
In other words, from A+B we derive the following point:

C. we accept to release water from the lakes to sustain short-term low-flows of a certain level k only until a certain target volume FC_X, below which we could not be able to face low-flows events in the future of a level more critical than k.

From this choice we derive a new framework to counterbalance low-flows step-costs and penalty to be under the FC.
To define the levels of low-flows severity k we can use the distance of the current flows from the different thresholds.
To define the target levels for the volume, FC_X, we developed some new FCs analogously to the original FC.
We started assuming that the FC is statistically constructed in order to sustain low-flows ensuring that the 1st vigilance threshold for low-flows $\overline{q^J_{vig.,LF}}$ at Paris can be satisfied for all the releasing season. This assumption is realistic and based on the philosophy of the statistical method developed in Bader [1992].
Since we have other three low-flows thresholds, we can derive other three different Filling Curves (FC2, FC3 and FC4) that can statistically ensure respectively the 2nd/3rd/4th thresholds (as the original FC ensures the 1st threshold).
To calculate these curves in a simplified way, we decided to follow the following procedure:

1. Integrate the differences in flows between the different low-flows thresholds at the extreme downstream station of Paris all over the releasing season. So we obtain the differences in volumes to be stored to ensure statistically the different thresholds for all the releasing season.
2. Allocate the volumes obtained at point 1 to the four lakes in function of their capacities (following the strategy of centralization of the controller) respect to the total capacity.
3. Subtract these volumes to the maximum volume of the original FC (FC1) of each lake obtaining the new maximum points of the curves FC2, FC3, FC4.
4. Transform FC1 in the new filling curves FC2, FC3, FC4, following the same structure of the original filling curve that is composed of 3 parts: releasing season, prolonged releasing period and filling season. For each part, to obtain the new curves we proceed as follows:
    a) For the releasing season, from July to November: we apply a constant contraction factor to the original FC equal to the ratio between the maximum point of the original FC and the maximum points of the new FCs calculated at point 3: max(FC1)/max(FC*i*), with *i=2,3,4*.
    b) For the prolonged releasing period, between November and January: FC4 is equal to the minimum exceptional volume of the reservoir. The other curves are obtained prolonging the volume at the end of the releasing season with an horizontal line until it meets the straight line that interpolates the point of the FC at the 1st of November and the point of FC4 at the 1st of January.
    c) For the filling season, between the end of the releasing season and July: we connect the points at the end of the prolonged releasing season, found in the point 4.b, with the maximum volume points of the FCs obtained at point 3.

The result of this procedure for the Marne lake is reported in the following figure, while for all the lakes it is reported in Appendix F.

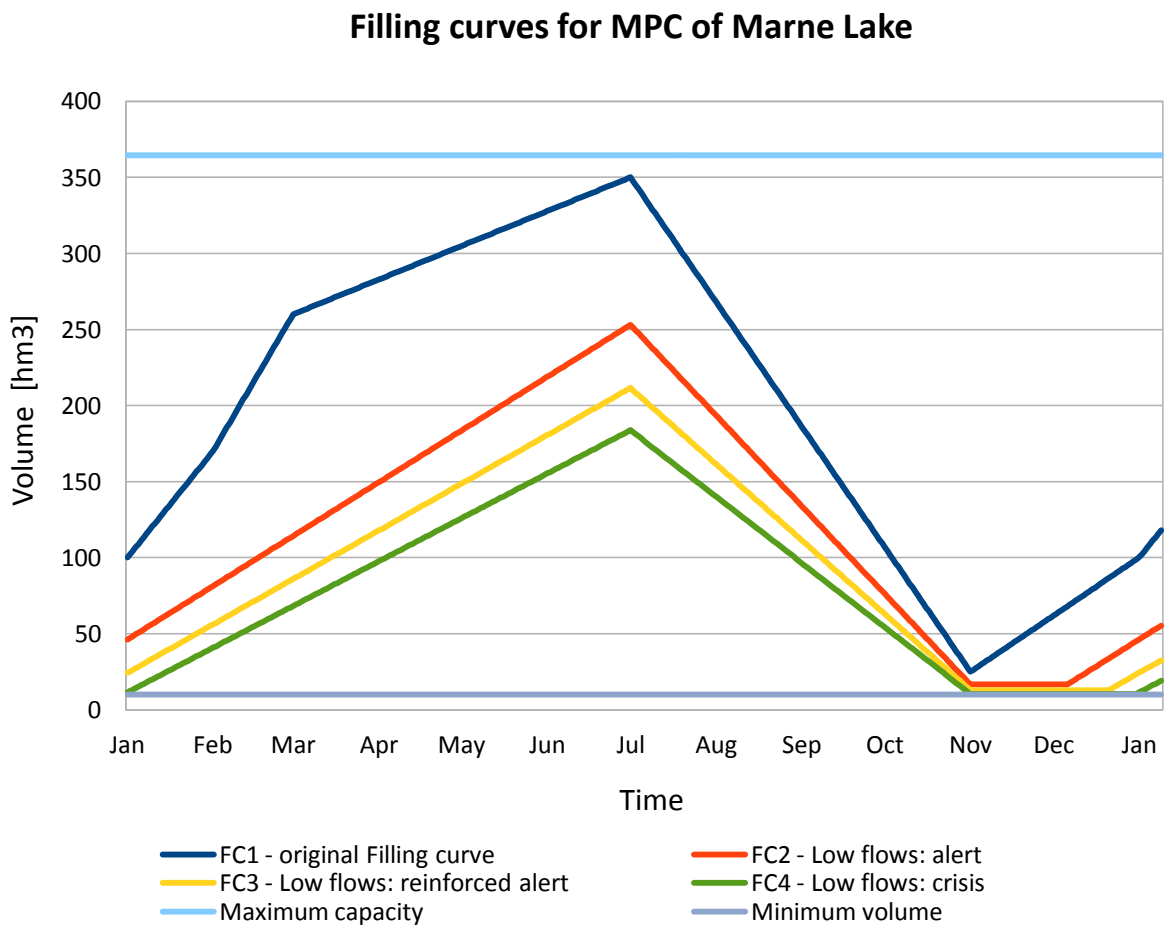


Figure 22. Filling curves defined for balancing low-flows step-costs and penalty to be under the FC for Marne lake.

At this point it's possible to use these new curves as reference levels to counter-balance the step-costs for low-flows with the penalty-cost on the final storage under the FC. To define the relative weight of these two costs we decided to follow the following assumptions:

1. The cost to be under the FC increases with a *piece-wise quadratic form* starting from zero at FC1. Each sub-function of the piece-wise function is defined for an interval of volume of the reservoir i between two close curves FC_j with j=1,2,3,4 and V_{MIN}.

$$\begin{cases} for\ j = 1,2,3: & g_h^{tot}(V_h) = w_{penalty_{FCj}} \cdot \sum_{i=1}^{4}(V_h^i - V_{FCj,h}^i)^2 \quad if \quad V_{FCj+1,h}^i \le V_h^i < V_{FCj,h}^i \\ for\ j = 4: & g_h^{tot}(V_h) = w_{penalty_{FC4}} \cdot \sum_{i=1}^{4}(V_h^i - V_{FC4,h}^i)^2 \quad if \quad V_{MIN}^i \le V_h^i < V_{FC4,h}^i \end{cases}$$

   where the coefficients of the quadratic function, $w_{penalty_{FCj}}$, change in each range [j, j+1] with j=1,2,3. Moreover: $w_{penalty_{FC1}} < w_{penalty_{FC2}} < w_{penalty_{FC3}} < w_{penalty_{FC4}}$.

2. As initial weight for pondering all the others, we set an arbitrary value for the sum of low-flows step-costs for being at $\overline{q^J_{LF\,2}}$ at all the stations for one time-step ($w_{lf\ T2-All\ stations}$). We remember that the low-flows step-costs are zero at $\overline{q^J_{LF\,1}}$, so, having fixed the value of $g_t^{j,LF}\left(\overline{q^J_{LF\,2}}\right)$ the first slope α of the step-cost function is determined.

3. We assume that the penalty cost associated to having the final storages of all the lakes at FC_j level is equal to the cost of having low-flows at level $\overline{q_{LF,J}}$ (with j=1,2,3,4) at all the stations s for all the horizon h (that is the maximum cost due to a drought of level $\overline{q_{LF,J}}$).

$$h \cdot \sum_{s=1}^{N_{stations}} g_t^{s,low-flows}\left(\overline{q^s{}_{LF\,J}}\right) = \sum_{i=1}^{Nlakes} g_h^i\left(V_h^{i,current} = V_h^{i,FC_j} \cdot\right)$$

This assumption translates the fact that between the penalty for having all the final storages under FC_j and the step-costs for low-flows between $\overline{q_{LF,J}}$ and $\overline{q_{LF,J+1}}$ our priority is the penalty on the final storages under FC_j (that leads to bigger costs). This is because we do not accept to go under FC_j and be anymore able to satisfy the j-th low-flow threshold in the future, just to satisfy it in the short forecast horizon (as stated in point C in this section). Moreover, it can be noticed that the fact of balancing the step-costs at all the stations with the penalty-costs at all the lakes is coherent with the centralization of the controller.
So, by using the equations of point 3, after defining the penalty cost function on the final storage under the FC (i.e. the values of $w_{penalty_{FCj}}$ with j $= 1,2,3,4$), the slopes of the low-flows step-costs function are consequently determined.

4.6.2.2 *Balancing high-flows step-costs and penalty to be over the FC*

At first glance, one could just give complete priority to floods step-costs because reservoirs are built to control floods. So we could weight more the step-costs for floods respect to the penalty-cost to have final storages over the FC. But how to define how 'weighting more' the step-costs? We might be tempted to take an extreme decision, setting the cost for a flood of level equal to the first threshold used to calculate the step-costs ($\overline{q^J{}_{hf\,1}}$) to be larger than the maximum cost to be over the FC (i.e. C(V=Vmax)) for all the lakes. This attempt is based on the same philosophy than the experiment done for low-flows with a very little weight on the penalty-cost respect to the low-flows step-costs. In fact it produces similar results satisfying the objective represented by the step-costs in the short horizon, but reaching a point from which this objective can't be anymore achieved in the future. For high-flows, this bad result means that the storing capacity of the lakes is already consumed after exceeding for some days the first less important threshold $\overline{q^J{}_{hf\,1}}$. So, when high-flows last more than a few days over $\overline{q^J{}_{hf\,1}}$ and increase to $\overline{q^J{}_{hf\,2}}$ or $\overline{q^J{}_{hf\,3}}$, there's not anymore the adequate capacity to store the excess flow in the reservoirs. This policy is clearly not sustainable.

To overcome this problem, we can adopt a similar solution to the one developed for low-flows in the previous section, defining some reference levels above the FC at which counter-balance the step-costs for floods at the different thresholds. For example, since the first threshold does not correspond to real floods but it's a vigilance threshold, we can limit the cost associated to overcoming this threshold with respect to the cost for being over the FC. In other words we want that the cost to be over the FC from a certain level ε over the FC is bigger than the cost for high-flows over the 1st vigilance threshold. Even for the other thresholds we can counterbalance the maximum cost associated to flows in a band between two flood thresholds with the cost to be at a certain level over the FC.
Following this strategy, we chose to set the cost for having the flows at all the control stations equal to the second threshold $\overline{q^J{}_{hf\,2}}$ for a period equal to the statistical average length of this event equal to the cost of being over the FC of a certain percentage ε of the free capacity. This equivalence is formalized in the following equation:

$$l_{hf\,2} \cdot \sum_{j=1}^{N_{stations}} g_t^{j,floods}\left(\overline{q^J_{hf\,2}}\right) = \sum_{i=1}^{Nlakes} g_h^i\left(V_h^{i,current} = V_h^{i,FC} + (V_{i,max} - V_h^{i,FC}) \cdot \varepsilon_{hf\,2}\right)$$

where:

- $V_h^{i,current}$ is the current volume of the reservoir i at time h;
- $V_h^{i,FC}$ is the volume of the objective filling curve of lake i at time h;
- $V_{i,max}$ is the maximum volume of the lake i;
- $\varepsilon_{hf\,2}$ is the percentage of the capacity at which the cost to be over the FC for all the lakes is equal to the cost of having all stations with flow at $\overline{q^J_{hf\,2}}$ for the average length $l_{hf\,2}$ of this event.

Similarly we can write the same kind of equations for the other thresholds. Then, by using these equations, after defining the penalty cost function on the final storage over the FC and the values of the parameters $\varepsilon_{hf\,k}$, with k=1,2,3, the slopes of the flood step-costs function are consequently determined. The values of the parameters $\varepsilon_{hf\,k}$ can be determined by sensitivity analysis. The values of the parameters used in the procedure described above ($l_{hf\,k}$, $\varepsilon_{hf\,k}$) are reported in Table 3.

In this way, the flood step-costs associated to flows in all the bands defined by the thresholds $\overline{q^J_{hf\,1}}$, $\overline{q^J_{hf\,2}}$ and $\overline{q^J_{hf\,3}}$ are balanced with the penalty to be over the FC.

4.7 Optimization problem

4.7.1 Optimization horizon and control horizon

The *optimization horizon (h)* is the horizon for which the control problem is solved.
As for calculating the optimization horizon it's necessary to take into account:

- The *information-control horizon* (T_{ic}), defined as the time span from the actual moment until the moment from which information (e.g. forecasted precipitation amount) does not influence the control actions anymore; it can be defined empirically (see Stive [2011]).
- The *forecast horizon* of the available predictions (equal to 9 days in our case), that is an upper bound for the effective usable optimization horizon.

In reservoirs management problems, from the stand-point of flood control the ideal optimization horizon can be defined as the time needed to create the storing capacity before the flood occurs. However, an indication on the effects of enlarging the horizon's length h can be derived by sensitivity analysis considering the perfect predictor, that can be simulated over historical periods for any length of the prediction horizon [Pianosi and Soncini-Sessa, 2009]. We remind to section 5.2.1 for the results of this sensitivity analysis.

The *control horizon* is the horizon for which a control decision is applied before changing it, i.e. the time span between two consequent decision-making steps. The control horizons can be taken time-varying along the optimization horizon, as it will be explained in section 4.7.1.3.

As for deciding the control horizon we have to take into account

- The *simulation time-step*, that is the time resolution used in the model of the controlled system (equal to 1 day in TGR model).
- The *forecasts time-step*, that limits how often we update the disturbances that affect the system (equal to 1 day in the forecasts available).

4.7.1.1 *Remarks on optimization and control horizon for floods control*

Optimization horizon: the ideal one for floods is the time needed before a flood to empty the reservoirs and create the storing capacity that will contain the flood [Pianosi and Soncini-Sessa, 2009]. For our case-study it is in the order of some days, about 5 days, because it is estimated to be the time needed to empty the reservoirs of a volume equal to the biggest historical floods of the period of data availability. The influence of the length of the optimization horizon h on the results will be tested performing a sensitivity analysis on the results with different h values (see section 5.2.1). For MPC with perfect predictor (using observations as forecasts) we can choose h as long as we want (since we have not the limit of the forecast horizon); for MPC with real forecasts and for TB-MPC we could overcome the limit of the forecast horizon available by using a "no-precipitation" scenario after the prediction horizon or other weather statistics.

Control horizon: a fine control horizon is useful for floods control, for the high-variance of inflows in periods of floods and the consequent need of a quick variation of the control. So our initial idea was to use a daily control horizon all along the optimization horizon. Then we have modified this choice, enlarging the control horizons, to reduce the number of variables to be optimized in order to deal with optimization problems, as it will be explained in sections 4.7.1.3 and 4.7.3.1.

4.7.1.2 *Remarks on optimization and control horizon for low-flows control*

Optimization horizon: the ideal horizon is the time needed to store enough water in the lakes to prevent low-flows (months) and satisfy the demand at downstream. This ideal horizon is more than the length of forecast horizon of ECMWF ensembles. We would need to have seasonal weather forecasts and demand forecasts. To take into account the water supply objective on a longer horizon than the forecast horizon, for the moment we use the penalty based on the final storages and FCs. The advantage of MPC can be to evaluate the convenience of a centralized hedging policy within the optimization horizon. In the future, to extend the decision horizon for low-flows, we could use "zero-precipitation" after the forecast horizon (e.g. 30 days) or statistic scenarios. In this case to reduce the number of decisions and the calculation time we could enlarge the control horizons (for example one decision every 10 days).

Control horizon: since the variance of rainfall and inflows in low-flows periods is lower than in high-flows periods, there's no need of a quick variation of the control within the optimization horizon. So we can use larger control horizons in case of low-flows conditions, taking one or two decisions along the optimization horizon, using the framework of the control horizon grids (see next section 4.7.1.3).

4.7.1.3 *Control horizons grids*

At the beginning of this study we opted to work with constant daily control horizons, calculating a different decision for every day. The first results that we obtained show us that a large number of decisions to take is not sustainable for the optimization problems encountered: the optimal solution improves decreasing the number of decisions. There's a trade-off between taking some accurate decisions and a lot of rough decisions. For this reason we decided to use a limited number of decisions enlarging the control horizons. Firstly, it must be noticed that each control decision along the prediction horizon could have a different horizon of application. So we can define the temporal distribution of the decisions over the prediction horizon using a control horizon grid (CHG).
To define a CHG, we define and use the following notation:

$$CHG = [t_1\ t_2\ \dots\ t_n]$$

$$\text{s.t. } \begin{cases} t_1 = 1 \\ t_n < h \end{cases}$$

where the time-steps t_i indicated in the square brackets represent the time-steps along the optimization horizon when the decisions are changed.

For example, using a grid equal to [1 2 3 5] means that: a daily decision is calculated for the first two time-steps (days); after that, the decision of the third step is repeated also for the fourth and from the fifth day to the end of the optimization horizon the same decision is applied every day. A schematization of the control horizon grid [1 2 3 5] is represented in the following figure, showing the distinct decisions calculated over the optimization horizon of 9 days.

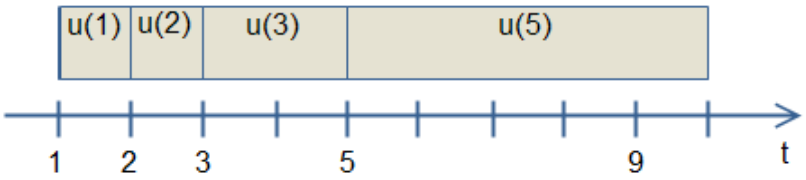


Figure 23. Schematization of the temporal distribution of the decision using a control horizon grid equal to [1 2 3 5] and applied over an optimization horizon of h=9 days.

Since the decision for the first time step of the prediction horizon is the only one that is effectively applied to the system, it's necessary to calculate it accurately and using a daily length for this first decision is preferable (i.e. fixing $t_2 = 2$). After that, since the system has a delay, the decisions for the first part of the optimization horizon are more important than the last ones, more distant in time. For this reason it's good to use a variable control horizon that expands going towards the end of the optimization horizon. Thanks to this idea, it's possible to reduce the number of decisions, in order to improve the convergence of the optimization and consequently reduce the computation time. The loss of performance of the optimal controls is limited because we are neglecting a detailed description of the control variations in the last part of the optimization horizon, that in part would have the most of the effects after the forecasts horizon for the presence of the delay in the hydraulic model. The use of a finer control horizon grid at the beginning of the prediction horizon allows to have more flexibility in changing decisions, that is very useful above all in case of floods control. So the problem is to find a compromise between these different necessities: a fine grid for flexibility in taking decisions and a large one for reducing the number of decisions (improving the convergence to the optimal solution). In case of low-flows, we can use a large grid all along the optimization horizon, since the inflows have a low variance in this case; for example we can take a distinct decision only for the first day and then the same decision for the rest of the horizon (CHG=[1 2]).
A sensitivity analysis was performed with different control horizon grids for high and low flows conditions and the results for high flows are provided in section 5.2.2.

4.7.2 Optimization algorithm

Since the model of the controlled system is not linear, the problem can't be directly solved in a linear-quadratic framework. If one would like to be able to apply this framework it would be necessary to use some stratagems to treat the conditions that make the model non-linear, in order to build the matrices of a linear model. Moreover, the constraints and objectives that we have just defined present a linear or quadratic piecewise form, conditioned to some *thresholds*, so they are not directly usable in the linear-quadratic framework, because the piecewise form and the presence of thresholds would require some tricks to manage it. We therefore excluded to follow the path of quadratic programming for these reasons. Moreover, at the beginning of this work there was at our disposal the code (programmed in *Scilab)* for the model of the watershed and for the simulation of the current management of the dams *(see section 4.9.1)*. So *direct-search optimization methods* were also preferable in terms of time required for implementation of the MPC controller, thanks to the possibility of using this code for the simulation of the system in the evaluation of the objective function. Among these methods, we decided to use the *Nelder-Mead* algorithm because it was well known as an efficient *derivative-free method for nonlinear optimization* [Nelder and Mead, 1965]. Moreover, Scilab provides different variants of this algorithm including a *constrained* version, called *Box Complex Method* [Box, 1965].

4.7.2.1 *Nelder-Mead (N-M) algorithm and the Box Complex method*

N-M is a *direct search* algorithm that:
– uses only objective-function values (*no gradient needed*);
– does not approximate the gradient.

The N-M algorithm is also known as *simplex method* because it uses a *simplex* that is a *set of n+1 vertices, in n dimensions space*. For this reason it should not be confused with Dantzig's simplex method for linear optimization. The simplex evolves in the search space subjected to geometric transformations of reflection, expansion, contraction and shrinkage. These operations are defined by the coefficients ρ (reflection), χ (expansion), γ (contraction), and σ (shrinkage).
The N-M algorithm follows the following steps [Baudin and Couvert, 2011]:

1. *Sort the vertices by function value*. Order the vertices: $f(1) \leq \ldots \leq f(n) \leq f(n+1)$
2. *Calculate the centroid of the simplex.* $B = (v(1) + \ldots + v(n))/n$
3. *Reflection*. Compute $R = (1+\rho)B - \rho v(n+1)$ and evaluate f(R).
4. *Expansion*. If $f(R)<f(1)$, compute $E=(1+\rho\chi)B - \rho\chi v(n+1)$ and evaluate f(E). If $f(E)<f(R)$, accept E and go to 1, else accept R and go to 5.
5. *Accept the new vertex R*. If $f(1) \leq f(R) < f(n)$, accept R and go to 1.
6. *Outside Contraction*. If $f(n) \leq f(R) < f(n+1)$, compute $Co=(1+\rho\gamma)B - \rho\gamma v(n+1)$ and evaluate f(Co). If $f(Co)<f(R)$, then accept Co and go to 1 else, go to 8.
7. *Inside Contraction*. If $f(n+1) \leq f(R)$, compute $Ci=(1-\gamma)B + \gamma v(n+1)$ and evaluate f(Ci). If $f(Ci)<f(n+1)$, then accept Ci and go to 1 else, go to 8.
8. *Shrink*. Compute the points $v(i)=v(1)+\sigma(v(i)-v(1))$ and evaluate $f(i)=f(x(i))$, for $i=2,3,\ldots,n+1$. Go to 1.

This algorithm is known to be a robust method to manage "noisy" functions, i.e. situations where the cost function is the sum of a general nonlinear function and a low magnitude function. This property to be less influenced by local optima could be helpful in our case study due to the complex form of the global objective function. The drawback of the choice of this algorithm is that it is known to be efficient for small problems, i.e. 30 variables at most, (see Baudin and Couvert, 2011) while the problem size of our case study is in general larger. In fact we have 8 variables for each decision-making day in the optimization horizon, so, for example, setting 4 decision-making steps in the optimization horizon, we have 32 variables for deterministic MPC and more for TB-MPC (a value function of the tree structure, see section 4.8).

The N-M component in Scilab provides three simplex-based algorithms for unconstrained and non-linearly constrained optimization problems.
We used the standard *Box Complex algorithm* [Box, 1965], i.e. a *constrained version* of the N-M method. This algorithm is based on a variable shape simplex, called *complex*, made of an arbitrary k number of vertices (k=2n is recommended by Box) that moves and evolves in the search space. The method requires that *an initial point in the search space* is given, that is assumed to be the first vertex.
Then, a set of *k-1 vertices around the initial solution* can be generated using different methods:

- the *standard method suggested by Box* that generates the other points using a random number and the bounds on the variables,
- the *Spendley method* that generates a regular simplex with a fixed length of the edges,
- the *Pfeffer method* that computes the simplex heuristically in the neighborhood of the initial guess.

Among these points, the worst solution is replaced by another point generated by the geometric transformations of the classic N-M algorithm. Firstly the algorithm performs a *reflection and expansion* of the worst point far from the *centroid* of the remaining points. Eventually a *contraction* into this centroid if the reflection with expansion is not successful, that is if the new trial point is worst or if it does not satisfy some constraints. The above procedure is repeated, as long as the simplex has not collapsed into the centroid, overcoming a tolerance on the simplex size, or as long as the function values in the simplex are

more different between them than a tolerance value or until the algorithm reaches a predetermined maximum number of iterations or function evaluations.

4.7.2.2 *Main features of the algorithm used for the case study*

Initialization of the decisions

The initialization of the decisions has a great influence on the optimization results. Since the convergence of the algorithm is resulted very slow, the farther the initial solution is from the optimal solution, the more difficult is to have a good final solution after the optimization process. So, it's better to start from an initial good solution to generate the initial simplex.
The choice we have made to initialize accurately the decisions is as follows Calling u_0 the vector of initial decisions for all the lakes at time t (i.e. inlet/outlet flows over the optimization horizon), u_0 is assumed to be equal to the last solution, u_{opt}, found by the optimizer at time step (t-1) and receded of one time-step ahead over the receding horizon (repeating twice the decision of the last day). This is a good initial solution since the weather conditions over the receding horizon change smoothly from a time-step to another and consequently the decisions should gradually change from a day to another.
For the very first time step of the MPC simulation horizon, T_i , a good initialization for the decisions can be made taking the decisions of the actual management simulation.

Start and stop criteria

Optimization is run only if the initial value of the cost function is bigger than a tolerance value ($T_{f0}=10^{-4}$). We added this *start condition* to speed up the algorithm.
During the optimization there are different *stop criteria* that are sequentially checked; there are some criteria activated by the default setting of the algorithm, others that we chose to activate.

We have activated the following criteria:

1. The maximum *number of iterations* (activated by default);
2. The maximum *number of function evaluations* (activated by default);
3. The maximum *number of restarts* (activated by default);
4. The absolute and relative *difference in the optimal decisions of two following iterations*;
5. The tolerance on the difference between the *simplex size* at the current and the first iteration;
6. *Impossible improvement* for reaching a minimum value (activated by default);
7. *Box termination option*: the absolute tolerance on *difference of objective function values in the simplex* between the best and worst vertices. If the maximum difference among the vertices is less than a tolerance value defined by the parameter *BTf* for a certain number of consecutive iterations *BNm* the algorithm is stopped. The value of *BTf* is essential for a good compromise between convergence and calculation time. To increase the speed of the algorithm we used the following expression for *BTf* that increases the value of *BTf* respect to the default value (10^{-5}) in function of the initial value of the objective f_0:

$$BTf = min(max(f_0 \cdot 10^{-6}, 0.05), w_{sc}^{Qres} * 10^{-6})$$

where there are two check conditions to verify that the tolerance value is at least equal to 0.05 and at most equal to $w_{sc}^{Qres} * 10^{-6}$ (where w_{sc}^{Qres} has an order of magnitude of 10^9).

Automatic default restarts of NM

The usual method for checking that the global rather than a local minimum has been found is to restart the algorithm from different points, and infer that if they all converge to the same solution then a global optimum has been found. In several dimensions, for a problem for which the feasible region of parameter space is small, the discovery of an alternative permissible initial point can present considerable difficulty. With the Complex method, there is no difficulty in using the same initial point with different pseudo-random number initiators to perform such a rough check as to whether the optimum is indeed global.

4.7.3 Improvements of NM algorithm for the case study

The application of the NM Box algorithm to the case study has encountered some problems due to a slow rate of convergence. The algorithm works minimizing the costs, but since the convergence is slow it is usually terminated by reaching the maximum number of iterations or restarts and so it doesn't provide the optimal solution but a sub-optimal one. Different experiments were performed to see if the solution could improve increasing the maximum number of iterations, but it seemed to us that the improvement was too little respect to the increase in computation time. The termination criteria reached the most of the time was still the maximum number of iterations, even when this parameter of the algorithm was set to a very high number (3000-5000 iterations) leading to high calculation time.
So, to improve the convergence of the algorithm, we developed some strategies that we present hereafter.

4.7.3.1 *Reduction of the number of decisions*

It was already noted that Nelder-Mead manages to explore better low-dimensional than high-dimensional search spaces: the efficiency of the algorithm is jeopardized with problems involving more than 30-40 variables *(see section 4.7.2.1)*. The reason is that the simplex-method could be not efficient dealing with a high-number of possible directions of exploration. So we hypothesized that the poor performance of the algorithm for our case-study could be due to the large number of variables to be optimized. In fact we have 8 variables for each decision-making day of the horizon h; so, for example, taking h=9 days and constant daily control horizons, there are 72 variables. For this reason, we developed two strategies to improve the convergence of the algorithm for the case study, reducing the number of decisions. These strategies consist in enlarging the control horizons, since the number of decisions is inversely proportional to the control horizons lengths.
The first strategy is to use a control horizon grid as explicated in section 4.7.1.3 and choosing this grid so that the number of decisions is limited, for example 40 variables at most. This condition means that the decision-making days should be 4 or 5 at most in the optimization horizon for deterministic MPC (and less for TB-MPC).
The second strategy consists in performing a loop of optimizations reducing iteratively the number of variables. We called this iterative procedure “clever restart”. The iterative reduction of the number of decisions is done by progressively enlarging the control horizons when the first control horizon grid doesn’t provide an optimal solution. When an optimal solution is found using an enlarged control horizon grid, this solution is used to restart the algorithm with the initial finer control horizon grid. In this way a better initial condition for the finer grid is provided and the final optimal solution is improved. The clever restart can be seen as a restart of the NM algorithm changing control horizon grids, starting from the last solution found with the previous control horizon grid readapted to the new one. When the control horizons are enlarged the average of the decisions of the finer control horizon grid is taken, while passing from a larger to a finer grid the same decision is repeated for different decision time-steps. If the algorithm reaches the best solution in one of the middle iterations, and not in the last one with the finest grid, then this solution is saved and taken as the best solution. Thanks to this procedure the results were improved: the termination criteria that indicates the convergence of the algorithm, as the size of the simplex, were encountered more times than before, indicating that the simplex was shrinking towards the global minimum of the function more easily. However, after the implementation of other strategies (in particular, the restart of the algorithm after expert-based initialization, see section 4.7.3.3) the improvement deriving by this procedure became negligible in the end. So, in the current implementation, used for obtaining the results of this thesis, this procedure is not used anymore, but we presented it because it is an original procedure that could be useful again for future applications of this work or for other similar problems.

4.7.3.2 *Normalization of the decision variables*

Since the problem includes decision variables of very different scales (different capacities of the discharge channels), we decided to normalize the variables to be optimized to their domain, to have the same amplitude for all decisions. By operating this normalization, the search space can be better explored by the optimizer in all the directions to the same extent.

The normalization operated for the component k of the decision vector is given by the equation:

$$\overline{u_t^k} = \frac{u^k{}_t - u^k{}_0}{u^k{}_{Max} - u^k{}_{min}}$$

Where:

- $u^k{}_t$ = current decisions vector component k at time t to be normalized;
- $u^k{}_0$ = initialization of the decisions vector component k at time 0, around which $u^k{}_t$ is normalized;
- $u^k{}_{Max}$ = maximum value for the decisions vector component k;
- $u^k{}_{min}$ = minimum value for the decisions vector component k.

In this way the new domain is of unitary dimension for all the decisions built around the initial solution. After normalization, the initial solution components correspond to zero values, while the minimum and maximum values of each decision correspond respectively to a negative value α and a positive value β, being the distance $|\beta - \alpha|$ unitary.

Analyzing the results of the optimization it was proved an improvement of the convergence and efficiency of the algorithm using the normalization option. For a comparison of the results with and without normalization we remind to section 5.2.3

4.7.3.3 *Restart of the algorithm after expert-based initialization of the decisions*

When the optimizer finds a sub-optimal solution, the solution found by the NM algorithm is compared with a solution proposed on the basis of expert-based rules. These rules are based on the conditions downstream the lakes to control flood and drought events and on the storages of the lakes to follow the filling curves in normal conditions. For example:

- **For low and high flows control**: if critical low or high flows are forecasted at the downstream stations, the decisions are calculated in order to release/store water from/into the reservoirs so that the thresholds are respected if possible.
- **For following the filling curves**: in normal conditions (when no critical events are forecasted), the expert-based solution proposes to reach the filling curves at the first day of the optimization horizon, if this is possible. This solution is proposed to overcome the problem that the MPC does not follow exactly the filling curves in the closed-loop simulation, using the penalty cost on the storages at the end of the optimization horizon. This happens because the MPC does not have to reach the filling curves the first day when the optimal decisions are effectively taken but only at the end of the horizon.

The expert-based proposed solution is then combined with the current "optimal" solution proposed by the NM algorithm. The combinations so obtained (in order of 10^2 combinations) are tested and compared among them. The best solution among these combinations is then taken as initial solution to restart the optimization algorithm.

4.8 Use of ensemble forecasts in TB-MPC problem

In this section we present the extensions in the formulation of the MPC problem presented so far in order to use the rainfall ensemble forecasts instead of a deterministic trajectory. What changes between the formulation of MPC and TB-MPC is the use of different trajectories for the disturbances and the structure of the decisions to be optimized, that is more complex in case of the TB-MPC.

4.8.1 Scenario reduction and tree generation from original EFs

It could be possible to generate the trees using the ensembles of rainfall forecasts or inflows forecasts, that can be obtained running the hydrological model using as inputs the scenarios of the ensembles for the rainfall. At this point the question arises whether it would be better to perform the tree generation on the rainfall or inflows predictions. Our answer is that it is better to generate the trees starting from the rainfall predictions, since the hydrological model would attenuate the differences among the scenarios at the first time-steps of the prediction horizon because of the delay of the rainfall-runoff process and the dependence on the same initial condition.
The scenario reduction algorithm reduces the number of scenarios in the ensemble to a predefined number. Stive [2011] provides useful information about the right amount of reduced scenarios, showing that using more than 8 scenarios does not yield any better performance, but only increases calculation time. On this basis, we chose a number of scenarios of 6, to reduce calculation time for our optimization problem, in which a big number of decisions is involved already using only a deterministic scenario. The marginal information loss of the scenario reduction from 8 to 6 was judged negligible.

An example of reduced ensemble obtained from the original one is reported in the following figure.

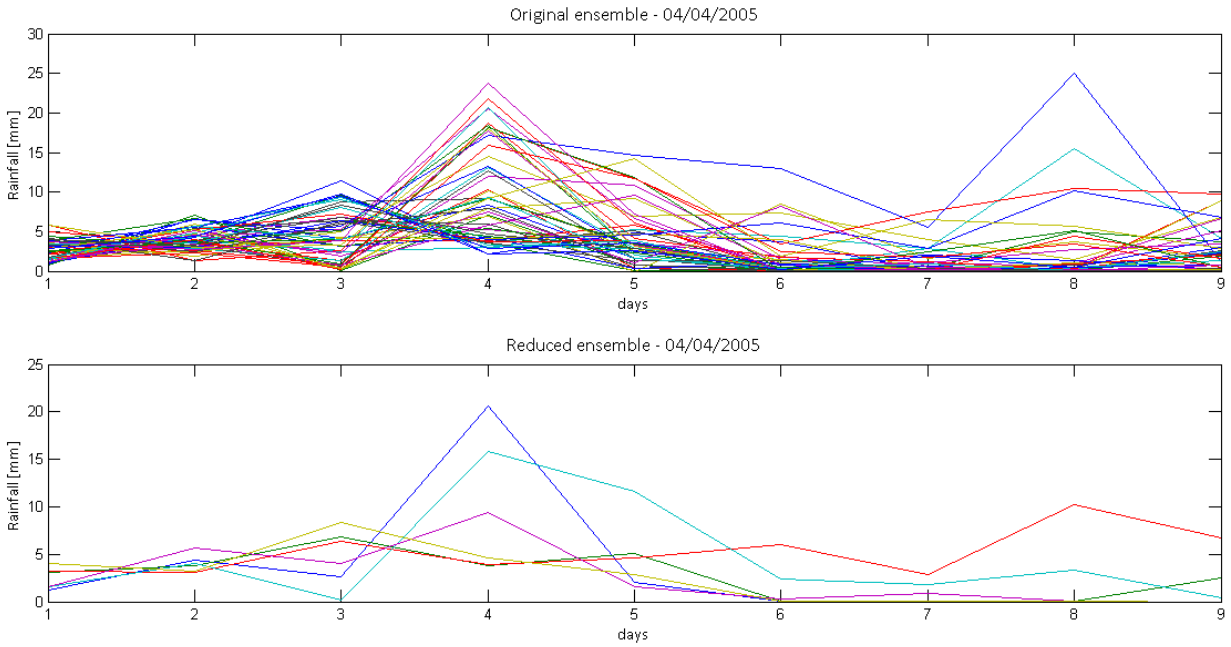


Figure 24. Example of reduced ensemble of 6 scenarios generated starting from the original ensemble of 50 scenarios for the day 04/4/2005.

In order to generate the tree structure starting from the reduced ensemble we must fix a value for σ, i.e. the standard deviation of the observational uncertainty. The parameter σ is used to determine whether two ensemble members are really distinguishable or their difference could be due to the uncertainty in the observational process. For the physical meaning of this parameter, a first idea tested was to take it equal to the standard deviation of the ensemble at the first day of forecasts. However, this solution showed some problems, for example when the variance of the ensemble at the first day was too little (no precipitation condition). So, we decided to take another value for σ taking into account the correlation between variance and average of rainfall forecasts and the variability of the forecasts along the horizon. The variance of the rainfall forecasts is correlated with the average value of forecasted rainfall and in particular it increases as the rainfall increases, as it is shown in the following graph in Figure 25.

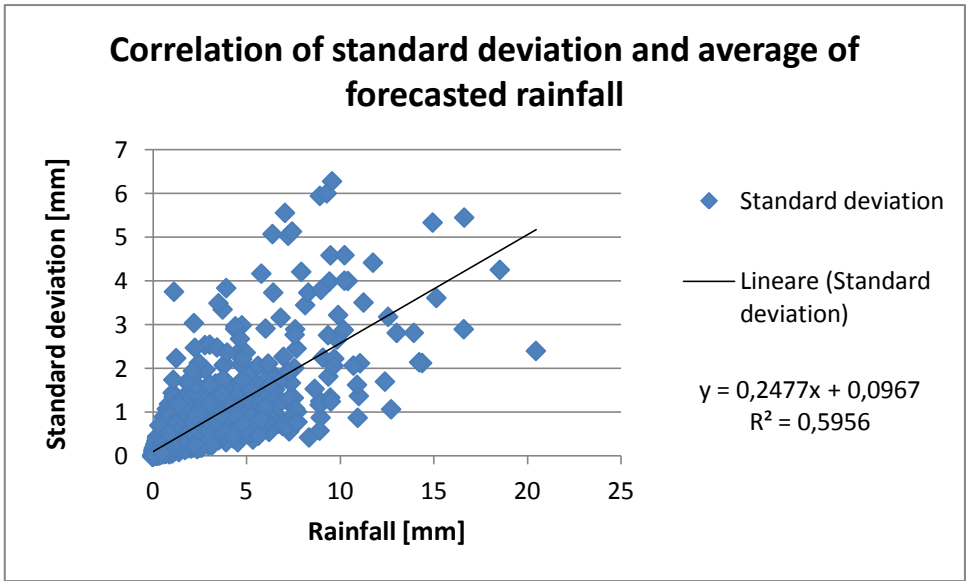


Figure 25. Correlation between observational uncertainty [mm] and average forecasted rainfall [mm] at the first day of the forecasting horizon.

The linear interpolation function shown in the figure above is used to calculate σ, at each time step of application of TB-MPC, in function of the average of the rainfall forecasts values all over the prediction horizon. This formulation provides also a minimum threshold for σ (equal to 0,0967) that is useful in order to avoid taking a more little value for this parameter. A too little value, close to zero, would lead to determine branching points even for branches with negligible difference, that instead should be bundled together.
A second parameter for the tree generation that must be fixed is the confidence level p^*, for which we took a value of 0,95.
The tree generated from the reduced ensemble shown in Figure 24 is reported in the following figure.

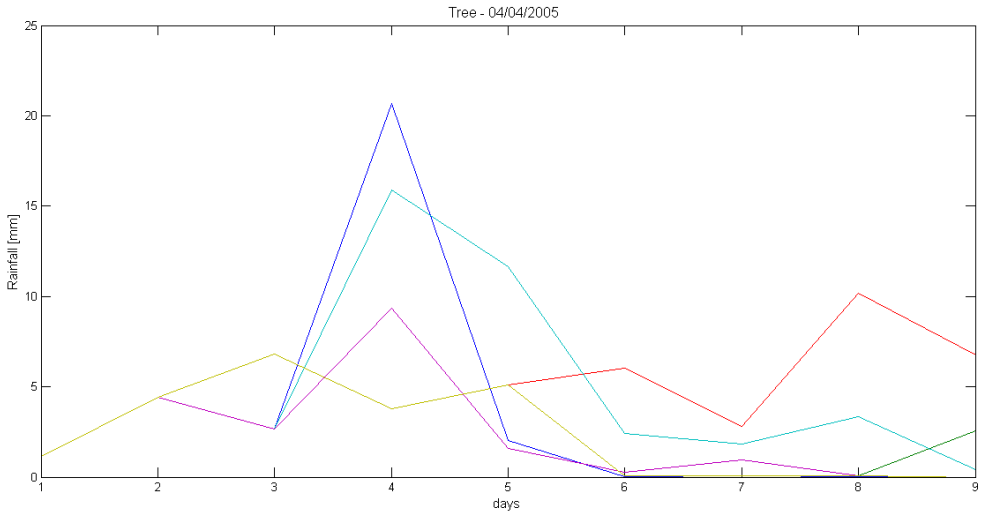


Figure 26. Example of tree of forecasts generated starting from the original ensemble for the day 04/4/2005.

It must be noticed that this tree of rainfall forecasts is not used directly in the TB-MPC problem, as the rainfall forecasts trajectories used are those of the reduced ensemble. The output of the tree structure generation that will be directly used in the TB-MPC problem is the scenario tree nodal partition matrix M,

that will say which branches of the tree of controls are associated to each member of the reduced ensemble (see section 3.2.3). The controls tree is assumed to have the same structure than the forecasts tree, as for example the one in the figure above.

4.8.2 Branching points adaptation to control horizon grids

The branching points definition must take into account the period of observation as discussed in Raso et al. [2012]. The decisions can be taken only with a time period multiple of the observation period, because we can re-solve the problem only when the new observations are available. In our case study, we have a time resolution of one day for the observations, so we can have a branching point at each day of the forecasting horizon. However, we structured the decisions along the prediction horizon using a control horizon grid, made up of variable control periods (see section 4.7.1.3). For this reason the branching points, that correspond to the instants where decisions can be changed must coincide with the time-steps where a decision can be effectively calculated by our optimizer, according to the control horizon grids. This is an original aspect emerged in the application of TB-MPC for our case-study, that was never treated before. There are different solutions that could be taken to solve this question. Following the idea in Raso et al. [2012] used for adapting the branching points to the observation period, we constructed the tree using all the forecasts with the initial fine resolution (constant time step of 1 day) and then we moved the branching points to the next closest time-step of the prediction horizon that belongs to the control horizon grid.

4.8.3 Hydrological model use

As we have already remarked in section 4.1, the hydrological model is not affected by the decisions operated on the four dams and so we can generate all the inflow forecast scenarios off-line (before starting the optimization process) for all the time-steps of the period of application of TB-MPC. For TB-MPC, we give in input to the hydrological model the members of the reduced ensemble of rainfall forecasts to obtain the correspondent inflows forecasts. This information flow is shown in the following flow chart in Figure 27. The rainfall and the inflows forecasts (P_f and I_f) can be represented by two hyper-matrixes that have as dimensions: the prediction horizon (h=9 days), the number of intermediate basins (N_{BVI}=32) and the number of members of the reduced ensemble (n=6). The initial conditions used by the hydrological model at each time-step t, s(t) and r(t), are calculated by calling the same model with past observations of rainfall and ETP (calculated by Penman equation using weather observations) available at time t.

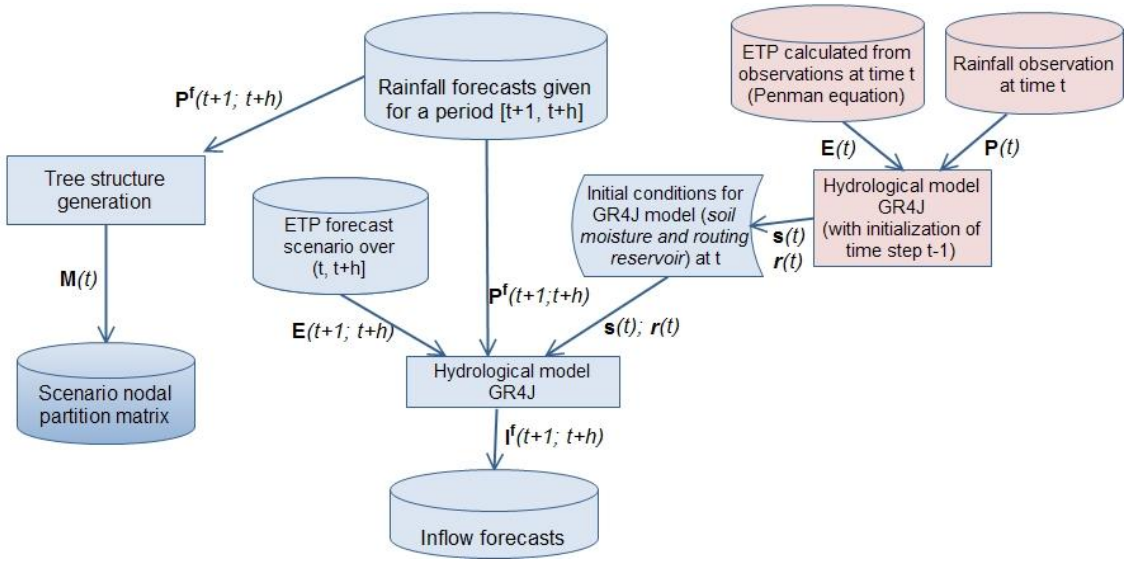


Figure 27. Scheme of the use of the rainfall forecasts ensemble in the control problem at time t.

4.8.4 Tree-Based optimal control problem

In the TB-MPC formulation the same cost-function than in the deterministic one is used in a loop to calculate the weighted sum of the costs for all the rainfall trajectories of the reduced ensemble. The tree structure defined in the scenario nodal partition matrix M is used to structure the controls tree as explicated in section 3.2.3. For the different ensemble members, different trajectories of the decisions are calculated after the branching points. The number of decisions to calculate is so influenced by:

- the number of scenarios of the reduced ensemble;
- the number of distinct branches, equal to max(M);
- the control horizon grid.

4.9 Code architecture

4.9.1 Notes on TGR model and current rules simulation code

At the beginning of this work the code of the TGR model was already available and that for the simulation of the current rules was being developed for the work presented in Dorchies et al. [2013]. This package of code was developed in *Scilab* for the most part and in *C* for some functions.
Scilab is a free and open source software for numerical computation providing a high-level, numerically oriented programming language. It is an interpreted programming environment, operating with matrices as the main data type using a syntax similar to Matlab. Scilab is an open-source alternative to Matlab released under the CeCILL license [www.scilab.org]. The C functions are interfaced with the Scilab platform thanks to a C compiler for Scilab (Mingw Compiler). The user stays in the environment of Scilab without using the C language. It's only necessary to define the interface for Scilab of each C function in C code. The reason of using C language for some functions is that it allows a drastic diminution of computation time, especially when using variables that require a large memory, storing long time-series. The Scilab version used is Scilab 5.3.3. The Mingw compiler version is Gcc 4.5.0-32.

The implementation of the TGR model (GR+LR) is organized in the following main scripts and functions:

- Loader.sce: loading environment for TGR model
- ReadData.sce*:* loading the data necessary for the simulation of the system: weather data, measures of river flows, all the parameters of the system (constraints);
- TGRM.sci, TGR.sce, implementing the call of the hydrologic and hydraulic models for the case study;
- c_GR4J.c, c_GRK.c, implementing the hydrologic models GR4J/GRK;
- c_LR.c, implementing the hydraulic model LR.

The implementation of the simulation of actual rules is organized in the following different scripts and functions:

- c_barrageP.c that calculates the inlet connections flows with the actual management rules and updates also the volume of the reservoir;
- c_barrageR.c that calculates the outlet connections flows;
- c_barragePR.c that calculates inflow and outflow of Pannecière lake

4.9.2 Code for MPC and TB-MPC management simulation

4.9.2.1 *Organization of the code*

The main interface for MPC functioning is represented by the following scripts:

- *Loader_user.sce*: loading personal setting for the simulation to be performed (possibility of choosing the period of application of MPC, the scenario and changing the MPC setting);
- *Config_MPC.sce*: definition of the setting for MPC;
- *Simulation.sce*: script to be called to run MPC management simulation;
- *Benchmark_costfunction.sce*: run some simulations with different values of some parameters for sensitivity analysis.

The implementation of the MPC simulation is organized in the following scripts and functions:

- *CalcQGR4J4MPC.sci*: calculates the inflows by calling the GR4J model (c_GR4J); to use GR4J model for MPC it's necessary to use a *warming up period to realign some hidden states of the model*; it was found that one year of warming up period it's sufficient (with a residual error in the inflows less than 10^{-4}); the warming up period is removed in this function;
- *c_LR_MPC.c:* it's the hydraulic model modified to be used in the MPC method; the modifications respect to c_LR were necessary to use the model with a receding horizon; the interface of this C function with Scilab is defined in c_intTGRM.c;
- *reservoir_model.sci*: performs the simulation of the four reservoirs models;
- *sim_MPC.sci*: performs the simulation of the whole system calling the hydraulic model and the reservoirs model;
- base_MPC: calculation of inflows series off-line by calling the GR4J model (with observed rainfall); definition of optimization setting; initialization of the MPC environment (volumes of the lakes, river flows and decisions);
- *Deterministic_MPC*.sci: performs the deterministic MPC optimization and simulation for the whole period of application;
- *NM_optimization.sci*: defines optimization setting for NM method and performs optimization;
- *costfunction.sci*: calculates the global cost-function value, calling all the sub-objective functions;
- *stepcosts.sci*: includes the different sub-objectives functions: stepcosts, softconstraints and penalty;
- *reinitialize_u0*.sci:

The implementation of the TB-MPC simulation is based on the code architecture of the MPC simulation using some additional functions and scripts:

- *AdaptTreesToCsPatterns.sci*: this function adapts the tree structure to the control horizon grid, so that all the branching points fall at a decision step; if a branching point is not in a decision step, it's shifted to the following decision-step;
- *TreeBased_MPC.sci*: can be used to run the optimization and simulation of the Tree-Based-MPC management; it is called by the script "Simulation.sce";
- *TB_costfunction.sci:* calculates the weighted sum of the cost-values relative to all the members of a tree of forecasted inflows; for each member of the tree the cost-function is called (*costfunction.sci*, the same used by deterministic MPC).

Moreover for the tree generation a Matlab toolbox (implemented by L. Raso, 2012) is used to generate the trees starting from the ensemble forecast. This toolbox implements the scenario reduction and the "information based" method for the tree structure generation [Raso et al., 2012].

4.9.2.2 *Functioning of the MPC management simulation*

The user who wants to run a simulation of MPC management must:

1. execute in Scilab the script builder.sce that compiles the C functions (only before the first run);
2. define the desired options in Loader_user.sce: set the value of the flag variable "UseMPC" for choosing MPC (UseMPC=1) or TB-MPC (UseMPC=2) method, define the simulation horizon and other important options, as the use of a perfect predictor or real forecasts for MPC;
3. execute Loader.sce for the loading of all the data and settings (including the ones in Loader_user.sce);
4. run the simulation executing the script Simulation.sce.

In the comments of our code is possible to find more details about the functioning of the simulation.

5. Analysis of the results

In this chapter the results will be presented and analyzed. It is possible to evaluate the performances of the Real-Time Controllers implemented using the different climate forecasting models introduced in section 4.1 (perfect predictor, deterministic and ensemble forecasts by ECMWF). The perfect predictor is taken as the baseline case used for running simulation experiments for sensitivity analysis on some parameters of the control problem: the length of the receding horizon h, the control horizon grids and the normalization of the decisions variables. The optimal setting found with this analysis is then used in the further analysis. Several simulation experiments were performed over past scenarios using the historical realizations of precipitations on the watershed. Historical flood and drought events were chosen to test the ability of the MPC method in tackling them. Firstly, the results of MPC with perfect predictor will be compared with the results of current management to evaluate the maximum improvement expected by integrating the use of real-time information (forecasts) in the control. Then the results of simulations using real deterministic and ensemble forecasts will be presented and the performances of MPC and TB-MPC will be compared. The results will be analyzed by calculation of the objective functions values in closed loop simulations. Moreover the results will be graphically presented, by plotting the volumes of the lakes in time and the hydrographs of the river flows at some key control stations downstream the lakes.

5.1 Performance indicators in closed loop simulation

The performance of the real-time controllers can be evaluated analyzing the results of the simulations in relation to the objectives set for the case study. The performance can be assumed to be the value of the objective function measured in *closed loop simulations*. The closed loop simulation is the simulation that results from the application of the receding horizon strategy over the period of simulation. In other words at each time-step only the first control of the sequence of optimal controls is applied and then the same problem is repeated after the evolution of the system under the realized disturbances (observed rainfall). In order to compare the performances of different controllers, it's necessary to calculate the objective function values over a certain time-span, called *simulation horizon* (T_s). Over this horizon the step-costs for floods or low flows (g_t^{HF} or g_t^{LF}) are calculated and aggregated (with the sum), to obtain the indicators of the performance of the controlled system (I_{floods} and $I_{low\ flows}$):

$$I_{floods} = \sum_{t=1}^{T_s} g_t^{HF,tot}(\boldsymbol{u}_t) \quad [-]$$

$$I_{low\ flows} = \sum_{t=1}^{T_s} g_t^{LF,tot}(\boldsymbol{u}_t) \quad [-]$$

where the step-costs at time t, $g_t^{HF,tot}$ and $g_t^{LF,tot}$, are the weighted sum of the step-costs at all the control stations and depend on all the decisions for time-step t (vector $\boldsymbol{u}_t$). It must be remembered here that the step-costs are expressed in *dimensionless* units, because they have been normalized (see section 4.6.1).
It is important to remark that even if the cost functions that we built are not represented in a certain monetary or physical unit, this is not a problem for our performance analysis. The costs are strongly correlated with the objectives set by the stakeholders, being a linear function of the flow rates over (and

under) the critical thresholds at the control stations. So, they are good indicators for the analysis, even if they don't reflect a realistic order of magnitude. Moreover for comparisons analysis it is sufficient to dispose of relative values.

5.2 Sensitivity analysis

In this section we present the sensitivity analysis on some key parameters of the control problem: the length of the receding horizon h, the control horizon grids and the normalization of the decision variables. This analysis has been carried out by using a perfect predictor as forecasting model. The optimal setting found with the sensitivity analysis is then used in the further simulation experiments also using the other forecasting models. It must be noticed that the number of parameters of the MPC problem is very high and so it's not possible to present here a sensitivity analysis on all the parameters. However, to define the standard values of the parameters used in the simulation experiments presented, we tested the sensitivity of the results on other parameters, including the parameters used to define the objective function and balance the sub-objectives.

5.2.1 Sensitivity to the optimization horizon (h)

For MPC with perfect predictor (using observations as forecasts) we can choose the length of the optimization horizon h as long as we want, since we have not the limit of the real prediction ability. So we can test different lengths of the receding horizon h considered in the on line optimization problem. It's natural that we expect that the performance improves enlarging the prediction horizon since we include a longer time-series of rainfall forecasts. However it must be remembered that the forecasts available have a prediction horizon of 9 days, so this is the maximum length that we could use for the MPC and TB-MPC using real forecasts. So the question that arises is to what extent it could be useful to extend this horizon until h=9 days.
It must be noticed that it would be possible to use climate statistics beyond the 9 days of forecast but we chose to fit the maximum horizon length to the forecasts length for ease and for the necessity of keeping low the number of decisions to optimize. In fact the benefit of enlarging the horizon length h is reduced by the limitation on the number of decision variables. We have seen that for the optimization problems (sub-optimality) it is better to take some accurate decisions than a lot of rough decisions. This limitation can be formalized as a limitation on the control horizon grid, keeping low the number of decision time-steps (for example 2 or 3 decision time-steps). This restricts the variability of the control actions in the optimization horizon and so the usefulness of a long optimization horizon.
As we have noticed in the remarks in section 4.7.1, as for the ideal optimization horizon, from the standpoint of flood control it should be in the range of some days, while for low flows it should be much longer. For low flows control we would need to have seasonal weather forecasts and demand forecasts, that we don't have. Alternatively, for the time beyond the forecasts horizon, we could use climate statistics and a longer horizon (which is difficult because of the optimization problems). However, it must be remarked that we already used the climatology statistics with the new filling curves calculated for MPC for balancing low-flows step-costs and penalty on the final storages. For these reasons we focus this sensitivity analysis only in the point of view of flood control and we will use the same horizon in low-flows conditions.
In the following figure we report the results of the closed loop simulation using different optimization horizon lengths in terms of cumulated costs over the simulation horizon 1/12/1981 - 1/3/1982 (including a flood event). The control horizon grid is taken fixed to [1 2 3] when the horizon length is larger than 3 days and to [1 2] when the horizon length is equal to 3 days (for a bug in the code that prevents the choice of a decision step at the last simulation step of the optimization horizon).

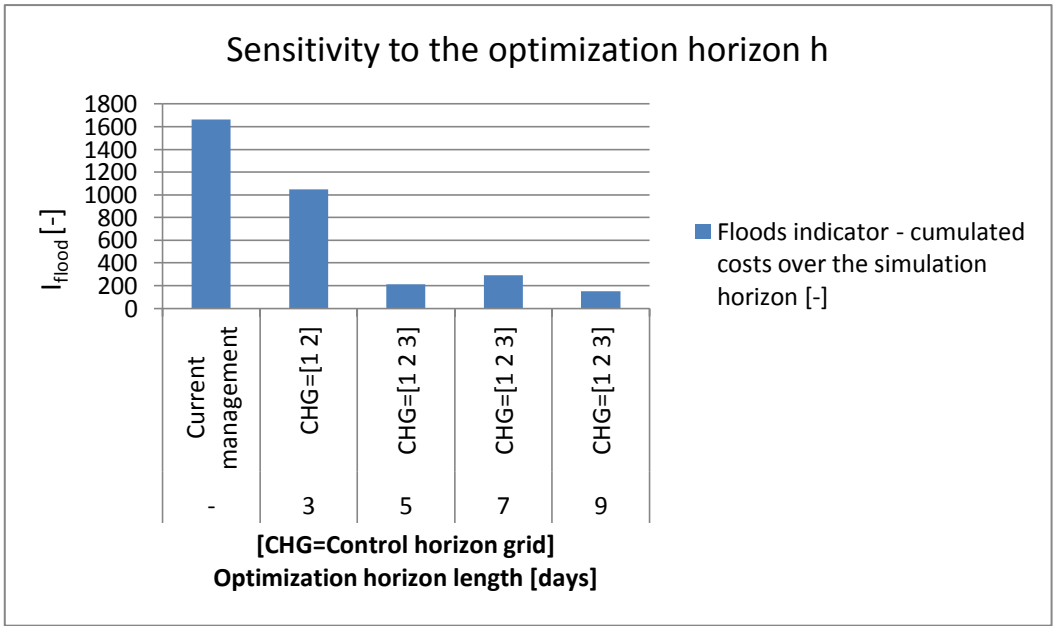


Figure 28. Cumulated flooding step-costs over the simulation horizon 1/12/1981 - 1/3/1982 (including a flood event) with current management and MPC using different lengths of the receding horizon h.

As shown in the histogram above, the performance indicator for high-flows (I_{floods}) confirms that the performance improves with the horizon's length. So, we chose a value of 9 days for the optimization horizon.

5.2.2 Sensitivity to the control horizon grid for high-flows

In section 4.7.1.3 we have introduced the idea of using a control horizon grid with different lengths of the control horizons that increase along the optimization horizon. We decided to use two different control grids for high and low flows conditions because a finer control horizon is more useful for floods control, for the high-variance of inflows in periods of floods and the consequent need of a quick variation of the control, while this accuracy is not necessary for normal conditions or low-flows.
In Figure 29 we report some results of the simulation experiments with different control horizon grids for high flows, over the simulation horizon 1/12/1981 - 1/3/1982 including a flood event. The length of the optimization horizon is taken equal to 9 days as fixed after the sensitivity analysis presented in the previous section. For the notation of the control horizon grids (numbers series in square brackets) we remind to section 4.7.1.3.

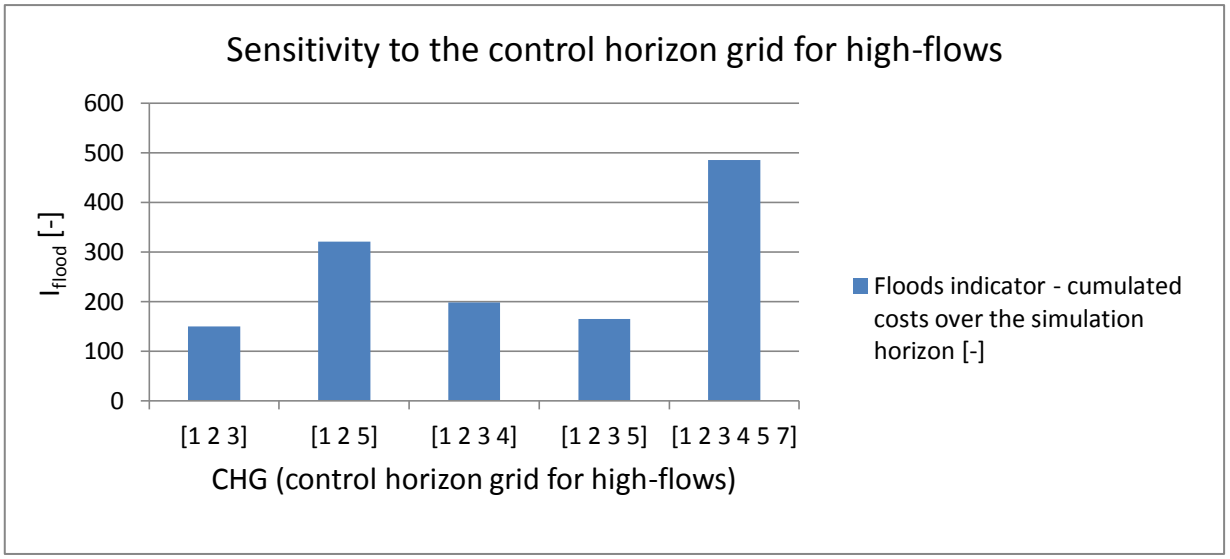


Figure 29. Cumulated floods step-costs over the simulation horizon 1/12/1981 - 1/3/1982 (including a flood event) with current management and MPC using different control horizon grids for high-flows.

In the histogram presented above it can be noticed that the two best control horizon grids for high-flows are [1 2 3] and [1 2 3 5]. So, we chose to take the control horizon grid equal to [1 2 3 5], even if slightly worse in performance than [1 2 3], because it allows to take more into account the variability of the predictions in TB-MPC along the forecast horizon and to include the branching points in the control grid. This would be difficult with the grid [1 2 3] since the ensemble members typically diverge more after the third day of the forecasting horizon. For deterministic MPC, it could be better to take the grid [1 2 3], but for consistence in the comparison of the results between MPC and TB-MPC, we took the same control horizon grid [1 2 3 5] for the both methods.
In case of normal or low-flow conditions the control horizon grid has been fixed to [1 2] after performing analogous sensitivity analysis experiments.

5.2.3 Sensitivity to the normalization option

In section 4.7.3 we have introduced the normalization of the decision variables, option that we developed to improve the performance of the optimization algorithm for the case study. In the following figure the results of the MPC optimization with and without normalization are reported, using the same simulation horizon of the flood event of January 1982 than the previous analysis. The configuration of the algorithm is the same in the two cases, changing only the normalization option (with or without). The results confirm an improvement by using the normalization option.

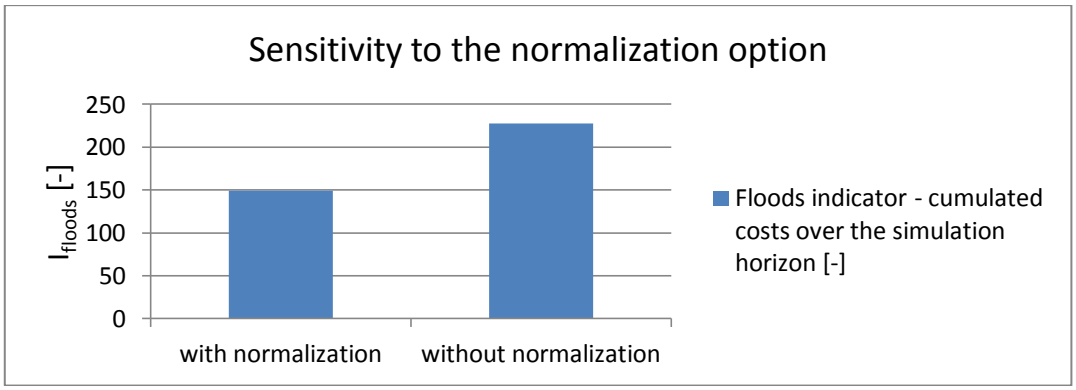


Figure 30. Cumulated floods step-costs over the simulation horizon 1/12/1981 - 1/3/1982 (including a flood event) with current management and MPC with and without normalization of the decision variables.

5.2.4 Standard configuration of MPC

The simulation results presented in this chapter are produced using a standard set of parameters, called standard configuration. This standard configuration has been chosen after performing several simulation experiments for sensitivity analysis, that we presented only for some parameters. The values of all the parameters are reported in the following tables.

Parameter	***Standard value***	***Unit***
Optimization horizon (h)	9	days
Control horizon grid for high-flows	[1 2 3]	-
Control horizon grid for low-flows	[1 2]	-
Normalization of decision variables in the optimization algorithm	True	-

Table 2. Standard set of general parameters of the MPC problem chosen by sensitivity analysis and used in the simulation experiments presented in this thesis.

Parameters for the sub-objectives definition and weighting (see par. 4.4, 4.5 and 4.6)		
Parameter	***Standard value***	***Unit***
p	0,9	-
$w_{step-costs(lf/hf)}$	1	-
$w_{penalty\,(\frac{over}{under}FC)}$	1	-
w_{sc_V}	10^{15}	-
w_{sc_Qres}	10^{9}	-
w_{bypass}	10	-
$w_{lf\ T2-All stations}$	10	-
$l_{hf\ 1}$	40	days
$l_{hf\ 2}$	4	days
$l_{hf\ 3}$	2	days
$\varepsilon_{hf\ 1}$	0	-
$\varepsilon_{hf\ 2}$	0,75	-
$\varepsilon_{hf\ 3}$	1,5	-
ε_{Qref}	0,1	-

Table 3. Standard set of parameters for the objectives definition and weighting, used in the simulation experiments presented in this thesis (see Chapter 4 for the meaning of the parameters).

5.3 Deterministic MPC with perfect predictor

The “perfect predictor", as explained in section 4.1, uses climate observations as forecasts, as if we were able to have weather forecasts coinciding with the realized values with zero variance. This is a not realistic hypothesis, but it's useful because it allows to obtain an upper bound of the performance that one could obtain by using the real-time controller. In practical terms the experiment with a perfect predictor could give us an idea of the value of the uncertainty of the forecasts. In fact by comparison of these results with the ones using real forecasts it’s possible to understand which improvement we could expect by reducing the errors of the forecasting models. In this sense, the results could be used to investigate the interest to invest research resources to improve forecasting models, instead of using these resources in other directions.

In order to run simulations over past scenarios it must remembered that from 1958 to 1988 the lakes were not already all operating. Thus for the application of MPC it has been chosen to run all the simulations over the so-called "NTP" scenario that emulates the presence of all the lakes from the beginning of the study period (1958). This assumption in fact allows us to test the centralized MPC for all the lakes following the filling curves (as formalized by the penalty) with an easy implementation: the FCs are assumed to be the same for each year, even for the years when lakes were not already in function.

5.3.1 Performance on flood events

5.3.1.1 *Flood event of 1982*

In order to analyze the performance of the controller on flood events, the period December 1981 - March 1982 was chosen for running different simulations, since in January 1982 there's the biggest flood of the period of data availability. In this event the naturalized flow (the river flow obtained removing the influence of the dams from the observed flows) is over the third high-flow threshold at one control station (Nogent-sur-Seine) and over the second at almost all the others. We can see this situation in the following figure reporting the naturalized flow at Nogent-sur-Seine (most critical station) over this period.

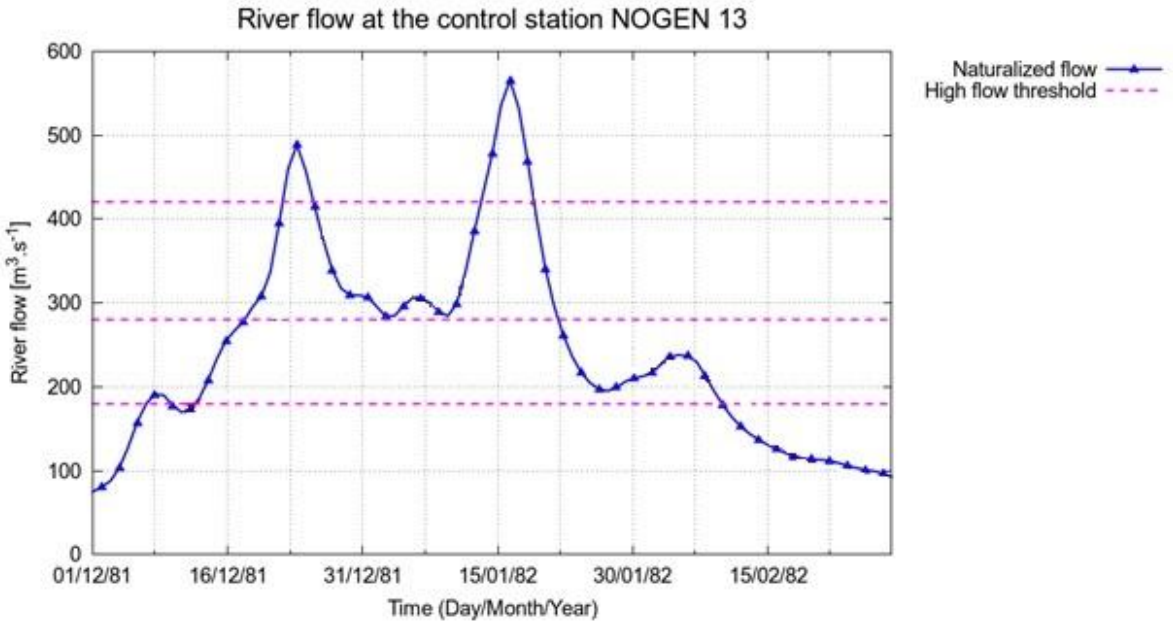


Figure 31. Simulated naturalized river flow [m3/s] at Nogent-sur-Seine from 1/12/1981 to 1/3/1982 (blue line). The 3 horizontal lines (dashed pink lines) correspond to the flooding thresholds.

The current rules manage to contain the flood at Nogent-sur-Seine just at the level of the third threshold (as we will show in Figure 33), that is a very critical level that should be avoided. So the question that arises is whether MPC can improve with respect to this performance.

5.3.1.2 *Simulated river flows*

The results of the simulation experiments with the MPC management using the perfect predictor on the flood event in January 1982 show a good performance in the control of this flood. Analyzing the simulated river flows at the control stations it can be noticed that the MPC management succeeds in its objective to contain the flood below the second threshold at almost all the control stations. For example, in Figure 32 the hydrographs at Paris resulting by MPC and current management simulation are shown. It can be noticed that when the latter one (curve in blue line with triangles) overcomes the second threshold it means that there was a flooding at the downstream station of Paris with the actual management. In

particular the river flow simulated with the current rules overpasses for two days (12-13th of January, 1982) the second flood threshold $\overline{q^J_{hf\,2}}$ equal to 1600 m^3/s and for 4 days the 1st threshold of security margin $\overline{q^J_{hf\,1}}$. Instead with the MPC management simulation (red line with squares) this flooding could be avoided, containing the peak just below the second threshold and overcoming the first threshold only for one day. The reader should remember here that the first threshold $\overline{q^J_{hf\,1}}$ is the level where the step-costs begin to increase starting from zero.

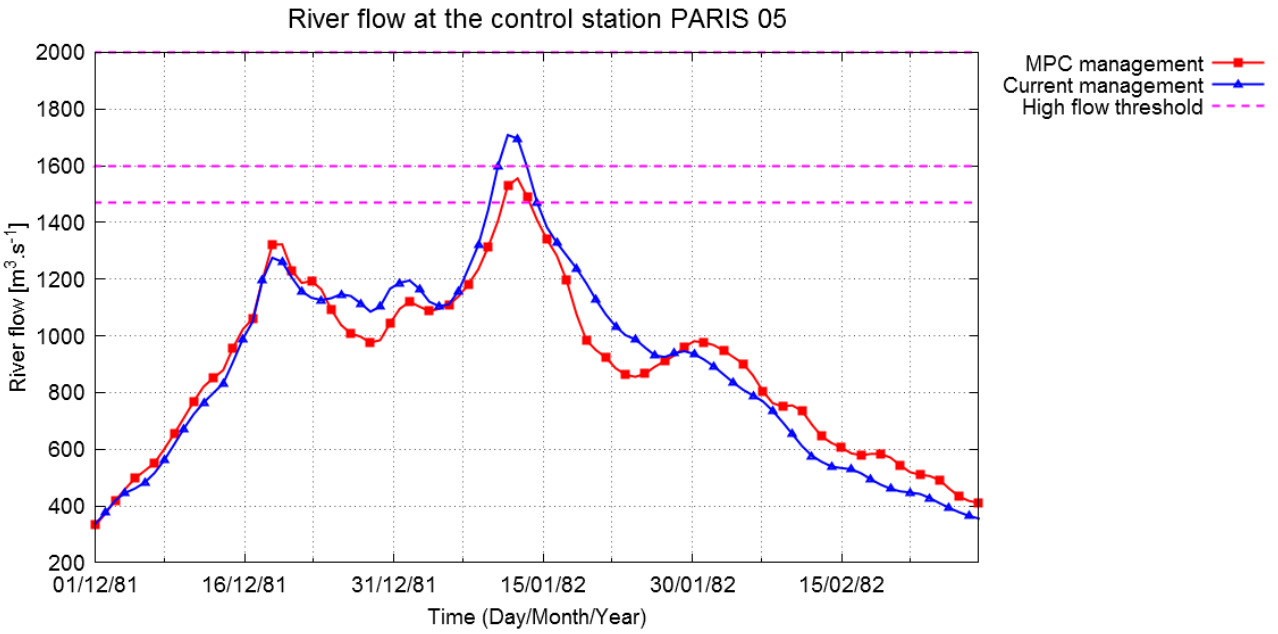


Figure 32. Simulated river flow [m3/s] at Paris from 1/12/1981 to 1/3/1982 with MPC (red line) and current rules (blue line).

When no flood is forecasted at any control station, the hydrograph at Paris follows the actual management simulated flow, as it happens at the beginning of the simulation horizon analyzed (first half of December 1981). Since the station of Paris is the extreme downstream of the system, if there are some flood problems at the upstream stations we see that the MPC hydrograph at Paris deviates from the current rules one. This is because the MPC is facing the problems upstream, deriving water from the river into the reservoirs. For example, this is what happens in the second half of December 1981 when the MPC is forced to derive water into the reservoirs for keeping the flow at Nogent-sur-Seine under the second threshold from the second half of December, as we can see in the following hydrograph. For this reason, also at the stations downstream Nogent-sur-Seine the river flow simulated by MPC in the second half of December is reduced respect to the current rules one.

Even at Nogent-sur-Seine, the most critical control station, the MPC management outperforms the current management. The hydrograph simulated by using MPC at Nogent overpasses the thresholds with a lower peak over the second threshold and for less time (4 days with MPC against one month with actual rules) as it can be seen in the following figure.

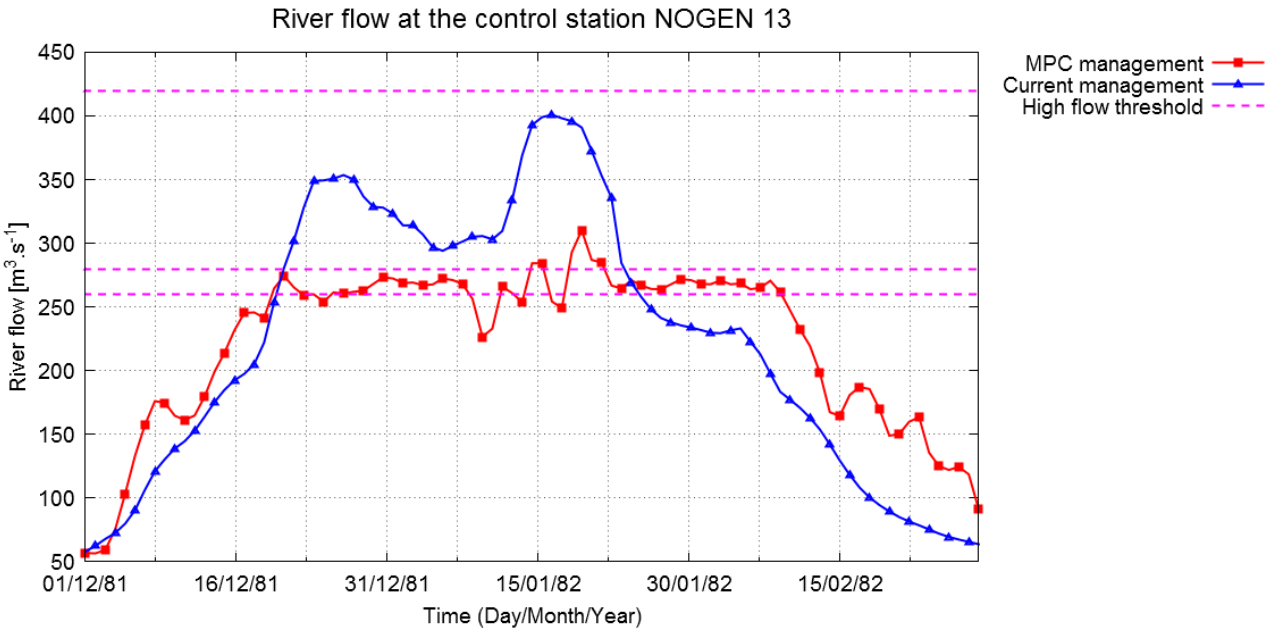


Figure 33. Simulated river flow [m^3/s] at Nogent-sur-Seine from 1/12/1981 to 1/3/1982 with MPC (red line) and current rules (blue line).

The same kind of improvement on the flow peaks simulated with the MPC respect to the current management can be noticed at all the control stations.

5.3.1.3 *Reservoirs volumes*

The peak reduction in the hydrographs is obtained by filling the four reservoirs more than the objective volume fixed by the FC when the flood event is forecasted, as we can clearly see in the following plot of the volume of the lake Marne over the simulation horizon.

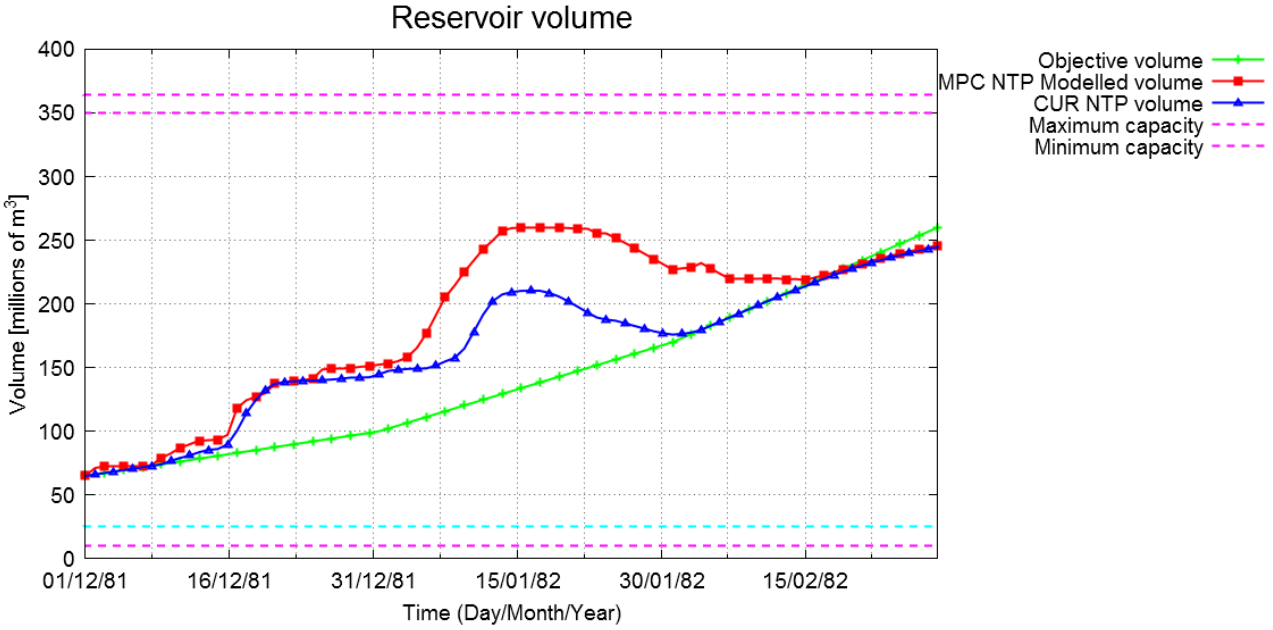


Figure 34. Volume [Mm3] of the Marne lake from 1/12/1981 to 1/3/1982 with MPC controller (red line) and current rules (blue line). The green line corresponds to the filling curve.

In the figure above it can be noticed that the volume of the lake Marne simulated with the MPC controller follows the current management at the beginning of the simulation horizon, deviating from the FC in the second half of December. Then the deviation of the MPC volume from the FCs increases at the beginning of January, when the flood (with the peak on the 12th-13th of January) begins to be included in the prediction horizon. This is a good example of the pro-active behavior of MPC: the MPC begins to fill the lake Marne about 9 days before the current management, as the length of the prediction horizon begins to include the forecasting of the flood.
The fact that the controller has to work to limit the flood damages at Nogent for a long period in winter makes the upstream lakes (Aube and Seine) be completely filled soon, consuming their capacity to laminate any eventual flood that occur later in time at the other downstream stations, as it can be seen in the following plots of the volume of the Aube lake. The evolution of the volume of the Seine lake is analogous.

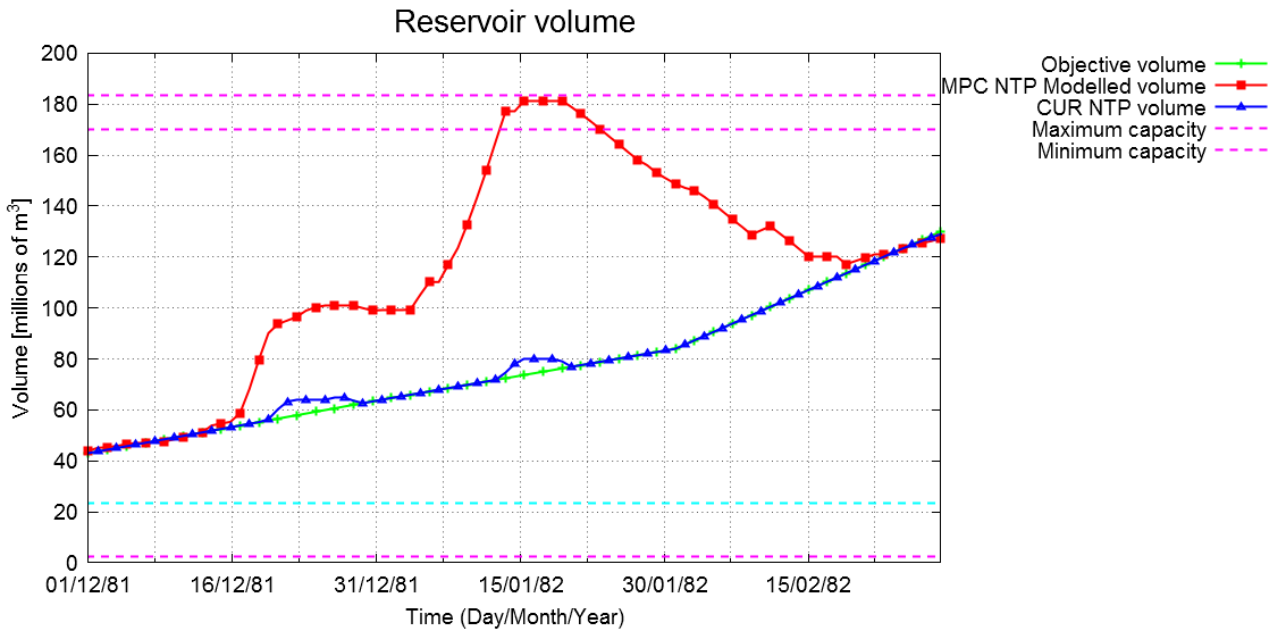


Figure 35. Volume [Mm3] of the Aube lake from 1/12/1981 to 1/3/1982 with MPC controller (red line) and current rules (blue line). The upper horizontal line correspond to the maximum exceptional capacity.

It must be remarked also that the final storages of the reservoirs are very close to the filling curves at the end of the simulation horizon, as it is shown in the figure above of the volume of the Aube lake and in Figure 34 for the Marne lake. The same behavior can be found for the other reservoirs. This means that the MPC management acts in a sustainable way, maintaining the capacity of the system to perform well in the future. This result is assured by the minimization of the penalty function on the final storage.

5.3.1.4 *Costs evolution*

It's interesting to analyze also the evolution of the optimal costs over the simulation horizon, as shown in the following figure for the flood event of 1982. This analysis allows us to understand why the MPC controller behaves in a certain way. It can be seen that when the problems on high flows (and reference flows) begin, in mid-December (time-step=15), the penalty cost to be over the filling curve begins to increase. This is because the controller fills the reservoirs paying the penalty cost on the storage to reduce the step-costs for high flows. These costs remain constant until the flood of the first half of January (time-steps=30-45), when the lakes are filled more and more to amortize the high flows costs growth. The two curves (high flows cost and cost to be over the FC) increase harmoniously, from time-step=35, until the limitations of the maximum capacities of the lakes are encountered. At this moment the cost for high-flows can no longer be contained.

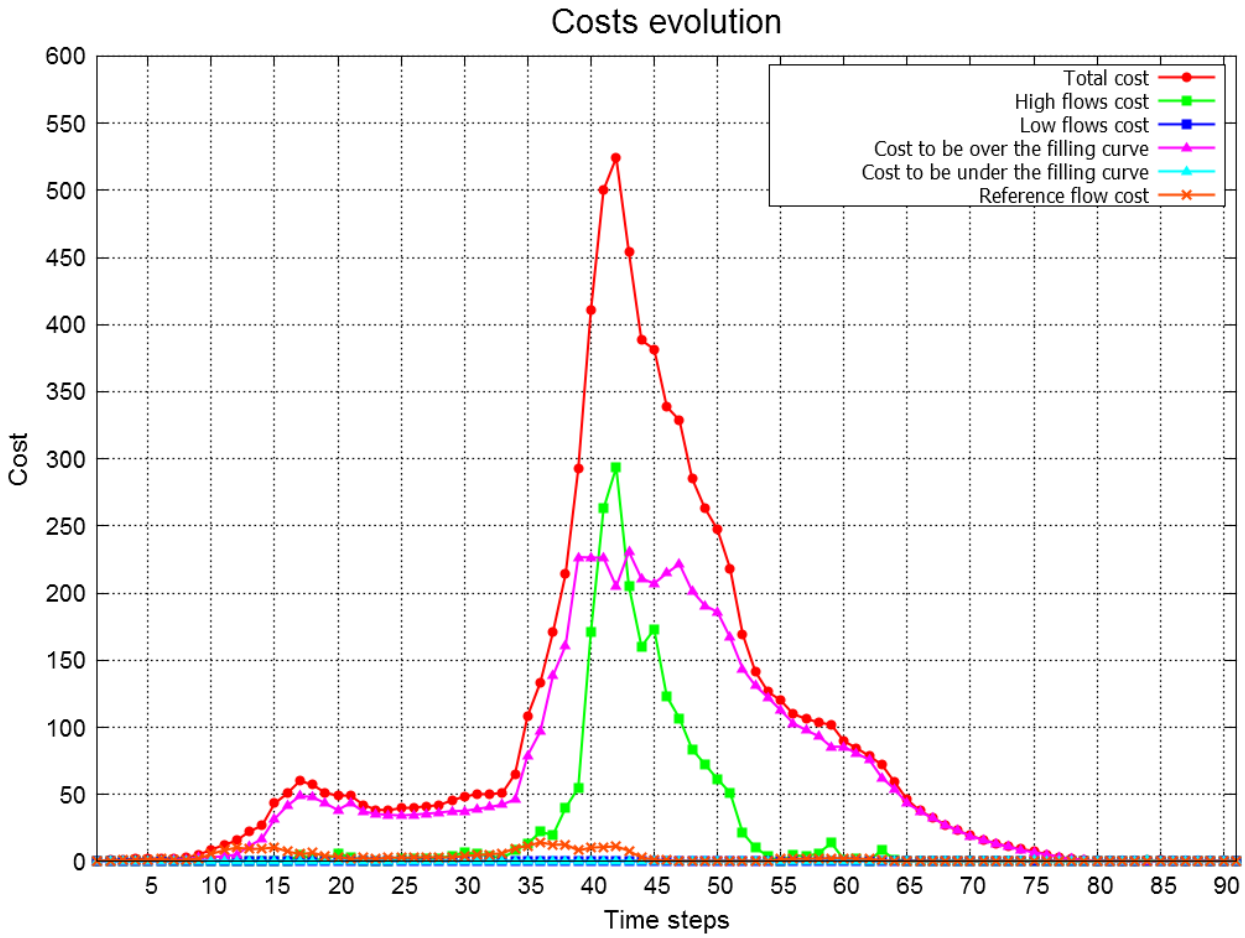


Figure 36. Optimal costs [-] evolution over the simulation horizon from 1/12/1981 (time-step = 0) to 1/3/1982 (time-step = 90) calculated over the prediction horizon at each time-step.

The result of the MPC simulation in terms of the floods performance indicator I_{floods} can be compared with those of the current rules and of the naturalized system, as reported in the histogram in Figure 37.

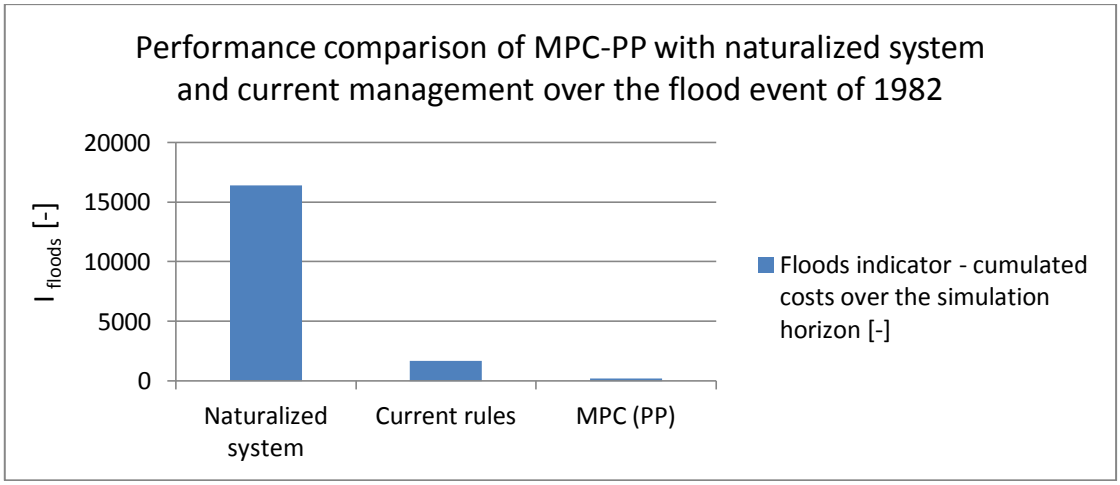


Figure 37. Floods performance indicator for the naturalized system, current management and MPC (with Perfect Predictor) over the simulation horizon 1/12/1981 - 1/3/1982.

Looking at the cumulated closed-loop costs for floods in the figure above, it can be noticed that the MPC outperforms the current management, that already brought a big improvement respect to the naturalized system.

5.3.1.5 *Some general comments (geographical and temporal conflicts)*

In general, from the point of view of flood control, the result that we expect is that at all the control stations the simulated river flow with the MPC controller must be:

- equal to the simulated river flow with the current management, if there are no problems at the control stations and/or in the storages of the lakes (FCs to recover);
- under the simulated river flow with the current management when this one overpasses the high flows thresholds and it's possible to avoid it deriving more water into the reservoirs than the current rules.

In other words, when with the current management the river flow exceeds the 2nd flood threshold (or the 3rd one) at the control station i, the flow peak must be cut by the MPC to be contained as most as possible under these thresholds, if the following conditions are respected:

- the reservoir(s) upstream the control station i has (have) a free storing capacity;
- the flow to be derived to get the river flow under the thresholds must be less than the capacity of the derivation channels.

In the several simulation experiments performed, it was found that the MPC manages to reduce the flow peaks over the 2nd threshold when the conditions mentioned above are verified. For example, the result for the flood event of 1982 shows that MPC manages to reduce the flood respect to the current management at all the stations until the maximum volumes of the reservoirs are reached. This problem of reaching the maximum volumes of the reservoirs expresses the temporal conflict of the objectives, that we have balanced using the framework explained in section 4.6.2.2.

There are some critical stations where the thresholds are more restrictive. The most critical control station is Nogent-sur-Seine. We can see that the thresholds for high flows defined at Nogent-sur-Seine are more restrictive than the thresholds for the other stations even immediately upstream/downstream. As it can be seen in the controlled stations network scheme reported in Appendix E, the station of Nogent-sur-Seine is just after the confluence between Aube and Seine. Looking at the values of the flow thresholds at this station and at the two immediately upstream (Arcis-sur-Aube and Mery-sur-Seine), we can see that the values at Nogent-sur-Seine are more restrictive (being less than the sum of the values at the two upstream stations). So, even when at the other stations the flows are not close to the second threshold, it can happen that at Nogent-sur-Seine there are flows very close to this threshold. This condition makes the MPC controller fill the reservoirs to avoid exceeding the flood threshold at Nogent. This problem limits the performance of the system for the possible flood events coming at other stations and it is representative of the geographic conflict of the objectives.

5.3.2 Performance on critical low-flows events

From the standpoint of drought attenuation, the advantage that MPC could bring respect to current rules lies in the feedback of the control relative to the downstream conditions. Moreover, the MPC management allows to control the lakes with more freedom respect to the filling curves management, thanks to the balance between the costs to be under the FC and low-flows step-costs. According to this balance, as explained in section 4.6.2.1, the MPC management releases more water than the current management when it's beneficial in terms of total costs (sum of costs to have a storage under the FC and of low-flows step-costs). This difference in behavior between MPC and current rules is of obvious importance for drought events of early or late occurrence along the year respect to the statistical occurrence used to

calculate the filling curves. For example, MPC is expected to face better droughts that occur in advance respect to statistics allowing to release water before the 1st of July. In order to analyze this behavior of the controller on low flows regimes, the long drought event of spring-summer 1976 was chosen for simulation experiments with MPC. This is the most critical and long drought event of the period of data availability: the low flows begin in the filling season (early spring). The naturalized flow is under the third and fourth low-flow thresholds at some control stations and under the second one at almost all the others. The simulation horizon chosen is from 1/3/1976 to 1/1/1977. The results show that the MPC outperforms the current management at all the stations keeping the river flow at the control stations over the second low-flow threshold for almost all the simulation horizon. This can be noticed for example in the following figure representing the hydrograph at Nogent-sur-Seine.

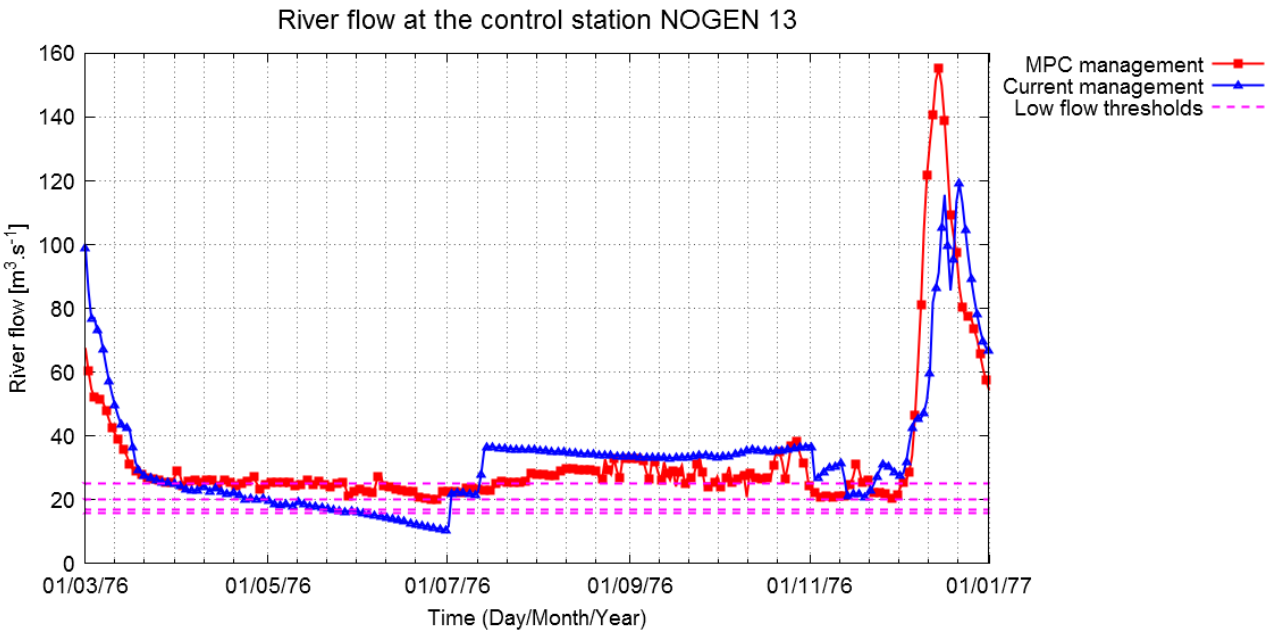


Figure 38. Simulated river flow [m^3/s] at Nogent-sur-Seine from 1/3/1976 to 1/1/1977 with MPC management (red line) and current management (blue line). The horizontal lines (dashed pink) correspond to the thresholds used to calculate the low-flows step-costs.

The bad performance of the current management until July 1976 is due to the absence of feedback in the control policy, that plans to continue to fill the reservoirs until July even if the river flows downstream are already under the low-flow thresholds. Instead the MPC controller begins to release water in April 1976, according to the balance of the costs to be under the FC and the low-flows step-costs. These two different behaviors can be noticed in the following Figure 39, representing the simulated volumes by using MPC and current rules.

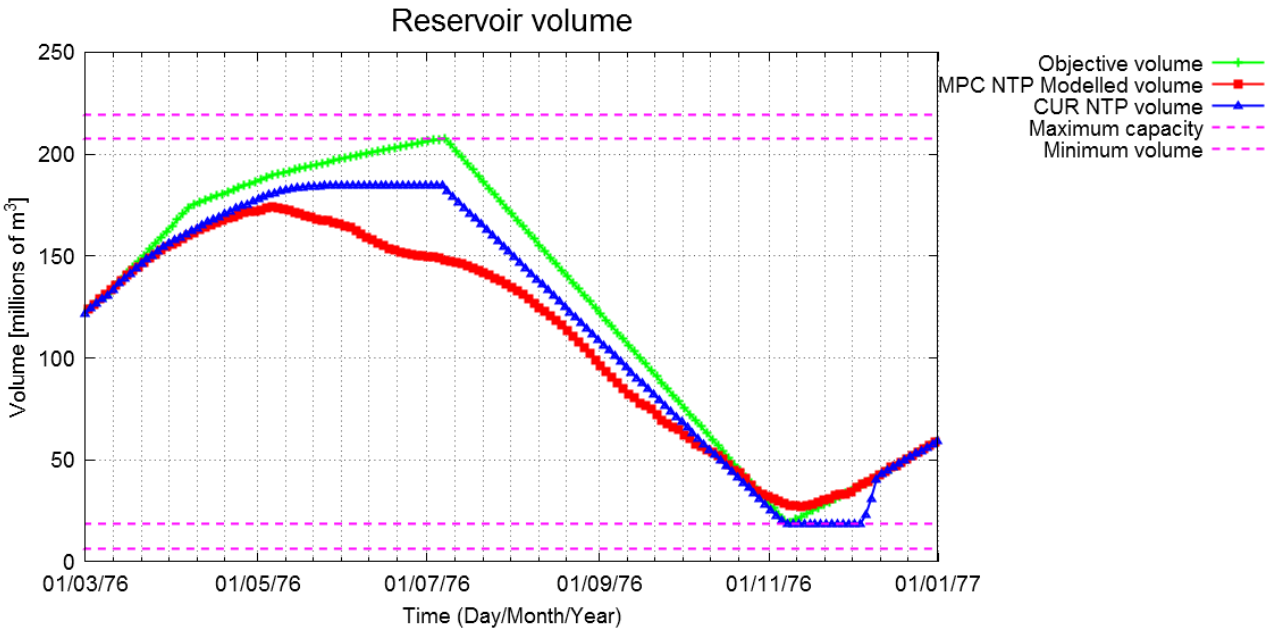


Figure 39. Volume [Mm3] of the Seine lake from 1/3/1976 to 1/1/1977 with MPC controller (red line) and current rules (blue line). The 2 lower horizontal lines correspond to the normal and exceptional minimum volumes.

The results of MPC, current rules and naturalized system in terms of cumulated low-flows step-costs in the closed-loop simulation are reported in the next figure. Comparing the values of the performance indicator we can see that the current management already carries a big improvement respect to the naturalized system, reducing the value of an order of magnitude. The MPC management brings a further slight improvement respect to current rules.

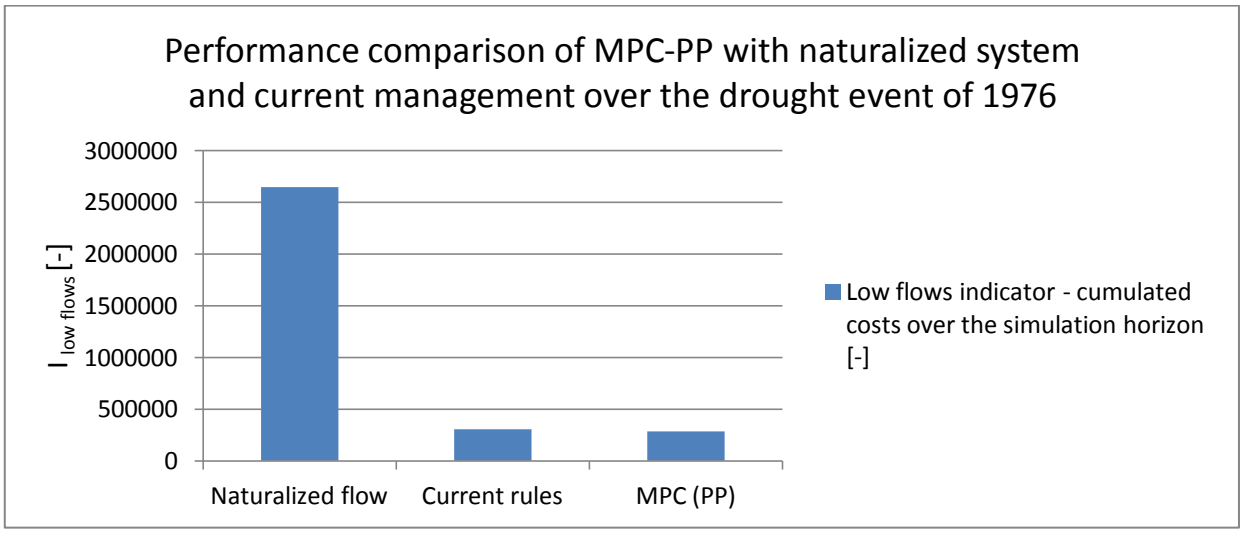


Figure 40. Low-flows performance indicator for naturalized system, current management and MPC (with Perfect Predictor) over the simulation horizon 1/3/1976 - 1/1/1977.

5.4 Deterministic MPC and TB-MPC with real forecasts

In this section we will present and compare the results of deterministic and stochastic MPC using real forecasts from the point of view of flood control. This choice is due to the fact that we can expect the bigger improvements by taking into account the uncertainty in the forecasts for high flows control, since this uncertainty typically increases with the inflows.

5.4.1 Constraining the thresholds for high and low flows

In the period of forecasts availability (11/3/2005 - 01/10/2008) there are not critical events of floods and droughts that jeopardize the actual management performance. So, it's easy to understand that our real-time controller would not encounter particular problems and would behave as the current manager.
In order to give a practical interest to our simulations using real forecasts, we decided to change artificially the thresholds for floods and low-flows used by MPC, so that they may become limiting for the hydrological regimes of the concerned period. For obtaining the new constrained thresholds we used a multiplicative coefficient calculated in order to have naturalized and current management flows over the high-flows thresholds at the control stations. The value of the multiplicative coefficient used for reducing the high flows thresholds for the simulations presented hereafter is $c_{hf} = 0{,}5$. Using the new thresholds so obtained (half of the original ones), some critical flood events can be found in the study period, for example in spring 2006 and 2007. In these two periods the naturalized and current management river flows overpass the new third threshold at some control stations and are above the second one at all the stations. So the MPC controller is expected to behave differently from current management to control these flood events. In Figure 41 we report the naturalized and current management flow at the downstream station of Paris over the period 11/3/2005 - 01/10/2008, and the new constrained thresholds.

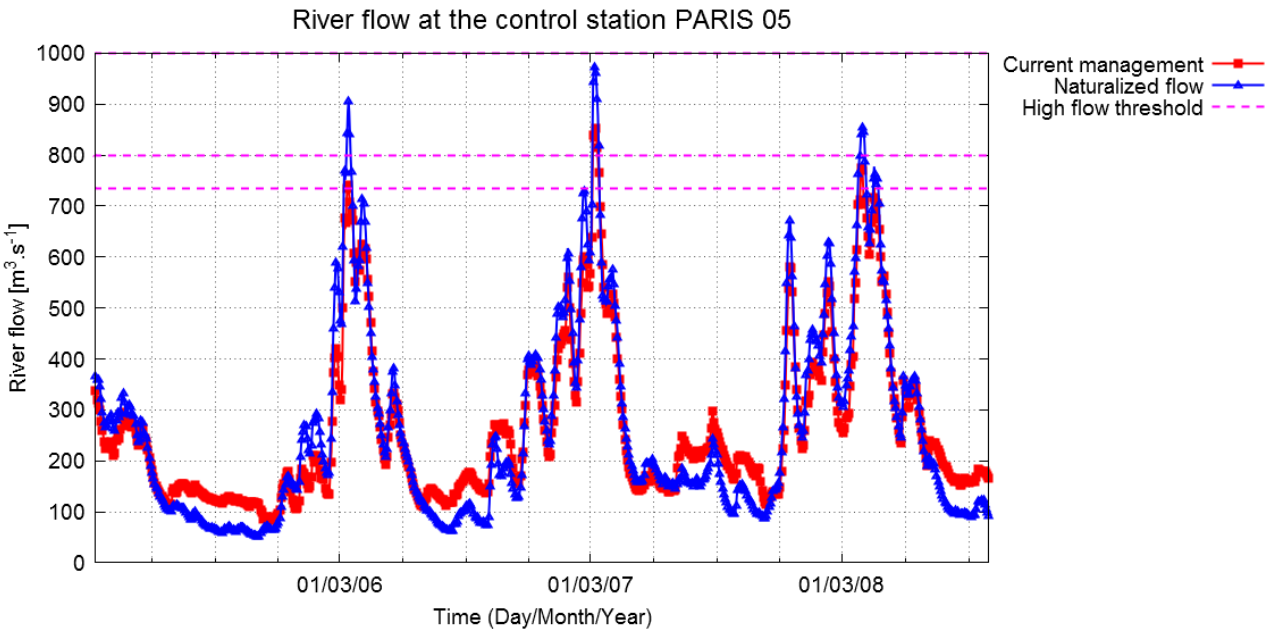


Figure 41. Simulated river flow [m^3/s] at Paris from 11/3/2005 to 1/10/2008 with current management (red line) and naturalized flows (blue line). The horizontal lines (dashed pink) correspond to the constrained thresholds ($c_{hf} = 0{,}5$) used to calculate the flood step-costs.

It must be noticed that the fact of using these new thresholds for MPC (and TB-MPC) makes comparisons with the current management not significant, because our real-time controllers do not consider anymore the same thresholds used so far for evaluating the current management. However, in Figure 41 we showed

the current management simulated flow because it was determinant to define the new constrained thresholds.

5.4.2 Performance on flood events

In order to analyze the performance of the controllers on flood events, we chose the period of spring 2007, since in this period there are the most critical high flows. The results presented hereafter are extracted from the results over the simulation horizon from 11/3/2005 to 1/10/2008.
The results of the simulation experiments with MPC and TB-MPC show a good performance in the control of the flood event in spring 2007. Analyzing the simulated river flows at the control stations it can be noticed that MPC and TB-MPC behave in a similar way, pursuing the objective of containing the river flows below the second flooding threshold at the control stations. In general both the controllers, MPC and TB-MPC, manage to stay under the second threshold or very close. The most critical stations in this simulation experiment are Paris and Noisiel, that present the most high peaks over the second threshold. We report in the following figure the hydrographs at Paris resulting by the simulation with MPC with perfect predictor, MPC with real deterministic forecasts and TB-MPC using ensemble forecasts. The performances ranking of the three methods at Paris is representative of the ranking at all the stations.

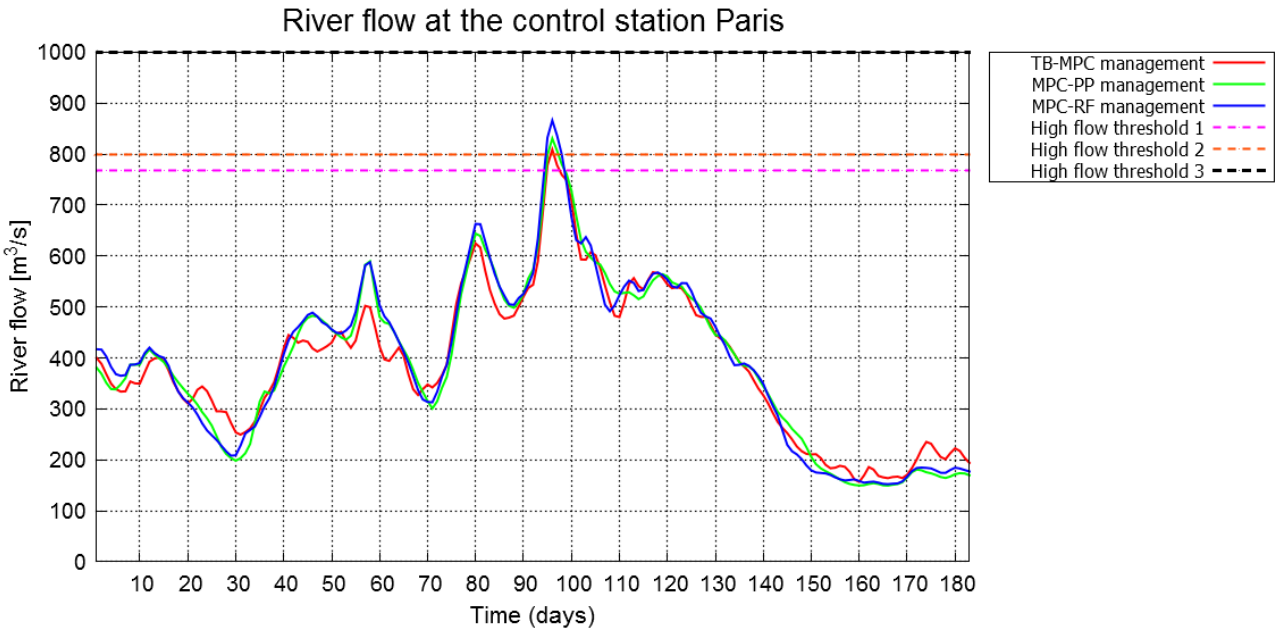


Figure 42. Simulated river flow [m3/s] at Paris from 1/12/2006 to 1/6/2007 with TB-MPC (red line), MPC with perfect predictor (green line) and MPC with real forecasts (blue line).

In the hydrograph above it can be noticed that TB-MPC does a little bit better than MPC with real forecasts and perfect predictor, reducing more the flow peaks over the second threshold, though the simulated flows with these different methods remain very close. A strange result is that MPC with perfect predictor ranks in the intermediate position between TB-MPC and MPC with real forecasts, while it was expected to be the upper bound of the methods. This is a not rational phenomenon because using uncertain forecasts should not lead to better results than a perfect knowledge of the future. This problem seems to be due to the sub-optimality of the solution of MPC, that has yet to be investigated and overcome.
The big flow peaks reduction operated by the centralized controllers is obtained filling the reservoirs more than the FCs, as already remarked in section 5.3 for the results of MPC with perfect predictor. For example, this behavior is shown in the following plot of the Aube lake volume.

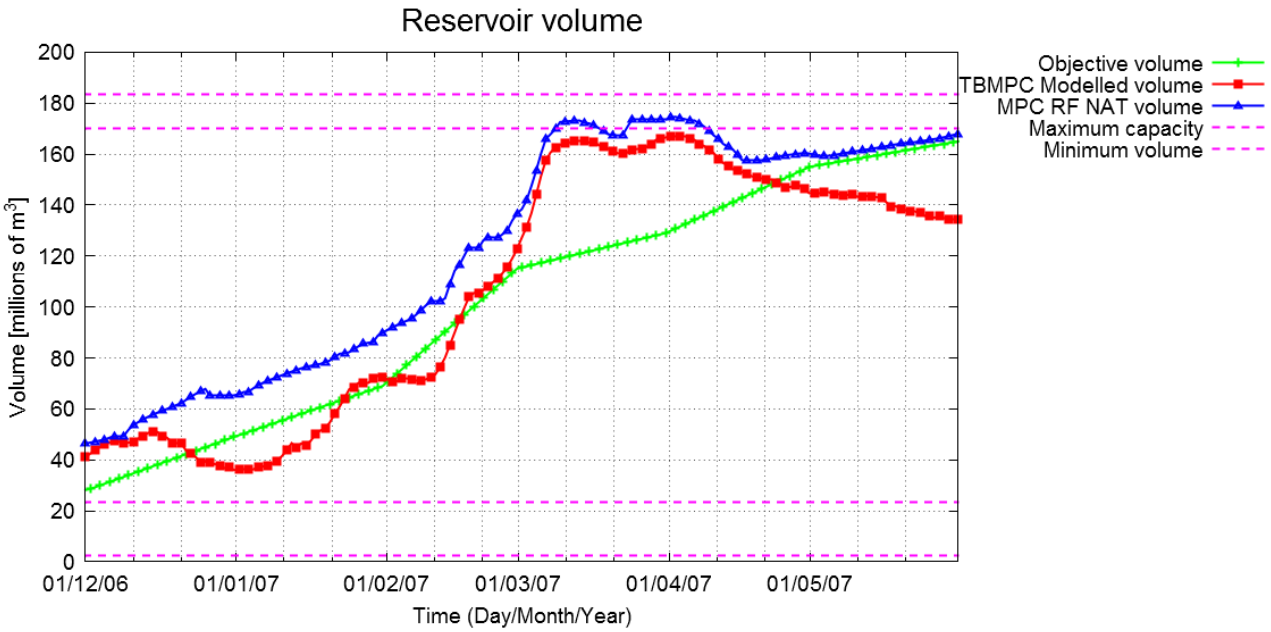


Figure 43. Volume [Mm3] of the Aube lake from 1/12/2006 to 1/6/2007 with TB-MPC (red line with squares) and MPC (blue line with triangles). The upper horizontal line correspond to the maximum exceptional capacity.

To compare the performances of MPC and TB-MPC we report the high-flows step-costs cumulated over the simulation horizon in Figure 44.

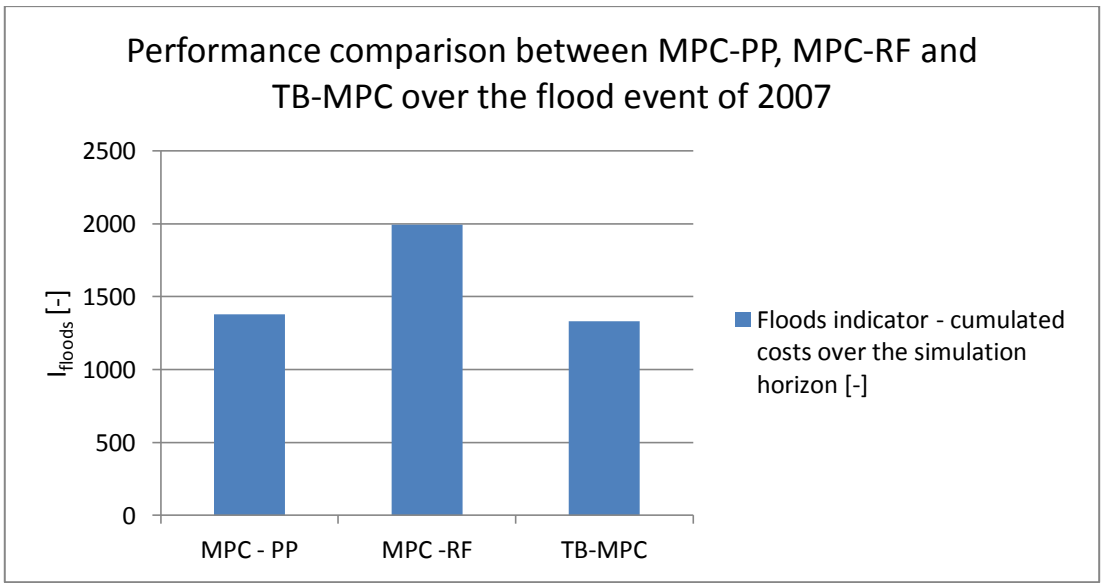


Figure 44. Costs of closed-loop simulation over the period 1/1/2007 - 1/6/2007 with MPC using a perfect predictor, MPC using real deterministic forecasts and TB-MPC using real ensemble forecasts.

Comparing TB-MPC and MPC in terms of cumulated flood step-costs, it can be noticed that TB-MPC clearly outperforms MPC with real forecasts, as already shown in the hydrographs above. As already discussed, the fact that TB-MPC does slightly better than MPC with perfect predictor is unexpected and probably due to optimization problems still unresolved, that have to be further investigated.

5.5 Testing MPC and TB-MPC over climate scenarios

One on-going direction of research is testing MPC and TB-MPC over future climate scenarios (GCM), in order to investigate their possible contribution in efficiency improvement to compensate the effects of climate changes. The simulation experiments are done using 7 different Global Circulation Models (GCM), forced by the moderate A1B - IPCC emission scenario, to predict the climate evolution over the period 2046-2065 (TF) and to produce past climate scenarios over the period 1961-1991 (TP). We can assume these GCM scenarios to be the future (and past) realizations of the meteorological stochastic process. In this way, each GCM model can be used as a deterministic perfect predictor. Afterwards, artificial deterministic and ensemble forecasts are generated with the same statistical properties of the errors of the available ones (ECMWF) over these scenarios. To perform this generation we used a *multivariate error model* to produce "*ensemble-dressing*" of the future (and past) climatic scenarios of P [Dhouioui, 2011]. The error model is calibrated on the error of real ensemble forecasts produced by ECMWF (available from 2005 to 2008), as schematically shown in the next figure.

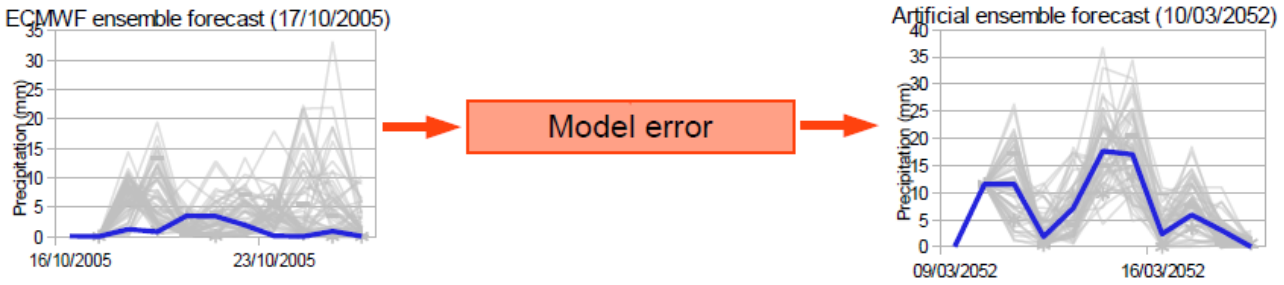


Figure 45. Outline of the artificial ensemble forecasts generation for future climate scenarios using an error model calibrated on the available ensemble forecasts for past scenario.

In the GCM scenarios, we have an increase of extreme events (floods and droughts) and so we can test the RTC on some interesting critical events. The results obtained so far over GCM scenarios regard the use of deterministic MPC with perfect predictor and TB-MPC with artificial ensemble forecasts; however, we must re-launch these simulations after the latest code improvements (and we still have to run the simulation of MPC with real forecasts). Anyway we present here the current results available of the simulations with TB-MPC over all the GCM scenarios in present time (TP - 1961/1991) and future time (TF - 2046-2065). Also current management simulation results are available, so we can compare the performance of TB-MPC with that of actual rules. These results on long simulation horizons (many years) can be analyzed looking at the distribution of the values of some efficiency indicators, already used for analyzing the current-management simulation results by Dorchies et al. [2013]. These indicators are based on two sets of conditions at the downstream control stations: the set of satisfactory cases, noted S, when flow at the control stations remains within the limits defined by the thresholds, and the set of failures, noted F, when flow is outside these limits. Here, we will analyze the following indicators:

a) The *failure events frequency*. A failure event is defined by the consecutive days when Q(t) is in the failure state [Hashimoto et al., 1982]. The *frequency of a failure event* is equal to $freq = \frac{365{,}25 \cdot j}{n}$, where: *freq* is the frequency in years^{-1}, n is the number of days of the study period and j the number of failure events during the period.

b) The *mean duration of a failure period* defined as: $d_{mean} = \frac{1}{j} \cdot \sum_{i=1}^{j} d(i)$, where d(i) is the duration of the i-th failure period.

c) The *failure rate*, that can be defined as the complementary of reliability. Reliability is the probability for flow Q(t) at time t to be in the satisfactory state S [Hashimoto et al., 1982].

d) The *mean vulnerability*. It is among the *vulnerability indicators* defined by Kjeldsen et al. [2004]. In case the reservoir management fails to maintain downstream flow under a high-flow threshold (or above a

low-flow threshold), vulnerability is defined as the volume that should have been taken (or, respectively, released) by the reservoirs to avoid this situation during the event considered. This can be calculated for each failure event. Statistical indicators can be calculated from the set of the vulnerability values computed for all the events, as the mean value, Vul1.

In Figure 46 we report the box-plots of the distribution of these indicators calculated over the results of TB-MPC and current management simulations, for high-flows and low-flows events.

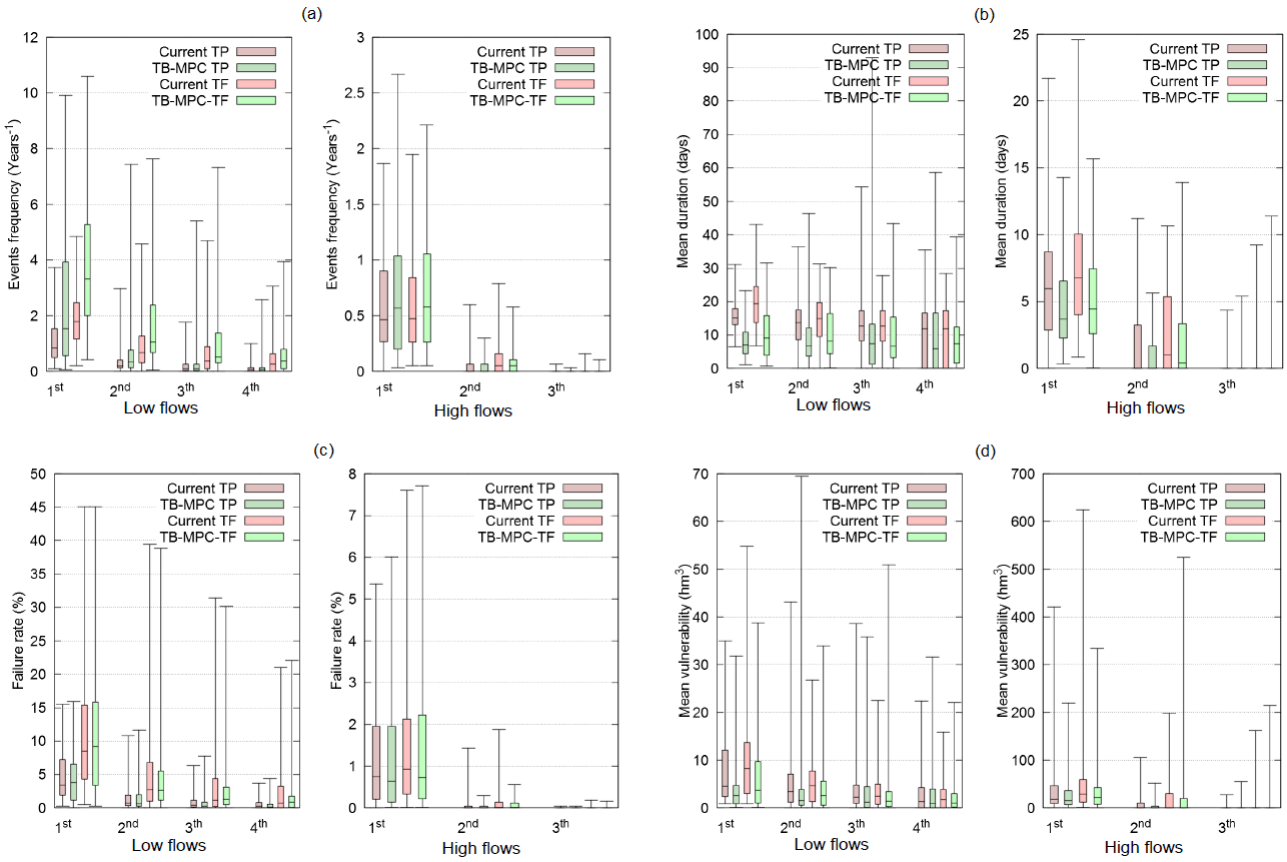


Figure 46. Performance benchmarks of TB-MPC and current management over the 7 GCM scenarios run for present and future time (TP and TF). Box-plots of: (a) events frequency; (b) mean duration; (c) failure rate; (d) mean vulnerability.

Comparing TB-MPC and current management, in the box-plots reported above, it can be noticed:

a) *Events frequency*: for high-flows there's not a clear trend; low-flows events are more frequent with TB-MPC. This result should not be interpreted negatively in itself, in fact the TB-MPC exceeds the thresholds for a larger number of times but for shorter periods than current management does. This behavior can be noticed in figures Figure 33 and Figure 38 where the simulated flow by using MPC oscillates many times near the thresholds, slightly exceeding them for some times, while the current management exceeds the thresholds only once but much more largely.
b) *Mean duration*: for both high and low flows the mean duration is less with TB-MPC.
c) *Failure rate*: the number of days over critical thresholds is always decreased by TB-MPC.
d) *Mean vulnerability*: on average, TB-MPC decreases the mean vulnerability.

From these results we can deduce that TB-MPC globally improves the performance of the reservoirs management. So the contribution of a centralized real time controller using ensemble forecasts can be determinant to compensate the effects of climate change.

6. Conclusions and perspectives

In this work we developed and tested a centralized Real-Time Controller for the reservoirs management on the Seine River using all data available in real-time, including deterministic and ensemble weather forecasts. Our goals were: (i) to investigate the effectiveness of MPC and TB-MPC to increase the efficiency of reservoirs management and to overcome the limitations of the current rules; (ii) to determine whether the stochastic formulation of TB-MPC improved compared to the deterministic MPC; (iii) to test the centralized RTC over future climate scenarios to assess its ability in dealing with the impacts of climate change on hydrological conditions.

(i) A centralized Model Predictive Controller is showing better results than the current decentralized rules as we have shown in the results presented in the previous chapter (see section 5.3).
The sources of improvement of the dams management efficiency by using MPC (and TB-MPC) can be identified in:

- *The centralization of the controller*: the natural advantage is that each lake contributes to the management of the flood and low flows in function of the remaining capacity and of the actual storage. For example, when a lake is full the other lakes supply their remaining capacity for controlling the flood at the downstream stations.

- *The feedback introduced in the control policy and the use of forecasts:* the benefit deriving from these elements is that the MPC management fills the lakes more than the FCs when a flood is forecasted, in order to reduce the flow peaks at the key downstream stations. In the same way, our controller releases more than the current rules when critical low flows are predicted.

(ii) The TB-MPC was expected to improve the performance respect to the deterministic MPC, above all in case of flood control, because in this case the uncertainty in the forecasts is bigger. As shown in the previous chapter (see section 5.4), the results of TB-MPC and MPC using real forecasts are very similar, even if TB-MPC performs a little bit better. However, as already noticed, the optimization encounters some problems because of the big number of variables to be optimized. So the results are affected by the uncertainty on the sub-optimality of the actual solution. This sub-optimality is expected to be more important in the stochastic formulation than in the deterministic one, in reason of the bigger number of variables. However, the results presented in this thesis are contradictory with this expectation, since the MPC with perfect predictor is slightly outperformed by TB-MPC. It seems that there's a problem in the MPC with perfect predictor solution that seems to be affected by sub-optimality more than TB-MPC. So it will be necessary to better assess the contributions of assimilation of the forecasts uncertainty in the improvement of the dams management, after overcoming these optimization problems, following the possible ways proposed in the end of this chapter as further research issue.

(iii) The results over future climate scenarios are preliminary, but seem to confirm that a centralized Model Predictive Controller can help to alleviate the impacts of climate change, improving the efficiency of the reservoirs management (see section 5.5).

Here below we analyze briefly the perspectives of further research for this work.

In the present work we used the same model for MPC and for representing the reality, so we analyzed only the uncertainty of the meteorological forecasts and not of the hydrological model. The uncertainty in the

hydrological model will be considered using the accurate model of the system implemented in SIC [J.P. Baume et al., 2005]. In Figure 47 the scheme of integration of the accurate model in the project is shown. At each time-step of application of MPC, the reduced model (TGR) is used as internal model to calculate the optimal controls that then can be used as inputs by the accurate model to update the states of the system. These states can also be used as inputs of the data-assimilation process to better calibrate the internal model. So at each simulation step, the internal model can be re-calibrated thanks to the accurate model that uses observations and optimal controls, to balance the uncertainty in the weather data and in the forecasting model.

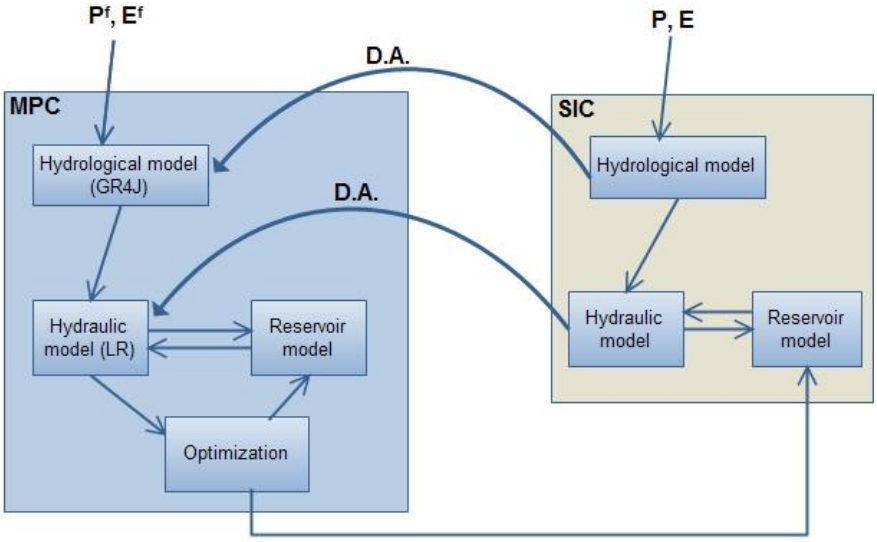


Figure 47. Flow-chart for the integration of the accurate model (implemented in SIC) and the data-assimilation (D.A.) process in the framework of the present work.

The accurate model and the data-assimilation are scheduled to be applied to the present work in the short term after the end of the research reported here.

In order to test TB-MPC and MPC-RF over past scenarios, before the period of availability of the real forecasts (before 2005), we will generate artificial predictions for the past as done for the future (see section 5.5). In this way we will be able to analyze the results on the observed period and compare TB-MPC and MPC, between them, and with the current management over a long period and without need to change the real thresholds.

In the results presented in the previous chapter, we have seen that the fact that the controller has to work to limit the high flows at the critical station of Nogent-sur-Seine for a long period in winter makes the lakes be completely filled soon, consuming their capacity to laminate any eventual flood that occur later in time at other stations. This geographical conflict among the objectives at different stations could be solved weighting less the cost at the critical station of Nogent-sur-Seine in order to maintain more capacity in the reservoirs to control the river flows at the other stations in the future. This aspect of better balancing the costs among the different stations could be further discussed with SGL, as a further research direction after the end of this work. As we have faced the temporal conflict, defining a penalty function on the final state of the lakes, it will be possible to formalize a balancing framework for the geographical conflict. A way could be the definition of different weights for the step costs at different control stations in relation to their importance, that should involve the manager and the stakeholders. For example, we can expect that a ranking among the stations for defining the weights could be correlated to the presence and importance of an urban area near the control station. In this sense it's reasonable that the most important weight should be assigned to the very downstream station of Paris.
The temporal conflict between the objectives was found to be a central aspect of the optimization problem. To face it, we developed a new framework for balancing the penalty costs to be over/under the

FCs and the costs for floods/droughts. For low-flows case, we used new filling curves under the first original one calculated to support the low-flows thresholds all over the releasing period, with a simplified method. This method is based on the assumption that the first FC is calculated to support the first low-flow threshold for all the releasing period. This hypothesis could be wrong. For this reason these new filling curves could be better constructed on statistical basis. For high flows, some curves over the first one were constructed as reference level to balance the high-flows step-costs. A first sensitivity analysis on the parameters used to calculate them (the storage levels used to counterbalance the costs) has been performed (testing some different values), but it could be refined.

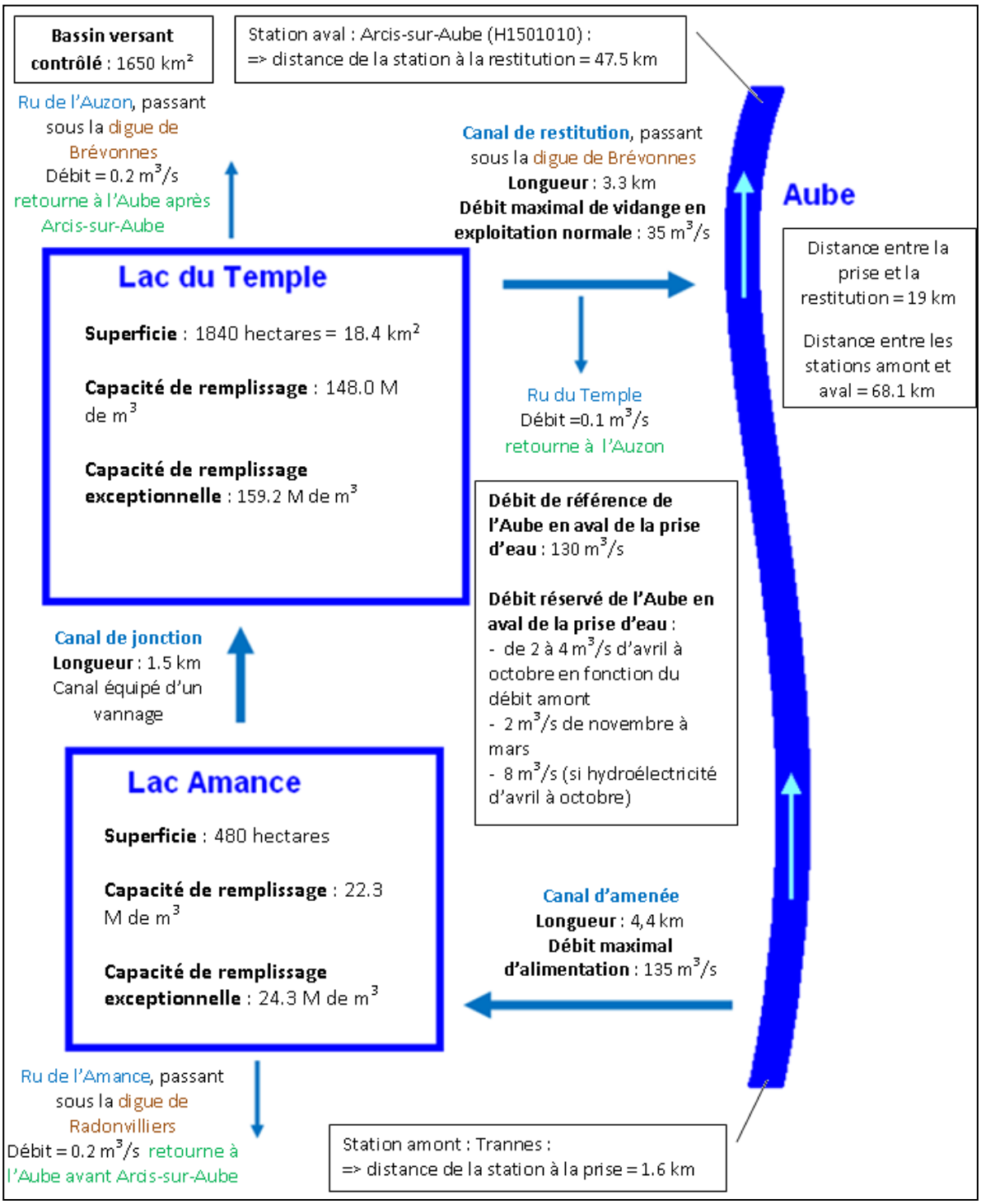
The most important limitation of this study is given by the optimization problems that we have already described in this thesis. The optimization algorithm of Nelder-Mead has proved to be not suitable for the case-study problem, despite all the efforts to improve its convergence. The uncertainty of the sub-optimality of the solution could be so large as to influence our comparison analysis between the solutions with MPC and TB-MPC. To make these analysis more robust, other optimization algorithms should be tested, for example LQ programming or evolutionary algorithms. The time required for these other implementations of the problem was too big to carry out them in the framework of the present work. Another direction of research to respond to the current sub-optimality of the solution could be the use of a decentralized control problem for each lake. In this way we could reduce the number of variables to be optimized and so obtain a better optimal solution. This first solution could be then used as initial solution for the centralized problem.

Appendices

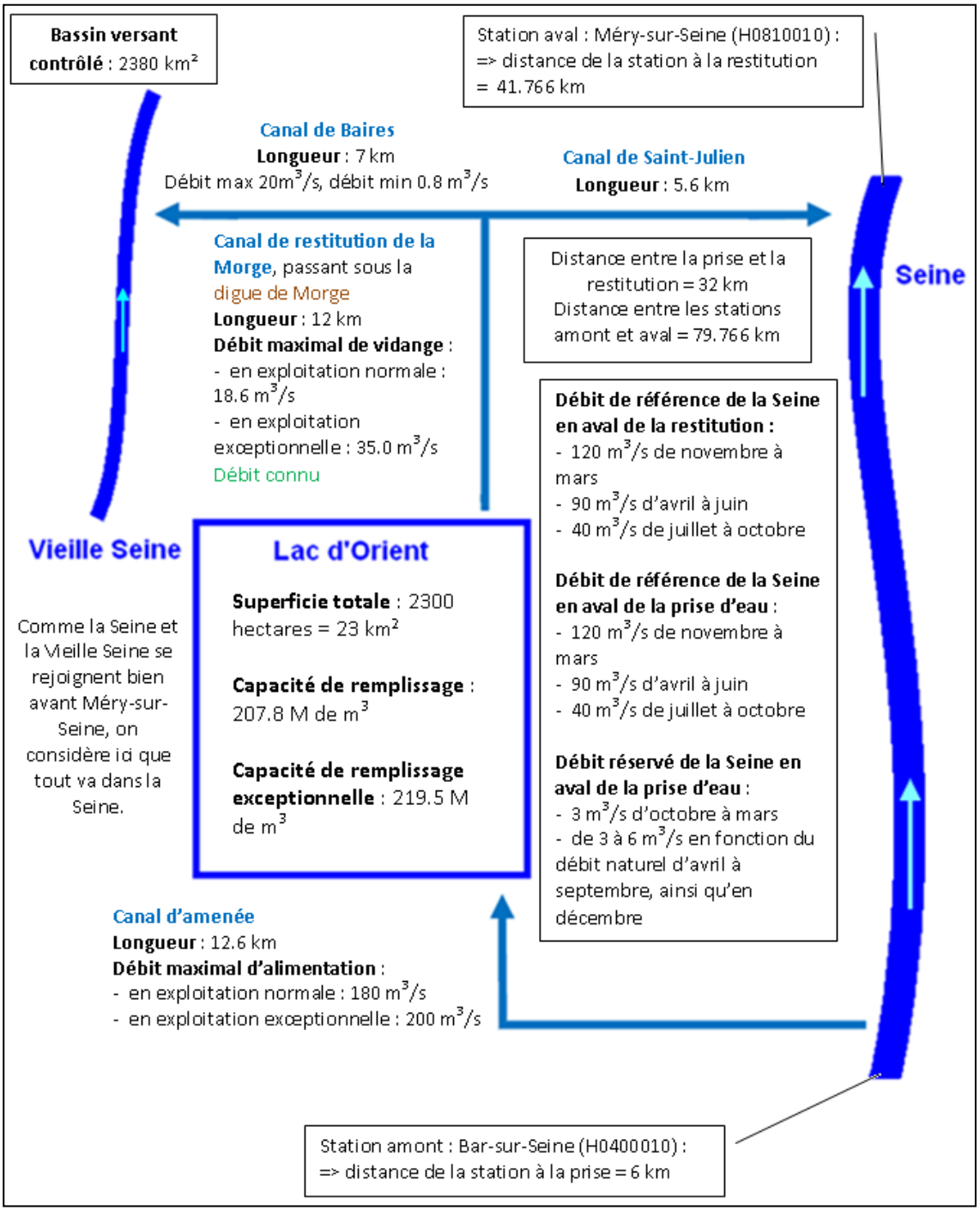
A. Technical schemes of the lakes

(Source: Dehay, 2012)

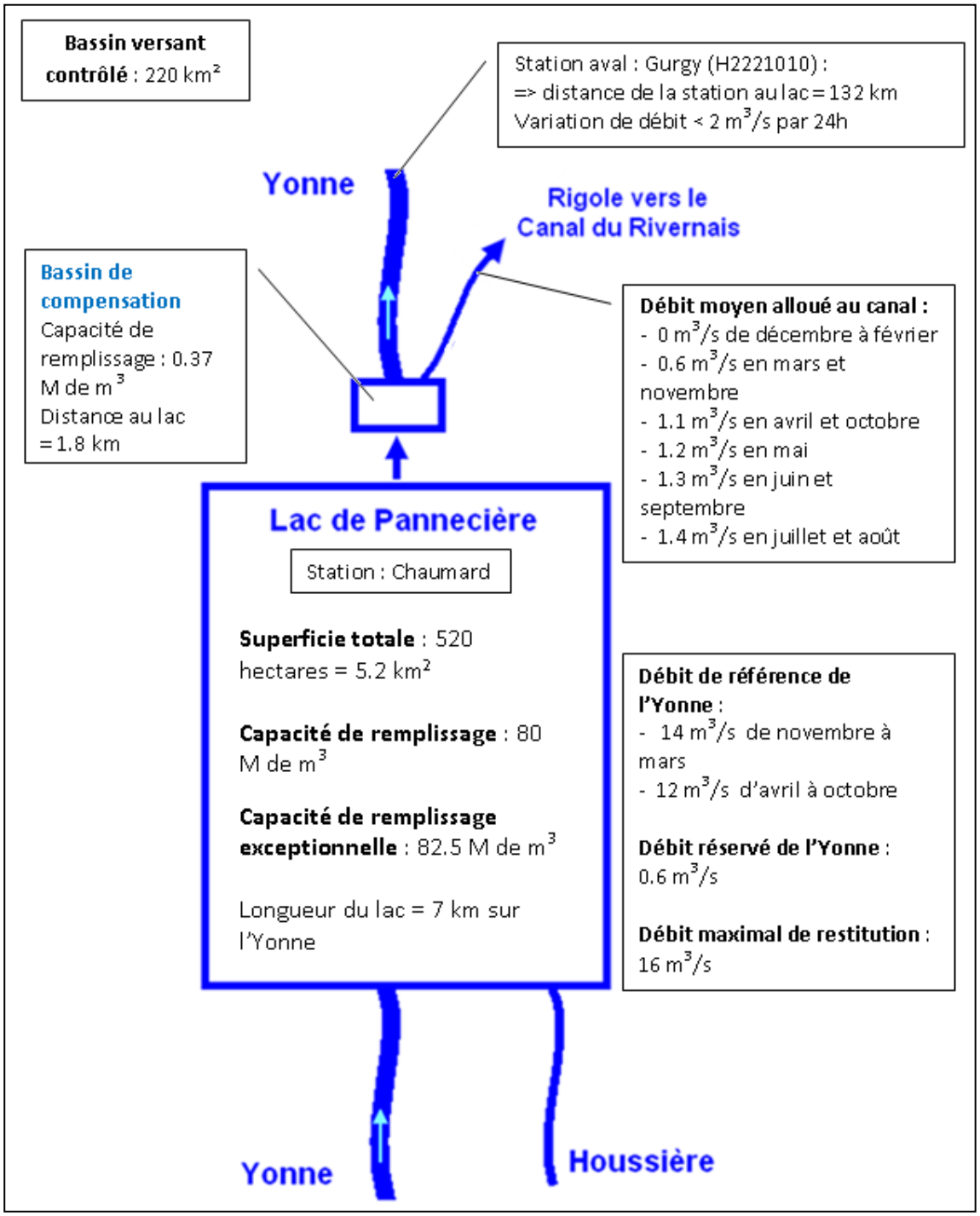
- **Aube**



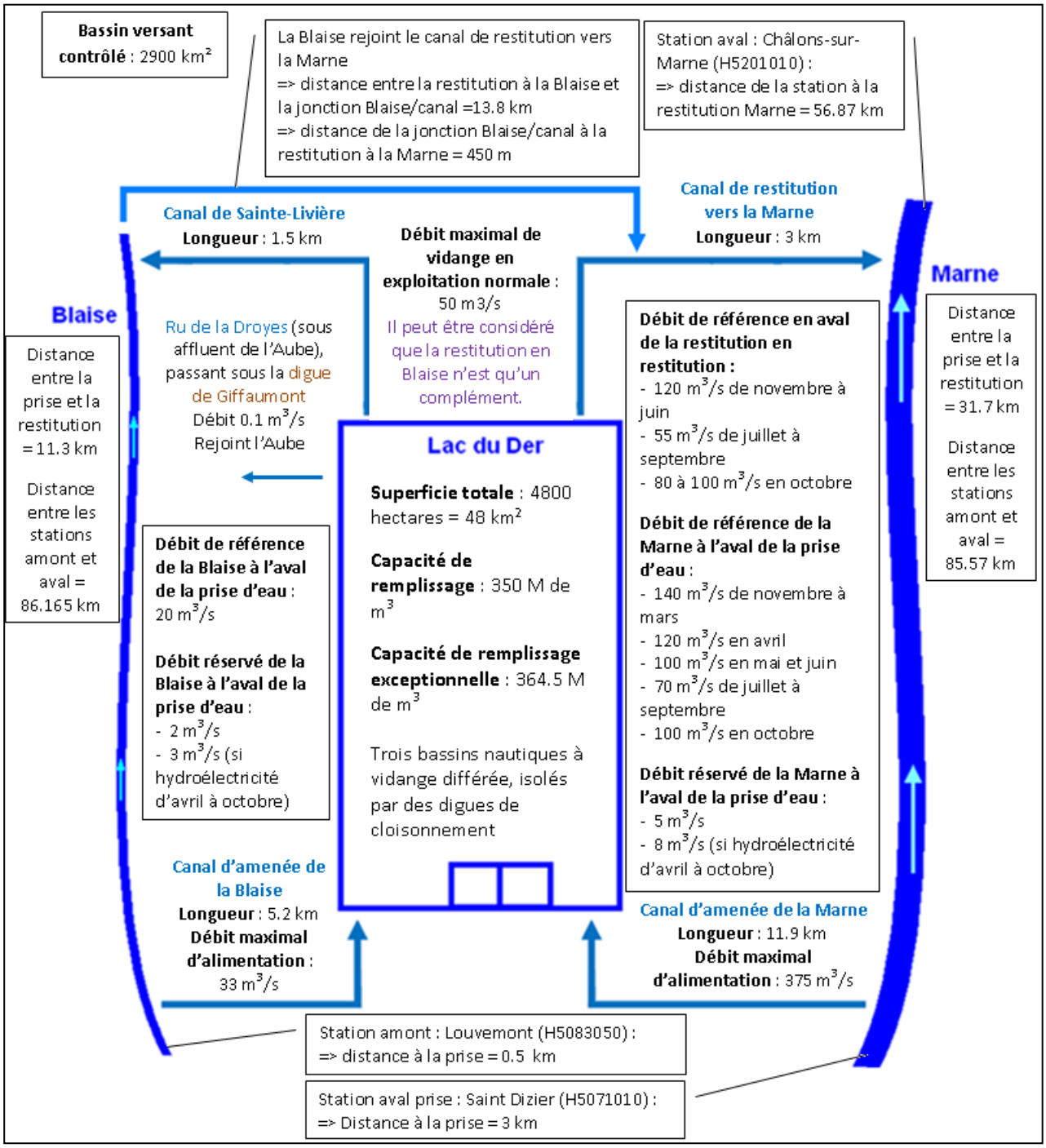
➢ **Seine**


Bassin versant contrôlé : 2380 km²
Station aval : Méry-sur-Seine (H0810010) : => distance de la station à la restitution = 41.766 km
Canal de Baires
Longueur : 7 km
Débit max 20m³/s, débit min 0.8 m³/s
Canal de Saint-Julien
Longueur : 5.6 km
Canal de restitution de la Morge, passant sous la digue de Morge
Longueur : 12 km
Débit maximal de vidange :
- en exploitation normale : 18.6 m³/s
- en exploitation exceptionnelle : 35.0 m³/s
Débit connu
Distance entre la prise et la restitution = 32 km
Distance entre les stations amont et aval = 79.766 km
Seine
Débit de référence de la Seine en aval de la restitution :
- 120 m³/s de novembre à mars
- 90 m³/s d'avril à juin
- 40 m³/s de juillet à octobre
Débit de référence de la Seine en aval de la prise d'eau :
- 120 m³/s de novembre à mars
- 90 m³/s d'avril à juin
- 40 m³/s de juillet à octobre
Débit réservé de la Seine en aval de la prise d'eau :
- 3 m³/s d'octobre à mars
- de 3 à 6 m³/s en fonction du débit naturel d'avril à septembre, ainsi qu'en décembre
Vieille Seine
Comme la Seine et la Vieille Seine se rejoignent bien avant Méry-sur-Seine, on considère ici que tout va dans la Seine.
Lac d'Orient
Superficie totale : 2300 hectares = 23 km²
Capacité de remplissage : 207.8 M de m³
Capacité de remplissage exceptionnelle : 219.5 M de m³
Canal d'amenée
Longueur : 12.6 km
Débit maximal d'alimentation :
- en exploitation normale : 180 m³/s
- en exploitation exceptionnelle : 200 m³/s
Station amont : Bar-sur-Seine (H0400010) : => distance de la station à la prise = 6 km

➢ **Pannecière**


Bassin versant contrôlé : 220 km²
Station aval : Gurgy (H2221010) :
=> distance de la station au lac = 132 km
Variation de débit < 2 m³/s par 24h
Yonne
Rigole vers le Canal du Rivernais
Bassin de compensation
Capacité de remplissage : 0.37 M de m³
Distance au lac = 1.8 km
Débit moyen alloué au canal :
- 0 m³/s de décembre à février
- 0.6 m³/s en mars et novembre
- 1.1 m³/s en avril et octobre
- 1.2 m³/s en mai
- 1.3 m³/s en juin et septembre
- 1.4 m³/s en juillet et août
Lac de Pannecière
Station : Chaumard
Superficie totale : 520 hectares = 5.2 km²
Capacité de remplissage : 80 M de m³
Capacité de remplissage exceptionnelle : 82.5 M de m³
Longueur du lac = 7 km sur l'Yonne
Débit de référence de l'Yonne :
- 14 m³/s de novembre à mars
- 12 m³/s d'avril à octobre
Débit réservé de l'Yonne : 0.6 m³/s
Débit maximal de restitution : 16 m³/s
Yonne
Houssière

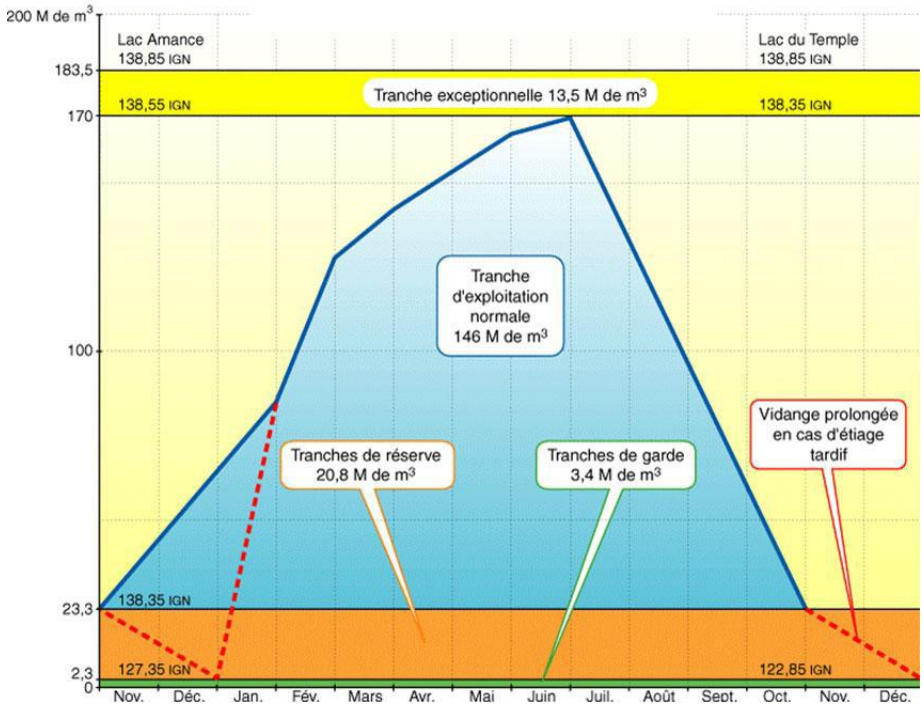
➢ Marne


Bassin versant contrôlé : 2900 km²
La Blaise rejoint le canal de restitution vers la Marne
=> distance entre la restitution à la Blaise et la jonction Blaise/canal =13.8 km
=> distance de la jonction Blaise/canal à la restitution à la Marne = 450 m
Station aval : Châlons-sur-Marne (H5201010) :
=> distance de la station à la restitution Marne = 56.87 km
Canal de Sainte-Livière
Longueur : 1.5 km
Débit maximal de vidange en exploitation normale : 50 m3/s
Il peut être considéré que la restitution en Blaise n'est qu'un complément.
Canal de restitution vers la Marne
Longueur : 3 km
Marne
Blaise
Distance entre la prise et la restitution = 11.3 km
Distance entre les stations amont et aval = 86.165 km
Ru de la Droyes (sous affluent de l'Aube), passant sous la digue de Giffaumont
Débit 0.1 m³/s
Rejoint l'Aube
Lac du Der
Superficie totale : 4800 hectares = 48 km²
Capacité de remplissage : 350 M de m³
Capacité de remplissage exceptionnelle : 364.5 M de m³
Trois bassins nautiques à vidange différée, isolés par des digues de cloisonnement
Débit de référence en aval de la restitution en restitution :
- 120 m³/s de novembre à juin
- 55 m³/s de juillet à septembre
- 80 à 100 m³/s en octobre
Débit de référence de la Marne à l'aval de la prise d'eau :
- 140 m³/s de novembre à mars
- 120 m³/s en avril
- 100 m³/s en mai et juin
- 70 m³/s de juillet à septembre
- 100 m³/s en octobre
Débit réservé de la Marne à l'aval de la prise d'eau :
- 5 m³/s
- 8 m³/s (si hydroélectricité d'avril à octobre)
Distance entre la prise et la restitution = 31.7 km
Distance entre les stations amont et aval = 85.57 km
Débit de référence de la Blaise à l'aval de la prise d'eau : 20 m³/s
Débit réservé de la Blaise à l'aval de la prise d'eau :
- 2 m³/s
- 3 m³/s (si hydroélectricité d'avril à octobre)
Canal d'amenée de la Blaise
Longueur : 5.2 km
Débit maximal d'alimentation : 33 m³/s
Canal d'amenée de la Marne
Longueur : 11.9 km
Débit maximal d'alimentation : 375 m³/s
Station amont : Louvemont (H5083050) :
=> distance à la prise = 0.5 km
Station aval prise : Saint Dizier (H5071010) :
=> Distance à la prise = 3 km

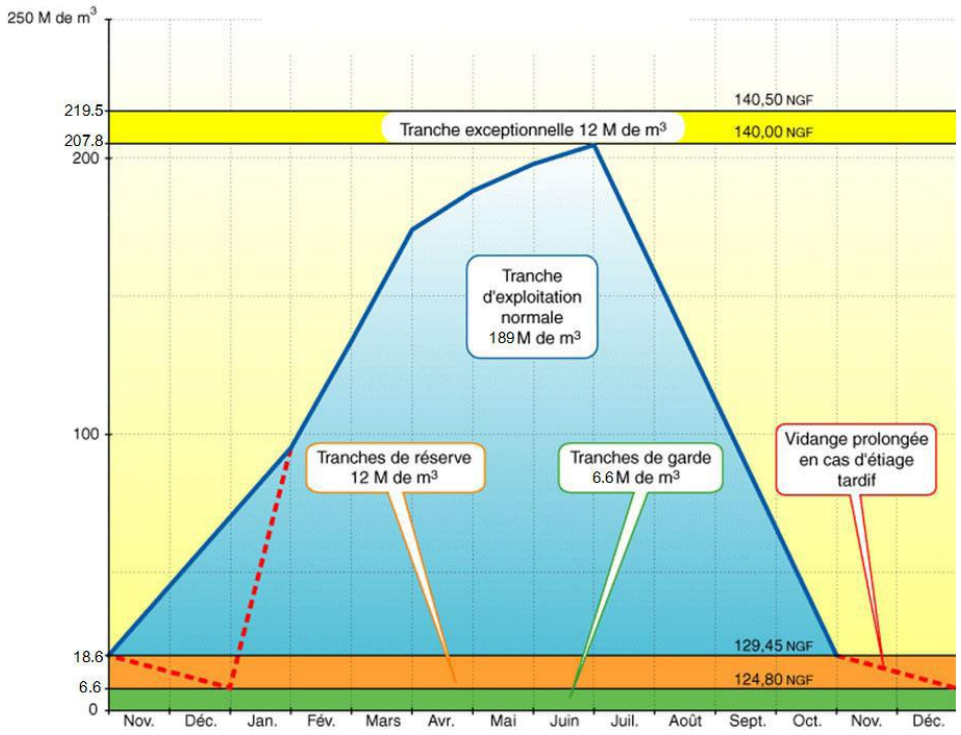
B. Management Filling Curves of the lakes

(Source: www.seinegrandslacs.fr)

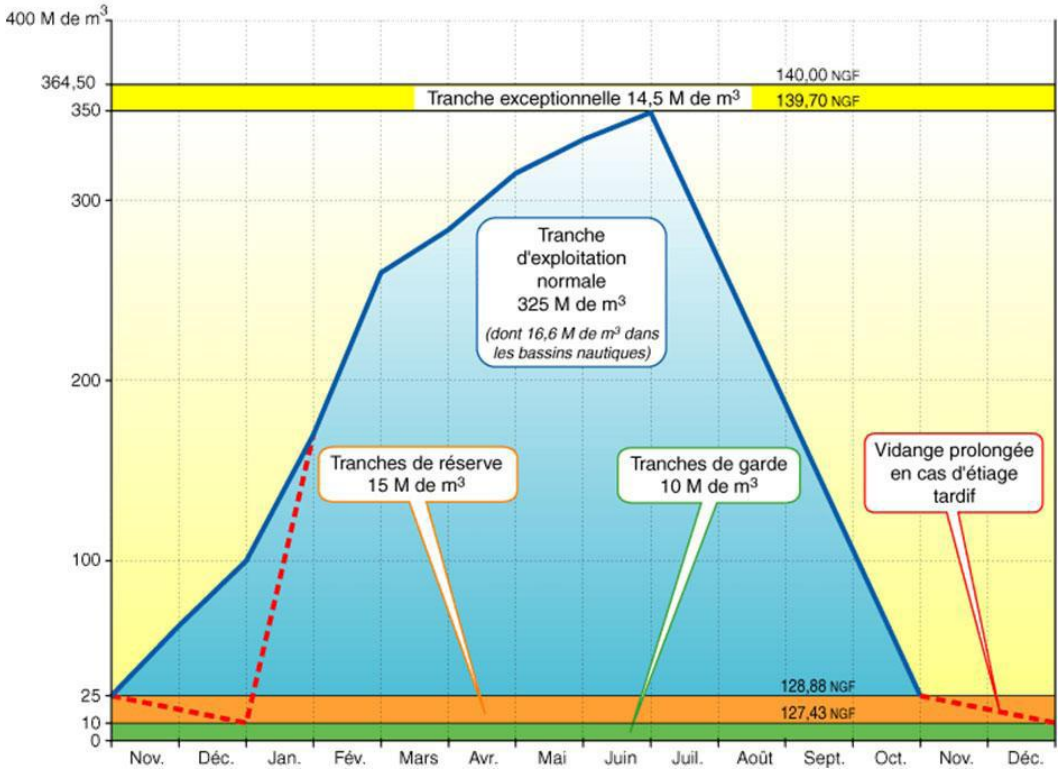
- **Aube**



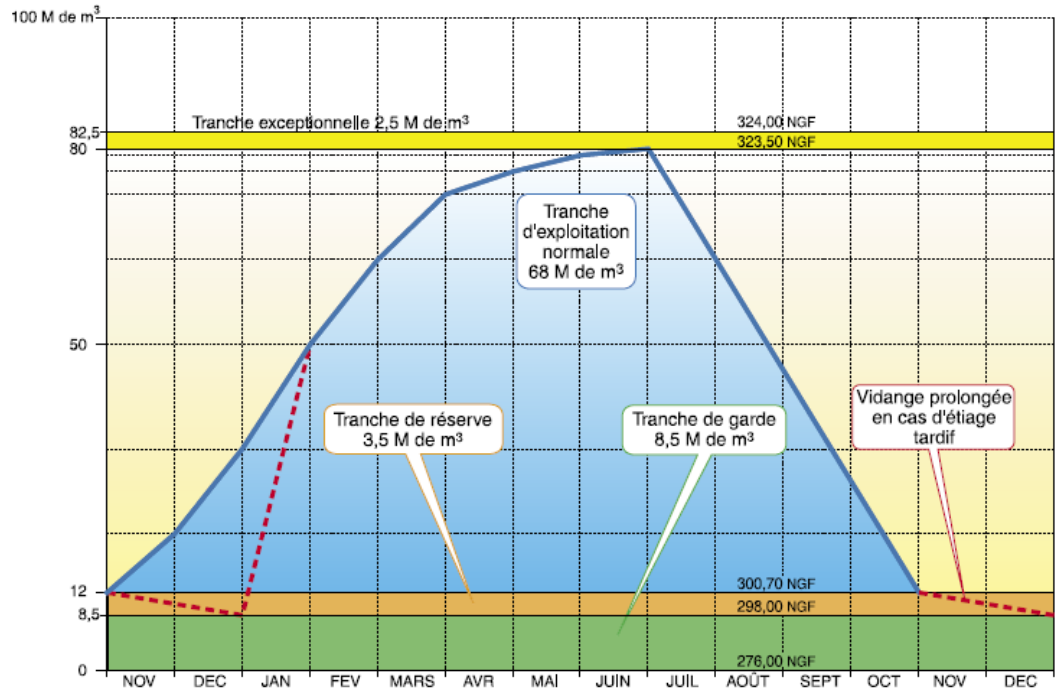
- **Seine**



➢ **Marne**



➢ **Pannecière**



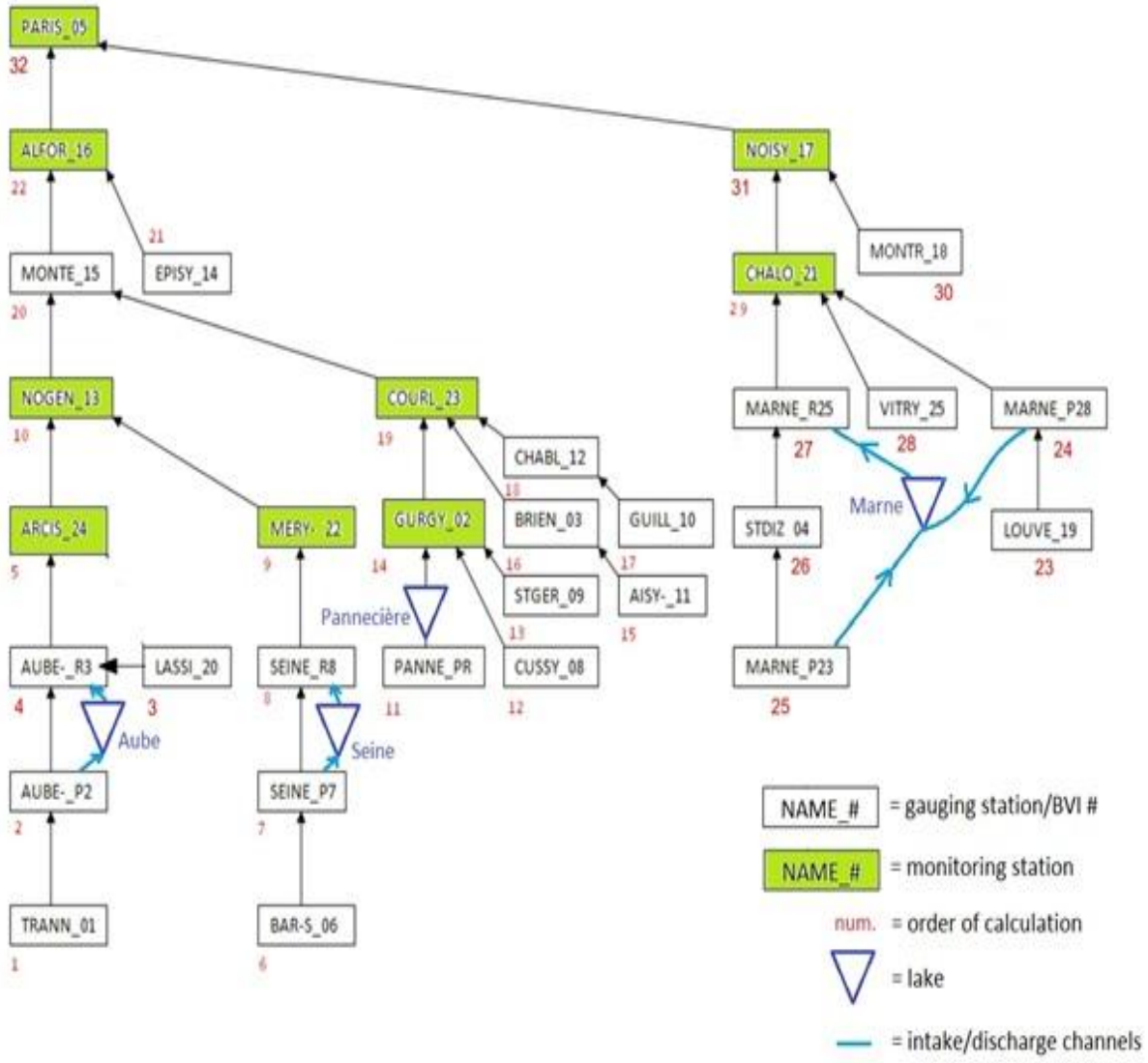
C. Map of the watershed division in BVI and gauging stations

(Source: Carte de situation des lacs-réservoirs. DREAL Champagne-Ardennes/I.I.B.R.B.S.. – Convention d'échanges d'informations relatives au risque d'inondation.)

D. List of the 25 gauging stations

Code_station	Code_HYDRO	Area (km²)	Location
TRANN_01	N/A	1557	L'Aube à Trannes
GURGY_02	H2221010	3819	L'Yonne à Gurgy
BRIEN_03	H2482010	2979	L'Armançon à Brienon-sur-Armançon
STDIZ_04	H5071010	2347	La Marne à Saint-Dizier
PARIS_05	H5920010	43824	La Seine à Paris
BAR-S_06	H0400010	2340	La Seine à Bar-sur-Seine
CHAUM_07	N/A	216	L'Yonne à Chaumard
CUSSY_08	H2172310	247	Le Cousin à Cussy-les-Forges
STGER_09	N/A	402	La Cure à St-Germain
GUILL_10	H2322020	488	Le Serein à Guillon
AISY-_11	H2452020	1349	L'Armançon à Aisy-sur-Armançon
CHABL_12	H2342010	1116	Le Serein à Chablis
NOGEN_13	N/A	9182	La Seine à Nogent-sur-Seine
EPISY_14	H3621010	3916	Le Loing à Épisy
MONTE_15	N/A	21199	La Seine à Montereau
ALFOR_16	H4340020	30784	La Seine à Alfortville
NOISI_17	H5841010	12547	La Marne à Noisiel
MONTR_18	H5752020	1184	Le Grand Morin à Montry
LOUVE_19	H5083050	461	La Blaise à Louvemont
LASSI_20	H1362010	876	La Voire à Lassicourt
CHALO_21	H5201010	6291	La Marne à Châlons-sur-Marne
MERY-_22	H0810010	3899	La Seine à Méry-sur-Seine
COURL_23	H2721010	10687	L'Yonne à Courlon-sur-Yonne
ARCIS_24	H1501010	3594	L'Aube à Arcis-sur-Aube
VITRY_25	H5172010	2109	La Saulx à Vitry-en-Perthois

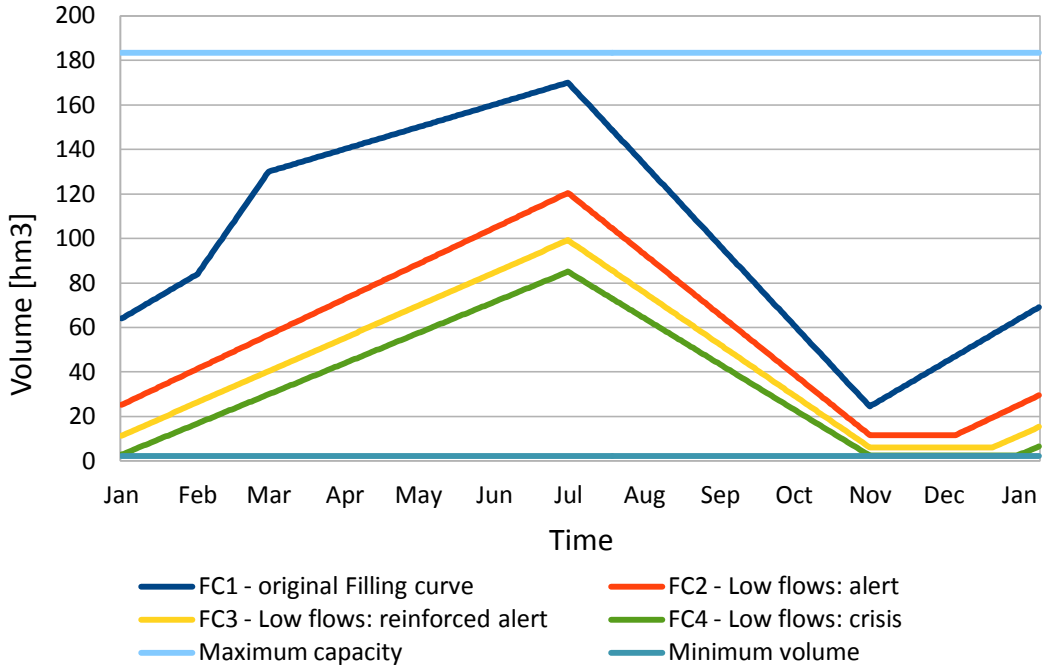
E. Scheme of the controlled system network



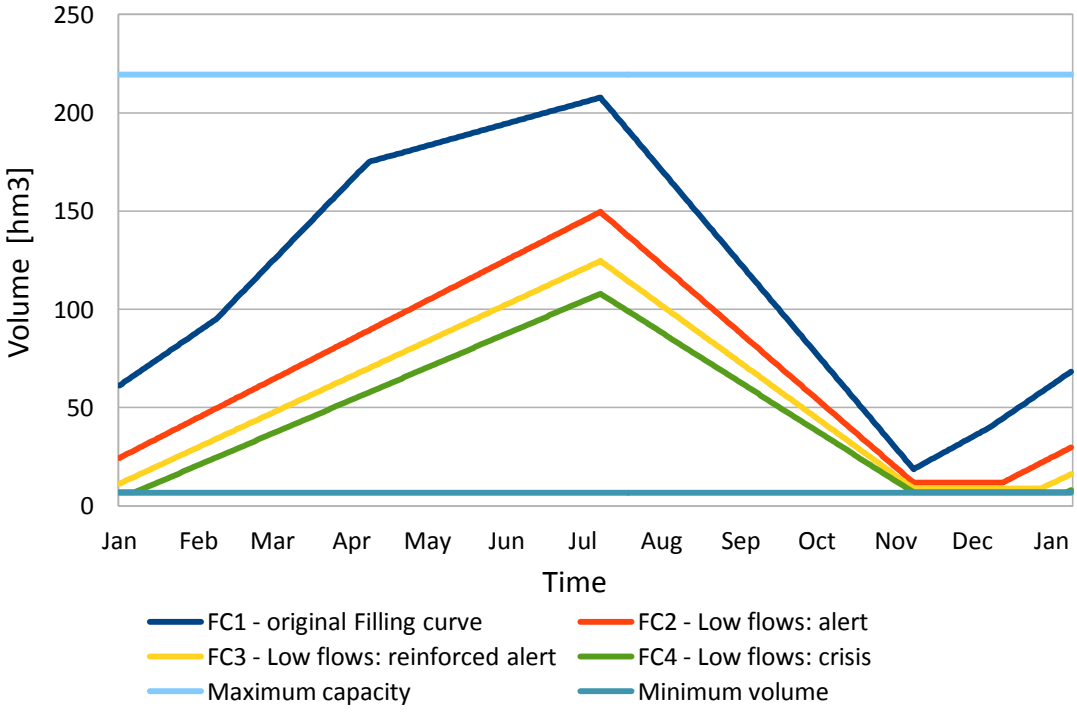
F. New Filling Curves FC1 - FC2 - FC3 - FC4

(FCs for balancing low-flows step-costs and penalty to be under the FC)

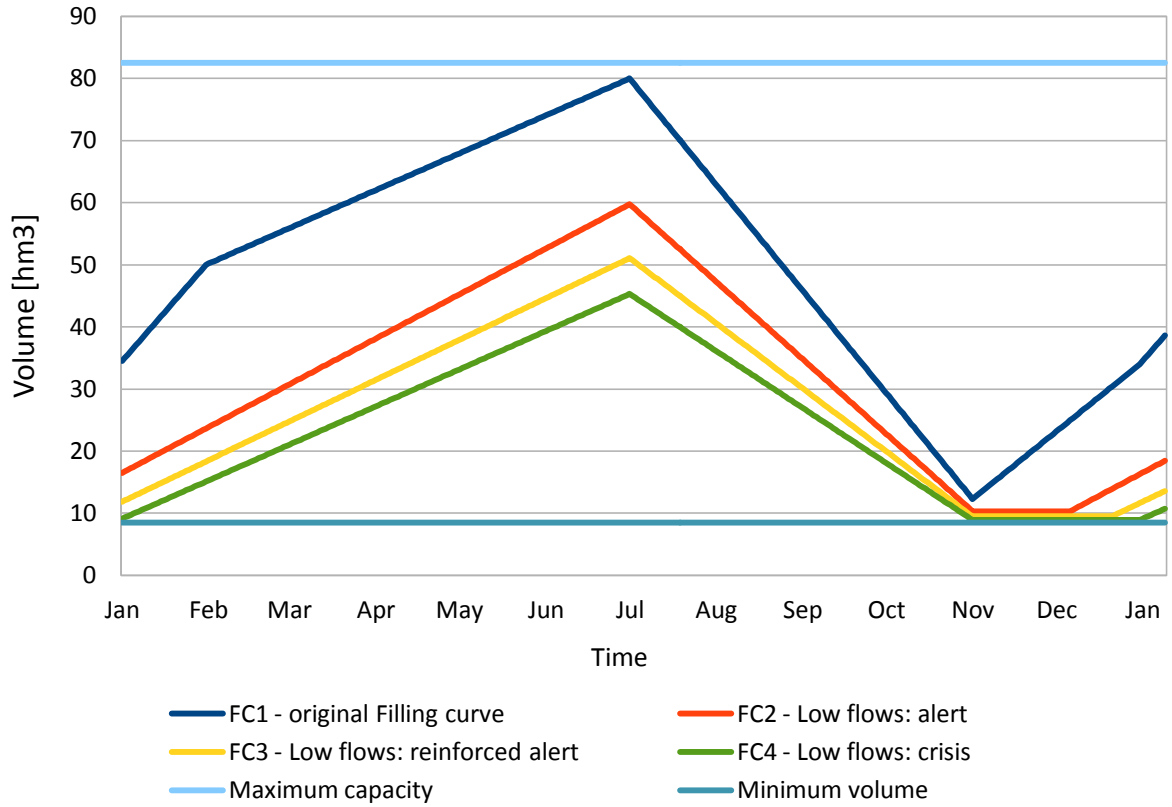
- **Aube**



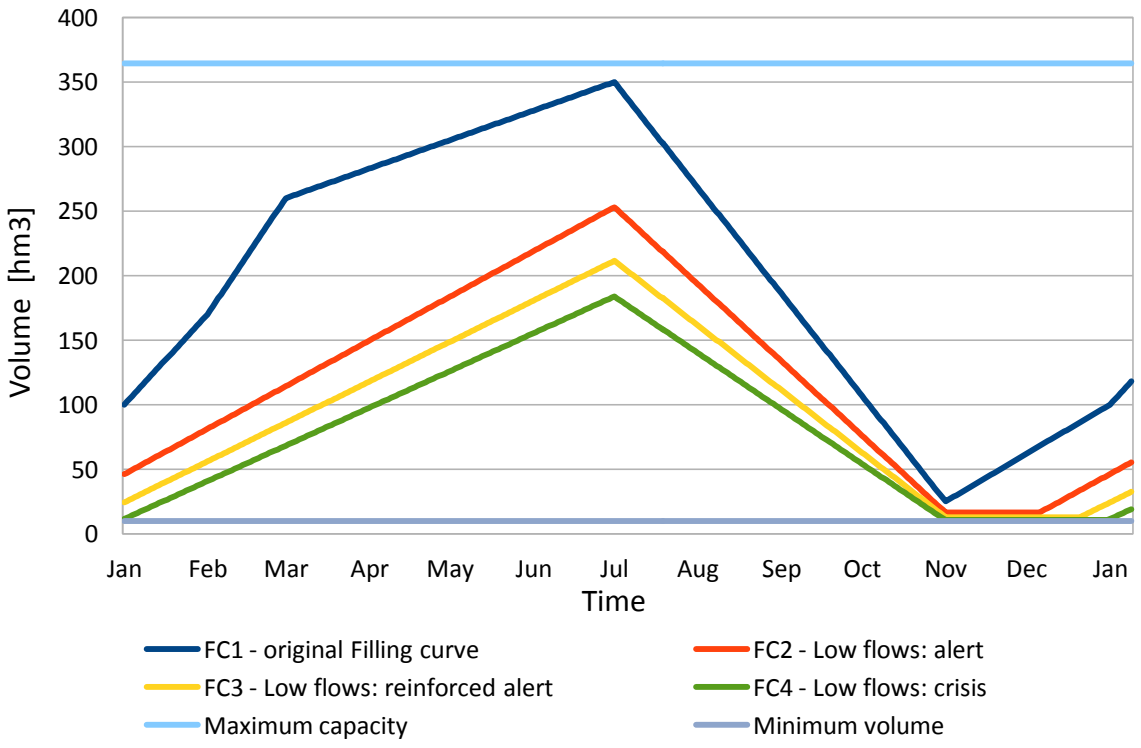
- **Seine**



➢ **Pannecière**



➢ **Marne**



Bibliography

Papers, books and thesis

Ambroise-Rendu, M. (1997), 1910 : La Seine en crue paralyse la capitale (1910: The flood of the Seine River paralyzes the capital city). *La Houille Blanche* (8): 40–44. doi:10.1051/lhb/1997075

Baume, J. P., Malaterre, P.-O., Belaud, G., and Le Guennec, B. (2005), SIC: a 1D Hydrodynamic Model for River and Irrigation Canal Modeling and Regulation. *Métodos Numéricos Em Recursos Hidricos,* 1 – 81.

Bellman, R. E. (1957), *Dynamic Programming*. Princeton University Press, Princeton, New Jersey.

Bertsekas, D. (1995), *Dynamic Programming and Optimal Control*. Athena Scientific, Boston, Massachusetts.

Bertsekas, D. P., and Tsitsiklis, J. N. (1996), *Neuro-Dynamic Programming*. Athena Scientific, Boston, Massachusetts.

Birge, J. R., and Louveaux, F. V. (1997), *Introduction to Stochastic Programming*. Springer Verlag, New York, NY.

Boè, J., Terray, L., Habets, F., and Martin, E. (2007), Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*. 27(12): 1643–1655. doi:10.1002/joc

Box, M. J. (1965), A new method of constrained optimization and a comparison with other methods. *The computer journal, 8(1), 42-52*. doi: 10.1093/comjnl/8.1.42

Castelletti, A., Galelli, S., Restelli, M., and Soncini-Sessa, R. (2010), Tree-based reinforcement learning for optimal water reservoir operation. *Water Resources Research*, *46*(9). doi:10.1029/2009WR008898

Dehay, F. (2012), *Etude de l'impact du changement climatique sur la gestion des lacs-réservoirs de la Seine (Study of the impact of climate change on the management of the reservoirs of the Seine River)*. Master thesis, ENGEES, Strasbourg, France.

Dhouioui, H. (2011), *Building multi-scenario forecasts from deterministic forecasts*. Master thesis, Paris VI, Paris, France.

Dorchies, D., Thirel, G., Jay-Allemand, M., Chauveau, M., Dehay, F., Bourgin, P.-Y., Perrin, C., Jost, C., Rizzoli, J.-L., Demerliac, S., Thépot, R. (2013), Climate change impacts on multi-objective reservoir management: case study on the Seine River basin, France. *The International Journal of River Basin Management, In press.*

Ducharne, A., Baubion, C., Beaudoin, N., Benoit, M., Billen, G., Brisson, N., Garnier, J., Kieken, H., Lebonvallet, S., Ledoux, E., Mary, B., Mignolet, C., Lice, X., Sauboua, E., Schott, C., Thery, S., Viennot, P. (2007), Long term prospective of the Seine River system: confronting climatic and direct anthropogenic changes. *The Science of the total environment*, *375*(1-3), 292–311. doi:10.1016/j.scitotenv.2006.12.011

Dupacova, J., Growe-Kuska, R., and Romisch, W. (2003), Scenario Reduction in Stochastic Programming: An Approach Using Probability Metrics. *Mathematical programming*, *95*(3), 493–511.

Growe-Kuska, N., Heitsch, H., and Romisch, W. (2003), Scenario Reduction and Scenario Tree Construction for Power Management Problems. In *Power Tech Conference Proceedings* 2003. *IEEE: Bologna*, vol 3; 7 pp.

Hashimoto, T., Stedinger, J. R., and Loucks, D. P. (1982), Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resources Research*, *18*(1), 14–20. doi:10.1029/WR018i001p00014

Kjeldsen, T. R., and Rosbjerg, D. (2004), Choice of reliability, resilience and vulnerability estimators for risk assessments of water resources systems. *Hydrological Sciences Journal*, *49*(5), 755–767. doi:10.1623/hysj.49.5.755.55136

Maciejowski, J. (2002), *Predictive control with constraints*. Pearson Education, Harlow, England.

Mayne, D. Q., Rawlings, J. B., Rao, C. V, and Scokaert, P. O. M. (2000), Constrained model predictive control : Stability and optimality. *Automatica*, *36:* 789-814.

Munier, S. (2009), *Modélisation intégrée des écoulements pour la gestion en temps réel d'un bassin versant anthropisé (Integrated flow modeling for real-time management of anthropized watershed.)*. PhD thesis, AgroParisTech, Paris, France.

Nelder, J., Mead, R. (1965), A simplex method for function minimization. *The computer journal*, *7*, 308–313.

Perrin, C., Michel, C., and Andréassian, V. (2003), Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, *279*(1-4): 275–289. doi:10.1016/S0022-1694(03)00225-7

Pianosi, F., and Soncini-Sessa, R. (2009), Real-time management of a multipurpose water reservoir with a heteroscedastic inflow model. *Water Resources Res.*, *45*(10), W10430. doi:10.1029/2008WR007335

Raso, L., Giesen, N. Van De, Stive, P., Schwanenberg, D., and Overloop, P. J. Van. (2012), Tree structure generation from ensemble forecasts for real time control. *Hydrological Processes*. doi:10.1002/hyp

Raso, L., Schwanenberg, D., van Overloop, P.J., and van de Giesen, N. (2013), Short-term Management of Water Systems using Ensemble Forecasts in Tree-Based Model Predictive Control. *Submitted to Advances in Water Resources*.

Soncini-Sessa, R., Weber, E., and Castelletti, A. (2007), *Integrated and participatory water resources management-theory*. Vol. 1., Elsevier Ed., Amsterdam.

Stive, P. M. (2011), *Performance assessment of tree-based model predictive control*. Master thesis, Delft University of Technology, Delft, The Netherlands.

Van Overloop, P.-J. (2006), *Model Predictive Control on Open Water Systems*. PhD thesis. Deft University of Technology, IOS Press, Delft, The Netherlands.

Van Overloop, P.-J., Weijs, S., and Dijkstra, S. (2008), Multiple Model Predictive Control on a drainage canal system. *Control Engineering Practice*, *16*(5), 531–540. doi:10.1016/j.conengprac.2007.06.002

Other documents

Baudin, M., and Couvert, V. (2011), *Optimization with Scilab*. Presentation in *ScilabTEC 2011, Technical Workshops.*

Bemporad, A., Morari, M., and Ricker, N. L. (2012), *Model Predictive Control Toolbox ™ User's Guide*. The Math Works, Inc., Natick, MA.

Carte de situation des lacs-réservoirs. DREAL Champagne-Ardennes/I.I.B.R.B.S.. – Convention d'échanges d'informations relatives au risque d'inondation.

Help documentation for Nelder-Mead optimization. Scilab 5.3.3.

Landrin, S. (2013, March 6), Comment Paris se prépare à la crue du siècle (How Paris is preparing to the flood of the century). *Le Monde*, France.

Websites

ClimAware - http://www.uni-kassel.de/fb14/wasserbau/CLIMAWARE/home/home.html

ECMWF - http://www.ecmwf.int

Seine Grand Lacs - http://www.seinegrandslacs.fr/