Model predictive approaches for real-time reservoir flood control under uncertainty a case study from South Korea

Koo, J.H.

DOI
10.4233/uuid:3b96e105-3643-4eae-9104-8f579d614dc6

Publication date
2025

Document Version
Final published version

Citation (APA)
Koo, J. H. (2025). *Model predictive approaches for real-time reservoir flood control under uncertainty: a case study from South Korea*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:3b96e105-3643-4eae-9104-8f579d614dc6

Ja-Ho Koo

Model Predictive Approaches for Real-Time Reservoir Flood Control Under Uncertainty: A Case Study from South Korea

MODEL PREDICTIVE APPROACHES FOR REAL-TIME RESERVOIR FLOOD CONTROL UNDER UNCERTAINTY: A CASE STUDY FROM SOUTH KOREA

Ja-Ho Koo

The cover illustration was generated using AI (DALL·E model via OpenAI)

MODEL PREDICTIVE APPROACHES FOR REAL-TIME RESERVOIR FLOOD CONTROL UNDER UNCERTAINTY: A CASE STUDY FROM SOUTH KOREA

DISSERTATION

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
and
in fulfilment of the requirement of the Rector of IHE Delft
Institute for Water Education, Prof.dr. E.J. Moors,
to be defended in public on
Monday, 3 November 2025 at 15:00 hours

by

Ja-Ho KOO

This dissertation has been approved by the (co)promotors.

Composition of the doctoral committee:

Rector Magnificus TU Delft	chairman
Rector IHE Delft	vice-chairman
Em.prof.dr. D.P. Solomatine	TU Delft/IHE Delft, promotor
Dr. E. Abraham	TU Delft, promotor
Dr. A. Jonoski	IHE Delft, copromotor

Independent members:

Prof.dr.ir. N.C. van de Giesen	TU Delft
Prof.dr. Z. Kapelan	TU Delft
Dr. S. Shariat Torbaghan	Wageningen University & Research
Prof.dr. M.D.C.M. de Oliveira Cunha	University of Coimbra, Portugal
Prof.dr.ir. B.H.K. De Schutter	TU Delft, reserve member

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE)

Published by IHE Delft Institute for Water Education
www.un-ihe.org
ISBN 978-90-73445-74-1

Acknowledgements

My PhD journey has been an amazing as well as stressful experience at the same time. One day, I felt that I could contribute to our world, but the next day, I was frustrated due to the feeling of ignorance and doubt. However, I could not have completed my PhD without the help and support of many important people in my life.

First of all, I would like to thank my supervisory team. I was incredibly fortunate to receive your support and guidance. At the beginning of my PhD journey, I had little idea of what was expected or how to proceed. Although the path was certainly challenging, your insightful discussions, encouragement, and collaborative exploration of research topics made it deeply rewarding.

To my promoter, Prof. Dimitri P. Solomatine, thank you for your consistent guidance and encouragement throughout this process. Your kind greetings and cheerful sense of humor were as admirable as your passion for research.

To my co-promoter, Prof. Andreja Jonoski, I am sincerely grateful for your invaluable feedback and for always being available when I needed support. Not only for your guidance in the research itself, but also for your kind support whenever I struggled with English, which helped me stay encouraged and never lose confidence.

To my promoter, Prof. Edo Abraham, your insightful guidance and thoughtful advice significantly broadened my perspective, introducing me to a completely new research field and enhancing my overall understanding of research methodologies. I am also deeply grateful for the kindness and patience you consistently showed whenever I hit a wall. The discussions we shared are among the most valuable and pleasant memories from my PhD journey. The guidance from all three of you has been key to my growth as a researcher.

I want to give special thanks to Ms. Kerry, my English teacher, who always believed I could successfully finish. Your words, "You can do it and you should do it now," helped me through the hardest times. Your teaching not only improved my English but also gave me the confidence to communicate with supervisors as well as colleagues.

My warm thanks go to Ms. Elke and Prof.dr. Geert Jan, who were more than just English teachers and neighbors. You helped me get familiar with life in the Netherlands and made me feel at home. The coffee and cookies we shared, our talks, and your help made my time there much easier and richer.

I also want to thank Dr.Ali for all the help with my second paper. Your helpful feedback and willingness to share your time made my work much better. I got more familiar with researching after our long discussion.

I am deeply thankful to Dr.Kim, who has been both a colleague and a fellow PhD student, and to his family. Having you on this journey made it less lonely, and our shared experiences created a special friendship. Our many talks and the way

we helped each other solve problems made both my research and my personal life better. Thank you for being such a good friend and inspiration.

I also want to thank my other colleagues and friends who shared this path with me. Our talks, work together, and support have added a lot to this work and to my growth as a person.

To my parents, I owe immeasurable gratitude. Your unconditional love, unwavering support, and prayers were a source of strength that carried me forward. This accomplishment is as much yours as it is mine, as I would not be who I am today without your guidance and love.

To my family, thank you so much for your love and support. You have kept me going through this hard journey, believing in me even when I had doubts. Having you there made me feel safe and comforted. Your presence gave me peace of mind and emotional strength when I needed it most. My beloved sons, Bonhyeong and Bonjun Koo, you were the reason why I decided to pursue a PhD in the Netherlands and why I made up my mind when I felt frustrated.

My wife, Gyeongah Cheon, even though we were having a hard time settling into new circumstances, I always felt you were with me. Thank you for understanding how much time and work this took and for the sacrifices you made so I could follow this journey. Finally, we dream the same future now.

Finally, I am grateful to IHE Delft and TU Delft for providing the support, resources, and good working environment for my research. In addition, I thank Korea Water Resources Public Corporation (K-water) for the sponsorship and the ICT Cooperative of Dutch Education and Research Institutions (SURF) for allowing us to utilize the Dutch national e-infrastructure with the support of the SURF Cooperative.

This achievement belongs to all of you who have been part of my journey. Thank you.

Summary

Climate change is making flood prediction and management increasingly important. In 2020, South Korea suffered severe flood damage in areas downstream of multipurpose reservoirs, primarily due to unexpected heavy rainfalls occurring in quick succession. All multipurpose reservoirs in Korea are managed by K-water, which follows an operating procedure that analyzes scenarios using real-time observation data and weather forecasts. Although there are existing operating rules, these rules are often based on common-sense practices that can be bypassed in critical situations, allowing decision-makers to adjust reservoir operations flexibly based on Flood Water Level (FWL), Flood season Restricted Water Level (FRWL), and Normal High Water Level (NHWL). This method works well when flood predictions are accurate or when experienced decision-makers are present. However, accurate rainfall prediction remains a challenge. Thus, systematic and model-based approaches are necessary to help determine water release amounts, especially when rainfall forecasts have high uncertainty.

Real-time reservoir flood control has traditionally relied on simulation models, but time constraints often prevent the review of extensive scenarios, potentially missing *optimal and explicitly risk-aware* options. While optimization approaches offer alternatives, practical implementation faces several challenges. First, operational objectives are rarely specified clearly in legal/operational guidelines, and when expressed as nonlinear formulas, the problem becomes computationally intractable. Second, operators' preferences regarding the relative importance of objectives change with flood conditions. Multi-objective optimization approaches that generate Pareto fronts could help visualize the trade-offs between competing objectives, but generating these Pareto sets at each time step requires optimizing the parameters that capture these dynamic preferences and system constraints. To address these challenges, a Model Predictive Control (MPC) framework incorporating practical objectives often overlooked in theoretical studies, such as minimizing the magnitude and frequency of changes in outflow schedules, is presented. We integrate a model-based learning concept for dynamic optimization of weights and parameters, which converts the originally intractable multi-objective nonlinear optimization problem into parameterized linear MPC problems.

However, system nonlinearity combined with these dynamic preferences results in optimization problems that are still computationally expensive and impractical during rapidly evolving flood events. Due to the computational challenges of real-time operation of Parameterised Dynamic Model Predictive Control (PD-MPC), we propose two data-driven approaches: (1) an explicit MPC using deep neural networks to directly determine optimal outflow schedules, and (2) a switched MPC that combines data-driven models that produce optimal weights based on hydrological conditions with linear MPC. Both methods leverage offline learning from the PD-

MPC framework to dramatically reduce computation time from approximately 10 minutes to less than one second, enabling prompt decision-making during rapidly evolving flood events. The explicit MPC demonstrates reliable performance for conditions similar to its training data, while the switched MPC maintains robustness across diverse scenarios due to its receding time horizon optimization process.

Although the above approaches work well as deterministic control tools, more advanced stochastic optimization-based MPC can offer ways to explicitly handle uncertainty. Uncertainty management with stochastic control requires the generation of a sufficiently large number of representative scenarios for inflows. However, traditionally used scenario generation models struggle to capture temporal dependencies or generate scenarios without requiring explicit probability distributions. From this perspective, a Bayesian Neural Network (BNN) model can successfully capture temporal dependencies in inflow time series with high accuracy for short prediction horizons, without requiring explicit probability distributions of variables to be known in advance. Although we can generate a large number of required scenarios with a BNN to assure we sufficiently represent the uncertainty in inflow, it would make stochastic optimization difficult as computational time scales nonlinearly with the number of scenarios. We therefore need scenario reduction approaches that preserve representativeness while reducing the number of scenarios significantly. However, existing scenario reduction approaches themselves lack appropriate distance measures optimally suited for hydrological applications. Therefore, this thesis investigates four distance measures with corresponding reduction algorithms, which are widely used in references: Manhattan with the K-median, Euclidean with the K-mean, Wasserstein with a one-step forward selection, and energy distances with a one-step forward selection algorithm. While the energy distance best preserves statistical characteristics of the original scenario set, the Euclidean distances have significantly lower computational costs. Additionally, the Manhattan and Euclidean distances retain extreme scenarios, which is crucial for flood control, in terms of a tailored performance measure that represents the size of the envelope of a scenario set (using l_1-norm), ensuring the reduced sets retain the range of maximum and minimum flows of the original scenarios.

Finally, this thesis combines these advanced scenario reduction methods to define a risk-constrained MPC problem using Conditional Value-at-Risk (CVaR) to reflect changes in operator risk-averseness by changing a confidence level. Traditional chance constraints in stochastic MPC only consider exceedance probability, while CVaR, which quantifies the expectation of exceedance, remains underutilized in reservoir flood control despite its proven value in risk minimization. By incorporating CVaR as soft constraints, operators can specify risk thresholds that reflect practical considerations rather than relying solely on physical limits that are rarely activated during typical flood events. A stochastic MPC with CVaR outperforms a deterministic counterpart in terms of reflecting the operator's risk-averseness and robustness to inflow uncertainty. Moreover, scenario reduction based on the Euclidean distance is more effective than energy distance-based reduction for real-time flood control applications, considering both closed-loop performance and computational

efficiency. Each proposed framework is validated through numerical experiments for the Geum River and Daecheong Reservoir in South Korea. Intractability in a nonlinear multi-objective optimization problem due to dynamic preferences can be effectively addressed by a PD-MPC framework and its explicit and switched extension based on data-driven models. In addition, stochastic MPC with CVaR incorporating scenario generation and reduction proves beneficial for managing hydrological uncertainty and operators' risk-averseness. We believe the proposed approaches offer sufficient flexibility to accommodate region-specific constraints and objectives, suggesting their potential utility for addressing water resource management challenges in diverse geographical contexts. We anticipate this research will contribute to expanding the application possibilities of real-time optimal reservoir flood control and lay the foundation for practical implementation. The presented methodology is not intended to replace manual operation but rather to provide tools for reducing operator stress in critical situations and ultimately enhancing decision-making capabilities.

Samenvatting

Door de klimaatverandering worden de voorspelling en het beheer van overstromingen steeds belangrijker. Zuid-Korea leed in 2020 ernstige overstromingsschade in gebieden stroomafwaarts van multifunctionele stuwmeren, voornamelijk als gevolg van onverwachte zware regenval die elkaar snel opvolgde. K-water beheert alle multifunctionele stuwmeren in Korea met behulp van een operationele procedure waarbij scenarios worden geanalyseerd aan de hand van real-time observatiegegevens en weersvoorspellingen. Hoewel er operationele regels bestaan, zijn deze regels vaak gebaseerd op gezond verstand en kunnen ze in kritieke situaties worden omzeild , waardoor besluitvormers het beheer van het stuwmeer flexibel kunnen aanpassen op basis van het overstromingswaterpeil (Flood Water Level, FWL), het overstromingsseizoenbeperkte waterpeil (Flood Season Restricted Water Level, FRWL) en het normale hoogwaterpeil (Normal High Water Level, NHWL). Deze methode werkt goed als de overstromingsvoorspellingen nauwkeurig zijn of als er ervaren besluitvormers aanwezig zijn. Een nauwkeurige voorspelling van regenval blijft echter een uitdaging. Daarom zijn systematische en modelgebaseerde benaderingen nodig om te helpen bij het bepalen van de hoeveel water die zal worden vrij gelaten uit het stuwmeer, vooral wanneer de neerslagvoorspellingen een hoge onzekerheid hebben.

Real-time overstromingsbeheer van stuwmeren is van oudsher gebaseerd op simulatiemodellen, maar door tijdsdruk kunnen vaak geen uitgebreide scenarios worden bekeken, waardoor mogelijk optimale en expliciet risicobewuste opties ontbreken. Hoewel optimalisatiebenaderingen alternatieven bieden, wordt de praktische implementatie geconfronteerd met verschillende uitdagingen. Ten eerste worden operationele doelstellingen zelden duidelijk gespecificeerd in wettelijke of operationele richtlijnen en als ze worden uitgedrukt als niet-lineaire formules, wordt het probleem rekenkundig onuitvoerbaar. Ten tweede veranderen de voorkeuren van de beheerders met betrekking tot het relatieve belang van doelstellingen naarmate de omstandigheden van de overstroming veranderen. Multi-objectieve optimalisatiebenaderingen die Pareto fronten genereren zouden kunnen helpen bij het visualiseren van de afwegingen tussen conflicterende doelstellingen, maar het genereren van deze Pareto sets voor elke tijdstap vereist het optimaliseren van de parameters die deze dynamische voorkeuren en systeembeperkingen weergeven. Om deze uitdagingen aan te gaan, wordt een MPC-kader (Model Predictive Control) gepresenteerd waarin praktische doelstellingen zijn opgenomen die vaak over het hoofd worden gezien in theoretische studies, zoals het minimaliseren van de omvang en frequentie van wijzigingen in het schema voor het vrijlaten van water uit het stuwmeer. Wij integreren een modelgebaseerd leerconcept voor dynamische optimalisatie van gewichten en parameters, waardoor het oorspronkelijk onuitvoerbare multi-objectieve niet-lineaire optimalisatieprobleem wordt omgezet in geparametriseerde lineaire MPC-problemen.

De niet-lineariteit van het systeem in combinatie met deze dynamische voorkeu-

ren resulteert echter in optimalisatieproblemen die nog steeds rekenkundig duur en onpraktisch zijn tijdens snel evoluerende overstromingen. Om de rekenkundige uitdagingen van de real-time werking van de Parameterized Dynamic Model Predictive Control (PD-MPC) het hoofd te bieden, stellen we twee datagestuurde benaderingen voor: (1) een expliciete MPC die diepe neurale netwerken gebruikt om direct het optimale schema voor het vrijlaten van water uit het stuwmeer te bepalen, en (2) een geschakelde MPC die datagestuurde modellen combineert om optimale gewichten te produceren op basis van hydrologische omstandigheden met lineaire MPC. Beide methoden maken gebruik van offline leren van het PD-MPC raamwerk om de rekentijd drastisch terug te brengen van ongeveer 10 minuten naar minder dan een seconde, waardoor snelle besluitvorming tijdens snel veranderende overstromingen mogelijk wordt. De expliciete MPC laat betrouwbare prestaties zien voor omstandigheden die vergelijkbaar zijn met de trainingsgegevens, terwijl de geschakelde MPC robuust blijft voor verschillende scenarios dankzij het optimalisatieproces met een afnemende tijds-horizon.

Hoewel bovenstaande benaderingen goed werken als deterministische besturingsinstrumenten, kan een meer geavanceerde MPC op basis van stochastische optimalisatie manieren bieden om expliciet met onzekerheid om te gaan. Het managen van de onzekerheid met de stochastische besturing vereist het genereren van een voldoende groot aantal representatieve scenarios voor de instroom naar het stuwmeer. Traditioneel gebruikte modellen voor het genereren van scenarios hebben echter moeite om temporele afhankelijkheden vast te leggen of om scenarios te genereren zonder expliciete kansverdelingen. Vanuit dit perspectief kan een Bayesiaans Neuraal Netwerk (BNN) model met succes temporele afhankelijkheden in instroomtijdreeksen met hoge nauwkeurigheid voor korte voorspellingshorizons vastleggen, zonder dat expliciete kansverdelingen van variabelen vooraf bekend hoeven te zijn. Hoewel een BNN een groot aantal vereiste scenarios kan genereren om er zeker van te zijn dat de onzekerheid in de instroom voldoende vertegenwoordigd is, zou dit stochastische optimalisatie bemoeilijken omdat de rekentijd niet-lineair schaalt met het aantal scenarios. Daarom hebben we scenarioreductiebenaderingen nodig die de onzekerheid in de instroom voldoende representeren. Bestaande benaderingen voor het reduceren van scenarios missen echter geschikte metrieken die optimaal geschikt zijn voor hydrologische toepassingen. Daarom worden er in dit proefschrift vier metrieken met bijbehorende reductiealgoritmen onderzocht, die veel in referenties worden gebruikt: Manhattan met de K-mediaan, Euclidean met de K-mean, Wasserstein met een eenstaps voorwaartse selectie en energie-metriek met een eenstaps voorwaarts selectiealgoritme. Terwijl de energie-metriek de statistische kenmerken van de oorspronkelijke scenarioset het best bewaart, hebben de Euclidische metrieken aanzienlijk lagere rekenkosten. Bovendien behouden de Manhattan- en Euclidische metrieken de extreme scenarios, wat cruciaal is voor overstromingsbeheer, wat inhoud dat een op maat gemaakte metriek het volledige bereik van de scenarioset weergeeft (met behulp van de l_1-norm), zodat de gereduceerde scenarios het bereik van maximale en minimale debieten van de oorspronkelijke scenarios behouden.

Tot slot combineert dit proefschrift deze geavanceerde methoden voor scena-

rioreductie om een risico-beperkt MPC-probleem te definiëren dat gebruik maakt van Conditional Value-at-Risk (CVaR) om veranderingen in de risicomijding van de beheerder weer te geven door het aanpassen van het betrouwbaarheidsniveau. Traditionele kansbeperkingen in stochastische MPC houden alleen rekening met overschrijdingskansen, terwijl CVaR, dat de overschrijdingsverwachting kwantificeert, onderbenut blijft voor overstromingsbeheer van stuwmeren ondanks zijn bewezen waarde voor risicominimalisatie. Door CVaR als zachte beperkingen op te nemen, kunnen beheerders risicodrempels specificeren die praktische overwegingen weerspiegelen in plaats van alleen te vertrouwen op fysieke limieten die zelden worden geactiveerd tijdens typische overstromingsgebeurtenissen. Een stochastische MPC met CVaR presteert beter dan een deterministische tegenhanger wat betreft het weerspiegelen van de risicomijdendheid van de beheerder en de robuustheid voor instroomonzekerheid. Op basis van de gesloten-loopprestaties als computerefficiëntie is de scenarioreductie gebruikmakend van de Euclidische metriek effectiever dan reductie op basis van de energie-metriek voor real-time overstromingsbeheertoepassingen.

Elk voorgesteld raamwerk wordt gevalideerd door middel van numerieke experimenten voor de rivier de Geum en het Daecheong Stuwmeer in Zuid-Korea. Een PD-MPC raamwerk in samenwerking met de expliciete en geschakelde uitbreiding op basis van datagestuurde modellen kan effectief de onuitvoerbaarheid van een niet-lineair multi-objectief optimalisatieprobleem als gevolg van dynamische voorkeuren benaderen. Bovendien blijkt stochastische MPC met CVaR, waarin scenarios worden gegenereerd en gereduceerd, gunstig te zijn voor het managen van hydrologische onzekerheid en risicomijdend gedrag van de beheerders. We zijn van mening dat de voorgestelde benaderingen voldoende flexibiliteit bieden om rekening te houden met regiospecifieke beperkingen en doelstellingen, wat suggereert dat ze nuttig kunnen zijn voor het aanpakken van uitdagingen op het gebied van waterbeheer in verschillende geografische contexten. We verwachten dat dit onderzoek zal bijdragen aan het uitbreiden van de toepassingsmogelijkheden van real-time optimale overstromingsbeheer van stuwmeren en de basis zal leggen voor praktische implementatie. De gepresenteerde methodologie is niet bedoeld om handmatige bediening te vervangen, maar om hulpmiddelen te bieden die de stress voor de beheerder in kritieke situaties verminderen en uiteindelijk de besluitvorming verbeteren.

Contents

1 Introduction

1.1 Motivation

Flood forecasting and overall flood risk management are becoming increasingly important in many countries, especially due to climate change, which is expected to increase such risks. In many locations, flood events are increasing in both their frequency and their severity. This is the case in Korea. In 2020, Korea experienced an unexpected severe flood. In fact, since 2000, when the construction of 16 multi-purpose dams that included flood protection storage was finalized, there has been little flood damage in the downstream areas of these dams. Flooding mainly occurred in small rivers or places where embankments had not been built yet. The main cause of such floods was regarded as a problem of failure of river maintenance, such as the construction or upgrade of embankments. Under these circumstances, severe flood impacts on the downstreams of some multi-purpose dams in the summer of 2020 were quite shocking to the reservoir operating organization (Korea Water Resources Corporation, K-water), downstream residents, and relevant experts. The cause of the flood damage actually seems simple – it is unexpectedly heavy rain pouring down upstream and downstream of the reservoirs at the same time. The reservoir operators tried to retain as much water as possible, but it was not enough to prevent flood impacts in the downstream area. Under such conditions, concerns were raised about what was needed to operate the reservoir more effectively under the growing localized and unexpected heavy rainfall, leading to the proposal of this research.

All multi-purpose dams (reservoirs) in Korea are run by K-water. In particular, flood control of multi-purpose reservoirs is being undertaken by K-water on behalf of the government. The operating procedure of multi-purpose reservoirs in Korea is as follows: At first, reservoir operators check real-time observation data and make a few scenarios based on rainfall forecasts published by the Korea Meteorological Administration (KMA). Analyzing multiple scenarios during a flood event is challenging; therefore, operators try to reduce the number of scenarios and select the most appropriate ones. In the next step, using a Decision Support System (DSS) including hydrological and hydraulic simulation models, operators analyse the scen-

arios and determine how much and when water will be released. Actually, operators only suggest the outflow schedule because it should be approved by the government, which is the Flood Control Office (FCO). However, the government usually approves the suggested schedules, otherwise gives some opinion and discusses the suggestion in consideration of the upstream and downstream situation. Moreover, the modification of outflow schedules within the approved amount and time is solely the responsibility of the operators. Therefore, the suggestion by K-water could be considered as *the decision*. Based on this approved decision, operators inspect the upstream and downstream rivers which could be affected by the outflow and give notice to the related local governments in advance. This is needed because local governments are responsible for the evacuation of residents, visitors and other related actions.

However, a decision made based on an observed hydrological condition at a certain time cannot be consistently true. There will be some differences between expected hydrological conditions and real conditions observed after a while due to the uncertainty of rainfall or inflow forecasts as well as the discrepancy between real hydrological systems and modelled ones. In addition, because forecasts are updated continuously, incorporating up-to-date forecasts into decision-making is important. Therefore, it is essential to update decisions following changes in hydrological conditions and forecasts. This is why an operator implements only one or two outflows in a decided outflow schedule and then updates the entire outflow schedule at the next time step.

Of course, there are operating rules for flood control for each multi-purpose reservoir in Korea. However, it stipulates only “common-sense" procedures. For example, if flooding is expected, the Reservoir Water Level (RWL) should be below the Flood season Restricted Water Level (FRWL); or, if inflow is within design amount, the release should be within the design outflow; or, an approval should be obtained in advance before starting release via spillways, and similar rules for actions under normal conditions. These rules could be ignored if a dangerous situation is expected. Consequently, decision makers can operate reservoirs very flexibly based on the Flood Water Level (FWL), the FRWL, the Normal High Water Level (NHWL), as reservoir level constraints or guidelines, as illustrated in Fig 1.1. In other words, the outflows (or outflow schedules) of multi-purpose reservoirs during flood events are determined by decision makers rather than by rules. This procedure has a great advantage if floods can be accurately predicted or if there are highly experienced decision makers. When a flood is expected, more storage capacity can be secured compared to design conditions by keeping the RWL below the FRWL in advance. On the other hand, when a small flood is expected at the end of the flood season, water levels may be maintained higher than the FRWL, increasing the chance of securing enough water for water supply and/or hydropower generation.

However, it is impossible to predict rainfall with unerring accuracy, especially in the case of a flash flood. Therefore, many experts have been trying to improve the accuracy of rainfall forecasts (Kim *et al.*, 2021). From the perspective of reservoir operation, the challenge is to identify better forecasts and follow-up procedures to

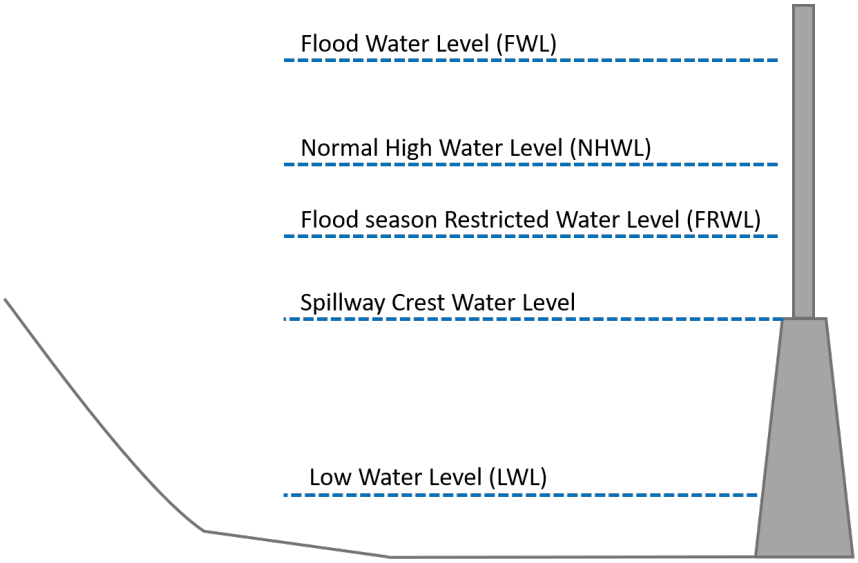


Figure 1.1 | The conceptual diagram of reservoir water levels. A spillway crest water level is generally under the FRWL in South Korea for the purpose of decreasing the water level to FRWL using spillway gates. However, in some cases, spillway crest levels are higher than FRWL. FRWL does not exist in some multipurpose reservoirs in South Korea, e.g., the Daecheong reservoir, with details presented in Chapter 2.

help operators (decision makers) decide how much water should be released, even when exploiting rainfall forecasts with high uncertainty. Personal experience and knowledge are important, but we need a more systematic and scientific way. This is the main motivation for this research.

1.2 Model Predictive Control

Many researchers have suggested optimization-based control methods, which can provide optimal outflows based on specified conditions, because flood control is, as its name implies, fundamentally a control problem (Castelletti *et al.*, 2023). Various optimization-based methodologies have been proposed to address challenges of real-time optimal flood control, particularly subject to uncertainty, either in rainfall and/or inflow forecasts (Jain *et al.*, 2023; Myo Lin *et al.*, 2020; Breckpot *et al.*, 2013a).

The receding horizon Model Predictive Control (MPC) is an optimal control framework that uses a system model to predict and evaluate the system's behaviour over a finite time horizon. The control input, which is the optimal outflow schedule in reservoir flood control, is calculated to optimise objectives and satisfy physical and performance constraints based on the current state and predicted behaviour. One of the two main characteristics of the receding horizon MPC approach is that only the first outflow or few outflows in the schedule are implemented, and the prediction and optimisation processes are repeated at each time step with updated information. This iterative process is generally referred to as receding horizon con-

trol (Van Overloop, 2006), as illustrated in Fig 1.2. In some applications, such an approach is also called a rolling horizon method (Wang *et al.*, 2014). A significant advantage of the receding horizon control is its robustness to disturbances, such as uncertain inflow forecasts and system uncertainty. This robustness stems from re-estimating predictions and recalculating optimal control inputs at each time step based on updated hydrological states, including revised inflow forecasts (Schwenzer *et al.*, 2021).

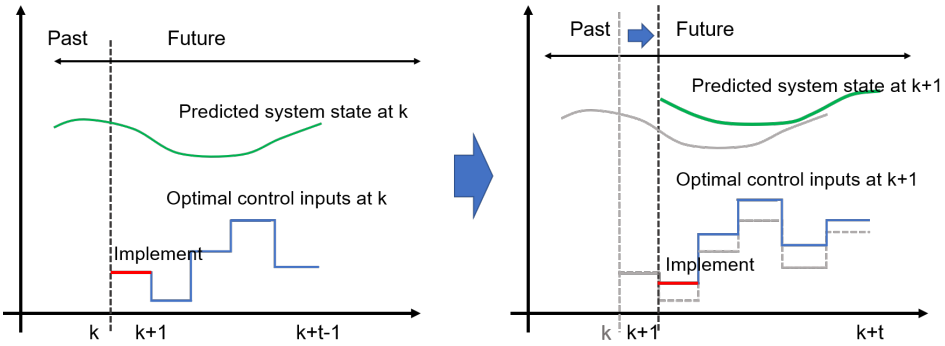


Figure 1.2 | The concept diagram of receding horizon MPC, in which the first one among the optimal control inputs generated at time step k is implemented. Then, the whole sequence of the control inputs is regenerated at $k+1$, reflecting updated states. (Modified from Alasali *et al.* (2020)).

The other characteristic is the combination of optimization with prediction using a system model. One approach to reservoir flood control that relies on optimization is the simulation-based heuristic approach, such as the Genetic Algorithm (GA) to find optimal outflow schedules. An MPC approach, on the other hand, transcribes the system dynamics as constraints and can solve the optimal control problem using mathematical optimization algorithms. Compared to heuristic approaches, an advantage is that such a method can consider physical and performance constraints explicitly. Additionally, mathematical optimization is computationally efficient and can guarantee that solutions are optimal when the problem is convex.

Based on the reservoir flood control procedure described in the previous section, i.e., generating an outflow schedule at each time step, and then updating it to reflect changes in forecasts and observed hydrological conditions, it appears that receding horizon MPC coincides with the practical flood control process.

1.3 Problem Description

1.3.1 Research gaps

Reservoir flood control has predominantly relied on simulation models in practice (Che and Mays, 2015), as illustrated in Fig 1.3. The primary rationale is that simulation models facilitate the incorporation of operators' various considerations, which may also be characterized as objectives or constraints. Through these

models, operators can flexibly evaluate diverse precipitation or inflow scenarios and compare outcomes across each scenario. Scenario-based decision-making enables the selection of alternatives aligned with the operator's preferences. Operators typically feel comfortable and confident with this approach, as their confidence often stems from the clarity of reasoning leading to a particular decision. However, it is almost impossible for operators to examine a large number of release scenarios in practice. In other words, the optimal discharge scenario could often be missed. The primary challenge is that considering additional scenarios requires more time, which can lead to delays in decision-making.

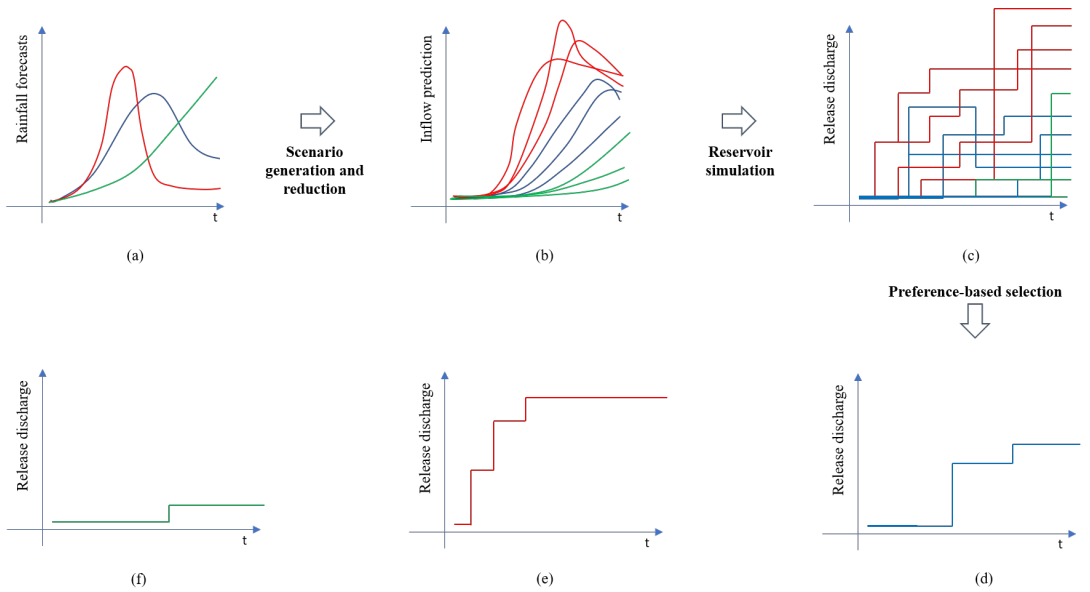


Figure 1.3 | The conceptual diagram of practical simulation-based flood control: (a) only a small number of rainfall forecasts, (b) a rainfall-runoff model produces a number of inflow scenarios based on different soil moisture, evapotranspiration conditions, and so on for each rainfall forecast, (c) an operator can make various potential outflow schedules based on a number of inflow scenarios (but, still only a few compared with all possible cases) and upstream/downstream conditions using a reservoir simulation model (e.g., one focusing on reducing reservoir water level to secure dam safety and another for reducing downstream flood risk). (d), (e), and (f) show examples of flood control decisions based on operator's preferences; (d) an outflow decision for mean inflow and similar risk for up and downstream, (e) a decision corresponding to high inflow scenarios and high risk averseness to downstream flood, (f) an outflow schedule with high risk tolerance.

In addition, quantitative precipitation or inflow forecasts typically cover periods shorter than the entire flood duration. For example, in Korea, no organization issues discharge forecasts for rivers, and the official quantitative precipitation forecasts published by the Korea Meteorological Administration (KMA) are limited to very short-term forecasts covering only the next 6 or 12 hours, updated hourly. Longer-term forecasts are provided solely in terms of precipitation probability or qualitative descriptors, such as "above average." Consequently, operators attempt to extend precipitation forecasts to encompass the entire rainfall event by consulting with meteorological experts and referencing radar imagery and other data sources. Inherent uncertainty remains a characteristic of precipitation forecasts, however.

Therefore, when developing operational scenarios, operators tend to incorporate precautionary margins to address potential emergency situations. For instance, even when sufficient storage capacity exists to retain inflows, increased spillway outflows may be initiated in anticipation of additional precipitation. Moreover, since precipitation forecast uncertainty is rarely quantified in the decision-making process, the determination of this precautionary storage margin relies entirely on the operator's judgment.

While optimization-based control approaches, including Model Predictive Control (MPC), have made substantial contributions in many applications (Schwenzer *et al.*, 2021), their practical application within reservoir flood control remains somewhat limited. This discrepancy can be mainly attributed to the disparities between research assumptions and the pragmatic necessities of real-time reservoir flood control. This can be primarily attributed to the insufficient consideration of the following factors.

1.3.2 Objective functions related to practical reservoir operation

Numerous researchers have proposed various objectives for reservoir operation (e.g., securing enough water to fulfil demands, maximization of hydropower generation, and protecting the ecosystem); however, during flood events, these objectives may not be prioritized. Minimizing outflows to reduce downstream flood risk and maintaining RWL between FWL and FRWL, which are stipulated by regulations mostly, have been the popular objectives for flood control. However, the practical objectives of operators for flood control, e.g., minimizing changes in outflow schedules and minimizing changes in outflows in an outflow schedule, are typically not specified in regulations or guidelines but are considered as tacit 'experience' or 'expert knowledge.' For example, even though an operator changes outflows frequently and urgently, it does not mean that the operator violates any regulations; however, the operator will be under significant pressure caused by many complaints from related organizations and residents, and needs to explain why those decisions are inevitable. Therefore, despite the lack of written regulations about these decisions, they should be considered important, and we can refer to them as practical objectives. Researchers have either paid insufficient attention to defining and incorporating these practical objectives into optimization processes or found them too complex to be included in such processes. Consequently, operators generally hesitate to adopt optimization tools, and researchers have tended to apply only a limited number of objectives or linear/quadratic objectives that are tractable for reservoir flood control optimization problems.

1.3.3 Operators' preferences in a multi-objective optimization problem

Preferences, defined as the selection criteria operators use to choose among control alternatives, are represented by weights and parameters in a multi-objective optimization problem for reservoir flood control. While many studies employing the receding horizon approach utilize weighted multi-objective optimization methods, generating a complete Pareto set at each time step is computationally prohibitive. Furthermore, using fixed weights to represent operators' preferences assumes that the relative importance of objectives remains constant throughout a flood event—an assumption that may not hold in dynamic situations. For instance, when a flood with low peak inflow and lasting only for one day is forecasted, an operator focuses on retaining inflow to maintain RWL within a safe range rather than reducing flood risk. However, if during the flood event, the forecast suddenly changes to predict a huge inflow within a short time and a longer duration, the operator is likely to switch the main point of view to circumvent flood damage.

1.3.4 Computational complexity

The complexity of reservoir flood control stems from system nonlinearity and/or intricate relationships among objectives. When dynamically changing preferences, i.e., reflecting shifting importance among objectives and parameters, are incorporated, computational complexity increases substantially. However, computational efficiency is critical for real-time implementation. While data-driven approaches offer valuable solutions to computational challenges in many fields, few studies have applied these methods to problems with numerous objectives and dynamically changing weights, especially for reservoir flood control. Moreover, dynamic relationships among objectives make it difficult for data-driven models to effectively learn the associations between system states and optimal control inputs or weights.

1.3.5 Uncertainty of reservoir inflow scenarios

Given that only a few covering short-term rainfall forecasts are available, the most effective approach for operators to address uncertainty, particularly when utilizing optimization processes, is to develop multiple quantitative scenarios that explicitly incorporate this uncertainty. Traditionally, statistical models, such as AutoRegressive Moving Average (ARMA) and AutoRegressive Integrated Moving Average (ARIMA), or historical data have been utilized. While data-driven models, such as Deep Neural Network (DNN), Bayesian Neural Network (BNN), and Quantile Regression Deep Neural Network (QRDNN), have been applied for scenario generation across various fields, but with limited application in flood control, they face two key limitations: difficulty in capturing temporal dependencies between variables and a predominantly deterministic operation that requires predefined input uncertainty to generate uncertain scenarios.

Stochastic optimization requires balancing the number of scenarios, i.e., too few inadequately reflect hydrological uncertainty, while too many create computational intractability. Thus, scenario reduction is essential for preserving key properties while improving computational efficiency. Although numerous reduction techniques exist based on various distance measures, limited research has identified which measures are most suitable for hydrological scenario reduction, with verification typically relying on numerical experiments based on stochastic optimization. However, the stochastic optimization results can be affected not only by scenario reduction but also by scenario generation, the formulation of the optimization problem, and so on; these factors comprehensively affect the results, making it difficult to identify which factor causes an improvement. Effective scenario reduction requires evaluation of both how well the reduced set preserves original scenario characteristics and how accurately it represents the specific dynamics of flood control systems.

1.3.6 Stochastic Optimization considering uncertain future inflow scenarios

Applying chance constraints in receding horizon Model Predictive Control (MPC) requires explicit probabilistic characterization of uncertainty, but these constraints enable probabilistic risk management when violations are inevitable by focusing only on exceedance probability rather than excess magnitude. Conditional Value-at-Risk (CVaR), which quantifies expected outcomes in worst-case scenarios, offers advantages through linearization capabilities and compatibility with Sample Average Approximation (SAA) without requiring predefined probability distributions for uncertainty; therefore, scenarios generated with attributed probabilities could be used directly.. However, how to apply CVaR in the reservoir flood control problem has not been addressed in the literature, to the best of the author's knowledge. Moreover, the physical constraints, such as Flood Water Level (FWL) or design discharge capacity, commonly used as constraint or objective thresholds, are designed for extreme conditions like Probable Maximum Flood (PMF), making them necessary but insufficient for typical short-term flood control operations where such limits rarely become active constraints.

1.4 Research Objectives

The general objective of this research is the development of an operator-oriented framework to generate the optimal outflow schedules for real-time reservoir operation during emergencies (flood events), which considers dynamically changing operators' preferences depending on changing hydrological conditions, under the explicit uncertainty.

To achieve the general objective, we formulate the specific objectives as follows:

1. Develop an Model Predictive Control (MPC) framework that incorporates dynamic preferences, integrating practical objectives for real-time flood control

under uncertainty.

2. Evaluate the potential of data-driven models to reduce computational complexity and enhance the practicality of an MPC framework under dynamic preference assumptions.

3. Evaluate the effectiveness of a Bayesian Neural Network (BNN) model for hydrological scenario generation and analyze the comparative advantages and limitations of various distance measures in scenario reduction methodologies.

4. Develop and validate Conditional Value-at-Risk (CVaR)-type soft constraints within a stochastic MPC framework for reservoir flood control to consider probabilistic risk and impact quantification of upstream/downstream flood.

1.5 Research Questions

On the basis of the problem description and research gaps, four research questions can be formulated. All of these derive from one general and important question: how to provide reservoir operators with the optimal control inputs, which is an outflow schedule at each time step, for flood control operation in real time under uncertainty.

1. Do operators' preferences change with hydrological conditions and potential impacts of projected inflows? Would it be appropriate to consider the set of preferences as dynamic?

2. How can we incorporate the dynamic nature of preferences within an Model Predictive Control (MPC) framework while maintaining computational efficiency suitable for real-time operation?

3. Regarding hydrological scenarios, what approaches are most appropriate for scenario generation and reduction, and how can their effectiveness be evaluated?

4. How can inflow and other uncertainties be effectively represented within stochastic MPC approaches where risk can be controlled explicitly? What is the best way to manage flood risk, reflecting an operator's risk tolerance in stochastic MPC?

1.6 Outline

Given the research motivation, problem description with research gaps, research objectives, and research questions already presented, the next six chapters are structured as illustrated in Figure 1.4. Note that each chapter has its own notation for equations.

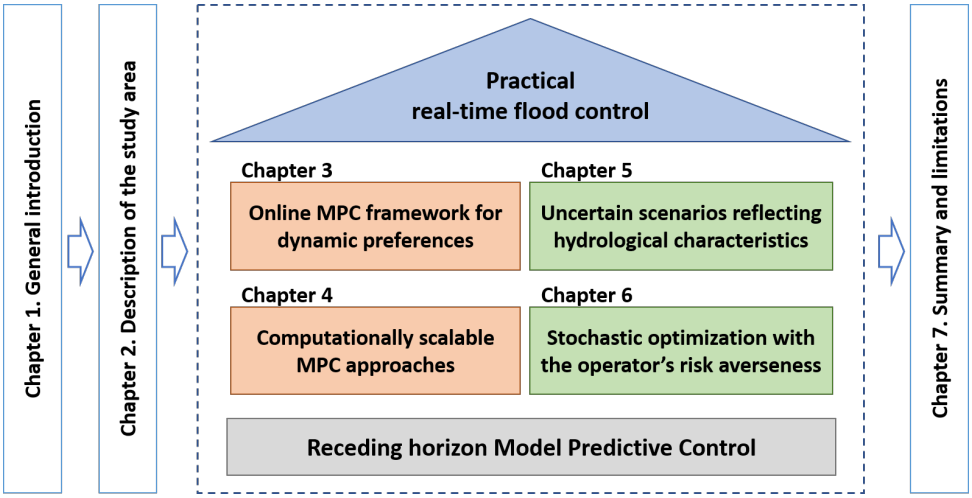


Figure 1.4 | The outline of this thesis and thematic organization of chapters from an introduction to conclusions.

Chapter 2 describes the details of the study area, the Daecheong reservoir and its downstream area in South Korea.

Chapter 3 discusses practical objectives and dynamic preferences. From flood control objectives in references, detailed practical objectives are presented, such as minimizing changes in outflow schedules, which is essential in practical flood control to support many organizations related to the flood control and reduce human errors. Given a number of nonlinear objectives, which makes the flood control optimization problem intractable, a Parameterised Dynamic Model Predictive Control (PD-MPC) framework is proposed, which converts the original intractable problem into parameterized linear MPC problems with dynamic optimization of weights and parameters. This is done by introducing a model-based learning concept. By numerical experiment, the dynamic nature of the operators' preferences as well as the effectiveness of the PD-MPC are demonstrated.

Chapter 4 develops tractable real-time flood control strategies that maintain performance while reducing computational complexity. Even though the PD-MPC is feasible for experimentation in this study, it is computationally very demanding since optimizing weights and parameters online requires a large number of iterations with the GA or other heuristics. To make it more efficient, two data-driven approaches based on MPC are presented: (1) an explicit MPC using deep neural networks to directly determine optimal outflow schedules, and (2) a switched MPC with data-driven models which produce optimal weights of objectives based on hydrological conditions. Both methods leverage offline learning from an online PD-MPC framework incorporating state-dependent weights.

Chapter 5 suggests a probabilistic data-driven model, specifically a BNN, as a scenario generation method to capture the temporal dependence of inflow time series. In addition, the applicability of the four distance measures widely used in the literature, i.e., the Manhattan distance, the Euclidean distance, the Wasserstein distance, and the energy distance, is evaluated from the perspective of computational

complexity, preservation of statistical characteristics, and inclusion of extreme scenarios. To assess the inclusion of extreme scenarios, which are critical in flood control, quantitatively, a tailored performance measure that represents the size of the envelope of a scenario set is proposed, ensuring the reduced sets retain the range of maximum and minimum flows of the original scenario.

Finally, Chapter 6 presents how to manage the reservoir flood (short-term) control risk under the uncertainty of hydrological prediction with a stochastic MPC framework. The CVaR constraints are applied, not as hard constraints but soft ones, in order to reflect the fact that the physical thresholds are only activated in extreme situations like a Probable Maximum Flood (PMF) condition. In addition, scenario generation and reduction methodologies proposed in Chapter 5 are evaluated in terms of stochastic MPC. In this chapter, the MPC estimates the downstream impacts of reservoir outflows explicitly, i.e., a downstream system model to predict water levels at a target point is integrated inside MPC.

2

Description of the Study Area: Daecheong Reservoir and its downstream area

The Daecheong Reservoir (hereafter referred to as DC) is a multipurpose reservoir located in the upper reaches of the Geum River, flowing through the central region of South Korea, as depicted in Fig 2.1.

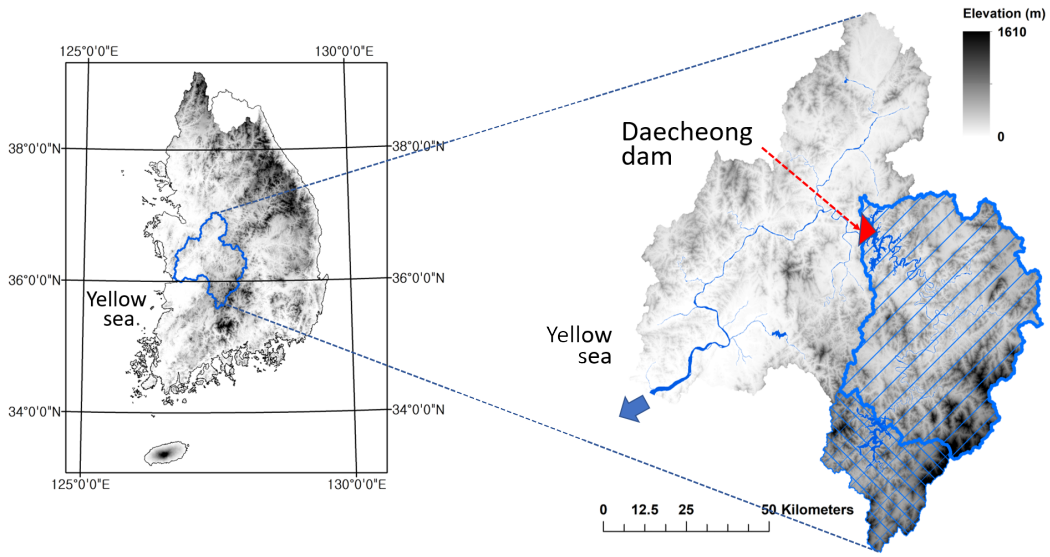


Figure 2.1 | Location of the DC. The left figure shows the location of the Geum river basin, and the right figure illustrates the location of the DC. The base maps are the digital elevation model (DEM) and the river shape file of Korea from the National Geographic Information Institute (https://www.ngii.go.kr) and the basin shape file from VWorld (https://www.vworld.kr) published by the Han River Flood Control Office (https://www.hrfco.go.kr).

The DC serves as a vital water source, delivering 922 000 m^3 of drinking water daily to approximately 2 million residents in surrounding areas, while also supporting irrigation, hydroelectric power generation, and flood control. Geographically, it

features a complex elongated structure with a dendritic drainage pattern, reaching a maximum width of around 1 km. The watershed flowing into DC covers 3204 km^2, and at full capacity, the reservoir holds 1490×10^6 m^3 of water, spans a surface area of 72.8 km^2, and reaches a maximum depth of 55 m (Chung *et al.*, 2009). Annual precipitation and water flow predominantly occur during summer months (June-September), with water levels fluctuating significantly, up to 15 m, during monsoon seasons. The hydraulic retention time averages about 145 days, characteristic of typical river-type reservoirs (Park *et al.*, 2021). Annual average rainfall is 1230 mm and average inflow is 102 m^3/s. The inflow for a five-year frequent flood is 5000 m^3/s, which is 50 times the annual average inflow.

Table 2.1 | Characteristics of the DC

Type	Value
Flood Water Level (FWL)	EL 80.0 m
Normal High Water Level (NHWL)	EL 76.5 m
Spillway crest level	EL 64.5 m
Low Water Level (LWL)	EL 60.0 m
Total storage	1490×10^6 m^3
Flood control storage	250×10^6 m^3
Spillway capacity (C_{spill})	11 680 m^3/s
Turbine capacity (C_{turb})	264 m^3/s

Although it impacts flood control in the Geum River basin significantly, DC has a relatively limited flood control capacity in comparison to its extensive watershed area. For instance, the 200-year flood inflow is almost 10 700 m^3/s, which can fill the flood control capacity in almost 6.5 hours. In addition, there is no restricted water level for the flood season in this reservoir. In essence, the DC is susceptible to flooding, necessitating rapid and well-informed flood control decisions by operators to avert extreme conditions. The characteristics of DC are provided in Table 2.1.

Hourly inflow data were collected for 24 years from 1997 to 2020, including 28 observed flood events from the National Water Resources Management Information System (WAMIS, https://www.wamis.go.kr) operated by the Ministry of Environment, Korea, and the Korea Water Resources Public Corporation (K-water), as presented in Table 2.2. These flood events are associated with a large amount of peak inflow and one to three peaks. To the best of our knowledge, most studies have primarily concentrated on floods lasting one to two days (Uysal *et al.*, 2018a; Breckpot *et al.*, 2013a; Delgoda *et al.*, 2013). In the case of short flood events, generating reliable outflows is relatively straightforward because the trade-offs among objectives are evident. For instance, it is clear that minimising outflows leads to an increase in reservoir water level, and vice versa. However, the situation becomes more complex over longer time frames. Minimizing outflows can ironically result in increased outflows if peak inflow occurs after the reservoir water level has already risen significantly. Therefore, our research examines floods lasting more than five days, featuring one to three peaks, to provide a more practical case study.

Additionally, inflow data were calculated based on measured outflows and changes in measured RWLs over time. K-water opens publicly RWLs and out-

Table 2.2 | Study flood events only for the DC

No.	From-To (YYYYMMDD HH)	Duration (hour)	Total rainfall (mm)	Peak inflow (m^3s^{-1})
1	19970630 21 - 19970711 22	266	304.7	5672
2	19980809 20 - 19980824 10	351	435.4	4993
3	19980929 01 - 19981005 10	154	183.9	4724
4	19990909 21 - 19990913 18	94	61.3	2562
5	19990922 20 - 19990926 16	93	102	3721
6	20000723 01 - 20000726 13	85	157.6	4234
7	20000824 02 - 20000903 07	246	242.8	3395
8	20000913 00 - 20000920 08	177	200.3	4012
9	20020830 19 - 20020905 21	147	207.8	10056
10	20030702 22 - 20030801 20	719	587.7	3689
11	20030818 02 - 20030826 10	201	206.2	3089
12	20030907 22 - 20030920 11	302	192.4	5407
13	20060708 19 - 20060724 10	376	427	4399
14	20060725 22 - 20060801 08	155	129.8	2379
15	20070828 22 - 20070909 23	290	303.8	4274
16	20070914 08 - 20070921 01	162	136.5	2462
17	20090711 06 - 20090720 10	221	230.84	4198
18	20110623 17 - 20110629 05	133	178.41	2067
19	20110707 20 - 20110718 22	267	304.96	6521
20	20110807 20 - 20110818 14	259	254.94	3557
21	20120812 04 - 20120821 08	221	239.77	3093
22	20120827 14 - 20120913 06	401	282.31	3430
23	20120913 06 - 20120923 08	243	140.28	4986
24	20160703 12 - 20160709 05	138	241.56	3655
25	20180825 21 - 20180908 16	332	379.93	2590
26	20200712 12 - 20200717 20	129	144.86	3779
27	20200719 22 - 20200817 19	694	614.79	6318
28	20200902 02 - 20200910 23	214	205.21	4468

flows on an hourly basis, allowing inflows to be calculated using the water balance equation, as presented in (2.1).

$$\frac{S_{t+\Delta t} - S_t}{\Delta t} = I_t - O_{total,t}. \tag{2.1}$$

where I_t and O_t denote an average inflow and outflow during Δt, the unit is m^3/s. For hourly data, Δt is an hour. Note that all published data have undergone quality control, meaning that this official data is free of missing values or outliers, as K-water has cleaned the data and verified its accuracy.

Despite the availability of sufficient data for the Daecheong reservoir, regarding water level observation data for upstream stations, we can collect hourly data from 2011 to 2020 due to missing data. There are a total of 22 water level stations in the upper stream, but only observed data from 16 water level stations are available for the study period, as presented in Fig 2.2. Other stations were either installed after 2011 or have missing data even in their officially published records. Unlike reservoir data, which contained no missing data or outliers, official water level data can contain missing data and outliers. This is reasonable because data quality regulations for reservoirs do not allow missing data, but for water level stations,

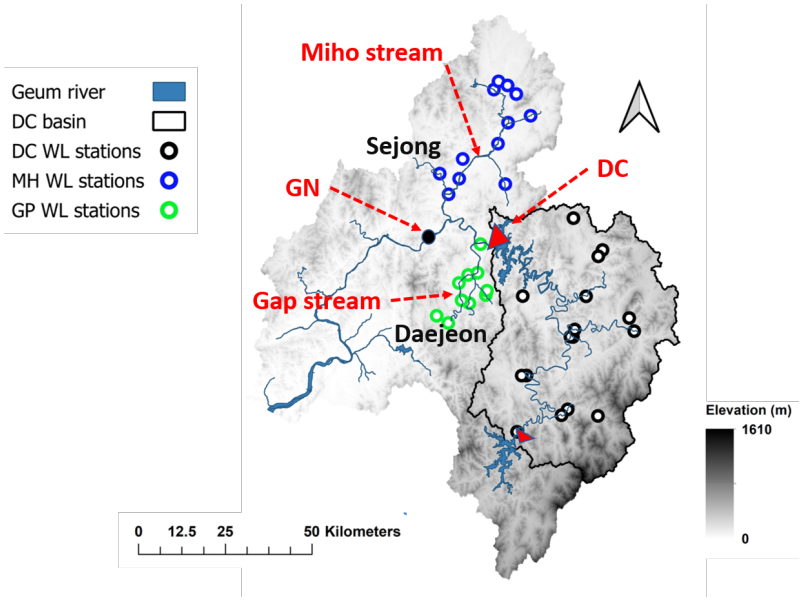


Figure 2.2 | Location of water level stations utilized in this study and tributaries. DC and GN refer to the Daecheong reservoir and Geumnam water level station, respectively.

data with missing values or outliers are allowed to be published even when they represent actual observations. Reservoir data are more strictly managed due to both their importance and the straightforward data production process. Further information is available in Jeong *et al.* (2021). In this study, instead of cleaning unofficial data ourselves, we only utilize data from stations without missing values or outliers. All observation data at upstream water level stations are obtained from the public data portal (www.data.go.kr) operated by the Korean government and the Korea Water Resources Public Corporation (www.kwater.or.kr).

Moreover, since Daejeon, a metropolitan city, and Sejong, an administrative city, are located directly downstream of the DC, flood control considering both the RWL and the downstream water level is required. The Geumnam station (hereafter referred to as GN), located within the administrative district of Sejong City, is close to DC and has observation data since 1993, as illustrated in Fig 2.2. The specifications of the GN in terms of flood control are shown in Table 2.3, respectively.

Table 2.3 | Flood control criteria at the Geumnam station (GN)

Attention level	Caution level	Alert level	Serious level
4.50 m	8.30 m	11.00 m	13.85 m

Despite sufficient data at the GN and DC, due to the limitation of observation data at Miho station (MH) and Gap station (GP) in the two tributaries, Miho and Gap streams, which is illustrated in Fig 2.3, we can only utilize a total of nine flood events from 2011 to 2020 when the downstream water levels need to be considered, as presented in Table 6.1. All data are obtained from the Geumgang Flood Control Office (FCO) in Korea through the public data portal and the Korea Water Resources Corporation. Note that the event numbers shown in Tables 2.2 and 6.1 may differ in subsequent chapters, particularly for testing and validation,

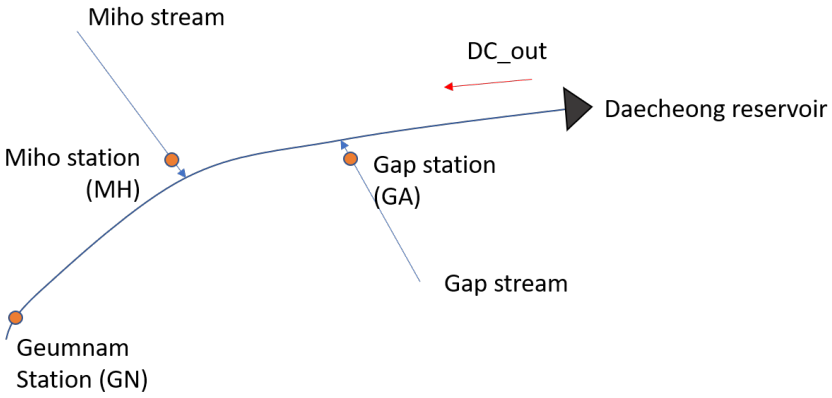


Figure 2.3 | Schematic diagram for the Geumnam station. The Miho station (MH) and Gap station (GP) are located just before the confluence of each tributary with the Geum river.

Table 2.4 | Study flood events including the downstream area

#	From - To (YYYYMMDD HH)	Duration (hour)	Peak inflow at DC (m^3/s)	Peak WL at GN (m)
1	20110807 10 - 20110819 00	279	3557	4.65
2	20120811 18 - 20120821 18	241	3093	3.41
3	20120827 04 - 20120913 16	421	3430	4.00
4	20120912 20 - 20120923 18	263	4986	2.01
5	20160703 02 - 20160709 15	158	3655	2.60
6	20180825 11 - 20180909 02	352	2590	4.53
7	20200712 02 - 20200718 06	149	3779	2.78
8	20200719 12 - 20200818 05	714	6318	7.35
9	20200901 16 - 20200911 09	114	1501	2.07

due to varying criteria used to split the events into training, testing, and validation sets according to each chapter's specific objectives.

3

Parameterized Dynamic Model Predictive Control[1]

Summary

Model Predictive Control (MPC) is an optimal control strategy suited for flood control of water resources infrastructure. Despite many studies on reservoir flood control and their theoretical contribution, optimization methodologies have not been widely applied in real-time operation due to disparities between research assumptions and practical requirements. To address this gap, we include practical objectives, such as minimizing the magnitude and frequency of changes in the existing outflow schedule. Incorporating these objectives transforms the problem into a multi-objective nonlinear optimization problem that is difficult to solve in real-time. Additionally, it is reasonable to assume that the weights and some parameters, considered the operators' preferences, vary depending on the system state. To overcome these limitations, we propose a framework that converts the original intractable problem into parameterized linear MPC problems with dynamic optimization of weights and parameters. This is done by introducing a model-based learning concept. We refer to this framework as parameterized Dynamic MPC (PD-MPC). The effectiveness of this framework is demonstrated through a numerical experiment for the Daecheong multipurpose reservoir in South Korea. We find that PD-MPC outperforms standard MPC-based designs without a dynamic optimization process for the objective weights and model parameters. Moreover, we demonstrate that the weights and parameters vary with changing hydrological conditions.

[1]This chapter is based on the publication: Koo, Ja-Ho, Edo Abraham, Andreja Jonoski, and Dimitri P. Solomatine., 2025. Balancing operator's risk averseness in model predictive control for real-time reservoir flood control. *Journal of Hydroinformatics*. 27 (4), 601–624. DOI: https://doi.org/10.2166/hydro.2025.019

3.1 Introduction

Many multipurpose reservoirs are designed and constructed with flood control as an important objective. They play a crucial role in mitigating flood risks downstream by retaining a significant portion of inflows in a given period. Additionally, the efficient operation of existing reservoirs is becoming increasingly important since costs and growing environmental concerns in society are making it challenging to build new reservoirs (Scudder, 2012). At the same time, climate change may lead to more severe and unpredictable flood events (Havens *et al.*, 2016), which are difficult for operators to manage efficiently using conventional simulation-based methods (Watts *et al.*, 2011).

Decisions regarding flood control operations for such reservoirs significantly influence basin flood conditions. Due to the importance of reservoir flood control, achieving *optimal* outflows of reservoirs has long been a focus (Giuliani *et al.*, 2021; Jain *et al.*, 2023). Various optimization methodologies, including Linear Programming (LP), Dynamic Programming (DP), and their variants, have been widely applied due to their capacity to guarantee optimal solutions (Labadie, 2004) for this type of problem. Another popular class of algorithms in the literature on reservoir optimization is randomised search techniques, mainly evolutionary algorithms and the Genetic Algorithm (GA). These have also gained traction, particularly for addressing complex problems that are not analytically expressed and less tractable to pose as mathematical optimization problems (Ahmad *et al.*, 2014). Some relatively recent literature has also focused on the approaches associated with control theory, primarily Model Predictive Control (MPC), with different optimization algorithms employed to solve the resulting optimal control problem (Castelletti *et al.*, 2023; Breckpot *et al.*, 2013a; Delgoda *et al.*, 2013).

Reservoir operators use only a limited horizon of rainfall and inflow forecasts, which are often uncertain. These forecasts are typically shorter than the whole length of a flood event (Breckpot *et al.*, 2013a). Operators must make decisions based on uncertain short-term forecasts. These decisions are not only about one outflow for the current time but also outflows for some time horizon, i.e., an outflow schedule. However, operators implement only the first (sometimes a few) outflows in the outflow schedule because, in general, forecasts are updated regularly, and new forecasts are considered less uncertain. Therefore, decisions should be updated repeatedly to reflect updated forecasts and hydrological conditions. In this sense, the receding horizon MPC concept, in which only the first control input is implemented and the optimization processes are repeated at each time step iteratively (Van Overloop, 2006), coincides with the reservoir flood control. While optimization-based control approaches, including MPC, have made substantial contributions in many applications (Schwenzer *et al.*, 2021), their practical application within reservoir flood control remains somewhat limited. This is mainly attributed to the practical objectives and preferences of operators, as follows.

Many researchers have introduced objectives for optimal reservoir operation (Tang *et al.*, 2019; Lin and Rutten, 2016; Malekmohammadi *et al.*, 2011; Reddy

and Kumar, 2006; Ko *et al.*, 1992), such as minimizing outflows via spillway gates, maximizing hydropower generation, and maintaining Reservoir Water Level (RWL) for water supply. However, during a flood event, most of these objectives may not be a priority.

What operators want in practice for flood control, i.e., operators' objectives for short-term planning, is not generally specified in the regulations and guidelines but is considered ambiguous 'experience' or 'expert knowledge'. Here, we call these 'practical objectives'. Note the distinction from mathematical objective functions, where mathematical equations, including system models and constraints, are used to calculate values of control objectives. Researchers seem to have paid insufficient attention to how to define and adopt these practical objectives in the optimization process (Ritter *et al.*, 2020; Teegavarapu and Simonovic, 2001). A recent survey conducted on water supply companies in England and Wales by Pianosi *et al.* (2020) has shown that factors, i.e., objectives and constraints, in decision-making were too complex to be included in an optimization process. Hence, operators generally hesitated to adopt optimization tools. Moreover, even if we successfully formulate practical objectives but with a number of highly nonlinear mathematical formulas, the problem then becomes intractable by the MPC approach (Berberich *et al.*, 2022; Allgower *et al.*, 2004). This is often a reason why many researchers apply only a limited number of objectives or linear and/or quadratic objectives, which are tractable even with several objectives, for a reservoir flood control problem (Uysal *et al.*, 2018b; Qi *et al.*, 2017; Breckpot *et al.*, 2013a) and so are deemed feasible to solve.

In addition, the preferences imply the selection criteria of operators among control alternatives, which are calculated from different weight sets and/or parameters. Therefore, selected weights and parameters in the multio-bjective optimization problem can be referred to as the *preference* (Wang *et al.*, 2017). In many studies focusing on reservoir flood control and applying the receding horizon MPC approach, weighted multi-objective optimization-based methods have been adopted (Hu *et al.*, 2014; Wang *et al.*, 2013). Hence, it is necessary to produce a Pareto set, which is a vector of non-dominant control inputs, using a weighted-sum method or applying one fixed weight representing the operators' preference, i.e., allocating high weights to the objective to which operators pay more attention. However, producing a Pareto set at every time step is computationally expensive (Peitz and Dellnitz, 2018). Additionally, we cannot be sure that the relative importance of objectives remains constant during a flood event.

To fill these research gaps, this chapter presents an MPC framework to generate an optimal flood control decision, which, in our opinion, would be practical and more acceptable to operators. The main idea is to explicitly define the practical objectives for reservoir flood control based on the first author's operational experience. We assume dynamic preferences, which we then formulate using parameterized linear MPC and dynamic optimization of weights and parameters of objectives to efficiently handle an otherwise intractable multi-objective nonlinear optimization problem. This methodology is tested for historical flood events with different fea-

tures, such as the number of inflow peaks and event lengths. Such verification is important because outflows and water level at the end of the first/second peak hugely influence the outflows of the next time steps.

This chapter is organised as follows. In Section 3.2, we present the objective functions for practical flood control and propose a framework to efficiently incorporate linear and nonlinear objectives under the assumption of dynamic preference. Section 3.3 describes the detailed MPC problem and a numerical experiment. The contribution and limitations of this research and the results of the numerical experiment are then presented in Section 3.4, followed by conclusions.

3.2 Method

3.2.1 Objective functions for practical flood control

The flood control of large reservoirs requires multiple practical considerations; one well-known conflicting objective is the need to reserve enough water to supply tocontractors at the end of a flood event. In this section, we propose additional important objectives and motivate their necessity based on the receding horizon MPC concept.

First, using the full capacity of control facilities such as spillway gates is not reasonable for large reservoirs. Instead, operators tend to desire less outflow generally. Second, it is preferable to limit the frequency of operations of spillway gates to prevent wear and malfunction. Furthermore, immediate and frequent changes in outflow schedules are not preferred because they can hinder the predictability of the flood situation of the downstream area for other flood control agencies. It is worth noting that even though we are focusing on flood-relevant objectives, other objectives for reservoir operation, such as maintaining minimum flow requirements or ensuring adequate flow for fish migration and downstream habitats, can be formulated similarly.

In defining the practical objectives and the subsequent methodology, we introduce a control input vector, $\mathbf{u}^k$, at time step k, defined as

$$\mathbf{u}^k = \{O^k_{total,k}, \ldots, O^k_{total,k+H-1}, O^k_{spill,k}, \ldots, O^k_{spill,k+H-1}\} \in \mathbb{R}^{2H} \tag{3.1}$$

where $O^k_{total,k+H-1}$ and $O^k_{spill,k+H-1}$ are the total outflow and spillway outflow decided at time step k for time $k+H-1$, respectively. H is the length of the prediction horizon. We define an augmented state vector that consists of the RWL, predicted inflow, and the outflows decided at the previous time step $k-1$ as $\mathbf{x}_k = \{S_k, I_k, \ldots, I_{k+H-1}, O^{k-1}_{total,k}, \ldots, O^{k-1}_{total,k+H-1}\} \in \mathbb{R}^{1+2H}$, where S_k and I_k are the storage volume and predicted inflow variables for time k, respectively. To avoid ambiguity, we define the term 'time step k' as the kth iteration of MPC and 'time t' as the exact hour at which a control input is supposed to be implemented. The detailed objectives for practical reservoir flood control are defined below.

minimizing the peak and total outflow via spillway gates

Operators typically prefer using turbines over spillways due to the two objectives of reducing the risk of flood damage in the downstream area and maximizing earnings from hydropower generation, respectively. Generally, these two objectives are not in conflict but rather complementary. For example, to reduce the peak outflow, it is essential to distribute outflows over time, which consequently leads to maximizing turbine outflows and minimizing the total volume of spillway outflows. Therefore, these objectives could be applied both together or separately; at least one of them appears to be necessary in any case. The objectives can be formulated as follows:

$$\tilde{J}1 \coloneqq \min_{\mathbf{u}^k} \max_{\forall t \in \{k,\ldots,k+H-1\}} O^k_{spill,t}, \text{ and} \tag{3.2}$$

$$\tilde{J}2 \coloneqq \min_{\mathbf{u}^k} \sum_{t=k}^{k+H-1} O^k_{spill,t}. \tag{3.3}$$

The first equation minimizes peak outflow in a prediction horizon (from k to $k+H-1$ at time step k), which is $\max_{\forall t \in \{k,\ldots,k+H-1\}} O^k_{spill,t}$, where $O^k_{spill,t}$ means a spillway outflow at time t decided at time step k. The second one minimizes the sum of spillway outflows, $\sum O^k_{spill,t}$, during a prediction horizon.

minimizing step-wise outflow changes in the prediction horizon

For large reservoirs, operators generally hesitate to change outflows drastically. Rather, it is preferred to implement smooth and deliberate adjustments, avoiding significant and sudden changes. This can be achieved by minimizing cumulative step-wise changes over a given receding horizon as:

$$\tilde{J}3 \coloneqq \min_{\mathbf{u}^k} \sum_{t=k}^{k+H-2} |O^k_{total,t+1} - O^k_{total,t}| \times w_t, \tag{3.4}$$

where the absolute value notation could be replaced with a quadratic of the step-wise changes. However, the use of absolute value is preferred to also penalise small changes in outflows. To avoid excessive penalization for every change, we can assign greater weight to changes occurring at shorter time intervals than the changes in outflows at longer intervals by introducing weights w_t in this objective.

It is noteworthy that $\tilde{J}3$ and $\tilde{J}2$ complement each other to generate an outflow sequence with less variability. When outflows need to increase, e.g., when RWL is expected to rise significantly due to substantial inflow, $\tilde{J}2$ leads to dropping the last outflow in the outflow schedule to minimize the total spillway discharge over the prediction horizon. This occurs because, in the discretised reservoir model, which will be detailed in Section 3.3.2, the last outflow does not affect RWLs within the same prediction horizon, as outflows and inflows at a given time, e.g., t, only influence the reservoir's hydrological condition at the subsequent time, $t+1$. $\tilde{J}3$ can prevent this undesirable schedule formation. Conversely, when outflows need to decrease, e.g., when inflows are insufficient to maintain RWLs for water supply

despite the earlier prediction of substantial inflows, $\tilde{J}3$ may impede the reduction of the final outflow, but $\tilde{J}2$ can enforce this reduction. This complementarity may be deemed less crucial when the prediction horizon is long, given the diminishing significance of outflow changes distant from the current time step. However, in scenarios with a short prediction horizon, such as the 6-hour case in our numerical experiment in Section 3.3, these objectives can be essential for making the optimal outflow practically applicable.

minimizing changes in outflows calculated at consecutive time steps

In receding horizon MPC, the optimal control inputs are generated at each time step for the prediction horizon. In many MPC applications, previous optimal control inputs are not taken into account at the next time step calculations. However, because the optimal control inputs decided at the previous time step are an *outflow schedule*, it is also crucial to minimize the changes in each control input over each prediction horizon to prevent sudden alterations in outflow schedules. This objective can be formulated as:

$$\tilde{J}4 \coloneqq \min_{\mathbf{u}^k} \sum_{t=k}^{k+H-2} |O^k_{total,t} - O^{k-1}_{total,t}| \times w_t, \tag{3.5}$$

where w_t are weights. The previous objective (3.4), which aims to minimize the step-wise changes within a prediction horizon, serves a similar purpose to (3.5) when outflows do not change much. To avoid excessive penalization for every change, the weight w_t is also applied for the same reason as in (3.4).

minimizing the RWL exceedance outside of the target range

While this objective has been a common consideration, researchers typically aim for RWL to be in close proximity to Normal High Water Level (NHWL) or within a specified target range (Delgoda *et al.*, 2013). Practical operators also seek to avoid RWL exceeding a certain water level to prepare for unexpected extreme events and maintain safety perceptions. Without this objective, optimal outflows often result in RWL approaching Flood Water Level (FWL).

$$\begin{aligned}
\tilde{J}5 &\coloneqq \min_{\mathbf{u}^k} \sum_{t=k}^{k+H-1} \delta S_U \times w_{S_U} + \delta S_L \times w_{S_L} + \delta S_H \times w_{S_H}, \\
\delta S_U &= \begin{cases} S_t - S_U & \text{if } S_t > S_U, \\ 0 & \text{otherwise,} \end{cases} \\
\delta S_L &= \begin{cases} S_L - S_t & \text{if } S_L > S_t, \\ 0 & \text{otherwise,} \end{cases} \\
\delta S_H &= \begin{cases} S_t - S_H & \text{if } S_t > S_H, \\ 0 & \text{otherwise,} \end{cases}
\end{aligned} \tag{3.6}$$

where S_t represents the reservoir storage at time t, and S_U and S_L denote the storage levels for the upper and lower boundaries of the target range, respectively. The parameter S_H corresponds to the highest water level that operators aim to avoid exceeding. We assign weights, denoted as w_{S_U}, w_{S_L}, and w_{S_H}, to these components. Generally, S_U and S_L have well-defined values that are less influenced by preference. S_U can be NHWL or the restricted water level during the flood season, while S_L may refer to the marginal storage reserved for water supply following a flood event. However, it is worth noting that S_H is often subject to the operator's perceived risk tolerance with respect to exceeding FWL.

Continuity of spillway gates condition

One well-known limitation of MPC is that it can generate myopic control inputs when using a short prediction horizon (Morari and Lee, 1999). In the context of reservoir flood control, this often leads to frequent opening and closing operations of spillway gates. This is not what operators usually want since it can increase the chance of actuators wearing out and human errors. Operators typically want to preserve the condition of spillway gates, i.e., when gates are already open, operators prefer not to close them until a flood event is certainly coming to an end. This objective can be formulated as follows:

$$\begin{aligned} \tilde{J}6 &\coloneqq \min_{\mathbf{u}^k} \sum_{t=k}^{k+H-1} \rho_t, \\ \rho_t &= \begin{cases} 1 & \text{if } O^k_{spill,t-1} \times O^k_{spill,t} = 0 \text{ and } O^k_{spill,t-1} + O^k_{spill,t} > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{3.7}$$

where ρ is a variable to check if spillway gates are operated. When the gate condition changes, ρ becomes 1. For example, when $O_{spill,t-1}$ is zero and $O_{spill,t}$ is not zero, then $O^k_{spill,t-1} + O^k_{spill,t} > 0$ and $O^k_{spill,t-1} \times O^k_{spill,t} = 0$. If the gate condition does not change, for example, both previous spillway outflow and current spillway outflow are zero, i.e., $O^k_{spill,t-1} + O^k_{spill,t} = 0$, or have positive values, i.e., $O^k_{spill,t-1} \times O^k_{spill,t} > 0$, and then ρ becomes zero.

This equation shares a similar purpose as $\tilde{J}3$ and $\tilde{J}4$, in that it penalises changes in outflow schedules. However, it holds a greater significance because it also regulates the opening and closing of gates. The reason is that $\tilde{J}3$ and $\tilde{J}4$, which penalise total discharge, can have the minimum zero values while both turbine and spillway gate states change in a schedule. Especially, $\tilde{J}4$ has zero values when the change of spillway gate states is planned in a previous time step. Adding the objective shown in $\tilde{J}6$ will ensure that changes in spillway gates are specifically avoided unless necessary. However, this objective is a penalty form of what is called a complementarity constraint (Powell *et al.*, 2016), so linearising this objective is difficult for MPC without adding binary variables representing ρ, resulting in a mixed-integer problem (Anitescu, 2000). In Section 3.2.3, this objective is included only in the Evaluator to get around the issue of optimization problem complexity.

Maintaining peak outflow under the peak inflow

Retaining inflow is one of the primary functions of reservoirs for flood control. Therefore, maintaining peak outflow under the peak inflow up to the current time step should be considered important. This can generally be achieved by $\tilde{J}1$. However, it is also worth considering this by adding another objective as follows:

$$\begin{aligned} \tilde{J}7 &:= \min_{\mathbf{u}^k} \gamma^k, \\ \gamma^k &= \begin{cases} \text{large value} & \text{if } \max_{\forall t \in \{k,\dots,k+H-1\}} O^k_{total,t} > \max_{\forall t \in \{0,\dots,k-1\}} I_t, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.8}$$

where I_t is the predicted inflow at time t.

Note that the peak outflow represents the maximum outflow within a prediction horizon, whereas the peak inflow denotes the maximum inflow up to the current time step, as shown in (3.8). This objective can be easily linearised, but we incorporate it into the Evaluator in Section 3.2.3 rather than the MPC formulation because it is optimized indirectly in the middle of minimizing the peak outflow in $\tilde{J}1$ in most cases.

Utilising turbine capacity prior to opening spillway gate

To ensure efficient utilization of turbine capacity before resorting to opening the spillway gate, a constraint is formulated as follows:

$$O^k_{spill,t} \times (O^k_{total,t} - MO_{turb}) = 0, \tag{3.9}$$

where MO_{turb} is the turbine capacity. This constraint is also reformulated as a soft constraint, serving as an objective to circumvent the complexities of a nonlinear constraint:

$$\begin{aligned} \tilde{J}8 &:= \min_{\mathbf{u}^k} \lambda^k, \\ \lambda^k &= \begin{cases} \text{large value} & \text{if } O^k_{spill,t} > 0 \text{ and } O^k_{total,t} <= MO_{turb} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.10}$$

This is also a complementarity constraint, which is hard to linearise. Therefore, in this research, we include this constraint as an objective in the evaluator using a penalty approach for the same reason as $\tilde{J}6$.

3.2.2 Dynamic characteristic of the preferences

As mentioned in Section 3.1, numerous studies aimed to find the best set of objective weights while assuming that the preference remains constant. This is because the relative importance of each objective typically does not change significantly when hydrological conditions remain stable. However, in the context of reservoir flood

control, operating preference can often shift with changes in hydrological states. This variability is evident from the objectives outlined in Section 3.2.1.

Some objectives may conflict with one another, while others may complement each other, and this dynamic depends on the current state. For instance, when RWL approaches FWL or S_H, the importance of $\tilde{J}5$ increases, necessitating a substantial increase in outflows with less importance for other objectives. Conversely, if RWL remains below S_U and spillway outflows are stable, $\tilde{J}3$ and $\tilde{J}4$ should be prioritized. When a significant increase in outflows becomes unavoidable, the highest weight should be assigned to $\tilde{J}1$, followed by $\tilde{J}5$. Consequently, we can conclude that the relative importance of each objective should vary depending on hydrological conditions.

Some parameters, like the target water level, have been considered static as well. However, the parameters such as S_U, S_L, and S_H in (3.6) directly impact the objective value of $\tilde{J}5$ and the optimal control inputs. Hence, we can not assume that these parameters remain static when the preference is not static.

3.2.3 Parameterized Dynamic MPC framework

Model Predictive Control (MPC) is widely used in reservoir flood control, leveraging system models to predict future states and determine optimal control inputs (Castelletti *et al.*, 2023; Breckpot *et al.*, 2013a; Delgoda *et al.*, 2013). When we assume dynamically changing objectives' weights and parameters, both control inputs and objectives' weights/parameters need to be optimized simultaneously. The problem becomes complicated due to their interdependency. Without additional criteria, this co-optimization problem would be intractable.

If we have the means to evaluate the derived optimal control inputs, which we call an evaluator here, we can simplify this problem. While traditional approaches typically generate a Pareto set, a set of alternatives, and select the optimal one using a single criterion, such as the possibility for system failure (Chen *et al.*, 2020), minimizing the worst-case impact (Yu *et al.*, 2023), or integrate simple but diverse evaluation criteria (Myo Lin *et al.*, 2020; Zhu *et al.*, 2018), computing the complete Pareto set becomes computationally intractable with multiple objectives (eight in our case), particularly for online optimization process of MPC.

To circumvent this complexity, we suggest a framework for searching for the best weights/parameters that optimize many objectives and satisfy constraints, accounting for the dynamic hydrological state. Instead of finding a Pareto set, this framework simultaneously searches the weight and parameter spaces on the basis of the evaluator in order to produce the optimal control inputs. We refer to this framework as Parameterised Dynamic Model Predictive Control (PD-MPC) as illustrated in Fig 3.1. Here, the disturbance, denoted as $\mathbf{d}_p^k$, represents an uncertain inflow vector at time step k, i.e., for time k to $k + H - 1$, where H is the length of the prediction horizon. The state vector at time step k is denoted as $\mathbf{x}^k$. $\mathbf{u}$ refers to the control inputs for a given weight and parameter vector, $\mathbf{z}$. E denotes the absolute performance evaluator.

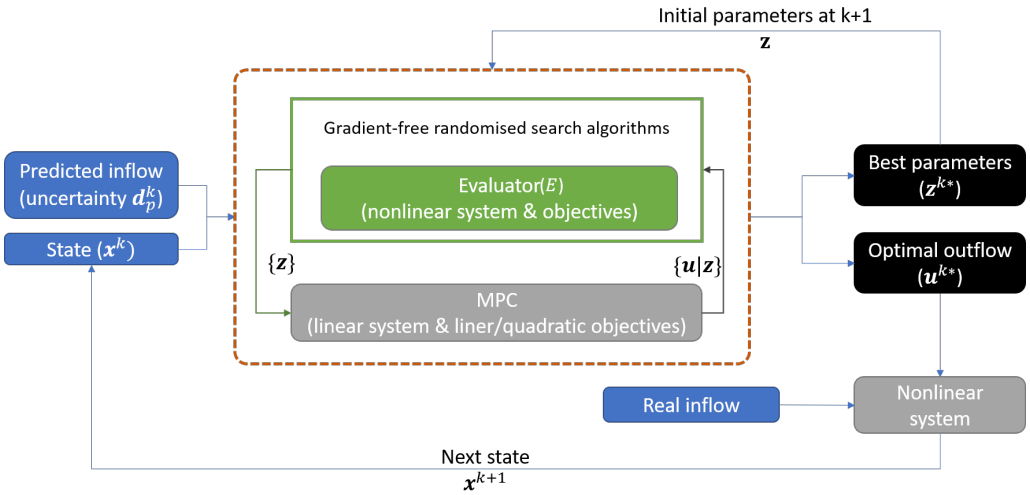


Figure 3.1 | The schematic diagram of PD-MPC approach, with $\mathbf{z}^k$ warm started with optimal values from the previous step.

The PD-MPC approach solves the optimal control problem at the k^{th} step of a receding horizon implementation:

$$\{\mathbf{z}^{k*}, \mathbf{u}^{k*}\} = \arg\min_{\mathbf{z}^k, \mathbf{u}^k, \mathbf{x}^k} E(\mathbf{z}^k, \mathbf{u}^k, \mathbf{x}^k | \mathbf{x}_0^k), \tag{3.11}$$

by alternatingly solving the two sets of problems:

$$\{\mathbf{z}^{k*}\} = \arg\min_{\mathbf{z}} \mathbf{E}(\mathbf{z} | \mathbf{x}^k, \mathbf{u}^k), \text{ and} \tag{3.12a}$$

$$\{\mathbf{u}^{k*} | \mathbf{z}^k\} = \arg\min_{\mathbf{x}^k, \mathbf{u}^k} \mathbf{J}(\mathbf{x}^k, \mathbf{u}^k | \mathbf{x}_0^k, \mathbf{z}^k), \tag{3.12b}$$

respectively, where $\mathbf{x}_0^k$ represents the initial state at time k, $\mathbf{J}(\cdot)$ is the objectives of MPC. In this equation, $E(\cdot)$ is not directly optimized; instead, it just evaluates $\mathbf{z}|\mathbf{x}^k, \mathbf{u}^k$, thus, $E(\cdot)$ can be formulated using highly nonlinear equations, e.g., conditional equations and exponential formulas, without limitation.

At each time step, we first solve the (3.12b) to obtain control actions for a specific weight set, then solve the (3.12a) to optimize both the weights/parameters and their corresponding control inputs through the evaluator, $E(\cdot)$. To solve this evaluation problem, we can employ heuristic optimization techniques, such as GA. Heuristic algorithms explore the weight and parameter space to find weights and parameters with the optimal value from the evaluator at each time step of the receding horizon recalculation. Therefore, this methodology enables us to decouple the optimization process into two separate tasks: finding optimal weights and determining control actions, which simplifies the co-design problem, as illustrated in Fig 3.1.

In this research, we linearize all proposed practical objectives for a linear MPC and integrate objectives presented in the previous section to formulate the evaluator, without limiting the evaluator to linear and nonlinear objectives that are straightforward to implement in mathematical optimization. Moreover, to reflect real hydrological processes, the evaluator uses a nonlinear system simulation (i.e., a

nonlinear height-volume curve) to calculate water levels. This allows the PD-MPC framework, primarily optimizing a linear MPC problem to find a release schedule, to select an outflow schedule that actively aligns with the operator's preferences. The reason why all objectives are linearized for $\mathbf{J}(\cdot)$, instead of applying quadratic equations as presented in Section 3.2.1, is to penalize even minor exceedances and linearizing an absolute form is straightforward. For detailed explanations, including formulas, please refer to Appendix A for the evaluator and Section 3.3.3 for the objectives, weights and parameters.

In addition, we utilise GA as our heuristic optimization algorithm. The Genetic Algorithm (GA) is one of the most prominent and easy-to-implement gradient-free algorithms which imitate the natural selection process (see, e.g., Katoch *et al.* (2021)). It continuously develops the population (a set of vectors seen as potential solutions) over each iteration, employing reproduction, crossover, and mutation operators, aiming at preserving the vectors with lower objective values and iteratively recombining them.

3.3 Numerical Experiment

3.3.1 Study flood events

Three flood events with a large amount of peak inflow and one to three peaks are selected from historical data as presented in Table 2.2 in the previous chapter. In addition, our research examines floods lasting more than eight days, featuring one to three peaks, to provide a more practical case study. Although these events can present significant operational challenges, demonstrating our method's practical utility, more extreme floods, such as a 200-year return period flood and a probable maximum flood (PMF), would provide more rigorous validation cases. However, the Daecheong reservoir, our case study site, was constructed over 40 years ago, and to the best of our knowledge, the hydrographs for the 200-year flood and PMF are unavailable. Moreover, the 200-year flood and PMF typically span only 1-3 days with a single peak, representing less uncertainty and total inflow volume.

To simulate the uncertainty associated with inflow predictions, we apply a conventional method in which uncertainty follows a Gaussian distribution with zero means (Uysal *et al.*, 2018a; Li *et al.*, 2010), instead of an ensemble or other scenario generation methods, for simplicity. Thus, 'predicted' inflow data is generated by multiplying random numbers by real inflow, as (3.13). A standard deviation of random numbers is a. Predicted inflow should be positive; therefore, the lower boundary c is introduced. Forecasting uncertainty tends to increase as the forecast

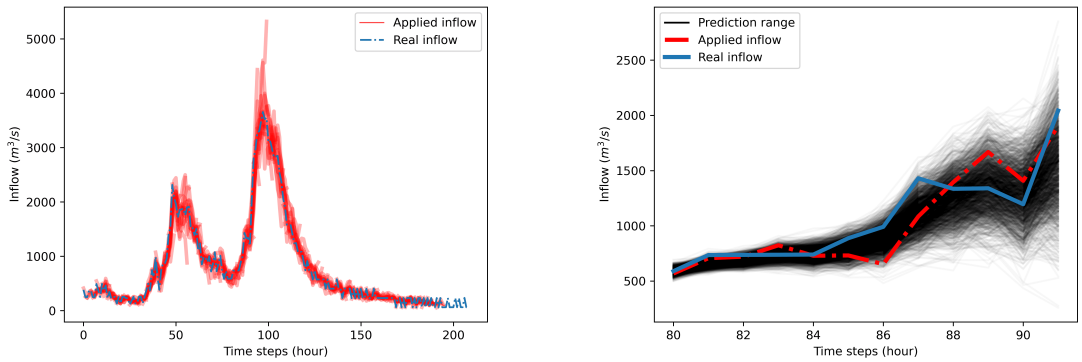
Table 3.1 | Study flood events

Event #	Periods	Duration	Peak inflow	Feature
1	from 2 July 2016 to 11 July 2016	197 hour	3,655 m^3/s	Double peak
2	from 25 August 2018 to 7 September 2018	311 hour	2,590 m^3/s	Triple peak
3	from 1 September 2020 to 12 September 2020	269 hour	4,468 m^3/s	One peak

period increases. To reflect this, we adjust the standard deviation by increasing it by b per each time step. Moreover, a moving average filter with a three-hour window is applied every time step. This filter helps prevent unreasonable high fluctuations in predicted inflow within the prediction horizon.

$$\begin{aligned} PI_t &= \frac{1}{3}\sum_{j=t-2}^{t} RI_j \times \omega I_t, \\ \omega I_t &= \begin{cases} 1+x, \text{ where } x \sim \mathcal{N}(0, (a+T\times b)^2) & \text{if } 1+x \geq c, \\ c & \text{otherwise,} \end{cases} \end{aligned} \tag{3.13}$$

where PI_t and RI_t are the predicted and real inflow at time t, while ωI_t denotes the uncertainty of inflow at time t. The variable T is the index of forecasting time t within the current prediction horizon (i.e., $T = 0 : N-1$), so the prediction uncertainty increases for a time farther into the future. The constants a, b, and c are set to 0.05, 0.03, and 0.1 - somehow arbitrarily here - to generate something that has roughly the same order of errors as in reality; the sample hydrographs of real and predicted inflow are demonstrated as follows this method is reliable.



(a) **(b)**

Figure 3.2 | Example of real and predicted inflow hydrographs for Event 1 when a prediction horizon is 12 hours: (a) the hydrograph for the whole event period, (b) an example of predicted inflow at time step 80.

We present three flood events, as depicted in Fig 3.2 and 3.3. In these illustrations, the 'applied inflow' is highlighted in red, representing the overlay of uncertain inflows generated individually for each time step within a prediction horizon. For instance, in Fig 3.2a, one red line representing an uncertain inflow starting from time step 0 to 11 is presented, and another uncertain inflow starting from time step 1 to 12 is presented. This pattern continues for each subsequent time step until the last time step. There are a total of 184 different uncertain inflows, all within a 12-hour prediction horizon, depicted in this figure. In Fig 3.2b, a total of 2000 black lines illustrate uncertain inflows by incorporating random errors into the actual inflow data.

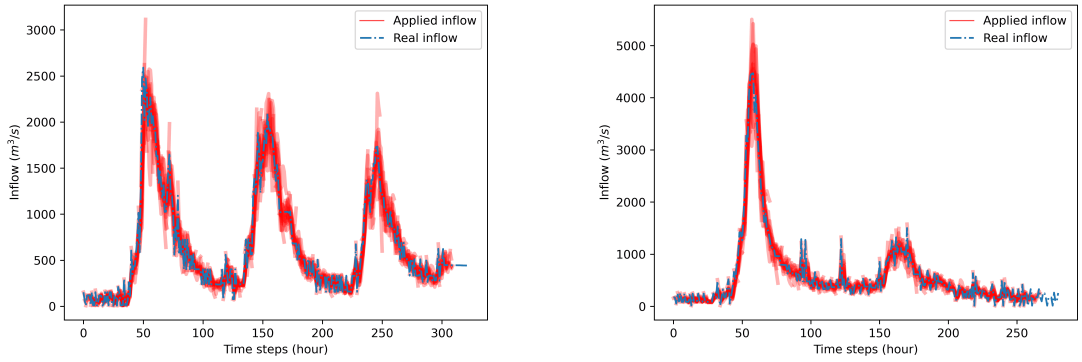


Figure 3.3 | Example of real and predicted inflow hydrographs with a 12-hour prediction horizon (a) for Event 2, and (b) for Event 3.

3.3.2 A reservoir flood control system

System model

A simple linear reservoir model can be expressed as

$$S_{t+1} = S_t + I_t - O_{total,t}, \tag{3.14}$$

where S_t is the current storage of a reservoir at time t, and I_t and O_t are the total amount of inflow and outflow for time t, respectively. Errors in measuring water level, calculating storage amount and outflow are not considered for simplicity. In addition, evaporation and seepage are not explicitly considered in (3.14), but these factors are already accounted for in the inflow data. This is because inflow is not directly measured but calculated using (3.14) from measured outflow and storage time series for the Daecheong reservoir.

Constraints

The operational constraints can be formulated as follows:

(1) The total outflow is more than the sum of the agreed amounts with water users and the in-stream flow, i.e., the downstream demand:

$$D_t \leq O_{total,t}, \tag{3.15}$$

where $O_{total,t}$ and D_t are the total outflows and the downstream demand at time t, respectively.

(2) RWL should be between FWL and LWL:

$$\text{LWS} \leq S_t \leq \text{FWS}, \tag{3.16}$$

where S_t is the reservoir storage at time t, while LWS and FWS are the reservoir storage at LWL and FWL, respectively.

(3) Outflows via the turbine and spillway gates are not able to exceed their capacity:

$$\begin{aligned} O_{turb,t} &\leq MO_{turb}, \\ O_{spill,t} &\leq MO_{spill}, \end{aligned} \tag{3.17}$$

where $O_{turb,t}$ and $O_{spill,t}$ represent the turbine and spillway outflow at time t, respectively. MO_{turb} and MO_{spill} are the outflow capacity of turbines and spillway gates. We take into account the spillway crest level in our analysis, but spillway outflow capacity depending on RWL is disregarded as explained in Section 3.3.1.

Objectives

By linearising the practical objectives described in Section 3.2.1 to penalise even minor exceedances, the objective function of MPC is set as below:

$$\min J1^k + J2^k + J3^k + J4^k + J5^k, \tag{3.18}$$

where,

$$J1^k \coloneqq \max_{t \in k,\ldots,k+H-1} O^k_{spill,t} \times w_1, \tag{3.19a}$$

$$J2^k \coloneqq \sum_{t=k}^{k+H-1} O^k_{spill,t} \times w_2, \tag{3.19b}$$

$$J3^k \coloneqq \sum_{t=k+1}^{k+H-1} \delta O_t^{in,I} \times w_{3,I} + \delta O_t^{in,D} \times w_{3,D}, \tag{3.19c}$$

$$J4^k \coloneqq \sum_{t=k}^{k+H-1} \delta O_t^{between,I} \times w_{4,I} + \delta Q_t^{between,D} \times w_{4,D}, \tag{3.19d}$$

$$J5^k \coloneqq \sum_{t=k}^{k+H-1} \mu S_t^1 \times w_{5,1} + \mu S_t^2 \times w_{5,2} + \mu S_t^3 \times w_{5,3}, \tag{3.19e}$$

subject to,

$$O^k_{spill,t} + O^k_{turb,t} - O^k_{total,t} = 0, \tag{3.20a}$$

$$O^k_{total,k} - O^{k-1}_{total,k} = 0, \tag{3.20b}$$

$$\delta O_t^{in,I} - \delta O_t^{in,D} + (O^k_{total,t} - O^k_{total,t-1}) \times w_{\delta O_{in,t}} = 0, \tag{3.20c}$$

$$\delta O_t^{between,I} - \delta O_t^{between,D} + (O^k_{total,t} - O^{k-1}_{total,t}) \times w_{\delta O_{between,t}} = 0, \tag{3.20d}$$

$$\mu S_t^1 + S_U - S_t \geq 0, \tag{3.20e}$$

$$\mu S_t^2 + S_t - S_L \geq 0, \tag{3.20f}$$

$$\mu S_t^3 + S_H - S_t \geq 0, \tag{3.20g}$$

$$\delta O_t^{in,I},\ \delta O_t^{in,D},\ \delta O_t^{between,I},\ \delta O_t^{between,D},\ \mu S_t^1,\ \mu S_t^2,\ \mu S_t^3 \geq 0, \tag{3.20h}$$

where J_1^k, for example, is an objective for the MPC formulation at time step k. All objectives are multiplied by their respective weights and then summed together to form the objective function, as shown in (3.18) and (3.19). Weights, i.e., w_1, w_2, $w_{3,I}$, $w_{3,D}$, $w_{4,I}$, $w_{4,D}$, $w_{5,1}$, $w_{5,2}$, and $w_{5,3}$, are multiplied by each objective.

Several positive slack variables are introduced to linearise objectives instead of applying absolute forms. These variables include $\delta O_t^{in,I}$ and $\delta O_t^{in,D}$, representing the step-wise increase and decrease in outflows in a prediction horizon; $\delta O_t^{between,I}$ and $\delta O_t^{between,D}$, representing the increase and decrease in outflow schedules at time step k and $k-1$; μS_t^1, μS_t^2, and μS_t^3, representing the storage violation above S_U, below S_L and above S_H, respectively. To account for varying penalties over time, we assign weights $w_{\delta O_{in,t}}$ and $w_{\delta O_{between,t}}$ to each change at time t. The idea is to penalise changes closer to the start of the horizon more than changes farther ahead in the horizon. In this work, we set $w_{\delta O_{in,t}}$ to $\frac{1}{(t-k+1)\times 3}$, and $w_{\delta O_{between,t}}$ is set as follows:

$$w_{\delta O_{between,t}} = \begin{cases} \frac{1}{t-k+1} & \text{if } t \leq k+3, \\ \frac{1}{(t-k+1)\times 2} & \text{otherwise.} \end{cases} \tag{3.21}$$

The target water levels, such as S_U and S_L, are defined as mentioned in Section 3.2.1. The weights of objectives in (3.19), as well as the allowed highest water level, S_H in (3.20g), will be found during the optimization process.

The important difference in the objective formulas presented in Section 3.2.1 is about $O_{total,k}^k$ in (3.20b) and $\tilde{J}4$ in (3.5). Note that $O_{total,k}^k$ is the first outflow in the optimal control inputs at time step k. First, changing the outflow for time k at time step k is practically impossible because it involves implementing the current outflow while calculating it. To address this, we separate this to ensure it, rather than assigning significant weight to the change in the first outflow in (3.20d). Hereby, the outflow decision is delayed for a time step.

3.3.3 PD-MPC design

Since the Korea Meteorological Agency (KMA) publishes 6-hour quantitative rainfall forecasts every hour, in the experiment, we employ four different prediction horizons: 6, 12, 18 and 24 hours. Starting from the 6 hours, the longer horizons allow us to explore the effect of horizon length on performance. The control horizon is the same as the prediction horizon. For the initial run at $t = 0$, we set the initial storage to the corresponding level of NHWL (EL. 76.5m), while the initial outflows via the turbines and spillway gates are set to 150 m^3/s, which is the average hourly outflow during the flood season from 2015 to 2020, and 0 m^3/s, respectively.

In this experiment, the GA optimizes 10 parameters, comprising of 9 weights for the objectives and 1 parameter for the highest RWL, denoted as S_H. To reduce the running time of GA, we impose search range limits for each weight and parameter, as outlined in Table 3.2. In $J5$ in (3.19), $w_{5,2}$ and $w_{5,3}$ are fixed at twice and twenty times of $w_{5,1}$, respectively, because maintaining RWL under S_H for dam safety is a

higher priority than maintaining it between target levels, additionally, in order to reduce the computational complexity by decreasing the number of weights which need to be optimized. Again, the reason why nominators for $w_{5,2}$ and $w_{5,3}$ are larger than others is to reduce the number of dynamic weights and to reflect the fact that the objectives corresponding to $w_{5,2}$ and $w_{5,3}$ are more important than the objective for $w_{5,1}$. Here, 40 and 400 are selected somewhat arbitrarily because precise values are not critical due to the PD-MPC, which finds optimal weights dynamically. Similar to the value of 20 for other multipliers, the values 40 and 400 are selected to prevent the solver from neglecting objectives when objective values are divided by a substantial value, e.g., FWS for $w_{5,1}$, $w_{5,2}$, and $w_{5,3}$. Instead of fixing these nominators and weights, we could extend the search range from 1-20 to 1-400 or 1-500, and so on. However, the results are the same, although it takes considerably more time. This is why we fixed the nominators instead of extending the search range while setting the same value for all nominators. Therefore, the number of weights and parameters explored becomes 8 from 10. All weights share the same length of the search range, except for $J2$. This is because the weight of $J2$, i.e., minimizing the total outflows via spillway gates, can be relatively small due to its similarity to $J1$, i.e., minimizing the peak outflows via spillway gates, and we want to emphasize $J1$. To ensure $J5$ is not neglected, the search ranges for $w_{5,1}$, $w_{5,2}$, and $w_{5,3}$ start from 1, not from 0.

Moreover, to simplify the optimization process and work with integer values in GA, we introduce multipliers based on these MAVEs. Because MAVE represents the largest anticipated magnitude of each objective, it is often adopted to normalise objectives to maintain numerical balance across different objectives. Each weight is then calculated by multiplying a selected integer value by its corresponding multiplier. This approach ensures that all objectives contribute as a proportion of the selected values in the search range, regardless of their natural scales, while allowing GA to work with simpler integer values during the search process.

We prepare MPC baselines with fixed weights and a fixed parameter set, denoted as 'Fixed', which would be used in standard operation, as presented in Table 3.2. Our comparative analysis aims to demonstrate that operators' decisions can be improved by dynamically optimizing weights in real-time. This approach allows for adapting risk preferences depending on the system state. To ensure the feasibility of the problem, two weight/parameter sets, which emphasise minimizing the changes in outflow schedules and minimizing peak outflow, are selected from the best weight/parameter sets generated from various PD-MPC tests and verified through trial and error (Uysal *et al.*, 2018a). Throughout this chapter, we refer to the value before being multiplied by the multiplier as the weights unless this would cause confusion otherwise.

In Table 3.2, a searching range of S_H is the storage where RWL is in {EL. 78.5m, 79.0, 79.5m} and S_H of the baselines, i.e., RWS_F (Reservoir Water Storage for Fixed weights/parameter set), which is the stored amount of water at EL.79.0m, as illustrated in Fig 3.4. A computer code is developed using Python. In detail, pyomo (Hart *et al.*, 2011), GLPK solver (Makhorin) and pyGAD (Gad, 2021) pack-

Table 3.2 | The possible range for PD-MPC and fixed preferences for the baseline MPC

P	w_1	w_2	$w_{3,I}$	$w_{3,D}$	$w_{4,I}$	$w_{4,D}$	$w_{5,1}$	$w_{5,2}$	$w_{5,3}$	S_H
Multiplier	$\frac{20}{MO_{spill}}$	$\frac{2}{MO_{spill}}$	$\frac{20}{MO_{spill}}$	$\frac{20}{MO_{spill}}$	$\frac{20}{MO_{spill}}$	$\frac{20}{MO_{spill}}$	$\frac{20}{FWS\times F}$	$\frac{40}{FWS\times F}$	$\frac{400}{FWS\times F}$	-
Search range	0-19	0-2	0-19	0-19	0-19	0-19	1-20	←	←	RWS
Fixed-1	3	1	3	3	20	20	15	←	←	RWS_F
Fixed-2	20	5	3	3	3	3	15	←	←	RWS_F

[1] Multipliers are introduced to normalise objective values using MAVE as the denominator of each objective, and nominators are set to prevent objective values from approaching zero in any case as well as to reduce the range of the search space.
[2] ← indicates the same value as in the previous column.
[3] Fixed-1 and Fixed-2 represent MPC baselines with fixed weights and a fixed parameter set.

ages were applied to implement the numerical experiment of PD-MPC.

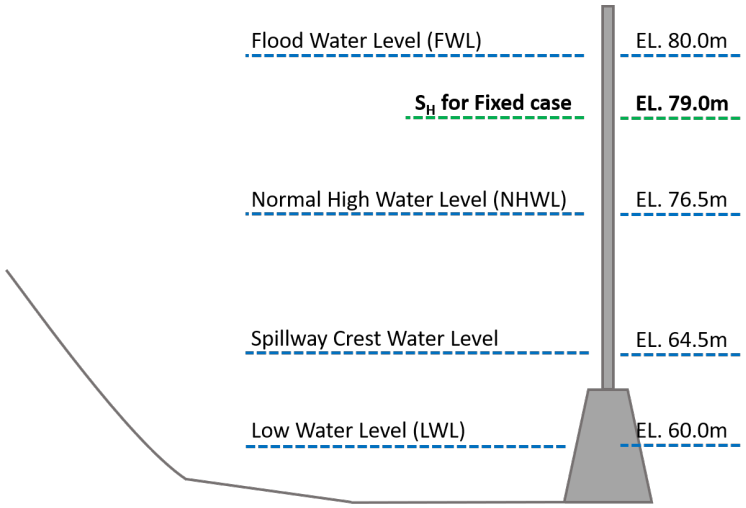


Figure 3.4 | The conceptual diagram of reservoir water levels and S_H for the Fixed case.

3.4 Results and Discussion

3.4.1 PD-MPC results

Parameterised Dynamic Model Predictive Control (PD-MPC) delivers reliable results across all events and prediction horizons. The results show only a few changes in outflow schedules, and all peak outflows remain below the peak inflow.

Fig 3.5 illustrates the optimal outflows with uncertain inflow and RWL for Event 1 across different prediction horizons. The generated optimal outflows, denoted as $\mathbf{u}^k$s, are represented by dashed lines of various colors, overlaid with the red line indicating the implemented total outflows in the figures. For instance, in Fig 3.5d, where outflows are decreasing, a 24-hour prediction horizon leads to numerous changes in outflows (see also Table 3.3). The figures show only a few changes in outflow schedules, and all peak outflows remain below the peak inflow.

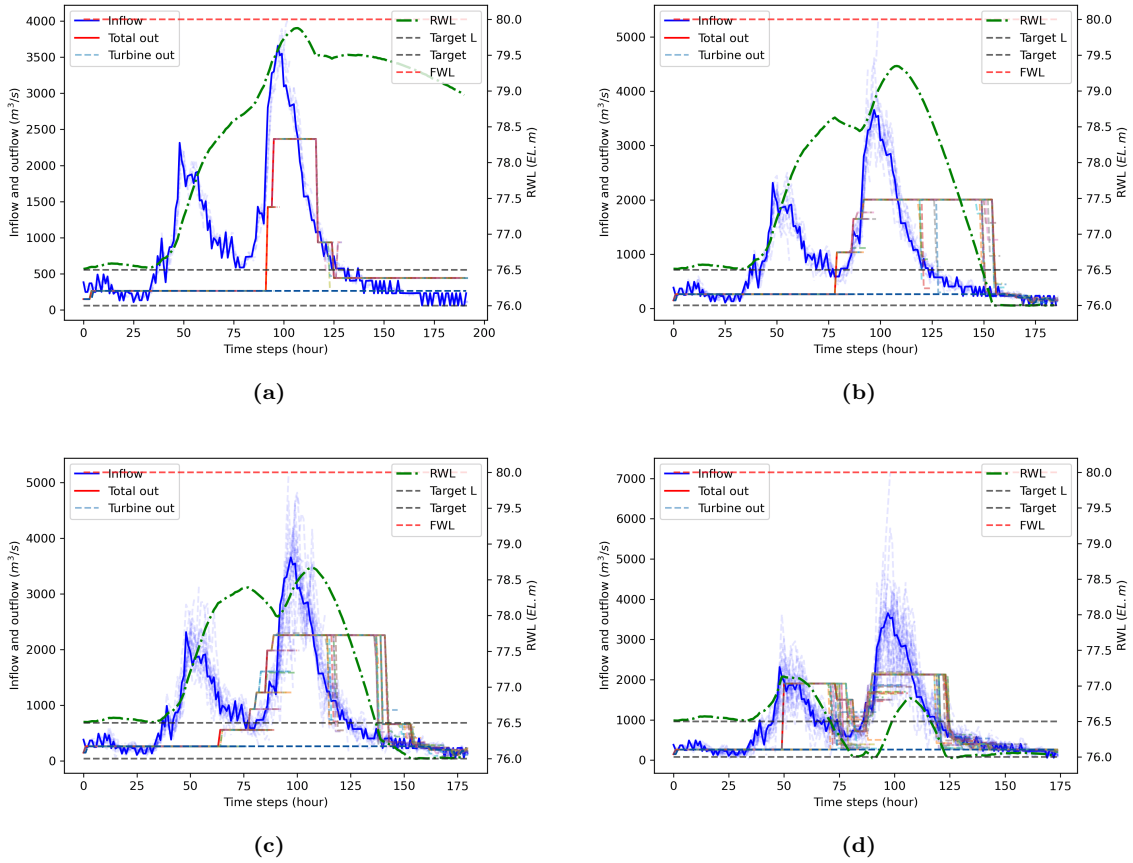


Figure 3.5 | Hydrographs under uncertain inflow and different prediction horizons for Event 1: (a) 6 hours, (b) 12 hours, (c) 18 hours, and (d) 24 hours.

The detailed result of the numerical experiment is presented in Table 3.3 for uncertain inflow and Table 3.4 for certain inflow. As expected, the performance of MPC under certainty generally surpasses the results obtained under uncertainty. Peak outflows and RWLs from uncertain inflow exceed those from certain inflow, and the reservoir should change the outflow schedule more often. Additionally, under uncertain inflow conditions for Event 1, a comparison of the results between PD-MPC and the Fixed cases is presented in Table 3.5. PD-MPC outperforms the Fixed cases for all items in the table.

The performance of this framework can be presented clearly when compared to the Fixed cases, as shown in Fig 3.6 under uncertain inflow. The Fixed-1 case assigns a relatively high weight to minimize the changes in outflow schedules, while the Fixed-2 case prioritizes minimizing peak outflow, as illustrated in Table 3.2. The results challenge our expectations. The Fixed cases do not consistently outperform PD-MPC, even in their main targets. It is impossible to assert that the Fixed-1 case, having a high weight for minimizing the changes in outflow schedules, is always superior to the other two cases, PD-MPC and Fixed-2, even concerning the changes in outflow schedules, as shown in Fig 3.6a and 3.6d. Similarly, peak outflows of the Fixed-2 case are not certainly less than the other two cases, as demonstrated

Table 3.3 | The detailed result of PD-MPC under uncertain inflow

Prediction horizon	Event	Peak outflow (m^3/s)	Peak RWL (EL. m)	Lowest RWL (EL. m)	Changes between time steps
6	1	2,367	79.88	76.50	9
12	1	2,005	79.35	76.00	9
18	1	2,296	78.67	76.00	11
24	1	2,128	77.13	75.98	9
6	2	1,919	79.82	76.13	20
12	2	1,591	79.04	76.00	8
18	2	1,273	77.56	76.00	21
24	2	1,915	76.81	76.01	30
6	3	2,105	79.81	76.34	13
12	3	1,323	78.37	75.99	10
18	3	1,245	78.19	76.00	18
24	3	1,273	78.07	76.01	38

Table 3.4 | The detailed result of PD-MPC under certain inflow

Prediction horizon	Event	Peak outflow (m^3/s)	Peak RWL (EL. m)	Lowest RWL (EL. m)	Changes between time steps
6	1	2,332	79.89	76.50	9
12	1	2,153	79.27	76.00	4
18	1	2,290	78.60	76.00	11
24	1	1,873	77.13	75.93	4
6	2	2,051	79.84	76.19	9
12	2	1,499	79.09	75.99	8
18	2	990	77.38	76.00	11
24	2	1,870	76.92	76.00	17
6	3	2,078	79.82	76.32	8
12	3	1,051	78.59	76.00	11
18	3	1,104	78.33	75.95	9
24	3	1,374	77.91	76.00	18

in Fig 3.6b and 3.6e, even though it has a high weight for minimizing peak outflow compared to the other cases. This suggests that fixed weights fail to consistently represent the operators' preferences. This can be attributed to the complexity of trade-offs for each time step and the relatively shorter prediction horizon length compared to the length of a flood event. For example, we often assume that a lower peak outflow leads to a higher peak RWL. However, the results show that this assumption does not always hold true, as demonstrated in Fig 3.6c and 3.6f. This is because the *myopic* outlook on optimal outflows at each time step could lead to suboptimal outcomes when viewed in the context of the entire flood event (Bøhn *et al.*, 2021; Morari and Lee, 1999). In our case, the evaluator effectively makes MPC have a long-term perspective compared to the Fixed weight cases. In addition, PD-MPC shows smaller peak outflow as well as lower peak RWL, especially in Fig 3.6b and 3.6c. This implies that PD-MPC utilised the storage capacity more effectively than the Fixed cases for this flood event.

In Fig 3.7a, the maximum penalty value of PD-MPC is less than 40 in any case. Additionally, the sums of penalty values for the entire flood event using PD-MPC across all prediction horizons show the lowest values. In Section 3.2.1, we described

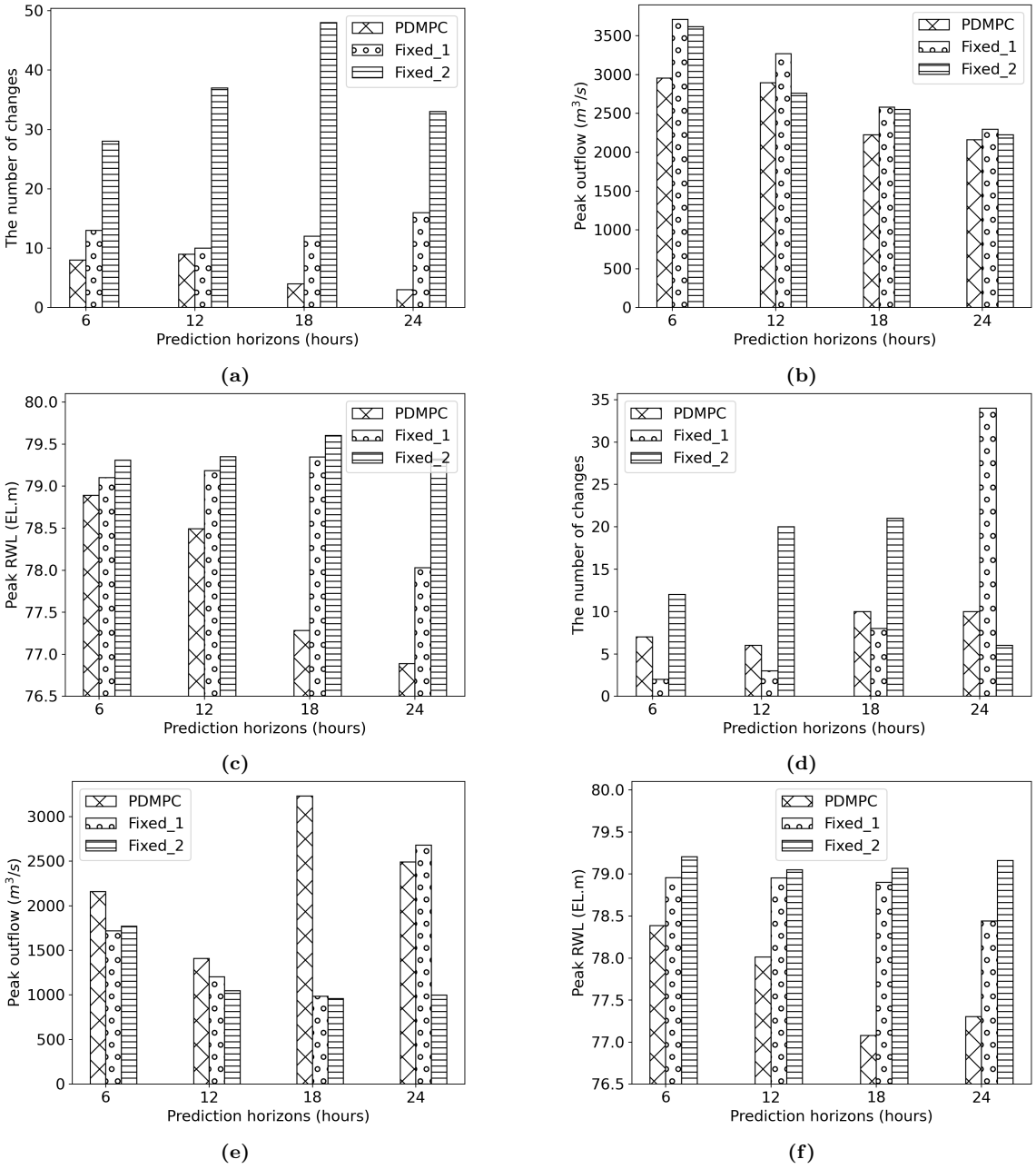


Figure 3.6 | Comparison with the Fixed cases under uncertain inflow: (a) the number of changes between consecutive outflow schedules for Event 1, (b) peak outflow for Event 1, (c) peak RWL for Event 1, (d) the number of changes between consecutive outflow schedules for Event 3, (e) peak outflow for Event 3, and (f) peak RWL for Event 3.

Table 3.5 | The comparison between PD-MPC and the Fixed cases for Event 1 under uncertain inflow

	Prediction horizon	Peak outflow (m^3/s)	Peak RWL (EL. m)	Lowest RWL (EL. m)	Changes between time steps
PD-MPC	6	2,367	79.88	76.50	9
Fixed 1	6	3,710	79.10	75.97	13
Fixed 2	6	3,617	79.31	76.31	28
PD-MPC	12	2,005	79.35	76.00	9
Fixed 1	12	3,267	79.19	76.37	10
Fixed 2	12	2,759	79.35	76.34	37
PD-MPC	18	2,259	78.67	76.00	11
Fixed 1	18	2,578	79.35	76.34	12
Fixed 2	18	2,551	79.60	76.31	48
PD-MPC	24	2,128	77.13	75.98	9
Fixed 1	24	2,259	78.03	75.95	16
Fixed 2	24	2,223	79.32	76.28	33

undesirable conditions, such as the peak outflow exceeding the maximum inflow up to the current time steps because retaining inflow is one of the reservoir's fundamental roles for flood control, RWL reaches FWL. We can see this also in Fig 3.7a. Since we assigned the 'large value' as 1,000 in (3.8) in Section 3.2.1, under normal operating conditions (i.e., without these undesirable conditions), the maximum penalty values are constrained below 1,000. Notably, both fixed cases exhibit these undesirable conditions many times, contrary to the fact that it has never occurred in PD-MPC.

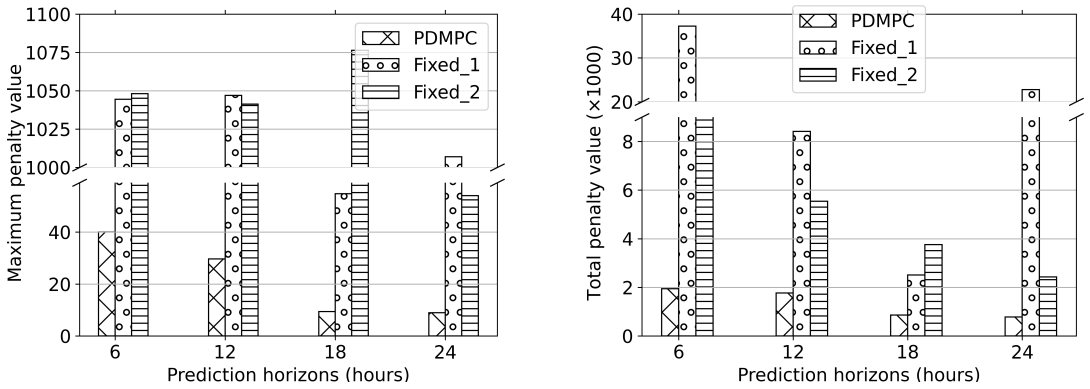


Figure 3.7 | The penalty values from the evaluator for Event 1 under uncertain inflow: (a) the maximum penalty, and (b) the total penalty (sum of penalty values).

To see the impact of the evaluator, we systematically adjusted the importance of the changes in outflows calculated at consecutive time steps and compared them with the previous one. The result is straightforward. Fig 3.8, 3.9, and 3.10 show the effect of the evaluator by changing the weight of the objective, which minimizes the changes in outflow schedules, i.e., $\tilde{J}4$ in Section 3.2.1, in the evaluator. When the evaluator assigns a higher weight to $\tilde{J}4$ marked as 'Higher' in figures, it leads to fewer changes in outflows but comes at the expense of peak outflow and RWL. PD-

MPC with the low $\tilde{J}4$ case marked as 'Lower' in the figures shows a greater number of changes. Given the unambiguous response from PD-MPC, it can be inferred that the evaluator effectively affects the selection of the optimal weight/parameter set.

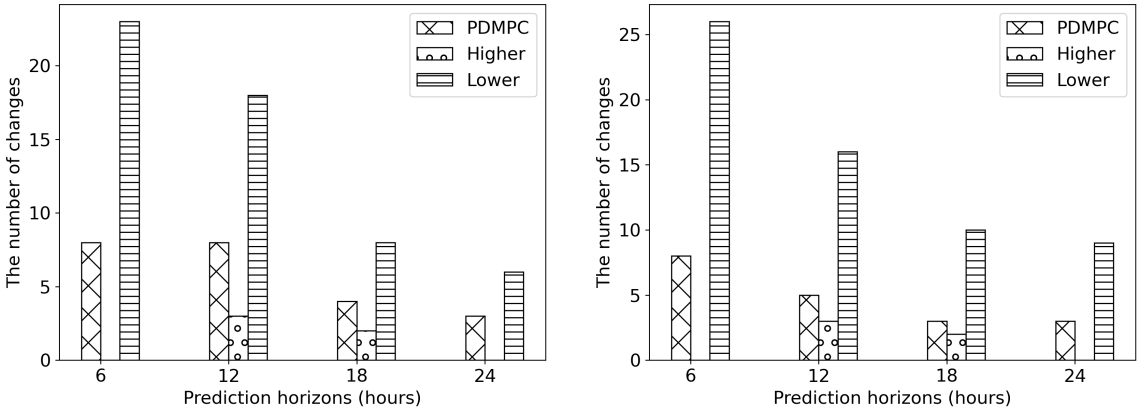


Figure 3.8 | The number of increases with different evaluator settings for Event 1: (a) under uncertain inflow, and (b) under certain inflow.

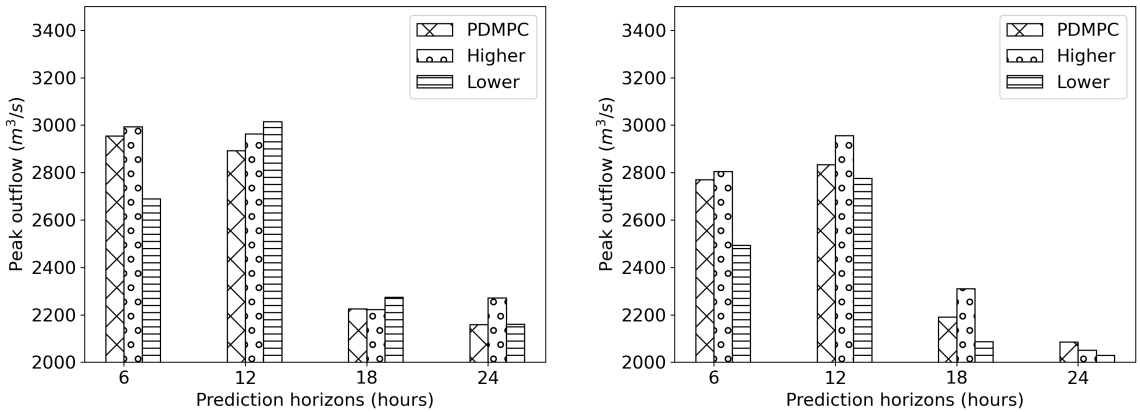


Figure 3.9 | The peak outflow with different evaluator settings for Event 1: (a) under uncertain inflow, and (b) under certain inflow.

3.4.2 Parameters as elements of the preference

In Section 3.1, we discussed that some researchers had focused on finding appropriate weights of objectives when defining preferences in a multi-objective setting. To demonstrate the parameters should also be regarded as important components of the preference, we conducted an experiment for Event 1, where we fixed S_H to the storage level at EL. 79.0m and compare it with the results of PD-MPC where S_H is a dynamic parameter to be optimized.

PD-MPC with varying S_H, outperforms PD-MPC with fixed S_H, which generally considers only the combination of weights as the representative of the preference, as

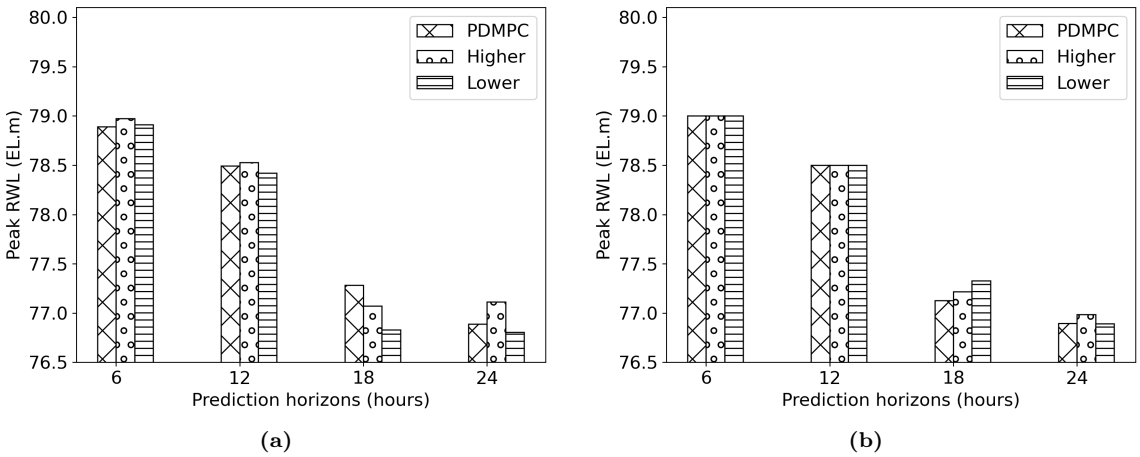


Figure 3.10 | The peak RWL with different evaluator settings for Event 1: (a) under uncertain inflow, and (b) under certain inflow.

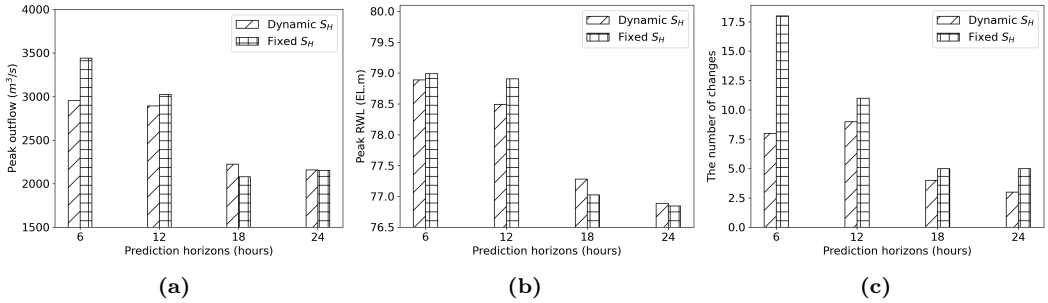


Figure 3.11 | Comparison between dynamically changing S_H and fixed S_H under uncertain inflow: (a) peak outflow, (b) peak RWL, and (c) the number of changes between consecutive outflow schedules.

illustrated in Fig 3.11. PD-MPC shows low peak outflows and RWL with the lowest number of changes simultaneously. It seems reasonable to assume that operators would prefer the evaluator to have a lower S_H because it results in a lower peak RWL. However, this is not the case because a lower S_H can lead to a higher peak outflow. PD-MPC, by varying S_H, effectively achieves lower peak outflows and RWLs simultaneously. This means an adaptively changing S_H helps operators utilise the reservoir storage more efficiently. For instance, in PD-MPC with the parameter S_H, spillway outflow tends to commence earlier compared to PD-MPC without S_H, achieved by maintaining a low S_H before the water level rises significantly. Consequently, there is an impact on reducing both the peak outflow and the peak RWL. In addition, when RWL is close to S_H, and there is a sudden increase in inflow, fixing S_H would result in a substantial and abrupt increase in outflow, even if there is some storage available between S_H and FWL. In contrast, with adaptively changing S_H, PD-MPC chooses to increase S_H to utilise the remaining storage, instead of resorting to a drastic and sudden increase in outflow.

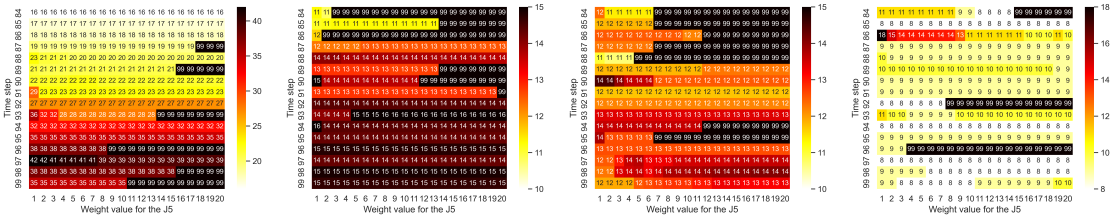


Figure 3.12 | The colour represents the penalty value calculated at each time step from $k = 84$ to $k = 99$, with different weight values for the $J5$ while inflow increases steeply for Event 1. The prediction horizons are for: (a) 6 hours, (b) 12 hours, (c) 18 hours, and (d) 24 hours.

3.4.3 The best weights/parameters vary with time

In Section 3.2.2, we discussed the complexity of trade-offs among objectives, which led us to assume that the operator preferences among them can be dynamic.

Fig 3.12 illustrates the variation in penalty values as a weight only for the objective function related to RWL ($J5$) changes within our search range, while other objectives' weights are the same as the weights from PD-MPC, resulting in Fig 3.5. For example, when a weight value for the $J5$ is two, it denotes that $w_{5,1}$ is $2 \times \frac{20}{FWS \times F}$, $w_{5,2}$ is $2 \times \frac{40}{FWS \times F}$, and $w_{5,3}$ is $2 \times \frac{400}{FWS \times F}$ (see Table 3.2). Bright colors in the figure represent relatively low penalty values. The value of 99 in black color does not mean the penalty value is precisely 99, but the actual value is much higher. For simplicity, we convert all high penalty values into 99, which indicates undesirable conditions such as when the outflow exceeds the peak inflow up to the current time step or the RWL approaches FWL. Since the evaluator computes a nonlinear objective with nonlinear system simulations (see Fig 3.1), the same penalty values, which also mean the same control inputs, are observed in many cases for different weight sets. Despite that, it is evident that the best weight set changes at every time step. For example, in Fig 3.12c, a low weight on $J5$ when $k < 90$ produces the lowest total objective value, but when $k = 90$, a high weight results in the lowest total objective value. This dynamic nature of optimal weight sets underscores the need for a flexible and adaptable approach in reservoir flood control. Additionally, it effectively conveys that the outputs from an MPC formulation with constant weights can, at best, be equal to those obtained through PD-MPC.

3.4.4 Discussion

It appears that the detailed design of practical objectives for reservoir flood control has not been thoroughly presented (Ritter *et al.*, 2020; Pianosi *et al.*, 2020), and it seems complicated to formulate practical objectives with only linear equations, as we described in Section 3.2.1. The receding optimal control of an MPC problem with numerous nonlinear objectives and constraints can become intractable to solve online (Berberich *et al.*, 2022; Allgower *et al.*, 2004). Given that preferences can be expressed as objective weights to represent their relative importance (Wang *et al.*, 2017), the optimal preferences (weights) should adapt to varying hydrological con-

ditions of the reservoir. However, it seems that the operators' preferences have not been adequately incorporated, and constant weights/parameters have been used to optimize reservoir flood control (Uysal *et al.*, 2018b; Qi *et al.*, 2017; Breckpot *et al.*, 2013a; Aydin *et al.*, 2022; Xu *et al.*, 2011; Van Overloop, 2006).

In this chapter, we harness the advantages of solving linear MPC problems by parameterizing the nonlinear and dynamic preferences around the different operating points. This allows us to optimize weights/parameters and control inputs simultaneously. Our approach is also in the spirit of multi-objective MPC methods for linear systems (Bemporad and De la Peña, 2009) where a time-varying, state-dependent decision criterion can be taken into account using parametric optimization. First, we present objectives for practical flood control in detail. Some of them have not been extensively covered in prior literature despite their significant importance in practice, such as minimizing the changes in outflow schedules. Subsequently, we categorize these objectives into linear and nonlinear ones following the parameterization of the operator's preference, i.e., the weights of objectives and parameters. We employ the Genetic Algorithm (GA) to optimize nonlinear objectives and constraints to derive the optimal weights/parameters of the linear MPC formulation at each time step. We refer to this framework as a Parameterised Dynamic Model Predictive Control (PD-MPC).

Through our numerical experiment, we demonstrated that PD-MPC shows robustness to the inflow uncertainty. It is important to note that the PD-MPC framework does not directly handle inflow uncertainty. Instead, we rely on the robustness of the receding horizon MPC approach to address uncertainty and produce reliable results (Schwenzer *et al.*, 2021; De Nicolao *et al.*, 1996). Our results reveal that PD-MPC outperformed MPC formulations with fixed weights/parameters, even when these fixed weights/parameters were specifically designed for individual objectives. We showed that the weights/parameters in MPC formulations and the weights of objectives should also vary dynamically to adapt to changing conditions. PD-MPC then effectively adapted to the changing hydrological conditions and continuously updated the weights/parameters. This adaptability allowed it to make optimal decisions in real-time, resulting in better overall performance in terms of peak outflows, reservoir water levels, and the number of changes in outflow schedules. This adaptability is the key factor in its superior performance and is also essential for generating optimal control inputs reflecting the dynamic characteristics of the operator preferences.

Nevertheless, this research has some limitations. First, the MPC formulation applied here does not consider the final states of the system. MPC with policy search algorithms (Song and Scaramuzza, 2022b) or a value function produced by a Reinforcement Learning (RL) model (Arroyo *et al.*, 2022a) can give a chance to find the approximation of the terminal cost such that our MPC framework can consider the whole period of a flood event. To the best of our knowledge, this approach has never been applied to a reservoir system for the purpose of flood control. The second limitation is related to uncertainty. We showed that PD-MPC can generate acceptable control inputs under uncertainty from the inherent feedback mechanism

of a receding horizon implementation; however, it is advisable to explore PD-MPC with stochastic/robust MPC (Saltık *et al.*, 2018) or Learning-based MPC (Hewing *et al.*, 2020a) to ensure robustness by considering uncertainty explicitly. We evaluate methodologies to generate uncertain scenarios and explicitly account for uncertainty in an MPC approach from a risk management perspective in Chapter 6. The third limitation pertains to our numerical experiments, which utilised three historical flood events. Our study events present a wide range of challenging operational conditions that demonstrate our method's practical utility, e.g., long event periods and two or three peaks. However, to establish broader applicability, numerical experiments under the 200-year flood and PMF are necessary, despite a limitation in obtaining the hydrograph of those extreme conditions.

Finally, we did not consider the downstream impact of reservoir outflows as well as the upper reservoir for simplicity. This should be considered important in practical operations. The downstream impact can be considered using routing models or various heuristics suggested in many studies (Hsu and Wei, 2007; Peng *et al.*, 2017; Le Ngo *et al.*, 2007). In chapter 6, the downstream impact is incorporated into a stochastic MPC approach. The entire reservoir system, including the upper reservoirs, needs to be explored, and we expect that PD-MPC could be applied to the joint optimization problem, though this may be a topic for further study.

3.5 Conclusions

This chapter addresses the limitations of existing reservoir control approaches from a practical point of view, highlighting the limitations of currently employed optimization approaches to take into account specific operators' preferences, which may change over time. Considering a multi-objective optimal control problem, we assume dynamic preferences of operators on the importance of different objectives. We then show that this translates to the dynamic characteristics of weights/parameters in the multi-objective optimisation. We then propose a PD-MPC framework as a parameterized linear MPC with dynamic optimization of weights/parameters via a model-based learning concept. We applied this methodology to the Daecheong reservoir, verified the dynamic-preference assumptions, and tested this framework's effectiveness. This chapter lays the foundation for developing more adaptive decision-making frameworks in reservoir flood control and other related fields, where the actual set of operators' preferences may be dynamic depending on the system state and anticipated external inputs.

Despite the mentioned limitations of this chapter, we think it allows for making one step towards the wider adoption of optimization approaches to real-time reservoir flood control. The presented methodology is, of course, not aimed at replacing manual operation but rather gives instruments for reducing operators' stress in critical situations and ultimately enhancing their ability to make better decisions.

Data Availability for this chapter

The data and codes for this chapter are available at https://doi.org/10.4121/9a6a0464-2981-470a-8d7a-48c7e7fff27d. All data can be used under the CC-BY-4.0 licence. The hydrological and operational data of the Daecheong reservoir were obtained from K-water's website (http://kwater.or.kr).

4

Learning-based Explicit and Switched MPC Approaches[1]

Summary

Parameterised Dynamic Model Predictive Control (PD-MPC) is effective for solving nonlinear multi-objective optimization problems with dynamic preferences. However, as stated in Chapter 3, it is computationally expensive when implemented online to find weights and parameters that represent dynamic operator preferences. The work in this chapter aims to develop tractable real-time flood control strategies that maintain performance while reducing computational complexity. We propose two data-driven approaches based on Model Predictive Control (MPC): (1) an explicit MPC using deep neural networks to directly determine optimal outflow schedules, and (2) a switched MPC that produces optimal weights of objectives based on hydrological conditions. Both methods leverage offline learning from an online Parameterized Dynamic MPC framework incorporating state-dependent weights. We tested these approaches on South Korea's Daecheong multipurpose reservoir using historical flood events with various patterns. The explicit MPC demonstrated reliable performance under conditions similar to its training data. However, it showed frequent changes in outflow schedules and constraint violations for scenarios outside training data. In contrast, the switched MPC maintained robustness across all test scenarios due to a linear optimization process in a receding horizon manner, though with slightly reduced performance compared to the explicit MPC under scenarios inside the range of training data. Most significantly, both approaches reduced computation time from approximately 10 minutes to less than one second, making real-time implementation feasible. This dramatic improvement enables computationally cheap and feasible decision-making during rapidly evolving flood events while maintaining near-optimal control performance.

[1]This chapter is based on the publication: Koo, Ja-Ho, Ali Moradvandi, Edo Abraham, Andreja Jonoski, and Dimitri P. Solomatine. (2025). Flood Control of Reservoir Systems: Learning-based Explicit and Switched Model Predictive Control Approaches. *PLOS Water*. **(accepted)**

4.1 Introduction

Floods are among the most destructive natural disasters, resulting in substantial economic and environmental damage across affected communities. Therefore, flood control is one of the main purposes of most reservoirs. Reservoir flood control, which significantly impacts watershed management, has attracted research attention for decades (Giuliani *et al.*, 2021; Labadie, 2004). However, rapidly changing hydrological conditions, highly uncertain forecasts, and changes in decision makers' preferences (Lee *et al.*, 2017; Cheng and Chau, 2001) make reservoir flood control challenging. Many researchers have suggested optimization-based control methods because flood control is a control problem, as its name implies (Castelletti *et al.*, 2023). Various optimization-based methodologies have been proposed to address challenges of real-time optimal flood control, particularly subject to uncertainty, either in rainfall or inflow forecasts (Jain *et al.*, 2023; Myo Lin *et al.*, 2020; Breckpot *et al.*, 2013a). The complexity of reservoir flood control arises from the system's nonlinearity, which stems from intricate relationships among different objectives that operators want to achieve, including minimizing changing outflow schedules, minimizing peak outflow, minimizing Reservoir Water Level (RWL), and minimizing opening and closing of spillway gates.

The flood control problem can be formulated as a multi-objective problem involving outflow schedules, RWL, and other factors. Various approaches have been proposed to effectively solve this multi-objective optimization problem of reservoir flood control. In classical Model Predictive Control (MPC) approaches, the relative importance of objectives is assumed to be constant throughout the flood event (Myo Lin *et al.*, 2020; Chiang and Willems, 2015). Therefore, optimal weights, as significant performance factors, should be determined properly and in advance. However, the relative importance of objectives can vary with the hydrological conditions in practice. Under this dynamic preference scenario, reflecting the operators' preferences in the weights of the objectives (Kong *et al.*, 2021), the optimal outflow sequence can be chosen from the Pareto front (Moridi and Yazdi, 2017). This approach is flexible in the decision-making process, either by operators or through methods such as multi-criteria decision-making at each time step. The performance of this method, however, depends on the effective derivation of the Pareto fronts in real-time. Nonetheless, generating a Pareto front for the real-time application can be intractable when a nonlinear and nonconvex multi-objective problem has to be solved. Because the search spaces for exploration can grow exponentially (Liu *et al.*, 2023a; Pan *et al.*, 2014), getting optimality guarantees for solutions can also be difficult, and it can take longer than the available decision window during flood events.

In the previous chapter, a Parameterised Dynamic Model Predictive Control (PD-MPC) framework to derive optimal outflow schedules in reservoir flood control problems has been investigated. The PD-MPC formulation explicitly considers the dynamic characteristic of the relative importance of objectives depending on the situation via time-varying relative weights for the different objectives. In this approach,

we separated the weight optimization and control action optimization in an alternating fashion until convergence. A Genetic Algorithm (GA) was employed to optimize objective weights and system parameters at the same time using the nonlinear system model and objectives, while the linear MPC problem was solved to optimality for a given weight set. The approach utilizes a single indicator, i.e., the absolute performance evaluator, to obtain optimal weights and parameters, along with optimal control inputs, at each time step. To the best of our knowledge, it was the first attempt to decide weights at each time step depending on hydrological conditions in real-time MPC-based reservoir flood control. Nonetheless, the approach still suffers from the inefficiency of online implementation due to the employed gradient-free algorithm requiring considerable time to achieve sufficient convergence. This can be limiting for the PD-MPC approach in real-time control.

In recent years, data-driven models like Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL) have emerged as effective tools to build efficient meta-(surrogate) models and thus reduce online computational complexity (Karg and Lucia, 2020; Nubert *et al.*, 2020; Hertneck *et al.*, 2018). These models can be trained offline to map the nonlinear relationships between inputs and desired outputs (Bemporad, 2023; Li *et al.*, 2021), enabling efficient online implementation in various fields. For example, an Artificial Neural Network (ANN) was trained offline using the input and output data of a nonlinear MPC controller for a power converter and a robot, allowing the ANN to be implemented on hardware as an efficient explicit MPC that directly maps states to optimal control inputs (Wang *et al.*, 2021; Hirose *et al.*, 2018). In Cai *et al.* (2021); Arroyo *et al.* (2022b), hybrid RL-MPC approaches are proposed where the original MPC non-linear program is truncated to be one step ahead, with cost-to-go functions of the MPC replaced with the value function of the following state from RL. In Song and Scaramuzza (2022a), Multi-Layer Perceptron (MLP) policies are learned offline from MPC to automatically choose hard-to-optimize decision parameters, and so to parameterize a linear MPC to be solved online.

In real-time reservoir operations, these data-driven approaches are particularly valuable because they can address a practical operational challenge, i.e., computational time under a multi-objective optimization problem. Despite many studies in various fields, it is difficult to find a study on adopting data-driven models for an optimization problem with many objectives and dynamically changing weights of objectives. Due to the complex relationship among objectives, dynamically changing weights can make it difficult to learn the relationship between states and control inputs or optimal weights in a data-driven model. The conventional approach, which collects training data directly from simulation environments and trains an ANN or RL model, is not suitable for this case. This is because, for example, there can be many weight sets which can produce an optimal control input.

In this chapter, we leverage data-driven approaches to avoid difficulties in dealing with changing weights and solving a highly nonlinear problem online. From a control point of view, in this chapter, we investigate explicit MPC and switched linear MPC approaches using data-driven models trained on the outputs of PD-MPC. The

switched MPC framework allows the MPC system models to switch based on system states and predefined switching rules (Liberzon, 2003), requiring prior knowledge of these rules (Zhang *et al.*, 2016). This allows simpler linear MPC problems to be solved online by selecting the appropriate model based on the initial state and the switching rules as a function of the states. In contrast, explicit MPC approaches compute control actions offline for all states by solving an optimal control problem for each state space partition (Alessio and Bemporad, 2009), establishing either a mapping table or piecewise affine functions to relate states and control actions.

For our reservoir optimal flood control problem, we develop explicit MPC and switched MPC approaches from the PD-MPC results computed under a large set of inflow scenarios. A DNN model is trained offline to directly map states to control inputs as an explicit MPC controller. Additionally, a data-driven model is used to learn a mapping from system states to optimal parameters (i.e., state-dependent switching rule for optimal objective weights and other system parameters to use in a linear MPC online); this latter approach can be formulated as a switched linear MPC problem.

The novel contributions of this chapter are: i) we can obtain a consistent optimal weight set for a specific state by adopting the ℓ_1-norm, ii) we present approaches for adopting explicit and switched MPC approaches using data-driven models, and iii) we demonstrate the advantages and disadvantages of switched and explicit MPC approaches for a multi-objective optimization problem with dynamically changing weights. To the best of the authors' knowledge, this is the first application and comparison of learning-based explicit and switched MPC for reservoir management.

This chapter is organized as follows. In Section 4.2, the PD-MPC framework and the study area are discussed. The control objectives and system constraints used in this research are also formulated. Section 4.3 presents the methodology and experimental setup. Focusing on approaches for data acquisition using PD-MPC, we describe the proposed data-driven surrogate models for explicit MPC and switched MPC. Section 4.4 presents results, discussion, assessing the efficiency and characteristics of the surrogate models based on the two considered approaches. Finally, conclusions are drawn.

4.2 Problem Formulation

4.2.1 PD-MPC design

As depicted in Fig 3.1, the PD-MPC approach dynamically optimizes weight parameters through the GA based evaluator, which explores potential weight and parameter spaces by mimicking natural selection and mutation processes. Although the GA was successfully used for the numerical experiments, another popular approach is Bayesian Optimization (BO) algorithms, such as the Tree-structured Parzen Estimator (TPE) algorithm (Bergstra *et al.*, 2011; Watanabe, 2023). This algorithm is considered efficient because the TPE algorithm explores solutions based on a probabilistic model and is easy to parallelize (Bergstra *et al.*, 2013).

In this chapter, we employ the TPE algorithm, a BO algorithm that makes use of tree-structured search spaces. At each time step, the TPE algorithm derives the optimal parameter vector by learning a surrogate probabilistic model for parameters and their associated objective values, i.e., $\mathbf{z}$ and E online. The TPE algorithm utilizes a Kernel Density Estimator, especially a Parzen estimator, as the surrogate model. It employs the Expected Improvement (EI) and weighted random sampling as the acquisition method to explore promising regions of parameters z and update the surrogate model. The exploration terminates when either the specified number of searches is reached, or the evaluation index falls into a predefined threshold. Thereafter, the best parameters and their associated evaluation value are returned. The corresponding control inputs, $\mathbf{u}^{k*}$, can then be computed. Therefore, the TPE algorithm is utilized at each time step in order to determine $\mathbf{z}^*$, minimizing the return (penalty) value from the absolute performance evaluator, E.

4.2.2 Reservoir flood control

A total of 13 linear and nonlinear objectives are suggested to describe the practical reservoir flood control problem, including minimizing peak spillway outflow, gate operations, and changes in outflow schedules in the previous chapter. In this chapter, we aim to approximate this rather complex PD-MPC problem derived from these considered objectives. To do so, the switched MPC and explicit MPC approaches are taken into account. To simplify the problem, three objectives for linear MPC and six nonlinear objectives to formulate the evaluator are then considered. This simplification is introduced because it may be difficult to show the performance of our approaches due to the highly complicated relationships among the objectives when considering all 13 of them. Moreover, it would be reasonable to focus on objectives only directly relating to flood control in practice, as other objectives, such as minimizing environmental effects and securing water for demands, naturally receive lower priorities during flood events.

The three objectives for linear MPC are related to RWL, peak outflow, and outflow schedule changes. Reservoir operators should maintain RWL within a target range not only to ensure the dam's safety but also to secure enough water to satisfy demands. Therefore, three target water levels are defined: (i) The lower target level, to secure water, so it constrains RWL from decreasing a certain level; (ii) The upper level, to secure the flood control capacity between this level and the FWL; and (iii) the highest level for dam safety, considering uncertainty. Furthermore, to reduce downstream flood risk, minimizing peak outflow is also an essential factor, as the high outflows directly affect downstream water levels. Meanwhile, because each outflow schedule is shared with other flood control organizations, the previously shared outflow schedule should be maintained consistently with the minimum number of changes possible. Moreover, limiting the frequency of operations of spillway gates is advantageous to prevent wear and malfunction. This can be achieved implicitly by minimizing changes in consecutive outflow schedules. Additionally, a soft constraint is added to constrain opening spillway gates before the total outflow exceeds the turbine capacity.

Therefore, the overall objective function can be written as follows:

$$\min_{\mathbf{u}} J^k = \min_{\mathbf{u}} z_1 J_1 + z_2 J_2 + z_3 J_3 + z_4 J_4, \tag{4.1}$$

in which

$$J_1 := \max_{t \in k,\ldots,k+N-1} O_{spill,t} \times N \times \alpha_1^{MAVE}, \tag{4.2a}$$

$$J_2 := \sum_{t=k}^{k+N-1} (\Delta S_t^1 + \Delta S_t^2 + \Delta S_t^3) \times \alpha_2^{MAVE}, \tag{4.2b}$$

$$J_3 := \sum_{t=k}^{k+N-1} (\Delta O_t^I + \Delta O_t^D) \times \alpha_1^{MAVE}, \tag{4.2c}$$

$$J_4 := \sum_{t=k}^{k+N-1} \Delta O_t^{com} \times \alpha_1^{MAVE}, \tag{4.2d}$$

in which z_1, z_2, z_3, and z_4 are the objective weights. To scale the different objective values, constants, i.e., α^{MAVE} here, based on the Maximum Allowed Value Estimate (MAVE) values of the associated objectives are applied (Aydin *et al.*, 2022; Xu *et al.*, 2011). The MAVE values are the spillway outflow capacity C_{spill} for J_1, J_2, and J_4, the storage amount between FWL and LWL for J_2, respectively. The optimization of these objectives is subject to the constraints:

$$O_{spill,t} + O_{turb,t} - O_{total,t} = 0, \tag{4.3a}$$

$$O_{total,t} - \mathrm{C}_{turb} + \Delta O_t^{com} - O_{spill,t} = 0, \tag{4.3b}$$

$$O_{total,k}^k - O_{total,k}^{k-1} = 0, \tag{4.3c}$$

$$\Delta S_t^1 + S_U - S_t \geq 0, \tag{4.3d}$$

$$\Delta S_t^2 + S_t - S_L \geq 0, \tag{4.3e}$$

$$\Delta S_t^3 + S_H - S_t \geq 0, \tag{4.3f}$$

$$\Delta O_t^I - \Delta O_t^D + (O_{total,t}^k - O_{total,t}^{k-1}) \times \sigma_t = 0, \tag{4.3g}$$

$$\Delta O_t^I,\ \Delta O_t^D,\ \Delta O_t^{com},\ \Delta S_t^1,\ \Delta S_t^2,\ \Delta S_t^3 \geq 0, \tag{4.3h}$$

in which k represents each time step for the defined receding horizon MPC, where t at time step k ranges from k to $k+N-1$. Therefore, the optimal control input vector obtained from the proposed linear MPC over the time span $t = k, k+1, ..., k+N-1$ can be written as follows:

$$\mathbf{u}^k = \{O_{total,k}^k, ..., O_{total,k+N-1}^k, O_{spill,k}^k, ..., O_{spill,k+N-1}^k, O_{turb,k}^k, ..., O_{turb,k+N-1}^k, \}, \tag{4.4}$$

where C_{turb} denotes the turbine outflow capacity, and $O_{total,t}^k$, $O_{spill,t}^k$, and $O_{turb,t}^k$ express the total outflow, the spillway gate outflow, and the turbine outflow at time t and time step k, respectively. We refer to $\mathbf{u}^k$ as an outflow schedule at time step k. The control horizon is N. For simplicity, we omit a superscript k, e.g., J_1^k at time step k denotes J_1, except when it is essential, e.g., in (4.3c) and (4.3g)

because these equations compare current outflows with outflows at previous time steps. ΔO_t^I, ΔO_t^D, ΔS_t^1, ΔS_t^2, ΔS_t^3, and ΔO_t^{com} are considered as slack variables. ΔO_t^I and ΔO_t^D express the amounts of increase and decrease between consecutive outflow schedules at time t, respectively. ΔS_t^1, ΔS_t^2, and ΔS_t^3 are also introduced to penalize the violations of target reservoir storage, i.e., the highest target water level, S_H, the upper target water level, S_U, and the lower target water level, S_L.

To address the complementarity between outflows via turbines and spillway gates, ΔQ_t^{com} is also taken into account. In general, this constraint is formulated by a nonlinear complementarity operator that can be written as Powell *et al.* (2016):

$$O_{spill} \times (O_{turb} - \mathrm{C}_{turb}) = 0. \tag{4.5}$$

This nonlinear constraint can be linearised by adding a slack variable ΔO_t^{com} in (4.3b) and using the penalty function approach to integrate constraint violations in the objective J_4. As an example, if $O_{total,k} \leq \mathrm{C}_{turb}$, then $O_{spill,t} = 0$ and ΔO_t^{com} is minimized by the difference between the total outflow and the turbines' capacity outflow. Conversely, if $O_{total,k} > \mathrm{C}_{turb}$, then $O_{turb,t} = \mathrm{C}_{turb}$ and ΔQ_t^{com} becomes zero. Since this is a soft constraint and this value is independent of other objectives, i.e., there is no trade-off with other objectives; we therefore deem J_4 as not dynamically responding to the relative preferences among objectives, and so z_4 is set to a constant.

To assign a higher penalty for changes in the near future, a time-dependent weight function, σ_t, is defined as follows:

$$\sigma_t = \begin{cases} 1/(t-k) & \text{if } t \geq k+1, \\ 1 & \text{otherwise.} \end{cases} \tag{4.6}$$

Even though there are nonlinear elements in the reservoir system, such as reservoir storage volume-water level relationship and water level-spillway outflow capacity relationship, we consider only the linear parts of the reservoir system by using reservoir volume directly as state instead of RWL and by assuming that errors from nonlinear factors can be included as additive errors within inflow uncertainty. These can be written as follows:

$$S_{t+1} = S_t + I_t - O_{total,t} \tag{4.7a}$$
$$\text{s.t. } 0 \leq O_{spill,t} \leq \mathrm{C}_{spill}, \tag{4.7b}$$
$$0 \leq O_{turb,t} \leq \mathrm{C}_{turb}, \tag{4.7c}$$
$$O_{min} \leq O_{turb,t} + O_{spill,t}, \tag{4.7d}$$
$$\mathrm{LWL} \leq \mathrm{RWL}_t \leq \mathrm{FWL}, \tag{4.7e}$$

where S_t, I_t, and O_t denote the reservoir storage, inflow, and outflow at time t, respectively. C_{spill} expresses the outflow capacity via spillway gates. The constraints are set for the spillway and turbine discharge capacity, the LWL, and the FWL. The minimum water supply should be more than the demands, O_{min}, as expressed in (4.7d). In fact, the total demand varies monthly; however, in this chapter, the annual average, $52\,\mathrm{m}^3/\mathrm{s}$, is used for all cases. Nonlinear elements, such as the relationship between RWL and storage volume, are used for the absolute performance evaluator.

Table 4.1 | Search spaces for weights in this research

	z_1	z_2	z_3
Search space	[0,1]	[0,1]	[0,1]
Search step	0.01	0.01	0.01

4.2.3 Online selection of the best weight set

Since the weight for the soft constraint does not need to change dynamically because its value is independent of other objectives and constraints, as mentioned earlier, we set z_4 to a constant with a reasonably small value for simplicity. Therefore, the optimal set to be determined consists of three weights, as follows:

$$\mathbf{z}^* = \{z_1^*, \quad z_2^*, \quad z_3^*\}. \tag{4.8}$$

Using TPE, we can explore search spaces, such as those in Table 4.1, at each time step of the receding horizon control implementation. The search space is quantized in the normalised range $[0, 1]$ with steps of 0.01 interval to sufficiently explore all the objectives' space. Reasonably, $\mathbf{z} = \{0, 0, 0\}$ is excluded from the search space. As mentioned, the evaluator plays an important role in providing the best weight set at each time step according to (3.12a) and Fig 3.1. We formulate the absolute performance evaluator consisting of six nonlinear objective functions as follows:

$$\min E = \min(E_1 + E_2 + E_3 + E_4 + E_5 + E_6), \tag{4.9}$$

where,

$$E_1 = \exp(p \times \max O_{spill,t}/\mathrm{C}_{spill}) - 1, \tag{4.10a}$$

$$E_2 = q \times [\max \mathrm{RWL} - \mathrm{NHWL}]^+ + q \times [\mathrm{RWL}_{t+N-1} - \mathrm{RWL}_t]^+ + v \times [\max \mathrm{RWL} - S_H]^+ + (q+v) \times [S_L - \min \mathrm{RWL}]^+ + q \times [\mathrm{RWL}_{t+N-1} - \mathrm{RWL}_t]^+, \tag{4.10b}$$

$$E_3 = \max \gamma_t \times \frac{w}{\exp(2 \times t)}, \qquad \gamma_t = \begin{cases} 0 & \text{if } |O^k_{spill,t} - O^{k-1}_{spill,t}| = 0, \\ 1 & \text{otherwise,} \end{cases} \tag{4.10c}$$

$$E_4 = \begin{cases} \exp(2 \times p \times \max_t(O_{spill,t+1} - O_{spill,t})) - 1 & \text{if } O_{spill,t} \geq \max I_t, \\ 0 & \text{otherwise,} \end{cases} \tag{4.10d}$$

$$E_5 = p \times \sum_t \kappa_t, \qquad \kappa_t = \begin{cases} 1 & \text{if } O_{spill,t} > 0 \ \& \ O_{spill,t+1} = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4.10e}$$

$$E_6 = \begin{cases} l & \text{if } \exists t \text{ where } O_{spill,t} \geq 0 \ \& \ O_{turb,t} < \mathrm{C}_{turb}, \\ 0 & \text{otherwise,} \end{cases} \tag{4.10f}$$

where $[.]^+$ is the ramp function, i.e., $[A]^+ = \frac{A+|A|}{2}$, and p, q, v, and w are constants to be used to normalize each objective value between zero and 10 over typical situations such as when RWL is lower than S_H and O_{spill} is within the peak inflow of the target flood, with values of 2.4, 10, 5, and 75, respectively. l should be very large, chosen here as 5000, to avoid opening the gates when the outflow capacity of turbines is not fully utilized. γ_t and κ_t are used for the values from conditional expressions.

Among the objectives in (4.10), E_1 to E_3 have the same meaning as J_1 to J_3, and E_6 as J_4 in (4.2), but they are more freely configured using exponential functions and conditional equations. Moreover, E_4 and E_5 are considered in the evaluator to account for preferences for i) minimizing outflow schedule changes inside a prediction horizon and ii) minimizing the number of gate closures in a prediction horizon after opening action. These objectives may not be as critical as E_1, E_2, and E_3 but are related to operators' preferences, as presented in Chapter 3.

It is worth noting that the objective values for J_1 and J_3 in the linear MPC can be normalized to be of the same scale by multiplying N for J_1. Additionally, changes are smaller than peaks, i.e., $\Delta O \leq \max O$. Furthermore, the effect of long-term changes is diminished by σ_t in (4.6). Therefore, the MAVE values can not scale the objective values of the linear MPC exactly. In contrast, the evaluator's objective E_3 in (4.10d) holds equal importance to the other objectives within the evaluator, as various exponential and conditional equations are used to scale objective values, ensuring they reach the same values in worst-case scenarios. Moreover, E_3 has the maximum value when there is at least one change in the outflow schedules.

4.2.4 Study flood events

Hourly inflow data presented in Chapter 2 is noise-corrupted due to measurement errors. To detect and remove noises, the observed inflow undergoes wavelet filtering and high-frequency components (first detail) are eliminated (Nourani *et al.*, 2014). For instance, Fig 4.1 illustrates inflow hydrographs before and after the wavelet filter. Fig 4.1a shows decomposition results for Event No.28, while Fig 4.1b presents the observed and filtered inflow hydrograph after applying the wavelet filter.

The uncertain inflow scenarios were generated by introducing an additive uncertain variable to the real inflow as described in Chapter 2.

4.3 Simulation Study

4.3.1 Data-driven models

Taking dynamical parameters into account based on real-time hydrological conditions used in PD-MPC is computationally expensive. For instance, with three dynamical weights and a search space of 100 quantized values per weight, each iteration of PD-MPC requires an average of 8.48 minutes to complete(CPU: AMD EPYC™ 9654). As the number of dynamical weights increases and the search space grows

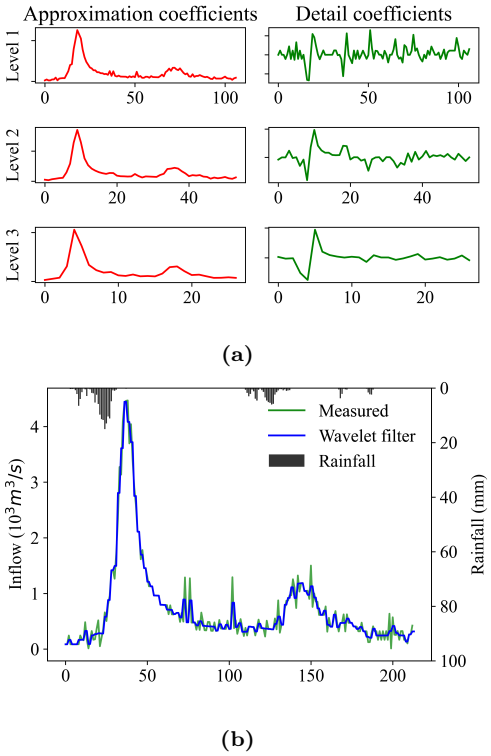


Figure 4.1 | For event No.28, (a) wavelet decomposition results, and (b) inflow hydrograph before and after a wavelet filter.

exponentially, the computational burden becomes too large. Therefore, two well-known methodologies, namely the explicit MPC and the switched MPC approach, are utilized to address this computational challenge of PD-MPC.

In conventional explicit linear MPC, it is assumed that for each predefined partition in the entire state space, there exists an affine function that maps the states to the optimal control inputs (Alessio and Bemporad, 2009). The state space may consist of several polyhedral partitions. Therefore, because of the high dimensionality of either state or control spaces, implementation of explicit MPC may be challenging (Alessio and Bemporad, 2009; Ferreau *et al.*, 2008). In conventional switched MPC approaches, the dynamic system states are first partitioned into predefined parts. Then, based on the given system states, control inputs are calculated by MPC with the corresponding dynamical system (Zhang *et al.*, 2016). Ensuring the continuity of control inputs in line with the well-defined optimized system partitions and switching rules in this approach is challenging (Zhu and Antsaklis, 2015).

For flood control, i.e., short-term operation, two specific challenges should be addressed: (i) unavailability of a comprehensive dataset including as many potential operational scenarios like flood events with several high peaks as possible; (ii) the smoothly changing states and the implicit relationship between the states and weights, which make it difficult to formulate switching rules for weights explicitly. Therefore, the implementation of explicit MPC and switched MPC is not straightforward.

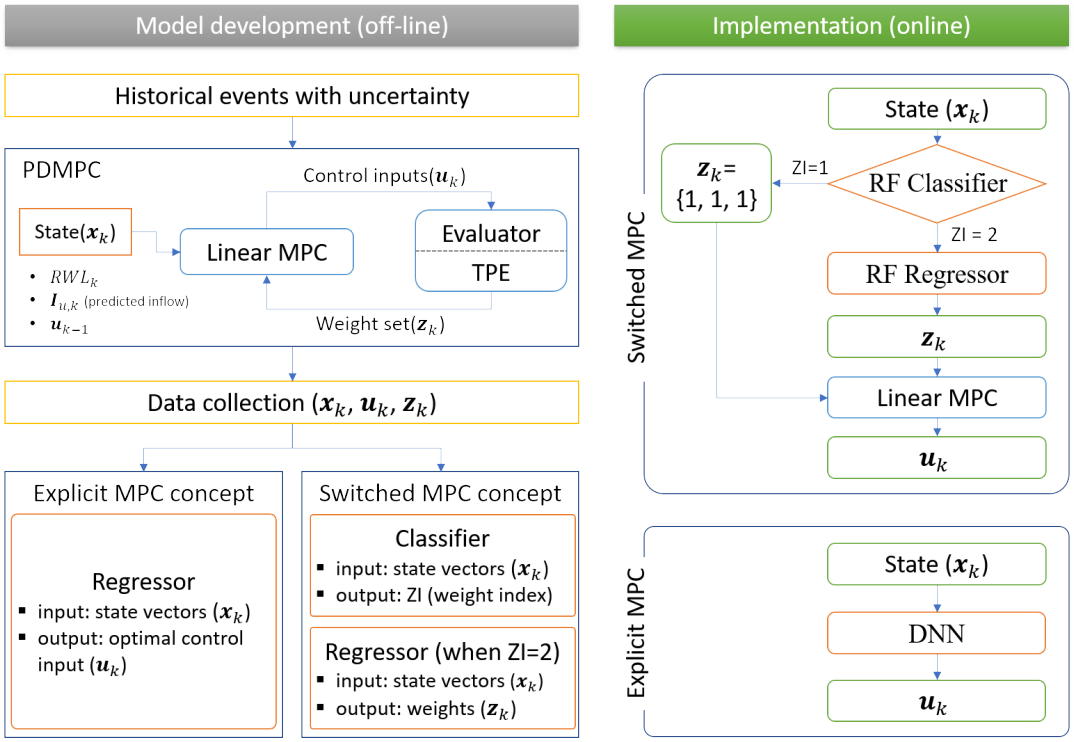


Figure 4.2 | A schematic diagram of the approach taken in this research. The left side illustrates the model development process, while the right side outlines its implementation.

In this work, we take a pragmatic approach, using a DNN model as a surrogate model for explicit MPC, trained based on PD-MPC results. In other words, the relationship between the states and the optimal control inputs is modeled using DNN based on the dataset obtained by PD-MPC, as shown in Fig 4.2. Similarly, for the switched MPC approach, the relationship between states and optimal weights from the PD-MPC results is modeled utilizing a Random Forest classifier and regressor as surrogate models for switching between different linear MPC controllers (i.e., switching between different weights). The detailed process illustrated in Fig 4.2 will be explained in the following section.

4.3.2 Data generation for surrogate models

Data consistency

At each time step of a receding horizon implementation, PD-MPC can propose a unique optimal outflow schedule $\mathbf{u}^{k*}$. However, there may be many combinations of the optimal parameter set, i.e. $\mathbf{z}^{k*} = \{z_1^*, z_2^*, z_3^*\}$, that could result in the same schedule. For instance, when a flood is going to occur, the RWL is between S_L and S_U, and flood control can be achieved by regulating turbine outflow itself, while some objectives are inactive (are zero) in this scenario, resulting in the same optimal outflow schedule with different weight sets corresponding to the zero objectives (i.e., inactive constraint violations). In such cases, PD-MPC would randomly select a set among several available candidates at that time step, as there would be no one-to-one

mapping between z^{k*} and u^{k*}.

This can be a problem for a surrogate to learn a mapping between states and optimal weights. Therefore, we would want to somehow project the weights corresponding to the objectives relating to inactive constraints to a lower dimension. An ℓ_1 regularization could, for example, be used to enforce these non-unique weights to have the same value (in this case, the mean of the weights for the objectives that are non-zero). An ℓ_1 regularisation is typically employed to prevent overfitting in DNN models by adding the ℓ_1-norm of model parameters to the original loss function. However, we incorporate the ℓ_1-norm for weights into the objective function of the TPE algorithm to construct preferences for the same weights. The objective function expressed by (4.9) is then rewritten as follows:

$$\min E + \rho \left\| z - \overline{z} \right\|_1 , \tag{4.11}$$

where ρ denotes the weight for the preference for the same weights and small parameter value, and $\|.\|_1$ expresses the ℓ_1-norm. To normalize a weight set, the average of weights $\overline{z}$ is subtracted.

It can be implied from (4.11) that, when there are multiple choices for weights, a set whose elements are as equal as each other should be chosen. Therefore, PD-MPC suggests $\mathbf{z}^*$ where its elements are the same, i.e., $z_1 = z_2 = z_3$, when $\mathbf{u}^*$ is not sensitive to $\mathbf{z}$ and it returns the unique optimal $\mathbf{z}^*$ when it is highly sensitive. It should be noted that if $\{z_1 = z_2 = z_3\}$, whatever each value is, then it is trivial that the optimal control inputs are always equal to the control inputs when $\{z_1 = 1, z_2 = 1, z_3 = 1\}$. Thus, we can set $\mathbf{z}^*$ to $\{z_1 = 1, z_2 = 1, z_3 = 1\}$ when $z_1 = z_2 = z_3$ without trying to find the exact value of each weight. Basically, in the case of large ρ, PD-MPC proposes a suboptimal $\mathbf{z}$ that tends to be closer to the equal weight preference than $\mathbf{z}^*$. On the other hand, in the case of small ρ, it always finds an optimal solution for $\mathbf{z}^*$; however, it may not distinguish situations where $\mathbf{u}^*$ is insensitive to $\mathbf{z}$, thereby failing to ensure data consistency. To select an appropriate ρ, a sensitivity analysis is conducted using five training flood events, including some extreme cases. A total of 10 candidates are considered for ρ, ranging from 1×10^{-5} to 1×10^{3} with an additional case where $\rho = 0$ (meaning $\rho = 1 \times 10^{-\infty}$). Moreover, for each event, five uncertain inflow forecasts are applied. To consider a sufficiently diverse set of scenarios, we examine three initial water levels, i.e., EL 76.0 m, EL 76.5 m, and EL 77.0 m, covering a total of 750 cases ($5 \times 10 \times 5 \times 3$).

The sensitivity analysis results are depicted in Fig 4.3. Each point represents the average or the maximum penalty of the five uncertain inflow scenarios. Points not marked in the figure, e.g., those for Event No.10 in Fig 4.3a with $\rho = 1 \times 10^{3}$, 1×10^{2}, and 1×10^{1}, indicate the cases where online PD-MPC iterations are infeasible, meaning that constraint violations occurred at least once. In this figure, as ρ decreases, the performance of PD-MPC is improved, i.e., penalty values decrease, and no further improvement is observed for smaller ρ than 1×10^{-3}.,

In Fig 4.4, ZI is an index for weights, where ZI= 1 corresponds to $\mathbf{z}^* = \{z_1 = z_2 = z_3 = 1\}$ and ZI= 2 is for all other cases. when ZI= 1, it also means that the same $\mathbf{u}^*$ is computed for various $\mathbf{z}$s, indicating that $\mathbf{u}^*$ is insensitive to $\mathbf{z}$. As

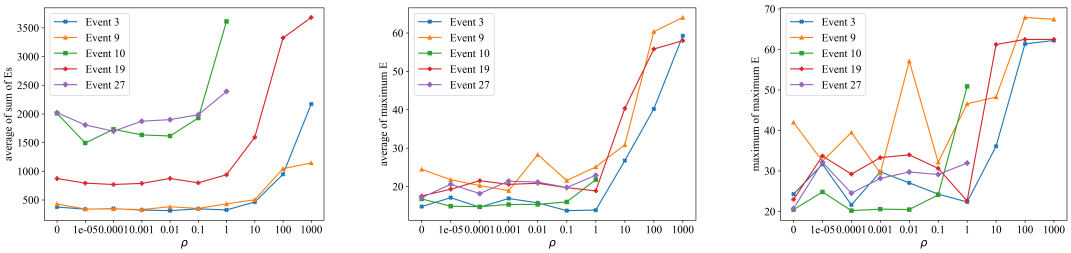


Figure 4.3 | Results of the sensitivity analysis in terms of the penalty values from the evaluator: (a) average values of the sum of penalty values for five different uncertain forecasts, (b) average values of the maximum penalty value of each uncertain forecast, and (c) maximum values of the maximum penalty value of each uncertain forecast.

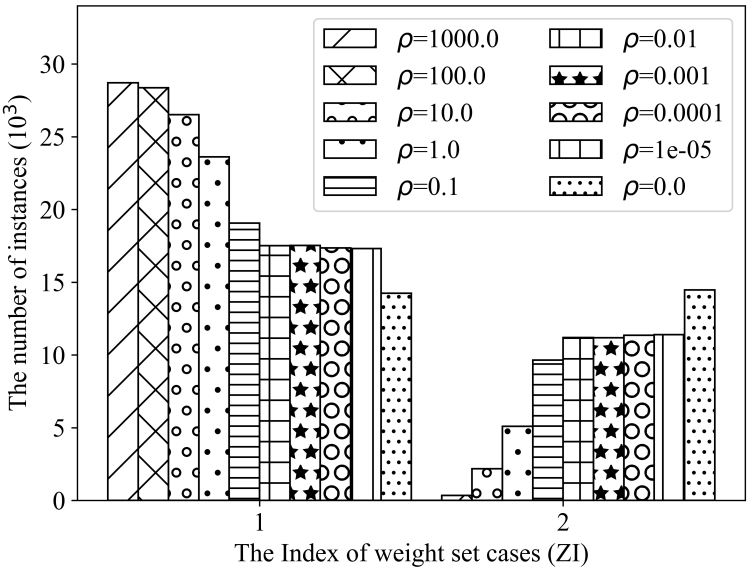


Figure 4.4 | The number of data for each ZI.

shown in the figure, in most cases, such as the early stage of a flood event, $\mathbf{u}^*$ is insensitive to $\mathbf{z}$, resulting in several instances with ZI= 1. Therefore, ZI= 1 and ZI= 2 are considered differently in this analysis. In this figure, as ρ decreases, a more diverse set of weights, i.e., ZI= 2 cases, appears. This indicates that even when the control inputs are less sensitive to weights, more weight sets closer to $\mathbf{z}^*$ can be obtained. Conversely, for larger values of ρ, ZI= 2 cases are only suggested when control inputs are highly sensitive to weights. When $\rho = 0$, most cases are ZI= 2, but it causes the inability to find the consistent weight set when the optimal control inputs are insensitive to weights, as discussed above. For instance, at time steps where multiple $\mathbf{z}$s lead to the same optimal control inputs and none of these $\mathbf{z}$s are in ZI= 1, with $\rho = 0$, the optimal weight set is randomly selected among ZI= 2 cases. However, for ρ greater than zero, a case with less difference between each weight is preferably chosen. Therefore, based on the sensitivity analysis and the discussion above, ρ can be set to 1×10^{-4}.

Table 4.2 | Study flood events

Purpose	Events	Cases	Event No.
Training	22	1210	All except validation and testing events
Validation[a]	3	165	#22−#24
Testing[b]	3	165	#1, #2, and #25

[a]one high ($\approx 5000\,m^3/s$, #23), two (#24) and three inflow peaks (#22).
[b]one high ($> 5000\,m^3/s$, #1), two high ($\approx 5000\,m^3/s$, #2), and three inflow peaks (#25).

Simulating multiple events for generating control learning data

Among the total of 28 flood events, 22, 3, and 3 events are assigned for model training, validation, and testing, respectively. These validation and testing events may not represent all practical cases, but we carefully choose them to cover various inflow patterns, such as multiple peak inflows, significant peak magnitudes, and so on. For this case study, we do not anticipate very different conditions compared to the range of extreme and average scenarios considered. For other applications where different conditions are anticipated, data-driven models can be trained using outputs of simulation models, which can, for example, produce future inflows based on changes in land use and land cover patterns, and climate change scenarios.

Data are collected through PD-MPC simulations under the following conditions. The initial water levels were set from NHWL-0.5 m to +0.5 m at intervals of 0.1 m, using five uncertain inflow predictions. Setting much higher or lower initial water levels than our initial levels allows PD-MPC to stabilize RWL quickly by increasing or decreasing outflows abruptly. Therefore, a narrow range of water levels around the NHWL is set as initial RWLs. Consequently, the training data consists of 1210 cases (21 events × 11 initial RWLs × five uncertain inflow cases), and each validation and testing data have 165 cases (three events × 11 initial RWLs × five uncertain inflow cases) as presented in Table 4.2.

It is worth noting that historically observed initial RWLs are lower than our simulation, e.g., Event No.3, which occurred near the end of the flood season, started with an RWL at elevation 72.2 m. This is because operators want to secure enough flood control capacity in case of unexpected huge inflows, even though lower RWLs increase risks to water supply. Thus, lower initial RWLs in historical data often result from the absence of an optimal flood control system, which is what this chapter aims to address. Therefore, we set the initial RWL in this research to the NHWL, representing the likely initial condition when operators have access to optimal reservoir flood control inputs at each time step.

4.3.3 Optimal outflow schedule by DNN Controller

The efficiency of DNNs can be affected by hyperparameter settings, such as the number of nodes and layers, the learning rate, the activation function, and the batch size. To find the best hyperparameters, a grid search based on mean squared errors is carried out (Huang *et al.*, 2022a). However, considering the extensive use of ramp functions in the evaluator, the Rectified Linear Unit (ReLU) is selected

Table 4.3 | Hyperparameters of derived DNN models

Learning rate	Layers	Nodes	Activation	Batch size	Epochs	Loss
0.0005	6	128	Relu	10	1000	MSE

Table 4.4 | Nash–Sutcliffe coefficient and RMSE for validation events

	$O_{total,t+1}$	$O_{total,t+2}$	$O_{total,t+3}$	$O_{total,t+4}$	$O_{total,t+5}$
NSC	0.998	0.986	0.982	0.980	0.963
RMSE (m^3/s)	42.1	105.5	119.2	124.2	167.6

as the activation function. In addition, the number of layers is also fixed to six based on a six-hour prediction horizon. We utilize the mean square error (MSE) loss function (Shelke *et al.*, 2023; Ahmad *et al.*, 2022; Maddu *et al.*, 2022) with the Adam optimizer (Kingma and Ba, 2014; Maddu *et al.*, 2022; Kao *et al.*, 2020; Le *et al.*, 2019), which is widely used in developing surrogate inflow prediction models. The MSE loss function is particularly suitable for flood inflow prediction due to the huge variance between peak and low inflows. Additionally, it helps in accurately estimating the peak values, which are critical when targeting flood prediction.

To reduce computational burden and searching time and to prevent overfitting during the process of finding optimal hyperparameters, 330 cases for six events among the training events are utilized along with the early stopping technique (Zhang *et al.*, 2020), and the five-fold cross-validation approach. Through the grid search method, we found that batch size had the greatest impact on model performance, followed by learning rate and the number of nodes, respectively. The best hyperparameters are presented in Table 4.3.

As the results shown in Fig 4.5 and summarized in Table 4.4 indicate, the DNN model, mapping the states to the control inputs, is able to mimic accurately the PD-MPC control. Note that the DNN model is trained only for total outflows, i.e., O_{total}, for simplicity because spillway outflows can be obtained by abstracting the turbine capacity from total outflows when total outflows are larger than the turbine capacity. The Nash-Sutcliffe model efficiency coefficient, which is commonly used to evaluate the prediction accuracy of hydrological models (Jain and Sudheer, 2008), is consistently more than 0.95, which demonstrates that the model is reliable enough to replace PD-MPC (Moriasi *et al.*, 2007). The Root Mean Squared Error (RMSE) also varies between $42\,m^3/s$ and $167\,m^3/s$. Considering that the initial inflows for the three flood events in the validation set are $227.6\,m^3/s$, $224.4\,m^3/s$, and $116.6\,m^3/s$, and their respective peak inflows are $4985.8\,m^3/s$, $3655.2\,m^3/s$, and $2590.4\,m^3/s$, we can say that the model is accurate. It should be noted that because O_t^k is fixed to the value determined in the previous time step, i.e., $k-1$, the DNN model output has $N-1$ values when the prediction horizon is N hours. As expected, predictions for shorter time horizons are more accurate, while performance gradually deteriorates as the prediction time increases.

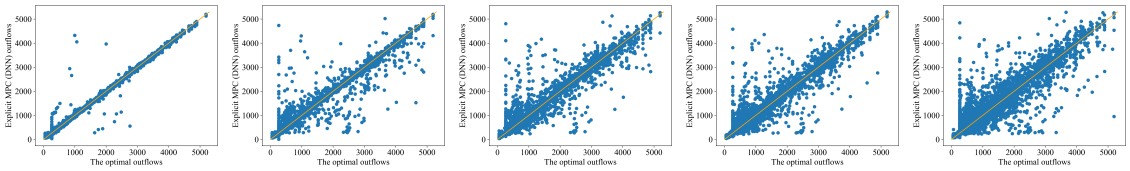


Figure 4.5 | Validation results of the DNN models for mapping states to outflow schedules: (a) $O_{total,t+1}$, (b) $O_{total,t+2}$, (c) $O_{total,t+3}$, (d) $O_{total,t+4}$, and (e) $O_{total,t+5}$. Performance diminishes with increases over time.

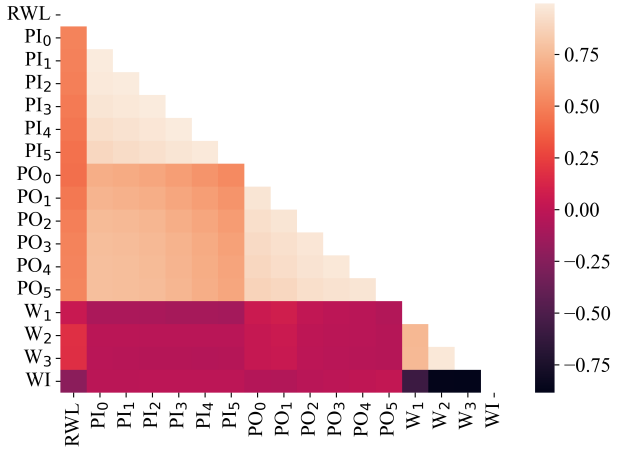


Figure 4.6 | The correlation among state elements and weights, where PI_1: predicted inflow at t+1, PO_1: previously decided outflow at t+1.

4.3.4 Training surrogate models for mapping states to weights of objectives

To simulate the changes in weights based on the states of the system, machine learning models are used to map them together. Given the predominance of ZI= 1 cases (shown in Fig 4.4) and the challenges of identifying a linear relationship between the states and the weights (shown in Fig 4.6), constructing a linear relationship to directly assign the best weights based on the states seems to be infeasible. Therefore, a binary classification model is employed to distinguish between the cases of ZI= 1 and ZI= 2. Additionally, a regression model is utilized to estimate the weights for ZI= 2.

Among the available criteria for selecting a binary classification algorithm, accuracy, precision score, and recall score are taken into account in this chapter. Accuracy represents the ratio of correctly classified instances to the total number of predictions. Precision score refers to the proportion of correctly classified instances for a specific class divided by the total instances classified as a given class by the model. Recall score measures the ratio of instances correctly classified as a particular class to the total instances of that class. When a classifier classifies a case with ZI= 1, the weight set is determined; however, when ZI= 2, the optimal weight set should be estimated using the regressor. In other words, a case needs to be classified as ZI= 2

Table 4.5 | Binary classification results on ZI

Classifier	Precision ZI=1	Precision ZI=2	Recall ZI=1	Recall ZI=2	Accuracy
Random forest	*0.952*	0.741	0.898	*0.865*	0.890
K neighbors	0.930	0.708	0.888	0.802	0.866
Support vector	0.941	0.676	0.864	0.839	0.857

when there is uncertainty and ZI= 1 when no uncertainty exists. Therefore, ZI= 1 and ZI= 2 are selected with a high precision score and a recall score, respectively.

Furthermore, to address the skewness of data in ZI, we adopt both under-sampling and over-sampling algorithms. The Edited Nearest Neighbors (ENN) is an under-sampling method that removes the nearest neighbors from the majority class data (Wilson, 1972). The Synthetic Minority Over-sampling Technique (SMOTE) (Chawla *et al.*, 2002) is an over-sampling technique where synthetic data are generated by randomly selecting one data point from the minority class and its nearest neighbor in the minority class. The synthetic data are classified by a classification model. Through the application of these techniques, the number of cases for ZI= 1 and ZI= 2 becomes balanced. The number of cases changes to 158134 for ZI= 1 from 187587 in the original data and 150783 from 89833 for ZI= 2.

Classifiers are constructed based on $k-$nearest Neighbors, Random Forest, and Support Vector algorithms using balanced data. As summarized in Table 4.5, the validation results indicate that the Random forest model is the best regarding precision and recall scores. For the implementation, a python library, imbalanced learning (LemaÃŽtre *et al.*, 2017) is used for the SMOTE-ENN algorithm, and the scikit-learn package (Pedregosa *et al.*, 2011) with default parameters for the classifiers.

To map the nonlinear relationship between the states and the weights when ZI= 2, a Random forest regressor is employed, which uses an ensemble approach with decision trees, which can effectively map the nonlinear relationship between inputs and outputs (Bemporad, 2023; Breiman, 2001). This model uses the average value at the final (decision-making) node for regression.

The parameters, i.e., the number of trees and the maximum depth of trees, are selected by grid search based on RMSE values for the validation events, as shown in Fig 4.7. The parameters are selected at the point where the RMSE curve starts to level off in order to simplify the model. The selected number of trees is then 150, and the maximum depth of each tree is 30. However, as can be seen in Fig 4.7, no significant difference is observed with different parameter values.

In addition, the Piecewise Affine (PWA) modelling approach, which is known for effectively approximating switched discrete-time nonlinear systems (Moradvandi *et al.*, 2023; Wang *et al.*, 2023), is also employed. They are widely used to strike a balance between simplicity and accuracy (Bemporad, 2023). This PWA approximation does not show considerable improvement in terms of the Peak RWL and the penalty values of the evaluator compared to the Random Forest regressor. The Random Forest regressor demonstrates even better performance in terms of the peak outflow and peak evaluator value. Linear Random Forest regressor (Ao *et al.*, 2019), which uses the Random Forest algorithm, but the final value is calculated by linear regression, showing similar results with the Random Forest regressor. The details on

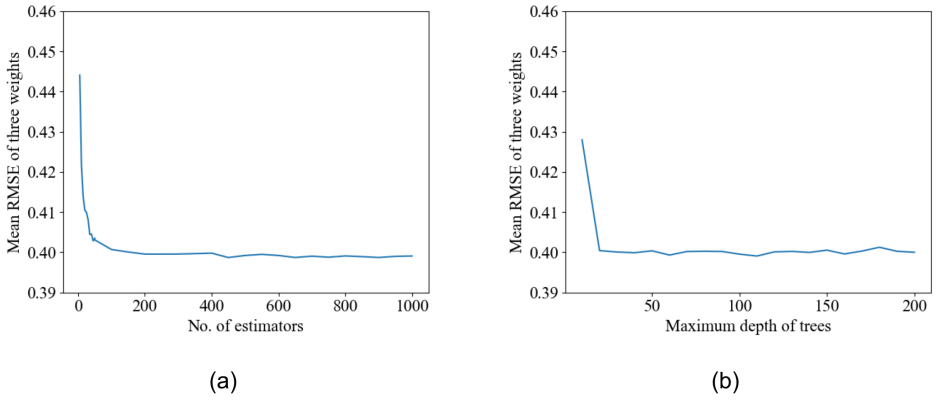


Figure 4.7 | Hyperparameter grid search results (to compare hyperparameters by the same scales, the scale of the y-axis is fixed to [0.40, 0.45]), (a) the number of trees (when the maximum depth of each tree is 30), and (b) the maximum depth of each tree (when the number of trees is 150).

the PWA model and Linear Random Forest regressor can be found in Appendix B.

4.4 Results and Discussion

In this section, the efficiency of data-driven models based on the two aforementioned MPC approaches on the test dataset for flood events is presented and discussed. In this regard, the flood events selected for testing, as shown in Table 4.2, include cases with two and three peaks of inflow and a large peak. Therefore, except for extreme cases like the Probable Maximum Flood for designing reservoirs, these events could represent situations similar to those encountered by operators during flood control operations.

4.4.1 Analyses of test events

The average of result indexes for each test event, i.e., the peak outflows, peak RWLs, minimum RWLs, the total evaluator values, and peak evaluator values, are presented in Table 4.6, Table 4.7, and Fig 4.8. 'Fixed 1', 'Fixed 2', and 'Fixed 3' refer to predefined weights, which are $\mathbf{z} = \{1, 1, 1\}$, $\{0.125, 0.625, 0.25\}$, and $\{0.0625, 0.625, 0.3125\}$, respectively. 'Fixed 2' and 'Fixed 3' represent the cases with a high weight for J_2 in (4.2), which emphasizes maintaining the RWL in the target water level. J_2 with a high weight forces the linear MPC controller to maintain the reservoir water level at the minimum possible level, ensuring operational stability even over extreme conditions. In addition, 'Fixed 3' highlights the importance of J_3, minimizing changes between outflow schedules.

'E-MPC' refers to the explicit MPC implemented using a DNN model, which is directly mapping the states and the outflow schedules. On the other hand, the switched MPC approach, which employs a random forest classifier and regressor as surrogate models for the weight vector, is referred to as 'S-MPC'.

Table 4.6 | Average results for all test events

	Peak outflow	Peak RWL	Minimum RWL	Total E	Peak E	No. changes
PD-MPC	4137.3	78.27	76.07	934.5	17.3	49.4
E-MPC	4167.1	78.09	76.03	3207.7	30.6	926.9
S-MPC	4650.1	79.61	75.97	2667.7	52.3	23.9

Table 4.7 | Average results for each test event

	Event	Peak out.	Peak RWL	Min. RWL	Total E	Peak E	No. changes
PD-MPC	1	5487.7	78.4	76.1	1007.9	18.7	38.9
Fixed 1	1	5304.8	79.9	76.0	2415.8	62.4	50.3
Fixed 2	1	5304.8	79.9	76.0	2408.7	62.4	49.3
Fixed 3	1	5474.9	80.0	76.0	2496.7	58.6	43.0
E-MPC	1	5340.0	78.4	76.0	2778.6	33.1	702.7
S-MPC	1	5225.7	79.5	76.0	1791.2	50.4	14.5
PD-MPC	2	5024.9	78.3	76.1	908.8	20.8	64.8
Fixed 1	2			Constraint violation			
Fixed 2	2			Constraint violation			
Fixed 3	2	4324.7	79.9	76.0	3454.7	57.6	46.8
E-MPC	2	5056.3	78.4	76.0	3392.5	29.6	1006.1
S-MPC	2	4154.6	79.4	76.0	2353.1	41.9	30.1
PD-MPC	25	1899.3	78.0	76.0	886.7	12.5	44.5
Fixed 1	25	1533.1	80.0	76.0	4571.4	59.4	35.0
Fixed 2	25	1533.1	80.0	76.0	4586.2	59.4	34.1
Fixed 3	25			Constraint violation			
E-MPC	25	2104.9	77.4	76.1	3452.2	29.2	1071.9
S-MPC	25	4570.1	79.9	76.0	3858.9	64.7	27.1

For cases with predefined weights, the system violates at least one constraint more than once, and the peak RWL reaches the FWL sometimes. For example, although Fixed 3 has a high weight to the water level, in which the weight for J_2 is ten times higher than J_1 and two times higher than J_3, it shows the highest peak RWL in Event No.1 and 25. This result implies that fixing the weights may not respond adequately, i.e., being sensitive enough, to changing hydrological situations, thereby failing to satisfy the constraints and produce reliable results from a long-term perspective. In contrast, E-MPC, which incorporates the DNN model, and S-MPC, which applies the Random Forest models, show stable operation for the test events. In this scenario, the maximum outflow capacity, minimum water supply, and RWL constraints are all satisfied. Even though the best performance belongs to PD-MPC, they also outperform the fixed weight cases with respect to the absolute performance evaluator value.

The Absolute Performance Evaluator quantifies what reservoir operators want to achieve during flood events. While conventional MPC approaches are limited to linear or quadratic objective formulations, making it challenging to express complex operational goals mathematically, the PD-MPC framework overcomes this fundamental limitation by enabling the use of diverse nonlinear equations, including exponential functions and conditional formulations. Therefore, this evaluator can serve as a definitive performance index, providing a single quantitative measure. As presented in Table 4.6, S-MPC shows a lower average penalty in terms of the evalu-

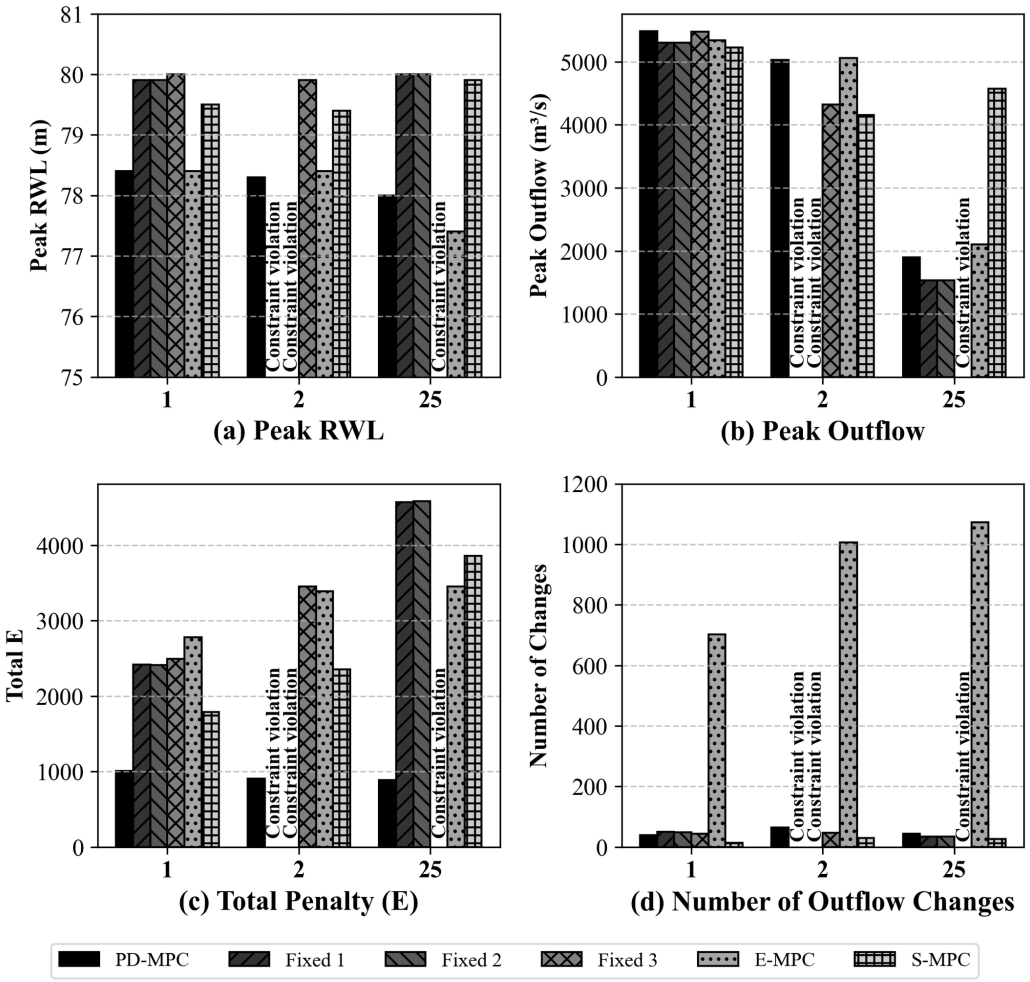


Figure 4.8 | Average results for each test event in terms of (a) peak RWL, (b) peak Outflow, (c) the sum of penalty values (Total E), and (d) the number of changes in outflow schedules.

ator compared to E-MPC. Fixed weight cases generally present higher penalties than both S-MPC and E-MPC, with one exception: in the event No.1, E-MPC exhibits the highest evaluator value. However, regarding the highest penalty during an event period, 'Peak E', E-MPC has a lower average penalty than S-MPC. The results indicate that the relative performance of S-MPC and E-MPC is contingent upon specific event characteristics. Note that 'Total E' in Tables 4.6 and 4.7 represents the sum of penalty values accumulated during an event period. Here, E is interpreted as a penalty since the evaluator's objectives are formulated as minimization problems.

Considering the peak RWL and peak outflow, E-MPC mimics PD-MPC results accurately. It can be concluded that E-MPC is able to respond adequately to changes in the states. However, E-MPC tends to change the outflow schedule frequently, leading to very high values for E_3 in (4.9). This is because even small changes in the outflow schedule yield an increase in E_3 value, and the DNN regressor is less

sensitive to minor deviations due to the effect of a loss function, which is the mean squared error.

On the other hand, S-MPC derives state-dependent weights, and the linear MPC, incorporating these weights, generates the optimal outflow schedule in real-time. As a result, in terms of the absolute performance evaluator, switched MPC shows lower penalty values and fewer changes in outflow schedules — an area emphasized by the evaluator, as mentioned in Section 4.2.3 — compared to E-MPC. Since the PD-MPC framework is designed to select optimal control inputs based on the evaluator, not the linear MPC's objectives, we can conclude that switched MPC better replicates the original multi-objective optimization results compared to E-MPC.

However, this improvement is obtained by compromising the peak outflow and peak RWL, since the number of changes, peak outflow, and peak RWL are correlated. This can be explained by errors in these two surrogate models, indicating that the models are not sufficiently sensitive to a given state.

Some errors in the classification of ZI and the regression of weights can be expected for the surrogate models. Firstly, if the classifier fails to identify situations where ZI= 2, i.e., when weights need to change, the model maintains RWLs between the target RWLs and outflows at the minimum amount. It yields subsequent higher RWLs and larger outflows. Furthermore, if the regressor fails to provide the optimal weights, linear MPC may struggle to accurately determine the appropriate adjustments to the outflow.

Moreover, when RWLs exceed EL 78.89 m, which is the maximum RWL in training data, switched MPC surrogate models need to provide optimal weights for states that fall outside the range of the training data. It can result in errors.

Despite errors in the surrogate models (RMSEs are approximately 0.45 in Fig 4.7) in optimal decision timings and outflows, the switched MPC approach enables linear MPC to produce robust results concerning constraints. Even if the surrogate models cannot present the optimal weights at a certain time step and the system state worsens in terms of the evaluator values, they can promptly respond to the slightly deteriorated new state due to the receding horizon control. Thus, linear MPC makes appropriate decisions with the weights provided by the surrogate model.

4.4.2 Comparison of explicit and switched MPC

For a deeper understanding of the characteristics of surrogate models within the frameworks of explicit MPC and switched MPC, we define scenarios that these models have not previously been trained on. In general, data-driven models often fail to produce reliable outputs when the inputs are beyond the range of the training data (Badillo *et al.*, 2020; Wang *et al.*, 2016). A scenario with an initial RWL of EL 75.0 m is defined, while, in the training data, the lowest water level is EL 75.95 m. This scenario is simulated for Event No.1 with one peak and a significant peak inflow.

The simulation results of the data-driven models for the aforementioned scenario are shown in Fig 4.9 and summarized in Table 4.8. As expected, PD-MPC outperforms other methods in this case. In addition, S-MPC does not violate any

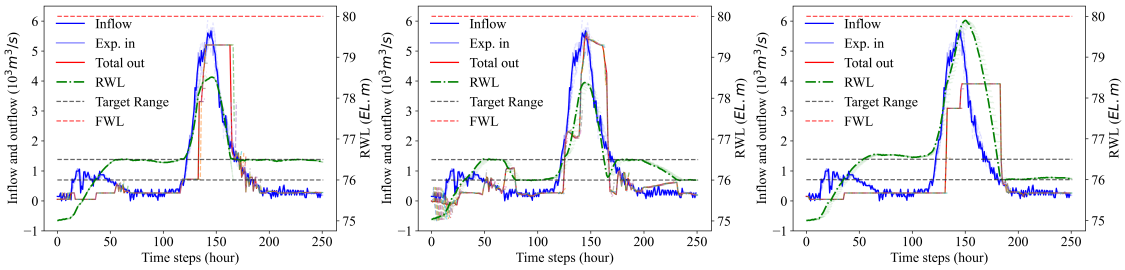


Figure 4.9 | Hydrographs for event No.#1 with unseen initial RWL, (a) PD-MPC, (b) Explicit MPC (DNN), and (c) Switched MPC (Random Forest classification and regression).

Table 4.8 | Results for unseen initial RWL for Event No.1

C	Peak out.	Min. out.	Peak RWL	Min. RWL	Total E	No. changes
PD-MPC	5202.0	52.0	78.51	74.99	1282.4	49
E-MPC	5473.0	**-844.0**	78.41	74.98	3091.2	714
S-MPC	3905.1	52.0	79.90	75.00	2377.6	20

constraints. However, E-MPC does not satisfy constraints compared to the other three cases, particularly when the RWL is low; E-MPC suggests negative outflow, e.g., $-844\,\mathrm{m}^3/\mathrm{s}$, to increase the reservoir level to the target range (from EL 76.0 m to EL 76.5 m), even lower than the minimum outflow for water supply, i.e. O_{min}. This is because S-MPC explicitly incorporates constraints through linear MPC, whereas E-MPC only maps the states and the outflow schedules.

4.4.3 Discussions

Considering the limitation of data-driven models to model complex relationships between states and outflow schedules, or states and weights, which are embedded in the training data, the PD-MPC framework outperforms in all cases. However, the explicit MPC and the switched MPC outperform in the fixed weight cases, demonstrating applicable results without violating constraints for the test events.

The DNN model based on the explicit MPC approach results in small and frequent changes in the outflow schedule, as shown in Table 4.7 and Fig 4.9b. This causes physically impossible outflows and constraint violations for the unseen cases. Despite these shortcomings, they could potentially be addressed by introducing heuristic rules based on an operator's experience. For instance, operators can design a simple heuristic algorithm in which a proposed outflow is replaced with the minimum outflow when the proposed outflow is less than the constrained minimum outflow. Negligible fluctuations between outflow schedules can also be ignored using these kinds of heuristic rules. Furthermore, as the number of training flood events increases and various uncertain inflow scenarios and initial RWLs are integrated, a surrogate DNN model can cover all potential states to avoid infeasible control inputs.

The model using the switched MPC approach also demonstrates reliable results

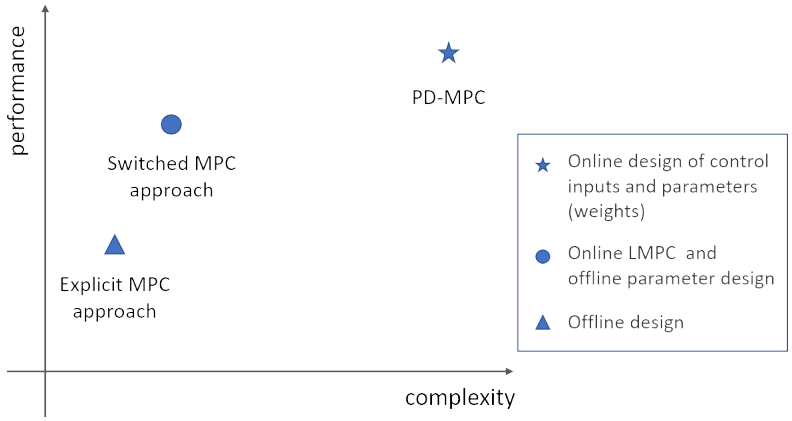


Figure 4.10 | Conceptual diagram of explicit and switched MPC approaches

for flood control. However, the complex relationship between the states and optimal weights poses a challenge due to the selection of a single weight set from multiple sets that produce the same optimal control inputs. As a result, using a conventional data-driven model often leads to significant testing errors. For instance, although not presented here, a DNN model does not outperform our proposed classifier and regressor. These errors cause an overall performance degradation, particularly delaying the timing of changes in the outflow schedule. Moreover, this degradation can cause a switched MPC surrogate model to encounter states far outside the training data. Consequently, the peak RWL is higher, and the peak outflow is larger than the PD-MPC results. The errors can also be mitigated by adopting knowledge-based heuristic rules. As an example, we can design different default weight sets for different states outside training data based on the operator's experience. The most crucial point to be considered is that online implementation of linear MPC can guarantee that the optimal outflow is stable, not violating constraints, even for unseen cases.

In summary, although PD-MPC outperforms other methods, it requires solving computationally complex nonlinear optimal control problems online for reservoir flood control, which makes it difficult to implement in practice. The switched MPC approach replaces the gradient-free randomized search algorithm, which was time-consuming for solving PD-MPC online. It can thus significantly reduce complexity. In addition, since linear MPC optimization is conducted every time step using updated information, results are robust to errors and avoid constraint violations. The explicit MPC approach has the lowest complexity for implementing online as it directly maps states to control, but a DNN model can present infeasible control inputs and needs more training data to cover every possible situation. The complexity and accuracy of the discussed methods are schematically described in Fig 4.10.

Even though we can convert an expensive online nonlinear optimization problem into expensive offline training and cheap online implementation, it will be better to make the online nonlinear optimization, i.e., the GA or TPE algorithm itself, cheap. This means the online implementation of the PD-MPC framework becomes more practical. The warm-starting strategy (Feurer *et al.*, 2015; Watanabe, 2023) can be

a good method to start. Adaptive learning techniques present promising alternatives to traditional data-driven models. For example, reinforcement learning (Li, 2017) and learning-based MPC approaches (Hewing *et al.*, 2020b; Yang *et al.*, 2020) have gained significant attention in recent research. These methods can effectively reduce the errors inherent in surrogate models, ultimately near optimality with lower computational complexity.

Another important consideration is how to scale this approach to a multi-reservoir system. In standard MPC approaches, some challenges in scaling the receding horizon MPC approach to complicated multi-reservoir systems have been reported (Myo Lin *et al.*, 2020; Lin and Rutten, 2016). In our framework, expanding the surrogate switched and explicit MPC approaches to a multi-reservoir system is technically straightforward, given that obtaining the optimal outflow schedules and/or the optimal weights is feasible. In such extensions, the number of system states will grow linearly with the number of reservoirs in the system. Since the switched approach uses linear MPC, the computational complexity of the real-time control will remain manageable. For the explicit MPC, it would be even more computationally efficient. However, the most complex task in such an approach is designing a sufficient number of scenarios for the PD-MPC to generate a sufficient range of control data sets for learning the explicit and switched models.

The uncertainty of inflow forecasts also needs to be considered. Even though we consider the uncertainty of inflows by adding noise to observed inflow data when generating the training data, we deal with it in a deterministic way. However, it is advisable to explore stochastic/robust MPC (Saltık *et al.*, 2018) to consider uncertainty explicitly. In addition, we acknowledge that we cannot present the comprehensive impacts of the inherent inaccuracy of data-driven models in reservoir flood control, despite our detailed explanations of how S-MPC and E-MPC respond to the errors of surrogate models within the studied flood scenarios. Many methodologies for uncertainty quantification of data-driven models, e.g., sensitivity analysis (Pianosi *et al.*, 2016), Bayesian-based approaches, bootstrap-based approaches, and the Monte-Carlo method, have been studied (Kasiviswanathan and Sudheer, 2013). However, incorporating these methodologies for comprehensive uncertainty analysis is beyond the scope of this research.

4.5 Conclusions

In this chapter, we propose data-driven surrogate models in the form of explicit and switched Model Predictive Control (MPC) approaches for a flood control problem with dynamically changing weights of multi-objectives, which is the first study based on our best knowledge. These approaches have the advantage of low computational complexity for online implementation compared to a Parameterised Dynamic Model Predictive Control (PD-MPC) framework proposed in Chapter 3. The PD-MPC formulation explicitly considers the dynamic characteristic of the relative importance of objectives depending on the situation via time-varying relative weights for the different objectives, which are optimized online at each time step of the MPC by

solving a highly nonlinear optimization problem. Through numerical experiments on the Daecheong multipurpose reservoir in South Korea, we demonstrate that both approaches outperform linear MPC approaches with predefined fixed weights and produce compatible results with PD-MPC. Especially, the switched MPC approach is robust in terms of model errors despite some constraint violations of the explicit MPC DNN model under previously unseen events. However, because the highly complex relationship between states and weights leads to the performance degradation of data-driven models for the switched MPC, testing these strategies for other case studies where there may be a larger size of training data may be useful.

Overall, we expect this chapter to contribute to widening the possibility of the application of real-time optimal reservoir flood control and lay the foundation for practical implementation in the field. The switched MPC approach, for example, offers a computationally efficient method for real-time reservoir operation to effectively balance multiple objectives, such as minimizing flood risk downstream, maintaining reservoir levels, and minimizing outflow schedule changes, even under uncertain inflow conditions we simulated. Additionally, the insights into the dynamic weighting of objectives based on real-time hydrological conditions and states can aid in developing more adaptive and robust flood control strategies.

Even though our research focuses on a single reservoir in South Korea, we believe our approach could be applicable to reservoirs with various hydrological conditions. The methodology presented in this chapter may be adaptable to different reservoir systems with their own unique characteristics and operational requirements, as the fundamental concepts of model predictive control and surrogate modelling potentially extend to diverse hydrological settings. The framework potentially offers sufficient flexibility to accommodate region-specific constraints and objectives, suggesting its potential utility for addressing water resource management challenges in different geographical contexts.

Data Availability for this chapter

The hydrological and operational data were acquired from Korea Water Resources Public Corporation's website (http://kwater.or.kr). The data and code of this chapter are available at https://doi.org/10.4121/b6dd9d97-118d-406e-867d-b821fb6d08d4 under the CC-BY-4.0 licence.

5

Scenario Generation & Reduction for Inflow Timeseries[1]

Summary

Uncertainty in future inflows is one of the key factors in reservoir flood control. We can consider this uncertainty explicitly in optimal control of reservoirs using scenario-based stochastic control approaches, where time series of inflows could be generated from probabilistic models representing different futures with their associated likelihoods. It then becomes essential to use a sufficiently large number of uncertain scenarios representing possible future states. Too many scenarios may lead to significant computational complexity for control optimisation, but too few scenarios may become less representative. Therefore, a method for preparing scenarios consists of two parts: scenario generation and reduction. Despite an extensive number of studies on these methodologies in various fields, to the best of our knowledge, there is little research tailored to reservoir inflow properties. In this chapter, we develop and apply a probabilistic data-driven model, specifically a Bayesian Neural Network (BNN), as a scenario generation method. The BNN model captures the temporal dependence of inflow time series and shows high accuracy, as measured by the Nash-Sutcliffe efficiency Coefficient (NSC) and Root Mean Squared Error (RMSE), for short prediction horizons. Accuracy decreases as the prediction time increases. For scenario reduction, an important element is the distance measure for how representative the reduced set is of future inflows. We evaluate the applicability of four distance measures: the Manhattan distance, the Euclidean distance, the Wasserstein distance, and the energy distance, which are widely used for scenario reduction in other applications. In terms of preserving the statistical characteristics of the original larger scenario set, we find that the energy distance is the best, followed by the Manhattan and Euclidean distance measures. However, regarding the inclusion of extreme scenarios, which are critical in flood control, the Manhat-

[1]This chapter is based on the publication: Koo, Ja-Ho, Edo Abraham, Andreja Jonoski, and Dimitri P. Solomatine. (2025). Comparison of scenario reduction approaches for reservoir inflow timeseries generated by a Bayesian Neural Network. *PLOS One*. **(under review)**

tan and Euclidean distances are found to be the best, as measured by a tailored quantitative index. This index represents the size of the envelope of a scenario set (using l_1-norm), which ensures the reduced sets retain the range of maximum and minimum flows of the original scenario. In addition, the results suggest that reduced scenarios can capture extreme scenarios adequately when $m \geq 30$, where m is the reduced scenario set size. In terms of computational complexity, the energy distance exhibits the highest computational cost, scaling quadratically with m, followed by the Wasserstein distance, which increases approximately linearly with m. Overall, considering the importance of extreme scenarios in flood control and the computational complexity of calculating the energy distance, the Manhattan and Euclidean distances with $m = 30$ can be recommended as practical measures for reducing inflow scenarios.

5.1 Introduction

Uncertainty in reservoir flood control operation is mainly due to uncertainty in the prediction of inflows (Zhao *et al.*, 2011). The inflow uncertainty can be expressed as a limited number of feasible scenarios (Séguin *et al.*, 2017a). Therefore, scenario generation aims to generate discrete scenarios that can represent this uncertainty. In case of employing stochastic reservoir optimization, the number of scenarios has a profound impact on the computational complexity, so it is desirable to reduce the number of scenarios, but at the same time preserving enough scenarios since they represent the inflow uncertainty (Löhndorf, 2016; Ziel, 2021).

Scenario generation relating to hydrological factors such as inflow or water level has conventionally been conducted based on statistical models such as autoregressive-moving average (ARMA), periodic ARMA (PARMA), and autoregressive integrated moving average (ARIMA), or historical observation data directly (Dires *et al.*, 2023). For example, hydrological scenarios can be generated by adding to historical time series a random value that follows a normal distribution (Li *et al.*, 2018) or by increasing the standard deviation of the normal distribution to longer time horizons to reflect the uncertainty increase over time. Historical data can be regarded as hydrological scenarios with equal probability (Xu *et al.*, 2017a), or scenarios with uniform distribution can be generated by adding a random value to the predicted value of the ARMA model (Haddad *et al.*, 2014).

The scenario tree is a frequently used scenario generation method (Kaut and Stein, 2003). The work in Pflug and Pichler (2016) proposed a nonparametric method to generate a scenario tree from historical data, which was utilized in the stochastic hydroelectric power optimization problem (Séguin *et al.*, 2017b,a). In the literature, the scenario tree has been widely used in various stochastic optimal control problems, but not in hydrological scenario generation. This is primarily because multi-step scenarios are generally required for reservoir flood control. This high dimensionality of the hydrological scenarios leads to the considerable complexity of a scenario tree, making it computationally infeasible in many cases.

Scenario generation by data-driven models has been applied in various fields,

such as autonomous driving and power generation. Ding *et al.* (2023) compared the generation methods of autonomous driving scenarios using deep learning models such as Bayesian Networks, Convolutional Neural Networks (CNN) and Long short-term memory (LSTM), and generative models such as Generative Adversarial Networks (GAN). In Sideratos and Hatziargyriou (2012); Vagropoulos *et al.* (2016), two radial basis function neural networks (RBFNN) and Deep Neural Networks (DNN) were applied to generate scenarios for power systems. Nevertheless, most data-driven models were used as deterministic predictive models. In this case, a Quantile Regression Deep Neural Network (QRDNN) provides an effective approach to model uncertainty by capturing the full conditional distribution of output variables rather than just point estimates. This method has been successfully applied to generate day-ahead electrical market price distributions as demonstrated in recent studies (Van der Heijden *et al.*, 2022, 2025). However, there is a limitation that a scenario composed of multi-time-step prices does not reflect the temporal dependence of prices. Conversely, probabilistic data-driven models such as Gaussian processes or Bayesian Neural Networks (BNN) can derive probabilistic outputs with deterministic inputs by directly deriving the probability distribution of parameters from the data (Jospin *et al.*, 2022; Gal and Ghahramani, 2016). Specifically, these probabilistic data-driven models are suitable for generating hydrological scenarios because they implicitly take into account the temporal dependence of inflows at different times.

A few scenarios can hardly reflect the uncertainty of hydrological processes, whereas too many scenarios can make the stochastic optimization problem that utilises them intractable (Dvorkin *et al.*, 2014). Scenario reduction is a technique to manage the computational efficiency without losing the properties of the original scenario set (Germscheid *et al.*, 2023). In Xu *et al.* (2012), the authors applied a k-means clustering algorithm to generate a scenario tree, and Dupačová *et al.* (2003) proposed a forward selection and backward reduction method based on the Wasserstein distance. Clustering is one of the popular scenario reduction methods. In Germscheid *et al.* (2024), the k-means clustering in the stochastic optimization problem for power demand response scheduling showed the highest risk-aversion in terms of cost and risk among scenario reduction methods such as probability distance-based scenario reduction (Heitsch and Römisch, 2003). In contrast, the work in Dvorkin *et al.* (2014) showed that the fast forward selection based on the Wasserstein distance was effective for the stochastic unit commitment problem in terms of operational cost by comparing the k-means clustering based on the Euclidean distance, backward scenario reduction, and importance-sampling scenario reduction.

Let an original scenario set be $Y(\mathcal{Y}, \mathbf{w})$, where $\mathcal{Y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$ with $\mathbf{y}_i \in \mathbb{R}^k$ with corresponding probabilities $\mathbf{w} = (w_1, ..., w_n)$, where n is the number of original scenarios, and a reduced scenario set be $X(\mathcal{X}, \mathbf{v})$, where $\mathcal{X} = (\mathbf{x}_1, ..., \mathbf{x}_m)$ with $\mathbf{x}_i \in \mathbb{R}^k$ with corresponding probabilities $\mathbf{v} = (v_1, ..., v_m)$, where m is the number of reduced scenarios. Then, scenario reduction is a technique that finds $X(\mathcal{X}, \mathbf{v})$ that well represents $Y(\mathcal{Y}, \mathbf{w})$, where $\mathcal{X} \subset \mathcal{Y}$ for discrete scenario reduction or $\mathcal{X} \not\subset \mathcal{Y}$ for

continuous reduction, when $m << n$ (Ziel, 2021; Rujeerapaiboon *et al.*, 2022). Here, 'representing' refers to minimizing the distance between sets. Therefore, adequately measuring the distance between scenarios is crucial. In general, the distance between scenarios ($\mathbf{y} \in \mathbb{R}^N$ and $\mathbf{x} \in \mathbb{R}^N$) is calculated by the Manhattan distance (l_1 distance) and the Euclidean distance (l_2 distance). For the distance between the two probability distributions ($\mathcal{Y}$ and $\mathcal{X}$), the Wasserstein distance (Pflug and Pichler, 2014; Rujeerapaiboon *et al.*, 2022) and energy distance (Székely and Rizzo, 2013) are widely utilized.

Despite the advances in scenario reduction techniques for various applications, significant gaps remain in the context of hydrological scenario reduction. Compared to the Euclidean distance embedded in the k-means clustering, which is relatively extensively studied, and Wasserstein distance, which is most widely used for scenario reduction in various fields (Ziel, 2021; Rujeerapaiboon *et al.*, 2022), the Manhattan distance and energy distance have limited use. In particular, few studies have investigated distance measures suitable for hydrological scenario reduction. In addition, reduction methods are often verified based on numerical experiments of the stochastic optimization problem (Germscheid *et al.*, 2024; Zhang *et al.*, 2023; Dvorkin *et al.*, 2014; Sharma *et al.*, 2013). However, performance may depend on the experimental sample. Moreover, if generated scenarios cannot accurately represent the real observation, it is difficult to distinguish whether the limitation of the stochastic optimization result is mainly caused by scenario generation or scenario reduction. Accordingly, it is essential to examine how much the reduced scenario set preserves the characteristics of the original scenario and how well it fits the characteristics of the flood control (Horejšová *et al.*, 2020; Hu and Li, 2019). While capturing extreme events is crucial for flood control due to their significant impact on hydrological flood conditions, it has not been fully explored which reduction technique best preserves the extreme events from the original scenarios.

This chapter aims to address the following objectives: 1) to develop and validate a data-driven probabilistic approach using Bayesian Neural Networks (BNN) for generating realistic reservoir inflow scenarios that capture the temporal dependence of hydrological time series, and 2) to systematically evaluate and compare the performance of different distance measures (Manhattan, Euclidean, Wasserstein, and energy distance) for reservoir inflow scenario reduction, with particular emphasis on their ability to preserve statistical characteristics and less likely extreme values, and 3) to provide practical recommendations on both the optimal number of reduced scenarios and the most suitable distance measures for reservoir flood control applications, considering the trade-off between computational efficiency, statistical accuracy, and preservation of extreme scenarios. To the best of our knowledge, this is the first comprehensive study that integrates a verification of a probabilistic data-driven approach for scenario generation and a rigorous evaluation of scenario reduction techniques for hydrological scenarios, especially considering flood scenarios.

This chapter is organized into four sections. Section 5.2 introduces the methodologies for scenario generation and reduction, focusing on Monte-Carlo dropout

Bayesian Neural Networks (BNN) and four distance measures. Section 5.3 details the case study area and presents the experimental results, including evaluating a reduced scenario set based on the l_1-norm to emphasize the inclusion of extreme scenarios. Finally, the conclusions summarize the key findings and implications of this chapter.

5.2 Method

5.2.1 Scenario generation by MC dropout BNN

From the stochastic programming perspective, the scenario approximates a continuous probability distribution of stochastic variables with a discrete distribution (Li *et al.*, 2020). Hydrological factors such as reservoir inflow and water level are continuously valued, and temporal dependence is critical. Therefore, the hydrological multi-period scenarios should be generated using a model in which temporal dependence can be captured. In this context, data-driven models such as Gaussian processes or BNN that can derive probabilistic hydrological factors are suitable for generating hydrological scenarios (Liu *et al.*, 2023b). However, compared to conventional Gaussian processes, which assume that the probability of uncertainty follows a Gaussian distribution, BNN can easily utilize arbitrary distributions and is known to be efficient for learning and evaluating large-scale data (Bao and Velni, 2023).

A BNN is an artificial neural network trained to derive probabilistic outputs based on Bayesian inference (Jospin *et al.*, 2022). In BNN, each parameter θ in (5.1a), which is traditionally treated as deterministic in DNN, follows a specific probability distribution, as follows:

$$\tilde{\mathbf{y}} = F(\mathbf{x}; \theta), \tag{5.1a}$$

$$\theta \sim p(\theta), \tag{5.1b}$$

where $p(\theta)$ is the probability distribution of each element of θ. BNN learns a probability distribution p of θ during the training process. However, estimating the posterior probability of parameters is computationally intensive and time-consuming (Gal and Ghahramani, 2016). As a result, Gal and Ghahramani (2016) proposed a Monte-Carlo (MC) dropout method. By randomly deactivating some nodes (dropout) and repeating random dropouts (Monte-Carlo simulation), we can obtain multiple estimated outputs, i.e., uncertain scenarios. The probability distribution of the outputs can be estimated. Thus, MC dropout BNN can be considered a BNN with a Bernoulli distribution for every θ (Theobald *et al.*, 2021; Gal and Ghahramani, 2016). Given the straightforward implementation of a dropout technique, this method has low complexity and enables fast approximation (Mae *et al.*, 2021; Sadr *et al.*, 2021).

5.2.2 Scenario reduction

The Manhattan and Euclidean distances are defined using the l_1-norm and l_2-norm between corresponding time series vectors of each scenario, respectively, as follows:

$$d_l(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_l = (\sum_i |x_i - y_i|^l)^{1/l}, \tag{5.2}$$

where d_l denotes the Manhattan distance when $l = 1$ and Euclidean distance when $l = 2$ between scenarios, $\mathbf{x}$ and $\mathbf{y}$, in which x_i and y_i are the i^{th} elements of each scenario.

Scenario reduction using the Manhattan and Euclidean distances utilizes the widely used clustering algorithm. The original scenarios are divided into m clusters, and the cluster centroids serve as reduced scenarios. K-median is a representative clustering algorithm using the Manhattan distance (Arya *et al.*, 2001), and k-means is a popular clustering algorithm using the Euclidean distance (Yadav and Sharma, 2013). Compared to the commonly used k-means algorithm, k-median is more effective for data with many outliers and asymmetric distributions due to the l_1-norm. Nevertheless, computational complexity is a disadvantage compared to k-means clustering.

An additional distinction is that the centroid in a cluster of k-median is the element-wise median of scenarios in a cluster, whereas the k-means' centroid is the element-wise mean. Every centroid is a reduced scenario. Therefore, a reduced scenario set contains a new set of time series chosen continuously in $\mathbb{R}^k$(i.e. $\mathcal{X} \not\subset \mathcal{Y}$), and so the method is also called continuous scenario reduction. Even though the continuous scenario reduction offers flexibility to find reduced sets from infinitely many options from a continuous set, it also has issues with respect to the validity of the new scenarios (Rujeerapaiboon *et al.*, 2022). For example, in hydrological scenarios, preserving the temporal dependence within each scenario is essential. Thus, selecting the element-wise median or mean of each scenario subset in clustering may weaken the preservation of the temporal dependence that was present in the original scenarios. Therefore, we modify the algorithms to select a reduction scenario from the original scenarios by identifying the one closest to each cluster's centroid (median or mean), measured by the corresponding distance measure, as follows:

$$\mathbf{x}_i = \arg\min_{\mathbf{y}_j \in C_i} \|\mathbf{y}_j - \mu_i\|_l, \tag{5.3}$$

where $\mathbf{x}_i$ is a reduced scenario of cluster i, μ_i is the mean of all scenarios in the cluster C_i. $l = 1$ is Manhattan distance and $l = 2$ for the Euclidean distance. As a result, this approach differs from conventional k-means clustering in that cluster means are not preserved.

The probability of a centroid of cluster C_i by k-median and k-means is defined as follows:

$$\Pr(\mathbf{x}_i) = \frac{|C_i|}{n}, \tag{5.4}$$

where $|C_i|$ is the number of scenarios in C_i and n is the total number of original scenarios.

Conversely, the Wasserstein distance and energy distance can be used to calculate the distance between two probability distributions. The Wasserstein distance can be expressed as the minimum (transport) cost required to make two probability distributions identical, as follows (Panaretos and Zemel, 2019):

$$d_{W,p}(X,Y) = \left(\inf_{\gamma \in \Gamma(X,Y)} \int_{\mathbb{R}^k \times \mathbb{R}^k} \|x-y\|^p \, d\gamma(x,y) \right)^{1/p}, \tag{5.5}$$

where X and Y represent the probability distributions of different scenario sets. $d_{W,p}$ is the p-Wasserstein distance, and $\Gamma(X,Y)$ is the joint probability distribution of X and Y.

The energy distance is defined as (5.6) (Székely and Rizzo, 2013; Ziel, 2021).

$$d_{E,p}(X,Y) = 2\mathbb{E}\|X-Y\|_2^p - \mathbb{E}\|X-X^{'}\|_2^p - \mathbb{E}\|Y-Y^{'}\|_2^p, \tag{5.6}$$

where $X^{'}$ is the independent and identically distributed (i.i.d.) copy of scenario distribution X. The energy distance effectively captures comprehensive distributional differences by considering both the inter-distributional and intra-distributional distances.

For these two distance measures, scenario reduction involves finding a subset X of the original scenario set Y that minimizes d_W or d_E, as follows:

$$X = \arg\min_{X \subset Y} d(X,Y), \text{ where } |X| = m, \tag{5.7}$$

where $d(X,Y)$ represents either d_W in (5.5) or d_E in (5.6), and m is the number of reduced scenarios.

The exact computation of the Wasserstein distance is computationally complex. However, (5.7) can be simplified to a (mixed-integer) linear program (Dupačová *et al.*, 2003; Rujeerapaiboon *et al.*, 2022). The work in Heitsch and Römisch (2003) suggested the optimal redistribution rule, where the probability of a reduced scenario equals the sum of the original probabilities of the unselected scenarios proximate to this reduced scenario. This allows us to calculate reduced scenarios and their probabilities without explicitly solving a mixed-integer linear program (Ziel, 2021; Heitsch and Römisch, 2003). However, the energy distance-based scenario reduction in (5.7) is computationally disadvantageous compared to the Wasserstein distance, as it requires solving a quadratic program with linear constraints (Ziel, 2021).

It is worthwhile to clarify that different algorithms are utilized for different distance measures. This is due to the different properties of each distance measure. The Manhattan and Euclidean distances are the distances between vectors, i.e., each scenario. However, the Wasserstein and energy distances can be measured between probability distributions, i.e., a distribution of original scenarios and a distribution of reduced scenarios. This difference leads to the utilization of different algorithms. We can use the distances between two scenarios to sort scenarios, i.e., clustering, and pick one representative scenario in each cluster. In contrast, we can compare the similarity between the original scenario set and the reduced set using the Wasserstein or energy distance (Heitsch and Römisch, 2007; Ziel, 2021). Thus, we can

either add scenarios to the reduced set (which starts empty) one by one, selecting the scenario that minimizes the distance between the two distributions (forward selection), or subtract scenarios from the reduced set (which initially includes all original scenarios). Alternatively, the reduced scenario set can be found directly by calculating all possible combinations of reduced scenarios and selecting the one with the smallest distance from the original scenarios.

If we apply the forward or backward selection method to the Manhattan or Euclidean distance, it is clear that only the scenario closest to the mean would be selected at each step. Furthermore, it is not possible to apply any clustering algorithms using the Wasserstein or energy distance. Therefore, it is evident that we only have limited feasible algorithms once a distance measure is selected. In addition, given that applying the same algorithm for all distance measures is not possible, it would be better to apply the 'best' or 'widely-used' algorithm for each distance measure to ensure a fair comparison. This is why we utilize the K-means for the Euclidean, K-median for the Manhattan, and forward selection for the Wasserstein and energy distance.

5.3 Results and Discussion

5.3.1 Study flood events

Since we aim to generate and reduce scenarios for flood events, it would be appropriate to use data exclusively from flood events. However, with only nine flood events available, the data are insufficient to train a BNN model. Accordingly, all hourly data of reservoir inflow and upstream water levels from September 2012 to 2020 are used to train and validate the BNN model. For the validation, 20% of the data is randomly selected. The model testing, however, is conducted on two specific flood events: a flood event in August 2011 (Event 1: 2011-08-07 to 2011-08-18, for 259 hours) and another in August 2012 (Event 2: 2012-08-12 to 2012-08-21, for 221 hours).

5.3.2 Scenario generation by BNN

An MC dropout BNN model for scenario generation has N output nodes corresponding to the prediction horizon N. Reservoir inflow, spatial mean rainfall, and the upstream water levels from 16 stations during autoregressive horizon B constitute the input features, so the number of input nodes is $B \times (1 + 1 + 16)$. While a smaller N can enhance the BNN model performance (Ghobadi and Kang, 2022), it compromises the effectiveness of reservoir flood control optimization in a receding horizon framework compared to a large N. Therefore, in this chapter, N is set to 12.

The performance of DNN models depends heavily on hyperparameters (Feurer and Hutter, 2019; Bergstra *et al.*, 2011). In this chapter, to optimize hyperparameters of the MC dropout BNN model, we employ the Tree Parzen Estimator

Table 5.1 | Hyperparameter optimisation result

B	dropout rate	nodes	hidden layers	learning rate	batch	activation
24	0.1	512	3	0.0005	64	Relu

(TPE) algorithm, which is a popular and widely validated method in hyperparameter optimization (Bergstra *et al.*, 2011; Watanabe, 2023). The Tree-structured Parzen Estimator (TPE) algorithm is a type of Bayesian optimization that models the conditional probability distributions of hyperparameters. In detail, it constructs two density distributions using Kernel Density Estimators, especially Parzen estimators here, as the surrogate models of the hyperparameter distributions associated with superior performance($l(x)$) and inferior performance($g(x)$), where x represents a hyperparameter set. The algorithm iteratively selects new hyperparameter sets by maximizing the expected improvement criterion, defined as the ratio $l(x)/g(x)$, thereby efficiently exploring the search space to identify optimal configurations. Here, the Root Mean Square Error (RMSE) of validation data is utilized as a performance index of hyperparameter sets. We implement this algorithm using Optuna (Akiba *et al.*, 2019), a Python library. To reduce the number of hyperparameters to be optimized, the early-stopping technique (Zhang *et al.*, 2020) is utilized, terminating the training process when validation loss shows no improvement over a specified epochs. Additionally, the number of past values B is also optimized at the same time. The optimal hyperparameters are presented in Table 6.3, and the architecture of the BNN is in Fig 5.1.

The MC dropout BNN model is evaluated using the Root Mean Square Error (RMSE) and the Nash-Sutcliffe model efficiency Coefficient (NSC), which are widely used to evaluate the reliability of a hydrological prediction model, as demonstrated in Fig 5.2 (Nash and Sutcliffe, 1970). For test events, the RMSE is 112.5 m^3/s, given that the mean inflow over the entire period is 106.2 m^3/s and the peak inflow is 3557.0 m^3/s. In addition, the NSC is 0.736, considering that 0.65 or higher NSC is generally judged as 'good' and 'very good' for over 0.75 in the hydrological prediction model (Moriasi *et al.*, 2007). Overall, the RMSE and NSC indicate the reliable performance of this model.

In particular, as shown in Fig 5.2, the BNN model shows a degradation of hydrological prediction performance with increasing prediction horizon, i.e., increasing the RMSE and decreasing the NSC. For example, the RMSE increases from 83.3 m^3/s (NSC: 0.885) for 1-hour predictions to 149.4 m^3/s (NSC: 0.432) for 12-hour predictions.

However, the prediction performance also degrades for the peak inflow, as illustrated in Fig 5.3. This limitation is particularly significant considering the importance of peak inflow during flood events for reservoir flood control. This degradation is mainly attributed to the limitation of training data, i.e., the lack of sufficient data with high inflow. While we utilize the data from the entire period for model training, the majority of cases have peaks that are much smaller than shown for Event 1 and Event 2 in Fig 5.3, where typical patterns have an upstream water level and reservoir inflow increasing to much lower levels for a short period and then

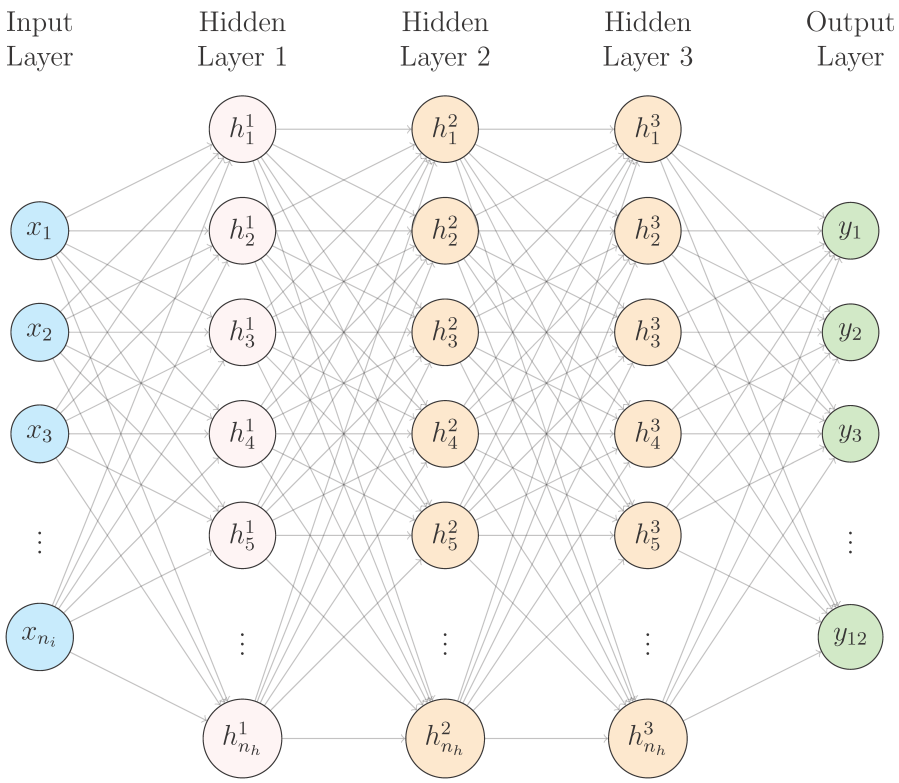


Figure 5.1 | Architecture of the Bayesian neural network (BNN) with three hidden layers. The network consists of $B \times 18$ input nodes ($n_i = B \times 18$), where B is the autoregressive horizon, including past inflow, rainfall and upstream water levels, three hidden layers with 512 nodes each ($n_h = 512$), and 12 output nodes. Connections between each layer are dropped randomly at a dropout rate of 0.1.

decrease smoothly. Training the BNN model only for peak inflows would result in improving peak inflow prediction accuracy but would substantially degrade overall model performance due to data insufficiency, compromising its practical utility.

5.3.3 Scenario reduction

Scenario reduction using clustering algorithms is implemented using the Python library, scikit-learn (Pedregosa *et al.*, 2011). For scenario reduction based on the Wasserstein distance and energy distance, we follow the method of (Ziel, 2021). As discussed, the final reduced scenarios $\mathbf{x} \in \mathcal{X}$, which are closest to centroids based on the distance measures, are selected from the original scenarios, to ensure $\mathcal{X} \subset \mathcal{Y}$. This approach differs from conventional clustering-based scenario reduction because it does not use the element-wise median (k-median) or mean (k-means) of each scenario cluster.

The original scenario set Y consists of 1000 scenarios to cover possible uncertain conditions, each represented by a 12-dimensional vector. These scenarios are generated at each time step. Considering the high dimensionality of each scenario and the substantial number of original scenarios, a simple 1-step forward selection is applied for scenario reduction with the Wasserstein distance and energy distance (Heitsch

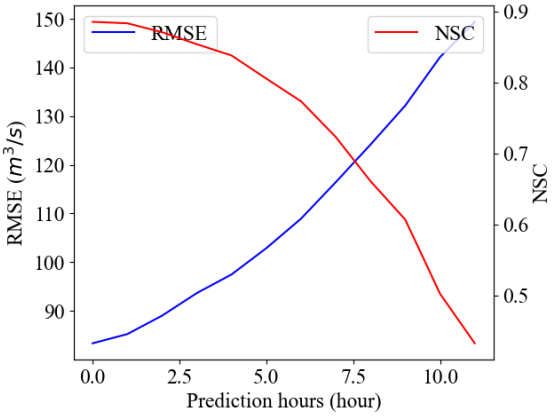


Figure 5.2 | BNN model performance with prediction length in RMSE and NSC metrics.

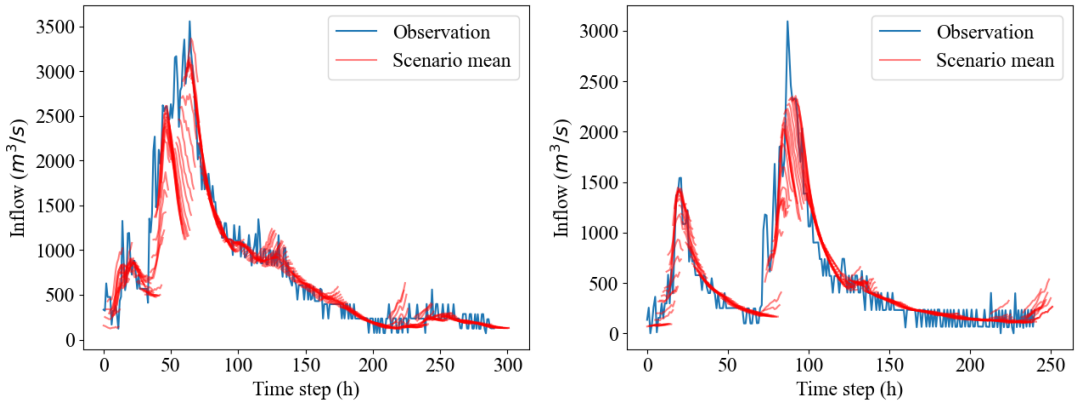


Figure 5.3 | Real and BNN prediction hydrographs for test events: (a) Event 1, (b) Event 2.

and Römisch, 2007; Ziel, 2021). In addition, the parameter p in (5.5) and (5.6) is set to one for both distance measures (Ziel, 2021).

Testing the scenario reduction for every time step of Event 1 and Event 2 is computationally intensive, particularly due to the computational complexity of the energy distance. Therefore, we examine three representative points from each event (case 1&4: increasing, case 2&5: decreasing, and case 3&6: stable), as shown in Fig 5.4, to evaluate the differences between distance measures..

When the number of reduced scenarios is $m = 10$, the scenario reduction results for each case are shown in Fig 5.5. In the figure, the solid grey lines refer to the original scenarios, the solid blue lines illustrate the reduced scenarios, and the thickness of the solid blue line represents the probability of each reduced scenario.

The water level and outflow from the reservoir during a flood event significantly impact the upstream and downstream flood conditions. Moreover, extreme events can have a considerable impact on dam safety. Therefore, it seems that a critical criterion for flood scenario reduction methods is whether extreme scenarios are included. From this perspective, the Manhattan and Euclidean distances demonstrate superior performance, followed by the Wasserstein and energy distance, as illustrated

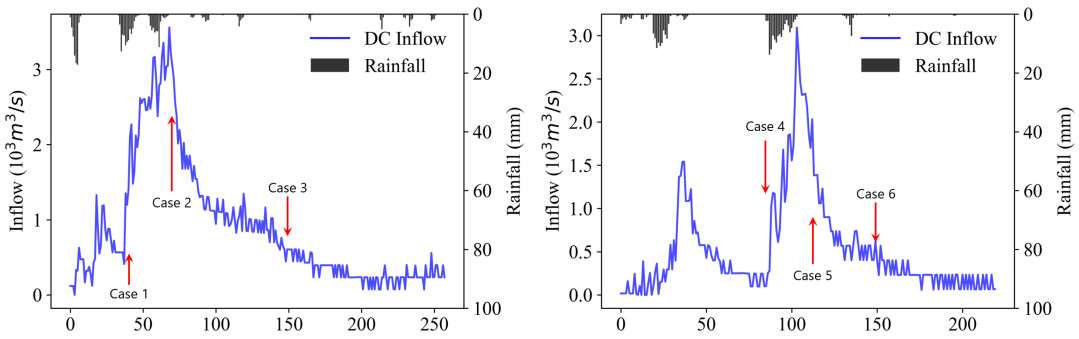


Figure 5.4 | Scenario reduction test cases in two events: (a) Event 1 (b) Event 2.

in Fig 5.5. Specifically, when using the energy distance and the Wasserstein distance, sparse extreme scenarios tend to be excluded.

Let's examine this in detail. To evaluate this, we need to estimate a scenario envelope and compare its size with the envelope size of the original scenarios. Although it is difficult to estimate the exact envelope due to the possibility of the existence of intersecting scenarios, we approximate the size of the envelope using l_1-norm as follows:

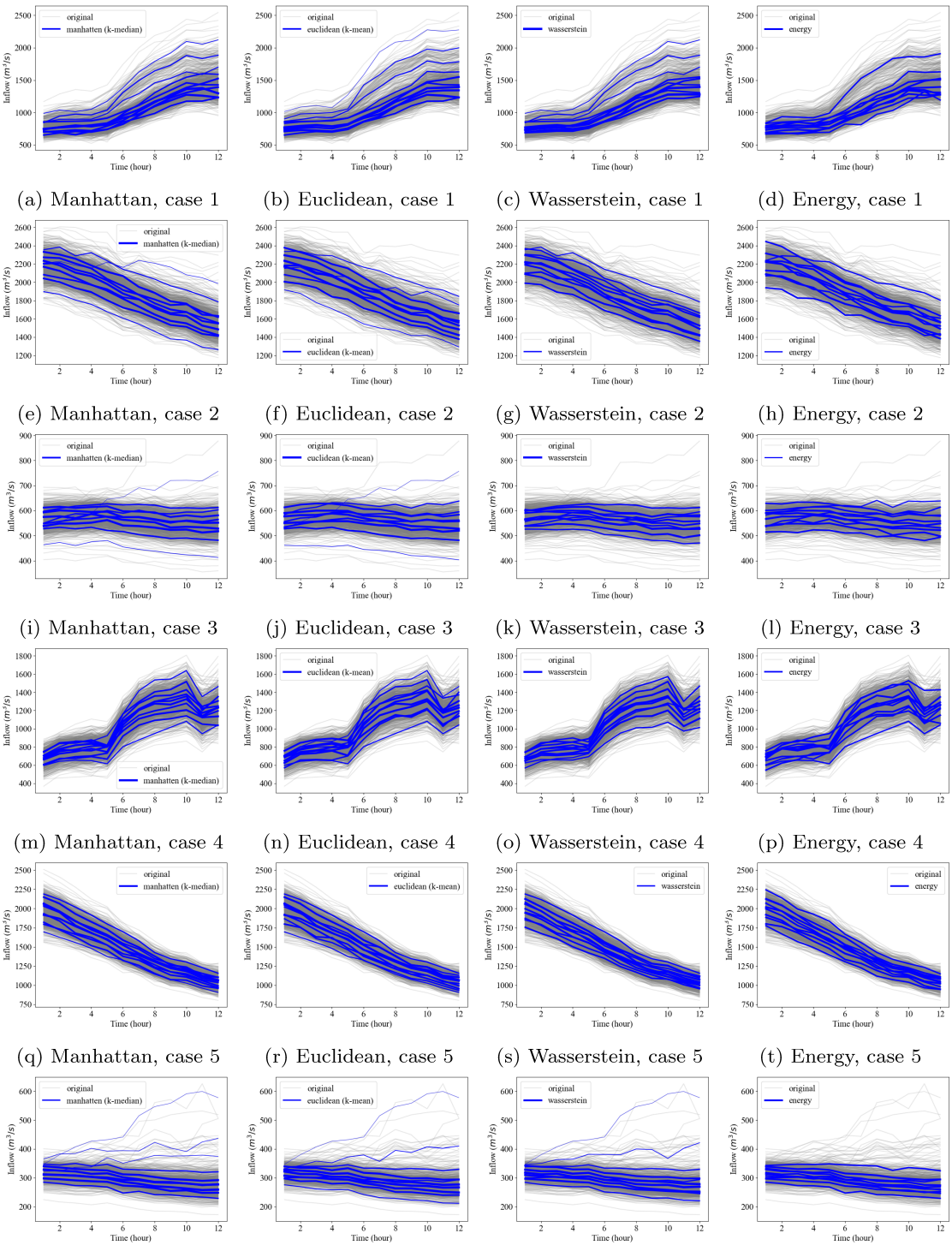
$$|\text{env}(Q)| = \max_{q \in Q} \|q\|_l - \min_{q \in Q} \|q\|_l, \tag{5.8}$$

where $|\text{env}(Q)|$ represents the size of the envelope of a scenario set Q, which can be an original or reduced set. Here, we set l to one.

The upper envelope encapsulates the maximum inflow coming in, and the lower envelope encapsulates the minimum inflow. These span the flood risk space as a function of water coming into the reservoir and are so important. Moreover, since accurately predicting the total volume of inflows is more significant than predicting each element of inflow scenarios precisely in terms of short-term flood scenarios, using l_1-norm, i.e., the sum of a scenario where all elements should have positive values, is reasonable. We can use this envelope volume, i.e., a span of flood risk in the scenario set, as a metric of goodness in assessing the reduced scenario set.

We analyze $|\text{env}(Q)|$, which is the difference between scenarios with maximum and minimum sums, for five different numbers of reduced scenarios ($m = 10$, 20, 30, 40, and 50), as illustrated in Fig 5.6. As described in Fig 5.5, the Manhattan distance and Euclidean distance preserve extreme scenarios even with small m values. The energy distance tends to focus more on the scenarios that frequently appear. All distance measures include more extreme scenarios as the number of reduced scenarios m increases. The result suggests that reduced scenario sets adequately capture extreme scenarios when $m \geq 30$. It is worth noting that this metric, $|\text{env}(Q)|$, is fair to the Wasserstein and energy distances because calculating l_1-norm in (5.8) is similar to the calculation of the Manhattan and Euclidean distance.

The mean (μ) and the standard deviation (σ) are the most widely used (Ziel, 2021) and sufficient metrics for comparing probability distributions of hydrological data (Laio *et al.*, 2009). This is reasonable because the uncertainty of inflow time



(a) Manhattan, case 1 (b) Euclidean, case 1 (c) Wasserstein, case 1 (d) Energy, case 1

(e) Manhattan, case 2 (f) Euclidean, case 2 (g) Wasserstein, case 2 (h) Energy, case 2

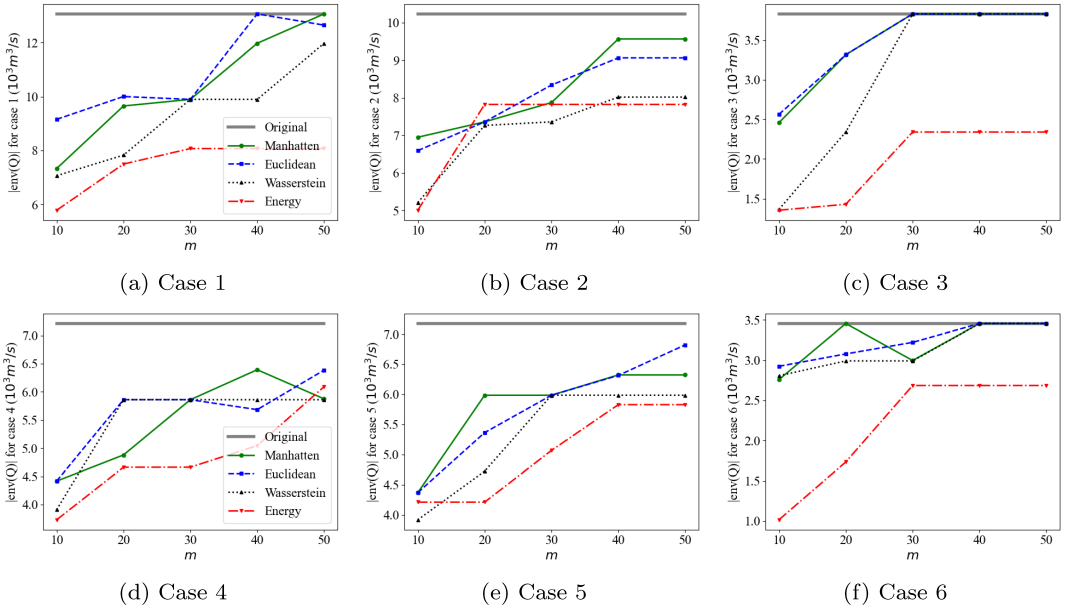
(i) Manhattan, case 3 (j) Euclidean, case 3 (k) Wasserstein, case 3 (l) Energy, case 3

(m) Manhattan, case 4 (n) Euclidean, case 4 (o) Wasserstein, case 4 (p) Energy, case 4

(q) Manhattan, case 5 (r) Euclidean, case 5 (s) Wasserstein, case 5 (t) Energy, case 5

(u) Manhattan, case 6 (v) Euclidean, case 6 (w) Wasserstein, case 6 (x) Energy, case 6

Figure 5.5 | Scenario reduction results when $m = 10$.



(a) Case 1 (b) Case 2 (c) Case 3

(d) Case 4 (e) Case 5 (f) Case 6

Figure 5.6 | The size of the envelopes of the original and reduced scenario sets for each case.

series, e.g., residuals for model uncertainty, has been modelled using a logistic distribution or a normal distribution (Coxon *et al.*, 2015; Montanari and Brath, 2004), whose statistics can therefore be completely described by the mean and standard deviation. We compare the mean and standard deviation between the original and reduced scenarios as shown in Fig 5.7. Regarding scenario means, the energy distance demonstrates the closest alignment with the original scenarios and the lowest Mean Absolute Errors (MAE) of means in Fig 5.7b. The Euclidean distance exhibits the largest differences, together with the Manhattan distance. Notably, the maximum difference in means is 3.5 m^3/s when $m = 10$ for the Manhattan distance, representing only 0.3% of the mean of 1045.3 m^3/s in the original scenarios, as shown in Fig 5.7a.

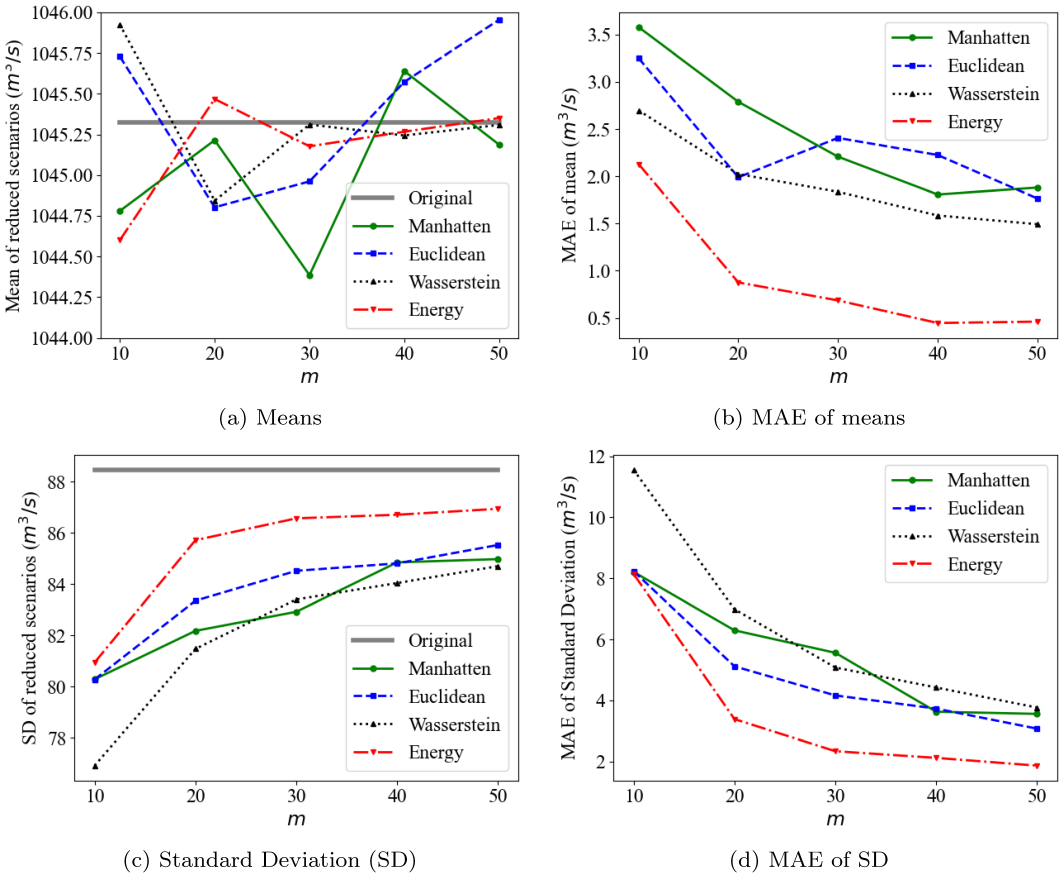
In terms of standard deviation, the energy and Euclidean distances demonstrate the minimum differences, as illustrated in Fig 5.7d. Notably, unlike the means of reduced scenarios, which show no clear relationship with m in Fig 5.7a, the standard deviations across all distance measures gradually converge to the original scenarios' σ as m increases, as shown in Fig 5.7c.

Regarding time-series scenarios, temporal correlation is one of the key characteristics, along with mean and variance (Gopikrishnan *et al.*, 2000). The Pearson correlation coefficient is a widely used statistical measure for quantifying linear correlation between two variables. The linear relationship between variables p and q is defined as follows:

$$r = \frac{\mathrm{Cov}(p, q)}{\sigma_p \sigma_q}, \tag{5.9}$$

where $\mathrm{Cov}(p, q)$ denotes the covariance and σ the standard deviation.

To examine how the temporal dependence structure of the predicted inflows is preserved across different distance measures, the element-wise correlations are



(a) Means

(b) MAE of means

(c) Standard Deviation (SD)

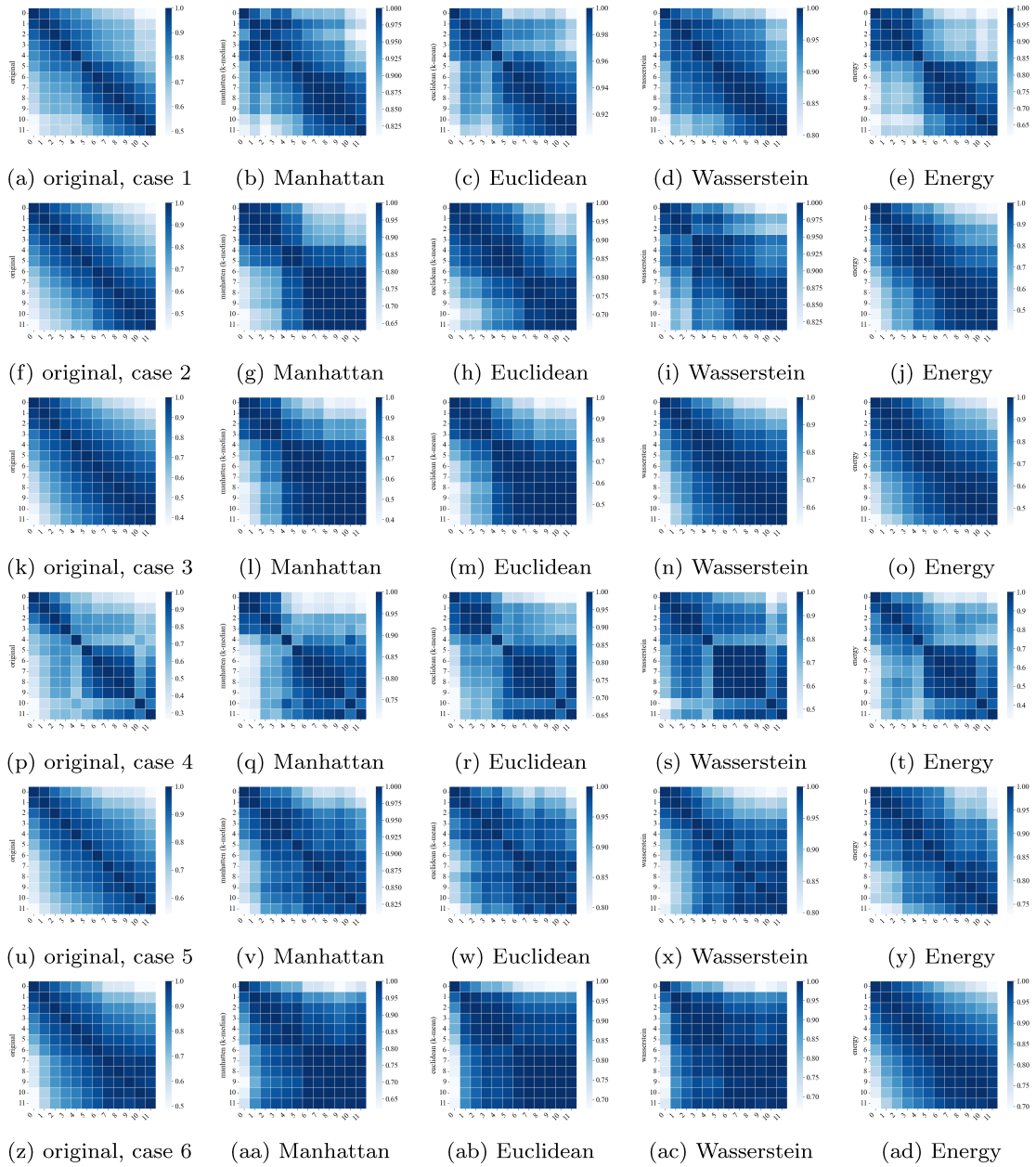
(d) MAE of SD

Figure 5.7 | Means and Standard deviations of original and reduced sets.

presented in Fig 5.8. The leftmost panel shows the element-wise correlation matrix of the original scenario set, while the others display the correlation matrices of the reduced sets obtained using the four distance measures. The correlation structure of the reduced set obtained using the energy distance exhibits the highest similarity to the original set, particularly evident in case 4 (Fig 5.8p to 5.8t). In contrast, the Wasserstein distance shows the most divergent correlation structure, and the Euclidean distance is similar to the Manhattan distance, as demonstrated in case 2 (Fig 5.8f to 5.8i).

For a detailed comparison, we calculate the mean absolute difference between the original scenarios' temporal correlation and the reduced sets' correlations for five different numbers of reduced scenarios ($m = 10$, 20, 30, 40, and 50). As depicted in Fig 5.9, the energy distance exhibits the smallest difference, while the Wasserstein distance demonstrates the largest deviation. As expected, the differences decrease as the number of reduced scenarios increases.

From the computational point of view, the energy distance exhibits the highest complexity and lowest efficiency. This is because $n - i$ quadratic problems must be solved to select the i^{th} scenario when applying a simple 1-step forward selection. The computational time increases linearly as the number of reduced scenarios and/or original scenarios increases. Because of the exhaustive exploration of all



(a) original, case 1 (b) Manhattan (c) Euclidean (d) Wasserstein (e) Energy

(f) original, case 2 (g) Manhattan (h) Euclidean (i) Wasserstein (j) Energy

(k) original, case 3 (l) Manhattan (m) Euclidean (n) Wasserstein (o) Energy

(p) original, case 4 (q) Manhattan (r) Euclidean (s) Wasserstein (t) Energy

(u) original, case 5 (v) Manhattan (w) Euclidean (x) Wasserstein (y) Energy

(z) original, case 6 (aa) Manhattan (ab) Euclidean (ac) Wasserstein (ad) Energy

Figure 5.8 | Pearson correlation between inflows at different times when $m = 10$.

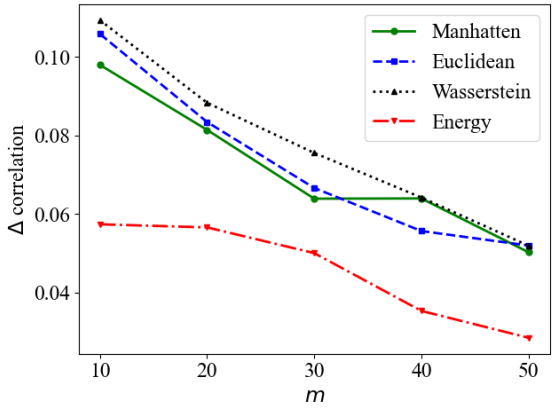


Figure 5.9 | Difference in correlations between original and reduced ones for five m values.

Table 5.2 | Computational times for four distance measures (seconds)

m	Manhattan	Euclidean	Wasserstein	Energy
10	4.2	4.0	11.8	72.5
20	3.8	4.0	23.8	308.7
30	3.8	4.7	39.0	712.0
40	3.8	5.0	62.7	1131.5
50	4.0	5.0	82.3	1787.3

potential combinations when incorporating the exact energy distance, the computational complexity would increase exponentially. The other distance measures are computationally efficient because of established efficient algorithms that do not need an explicit optimization process (Rujeerapaiboon *et al.*, 2022). For the Manhattan and Euclidean distances, the computation time remains relatively constant regardless of increases in m, as demonstrated in Table 5.2.

5.3.4 Discussion

Our research aims to verify the applicability of probabilistic data-driven models to generate uncertain scenarios considering the characteristics of hydrological scenarios. Here, the characteristics include both the temporal correlations and the accuracy degradation with increasing the prediction horizon. Our results suggest that the MC dropout BNN model effectively generates uncertain inflow scenarios. Without explicitly defining temporal correlations between elements in a scenario, the BNN model demonstrates reasonable performance (in terms of the RMSE and NSC). In addition, it reproduces the characteristic behavior of hydrological prediction, namely the degradation of performance with increasing prediction horizon (Fig 5.2).

However, our BNN model exhibited limitations in accurately predicting peak inflow (Fig 5.3). This limitation is primarily attributed to insufficient flood event data. The training dataset predominantly consists of 'normal' inflows characterized by small and short-term increases. Extreme values, such as peak flood inflow, are located at the tail of the data distribution, making accurate prediction difficult.

When the model is trained predominantly on lower inflow values, it struggles to predict extreme values, leading to overfitting on small inflows and underfitting on extreme events (Shi *et al.*, 2024; Ding *et al.*, 2019). We use all hourly data from 2012 to 2020 to train the BNN model. There are only one or two flood events each year, and there are only nine flood events during this period, as described in Section 5.3.1. This implies that most training data does not have considerable inflows. In addition, in most cases, the inflow decreases very soon without demonstrating significant increases, even when the inflow increases. Therefore, the model's inability to predict peak values can be attributed to insufficient exposure to extreme data.

We evaluate the advantages and disadvantages of existing scenario reduction methodologies, particularly from the perspective of distance measures. The performances are assessed by three criteria: preservation of statistical properties, inclusion of extreme events, and computational complexity. Scenario reduction using the energy distance demonstrates superior performance in preserving the statistical properties of the original scenario set generated by the BNN model. The original temporal correlation structure is best maintained by the reduced scenarios based on the energy distance (Fig 5.9). Although all distance measures effectively preserve mean values, the energy distance achieves the minimum MAE for both mean and standard deviation metrics (Fig 5.7b and 5.7d).

Consideration and analysis of extreme scenarios are crucial for flood control. The span between the upper and lower envelopes of original scenarios represents the flood risk space resulting from uncertain inflows. To take into account this span, we suggest the size of the envelope as a metric for assessing the reduced scenarios using the l_1-norm of scenarios. For this, the Manhattan and Euclidean distances demonstrate superior performance (Fig 5.6 and 5.5). Moreover, in terms of temporal correlation and standard deviation, both the Manhattan and Euclidean distances exhibit comparable performance, following the energy distance (Fig 5.9 and 5.7d). The Manhattan and Euclidean distances show that no substantial differences are observed in both statistical properties and extreme scenario preservation.

However, from the computational efficiency perspective, the use of the energy distance demonstrates significant limitations compared to other distance measures (Table 5.2). Even with only 10 scenarios, the computation time exceeds one minute, extending to approximately 12 minutes for $m = 30$. Given that the time horizon is usually less than an hour for flood control, a scenario reduction process requiring more than 10 minutes proves impractical (Breckpot *et al.*, 2013b). This is because the control inputs produced by an optimal control approach cannot be directly implemented in reservoir flood control due to their substantial impact on basin flood conditions. Additionally, the process requires time for decision-making and information sharing with relevant organizations.

Consequently, considering the importance of extreme scenarios for optimal flood operation, the Manhattan and Euclidean distances can be reasonable and practical choices for scenario reduction. However, when preserving the statistical properties, such as mean, standard deviation, and temporal correlation, is prioritized over including extreme scenarios, and computational time is not a critical concern, the

energy distance can be the best choice. Regardless of the distance measure selected, 30 reduced scenarios can be sufficient in terms of both inclusion of extreme events and preservation of statistical properties.

It should be noted that we employ a simple 1-step forward selection method for the Wasserstein and energy distances. Even though it has been widely applied due to its computational efficiency, this algorithm can not guarantee the selection of the optimal reduced set. To find the optimal reduced set closest to the original scenario set, all possible subsets should be explored. This is generally computationally intractable. However, the implementation of high-performance computing and/or parallelization techniques could potentially allow us to obtain the optimal reduced set based on the energy or Wasserstein distance with significantly reduced computational time. It remains to be determined whether this algorithm can improve the performance of these distance measures and whether reduced computational time is practical.

5.4 Conclusions

This chapter demonstrated that Bayesian Neural Networks can effectively generate probabilistic flood scenarios while capturing temporal dependencies in inflow timeseries, and provided a comprehensive evaluation of scenario reduction approaches. Our comparative analysis of four distance measures demonstrated that, contrary to its widespread adoption in literature, the Wasserstein distance offered no substantial advantages over the Manhattan, Euclidean, or energy distances for inflow scenario reduction. Each distance measure demonstrated distinct strengths and limitations in preserving statistical properties and extreme events that are critical for flood control applications.

From a statistical point of view, the use of the energy distance seems to be the preferred option, followed by the Manhattan and Euclidean distances. However, considering the importance of extreme scenarios in flood control and the computational complexity of the energy distance, the Manhattan and Euclidean distances are seen as effective and practical measures for reducing the hydrological scenarios.

Although scenario generation and reduction have been extensively studied in various fields, such as financial and electricity markets, their application to hydrological scenarios has received comparatively less attention. This chapter aims to contribute to and stimulate further research on hydrological scenario generation and reduction. The method presented here can be utilized both for generating probabilistic predictions of hydrological variables and for solving reservoir optimization problems.

One of the limitations of this chapter is that the importance of including extreme scenarios was not demonstrated through the use of the scenarios in flood control examples. As future work, it is interesting to give much more attention to this problem and demonstrate how two different reduction scenarios, based on the Euclidean and energy distance, influence the stochastic optimization problem solution.

Data Availability for this chapter

The hydrological data in this chapter were acquired from the public data portal (www.data.go.kr) and Korea Water Resources Public Corporation's website (http://kwater.or.kr). The data and code of this chapter are available at https://doi.org/10.4121/e343331b-496f-40ab-83eb-f546df6dffa6 under the CC-BY-4.0 licence.

6

Stochastic MPC with Conditional Value-at-Risk Constraints

Summary

This chapter presents a management framework for short-term reservoir flood control risk under hydrological prediction uncertainty with a stochastic Model Predictive Control (MPC) framework. An extensive set of uncertain scenarios for reservoir inflows and downstream water levels is generated through a Monte-Carlo dropout Bayesian Neural Network (BNN). To reduce the computational burden for stochastic MPC, we reduce scenarios by k-means clustering and the energy distance. The uncertainty represented in various scenarios is managed through the Conditional-Value-at-Risk (CVaR) constraint. We utilize CVaR_β soft constraints for which corresponding thresholds are selected by an operator so as to reflect the fact that physical thresholds are rarely activated during flood events. Our numerical experiment results on the Daecheong reservoir and downstream station (Guemnam) in South Korea show that CVaR_β soft constraints respond straightforwardly to the change in the operator's risk-averseness, β. In addition, a feasible solution can always be found that minimizes the threshold excess with a specified violation probability, even when constraint violation is inevitable. Finally, considering closed-loop performance and computational complexity, the scenario reduction based on the Euclidean distance is more effective and efficient than the energy distance-based reduction because it includes scenarios close to the upper and lower bounds.

6.1 Introduction

Conventional deterministic receding horizon Model Predictive Control (MPC) is somewhat robust to uncertainty due to receding horizon implementation, which continuously updates control inputs reflecting the latest state, but has limitations

in systematically dealing with probabilistic uncertainties (Schwenzer *et al.*, 2021). Robust MPC (RMPC) and Stochastic MPC (SMPC) provide a framework for applying the MPC approach to uncertainty (Heirung *et al.*, 2018). For all uncertain scenarios, robust MPC, which wants the constraints to be satisfied, has limitations in deriving overly conservative control inputs (Langson *et al.*, 2004). In contrast, SMPC directly utilizes the probabilistic nature of the uncertainty, which is already known or can be estimated in the MPC optimization process, in which system variables can be constrained with an acceptable violation or satisfaction probability. Therefore, it is possible to alleviate the conservatism of RMPC.

Applying a widely-used chance constraint to receding horizon MPC requires explicit probabilistic characteristics of the uncertainty (Farina *et al.*, 2016). Furthermore, since the purpose of a chance constraint in SMPC is to control the probability of constraint violation, there is a limit to controlling risk in case of inevitable constraint violation if the risk means the excess amount as well as the exceedance probability. Conditional Value-at-Risk (CVaR) is a metric to quantify and minimize the risk (Rockafellar *et al.*, 2000). In particular, CVaR is easy to linearize and has the advantage of not having to predefine an explicit probability distribution for uncertainty when applied in combination with the Sample Average Approximation (SAA) technique (Rockafellar and Uryasev, 2002).

Since the CVaR value can also be obtained simultaneously while minimizing the objective when using CVaR as an objective (Rockafellar *et al.*, 2000), many studies have applied CVaR objective(s) (Van Der Heijden *et al.*, 2023; Hakobyan *et al.*, 2019; Khodabakhsh and Sirouspour, 2016). For example, in the case of optimizing an investment portfolio, if we consider the probability-weighted sum of losses (or returns) whose probability is greater than the reference probability as an expected loss (return), then the probability-weighted sum of losses which have lower probability than the reference can be interpreted as the risk (Rockafellar *et al.*, 2000; Rockafellar and Uryasev, 2002). Therefore, it is reasonable to set the minimization of the CVaR as an objective. Many studies on stochastic reservoir operation have applied CVaR for the quantified (or further monetized) risk of floods or droughts as an objective (Khorshidi *et al.*, 2019; Xu *et al.*, 2017b; Huang *et al.*, 2022b; Ortiz-Partida *et al.*, 2019).

However, if limiting reservoir flood (i.e., short-term) control problems, the case where a target variable exceeds a threshold can be understood as a risk. For example, the expected water level where the water level exceeds the safety criteria of the embankment is the flood control risk of the river. In this case, an objective to minimize the CVaR for the water level will decrease the water level as much as possible, even when the water level is below the safety criteria. This can cause complex trade-offs with other objectives for flood control, e.g., it can hinder an upstream reservoir from increasing outflows to maintain the Reservoir Water Level (RWL) in the safety range. On the other hand, if the CVaR objective only minimizes the cases when the system state exceeds the threshold, it becomes less robust. Therefore, when it comes to reservoir flood control problems, it seems reasonable to apply CVaR as a constraint rather than an objective.

Physical limits are utilized in general as constraint thresholds. However, physical limits of large reservoirs, e.g., Flood Water Level (FWL) or design discharge capacity, are designed for extreme conditions such as Probable Maximum Flood. Therefore, constraints for physical limits are necessary but not sufficient in short-term flood control. In other words, it is highly likely that it will never be activated in typical flood control situations. Therefore, the constraint threshold should reasonably reflect the operator's preference in flood control. Here, preference means the threshold that an operator has through experience. For example, if the highest RWL in the last 10 years is an Elevation Level (EL) of 78.4 m, the operator would not want to exceed it as much as possible. In this case, EL 78.4 m can be the operator's threshold for the CVaR constraint for the RWL.

In this chapter, we propose the CVaR as a construct that takes into account the preferences of operators for short-term reservoir flood control. To validate CVaR constraints in stochastic MPC through a numerical experiment, we utilize methodologies from our previous research: a Monte-Carlo dropout Bayesian Neural Network (BNN) to generate an extensive set of uncertain scenarios and scenario reduction techniques to reduce computational complexity (Alarab *et al.*, 2021). In addition to demonstrating the feasibility of the CVaR constraints, we compare the performance of scenario reduction methods, which are based on the Wasserstein distance and Euclidean distance, from the perspective of closed-loop performance.

This chapter is organized into four sections. Section 6.2 describes the generation of uncertain scenarios and the application of scenario reduction methodologies. This section also develops a reservoir system model and a downstream water level prediction model, and presents the objectives and constraints of stochastic MPC, including CVaR-type constraints. Section 6.3 details the experimental results, comparing the performances of stochastic MPC against deterministic MPC and evaluating different scenario reduction methodologies. Finally, the conclusions summarize the key findings and implications of this chapter.

6.2 Method

6.2.1 System models for case study

Study flood events

Among the nine flood events from 2011 to 2020 presented in Chapter 2 for downstream area, six events are used for training system models for downstream water level, one validation event for hyperparameter optimization, and the other two events are used for numerical experiments of CVaR-based stochastic MPC, as presented in Table 6.1.

Table 6.1 | Study flood events

#	Purpose	From - To (YYYYMMDD HH, hour)	Peak inflow at DC (m^3/s)	Peak WL at GN (m)
1	Testing	20110807 10 - 20110819 00 (279)	3557	4.65
2	Testing	20120811 18 - 20120821 18 (241)	3093	3.41
3	Validation	20120827 04 - 20120913 16 (421)	3430	4.00
4	Training	20120912 20 - 20120923 18 (263)	4986	2.01
5	Training	20160703 02 - 20160709 15 (158)	3655	2.60
6	Training	20180825 11 - 20180909 02 (352)	2590	4.53
7	Training	20200712 02 - 20200718 06 (149)	3779	2.78
8	Training	20200719 12 - 20200818 05 (714)	6318	7.35
9	Training	20200901 16 - 20200911 09 (114)	1501	2.07

Reservoir Model

If the reservoir total discharge is $O_{total} \in \mathbb{R}$, the reservoir storage is $S \in \mathbb{R}$, and the total inflow is $I \in \mathbb{R}$, the reservoir system can be defined as follows:

$$\frac{\delta S}{\delta t} = I - O_{total}, \tag{6.1}$$

By discretizing with respect to time,

$$\frac{S_{t+\Delta t} - S_t}{\Delta t} = I_t - O_{total,t}. \tag{6.2}$$

where I_t and O_t denote an average inflow and outflow during Δt, the unit is m^3/s. We set the time step, Δt, to 1 hour based on the acquired data set.

Downstream Model

Reservoirs should be operated in consideration of the downstream water level. For example, when the downstream area is flooded, a decrease in outflows from upstream reservoirs is generally required. The water level of this downstream river can be estimated by using a hydraulic simulation model based on the Saint-Venant equation or a hydrological flood routing method such as the Muskingum method (Merkuryeva *et al.*, 2015). Both methods require accurate parameter values such as roughness, cross-sections, and tributary inflows, but the lack of measurement data and errors in measurement negatively affect the simulation results (Di Baldassarre and Montanari, 2009).

Recently, instead of physically-based models, studies using data-driven models such as AutoRegressive Exogenous (ARX) or Deep Neural Network (DNN) for river flood routing can be easily found (Khatibi *et al.*, 2011). DNN has excellent advantages in effectively mapping nonlinear relationships between input and output (Li *et al.*, 2021). However, there is a disadvantage in that a relatively considerable amount of data is required for training, and the model structure is generally complex. In contrast, ARX models are widely used in system identification due to their simple structure (Moradvandi *et al.*, 2023). Therefore, we applied a linear ARX model for a system model to model downstream water levels.

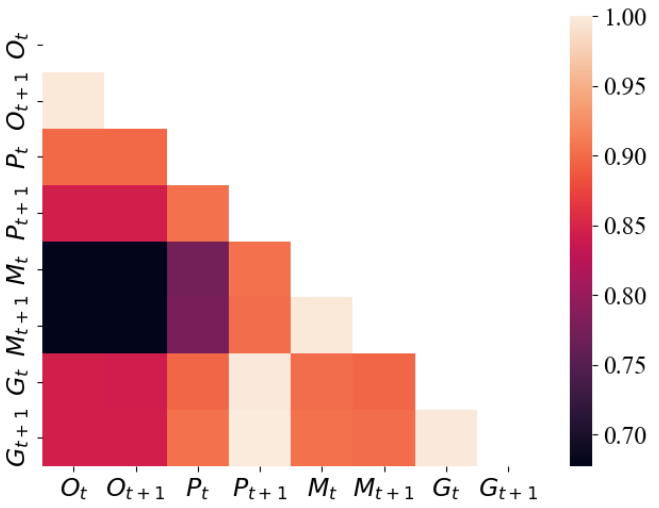


Figure 6.1 | Correlation between G_{t+1} and other input features.

The Linear ARX model assumes a linear relationship between input and output:

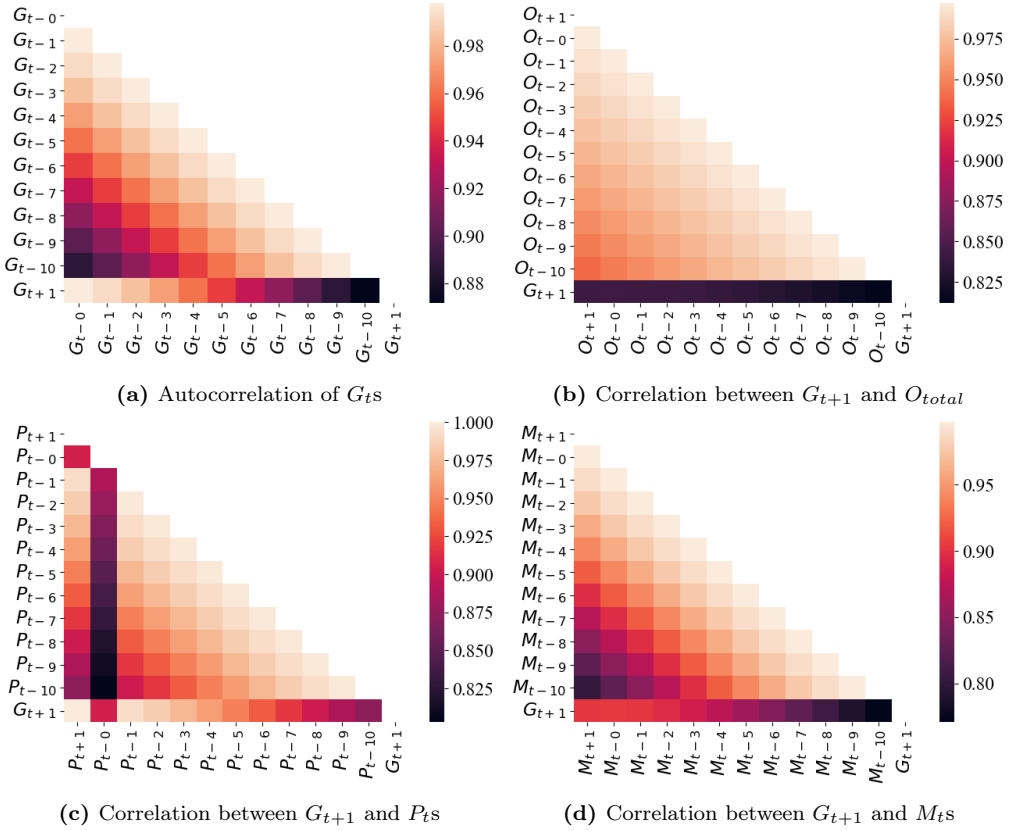
$$\tilde{\mathbf{y}} = \theta^T \mathbf{x}, \tag{6.3}$$

where input and output are $\mathbf{x} \in \mathbb{R}^{n_x}$ and $\tilde{\mathbf{y}} \in \mathbb{R}^{n_y}$, respectively. Here, $\mathbf{x}$ includes autoregressive and exogenous inputs. One thing to note is that our ARX model is a system model, not for prediction. That is, for both DNN and ARX models, the time horizon of exogenous inputs for GN water level at time $t+1$, G_{t+1}, can be up to $t+1$. This is because predicted values of exogenous factors for the entire prediction horizon are input at every time step.

First, covariance analysis was performed to determine the importance of the input factors. As a result of the covariance analysis at GN, as shown in Fig 6.1, GN water level at $t+1$, G_{t+1}, is affected in the order of G_t, and GP water levels, P_t, MH water levels, M_t, and reservoir outflow, O_t. Since O_t is the control input that needs to be optimized, G_{t+1} should be calculated inside the MPC optimization.

In general, a linear relationship exists between outflows at GN and inflows at GN, i.e., the sum of tributary discharges and DC outflows here, as well as between current and previous outflows at GN (Huang *et al.*, 2023; Chang *et al.*, 1983; Gill, 1978). The outflow at GN, then, is translated into a water level by a nonlinear discharge-water level relationship at GN, which is constructed using surveyed data. The correlation analysis results presented in Figs. 6.1 and 6.2 appear to properly reflect this theoretical relationship. However, the outflow from DC does not seem to significantly affect the water level at GN. Instead, the autocorrelations with previous GN water levels are considerably higher than the correlation with other variables. This means that the current water level at GN can be dominated by the previous water level.

The ARX model can be applied to stationary time series, and in many studies for predicting reservoir inflows, differencing is widely applied to make it stationary (Dires *et al.*, 2023). However, based on the results of the commonly used Augmented



(a) Autocorrelation of G_ts

(b) Correlation between G_{t+1} and O_{total}

(c) Correlation between G_{t+1} and P_ts

(d) Correlation between G_{t+1} and M_ts

Figure 6.2 | Autocorrelation and exogenous correlation where G_{t+1}, P_t, and M_t denote water level at GN, GP, and MH, respectively.

Dickey-Fuller test (Cheung and Lai, 1995), the p-value is 0.0002, which allows us to infer that our data is stationary at the 0.01 significance level (99% confidence)."

To implement MPC in a receding horizon manner, T linear ARX models can be utilized when the length of the prediction horizon is T. Each linear ARX model estimates the downstream water level, G_t, at time t. The numbers of past values for autoregressive and exogenous inputs of each ARX model, ρ and η, respectively, are selected from the range between one and T in such a way that the RMSE on the validation set is minimized. Therefore, the selected values of ρ and η for an ARX model at time t can be expressed as follows:

$$\rho_t^*, \eta_t^* = \arg\min \sqrt{\|\mathbf{y}_t - \theta_{\rho,\eta}^T \mathbf{x}_t\|_F^2}, \tag{6.4}$$

where $\|\mathbf{A}\|_F^2 = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2$, in which $a_{i,j} \in \mathbf{A}$ and $\mathbf{A} \in \mathbb{R}^{n \times m}$.

Alternatively, a linear ARX model where $t = 1$ in (6.4) for the one-step prediction can be recursively implemented as in Fig 6.3.

We compared both ARX models for every $t \in T$ and a one-step linear ARX model for recursive implementation with optimized autoregressive and exogenous input horizons. Both models show similar RMSEs, as shown in Fig 6.4a. Therefore, we utilize a simple one-step linear ARX model.

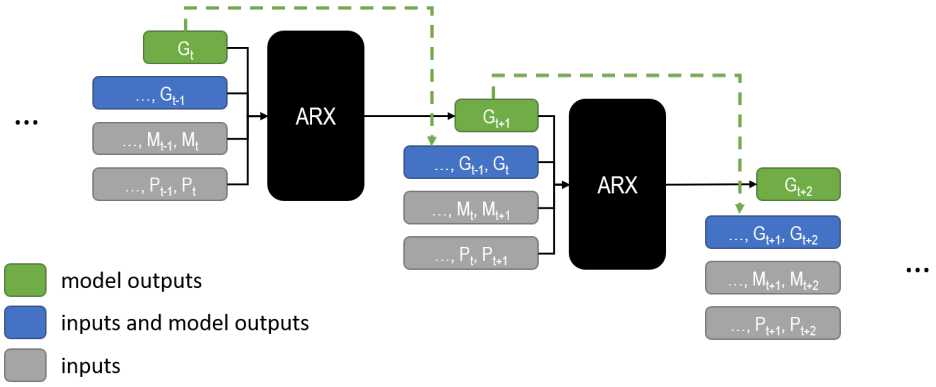


Figure 6.3 | A schematic diagram of the recursive implementation of a one-step ARX model.

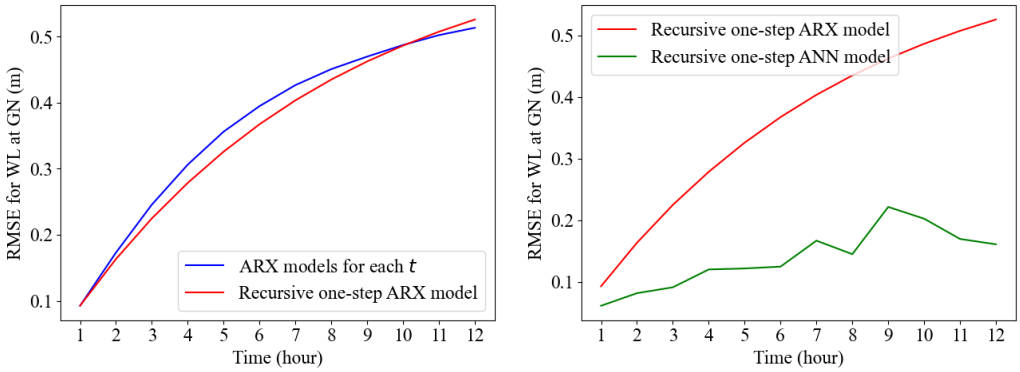
Our one-step linear ARX model is as follows:

$$\begin{aligned} G_{t+1} =& 2.72 \times 10^{-5} O_t + 1.55 \times 10^{-16} P_t + 2.73 \times 10^{-23} M_t + 0.90 G_t \\ &+ 1.16 \times 10^{-22} O_{t+1} + 0.30 \times 10^{-1} P_{t+1} + 0.60 \times 10^{-1} M_{t+1} \\ &+ 8.53 \times 10^{-26}, \end{aligned} \tag{6.5}$$

where O_t, P_t, and M_t are the elements of exogenous inputs, which are the reservoir outflow, GP water levels, and MH at time t, respectively. G_t is the water level at our target point, i.e., Guemnam station. Therefore, ρ is 1 and η is 2 in this research.

It should be noted that the target water level at GN is primarily influenced by its own previous water level, rather than by exogenous inputs, as illustrated in Fig 6.1. This suggests that controlling the water level at GN using this model can be challenging, as neither the previous water level nor the discharges from tributaries are directly controllable.

To investigate this issue further, we developed alternative downstream system models excluding autoregressive variables, which deviates from established hydrological practices. Notably, in the absence of the previous GN water level, the regression model exhibited a large intercept and low coefficients for exogenous variables, indicating that the downstream system is predominantly governed by autoregressive dynamics. Even though this moderate autocorrelation (determining an appropriate threshold for "moderate" coefficients is challenging; however, coefficients for autoregressive variables typically remain under 0.5 based on the Muskingum method (Elbashir, 2011), which is considerably small compared to 0.90 in our ARX model) accords with theoretical formulations like the Muskingum method (Chang *et al.*, 1983; Gill, 1978), this behavior is largely attributed to the backwater effect caused by the downstream weir, as well as the nature of the data, which represent controlled outflows from DC. Firstly, in 2014, a large weir was constructed downstream of GN, and it appears that the water levels at GN have since been influenced by its gate operations. The backwater effect induced by the weir, defined as the rise in upstream water level due to downstream hydraulic structures or flow resistance, likely contributes to the strong autocorrelation observed in the GN water level, outweighing the influence of other factors. Secondly, the reservoir reduced its outflows when the downstream water levels and/or expected discharges from tributaries were



(a) Two types of ARX implementation **(b)** One-step ANN vs one-step ARX model

Figure 6.4 | Comparison between three types of models: ARX models for each prediction time, one-step ARX model, and one-step ANN model.

Table 6.2 | Hyperparameter optimisation results for the DNN model

ρ	η	♯ nodes	♯ layers	learning rate	batch size	epochs	activation	RMSE
2	1	16	6	0.002	8	1100	Relu	0.05 m

high. This leads to underestimating the impacts of DC outflows. Addressing these issues would require a more sophisticated hydrodynamic downstream model capable of explicitly capturing both the backwater effect and the weir operations; however, developing such a model is beyond the scope of this chapter. Therefore, we employ the model presented in (6.5) in this chapter.

The DNN can be expressed by leveraging the nonlinear (or linear) relation F for a weight and bias matrix θ_i for hidden layer i as follows:

$$\tilde{\mathbf{y}} = F(\mathbf{x}; \theta_i) \quad \forall \quad i \in I, \tag{6.6}$$

where I represents hidden layers. Since the feedback system for the numerical experiment is applied outside of MPC optimization, we used the DNN model, which shows excellent performance in mapping nonlinear relationships, for the feedback system, even though the linear ARX model is applied as the downstream system model in MPC. Hyperparameters of the DNN model are optimized by the Tree-structured Parzen Estimator (TPE) algorithm (Zulfiqar *et al.*, 2022; Bergstra *et al.*, 2011). For hyperparameter optimization, ρ and η are also regarded as hyperparameters and optimized. The optimal hyperparameters are shown in Table 6.2. For the validation data, the DNN model shows lower RMSEs for all prediction horizons than the ARX model, as demonstrated in Fig 6.4b.

6.2.2 Scenarios generation and reduction

In stochastic MPC, scenarios refer to possible realizations of uncertain variables. In the previous chapter, we demonstrated the applicability of scenario generation using Monte-Carlo (MC) dropout BNN (Gal and Ghahramani, 2016) and compared four distance measures for scenario reduction of hydrological data.

Table 6.3 | Hyperparameter optimisation results for BNN models

	η	dropout rate	nodes	layers	learning rate	batch	activation
DC	24	0.1	512	3	0.0005	64	Relu
GP	12	0.1	512	3	0.0005	256	Relu
MH	12	0.1	512	3	0.0005	256	Tanh

Table 6.4 | RMSEs and NSCs of BNN models

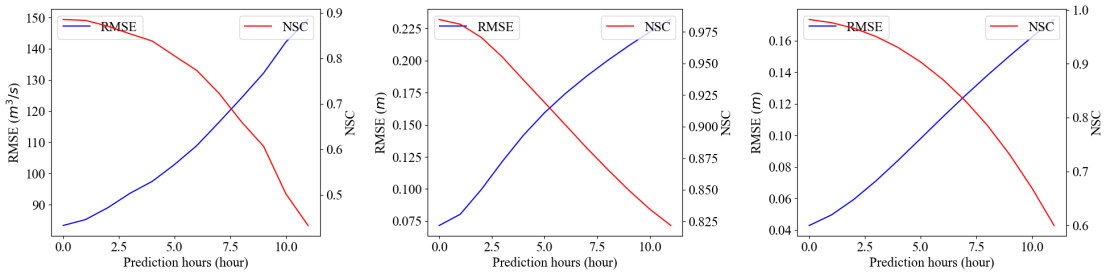
	mean RMSE	max RMSE	min RMSE	mean NSC	max NSC	min NSC
DC	110.39	149.38	83.31	0.77	0.89	0.43
GP	0.16	0.23	0.07	0.91	0.98	0.82
MH	0.11	0.17	0.04	0.85	0.98	0.60

To generate uncertain scenarios, MC dropout BNN models for GP water level, MH water level, and DC inflow need to be developed. Water level data at the target station for autoregressive horizon ρ and upstream water level data for exogenous horizon η are utilized to train BNN models. Data for a total of 11 upstream stations for GP and 12 stations for MH are collected. For simplicity, ρ is set to the same as η. In this chapter, we utilize the BNN model for DC inflow that was constructed in our previous research. Additionally, the BNN models for GP and MH water levels are developed using the same methodology and data for the same period as the BNN model for DC inflow. The optimized hyperparameters are shown in Table 6.3 and the average of generated scenarios for test data are illustrated in Fig 6.6. BNN models generate 1000 uncertain scenarios each time step.

The Root Mean Square Error (RMSE) and the Nash-Sutcliffe model efficiency coefficient (NSC) (Nash and Sutcliffe, 1970) are commonly used indicators to evaluate the reliability of hydrological prediction models. In particular, if the NSC is above 0.65, the performance of the model is 'good', and if it is above 0.75, the performance can be judged as 'very good' (Moriasi *et al.*, 2007). As demonstrated in Fig 6.5 and Table 6.4, RMSEs for DC, GP, and MH are 110.39 m^3/s, 0.16 m^3/s, and 0.11 m^3/s, respectively. NSCs are 0.89, 0.98, and 0.98, respectively, which can be called 'very good'. In addition, both indicators show that the performance deteriorates as the prediction time increases, as illustrated in Fig 6.5.

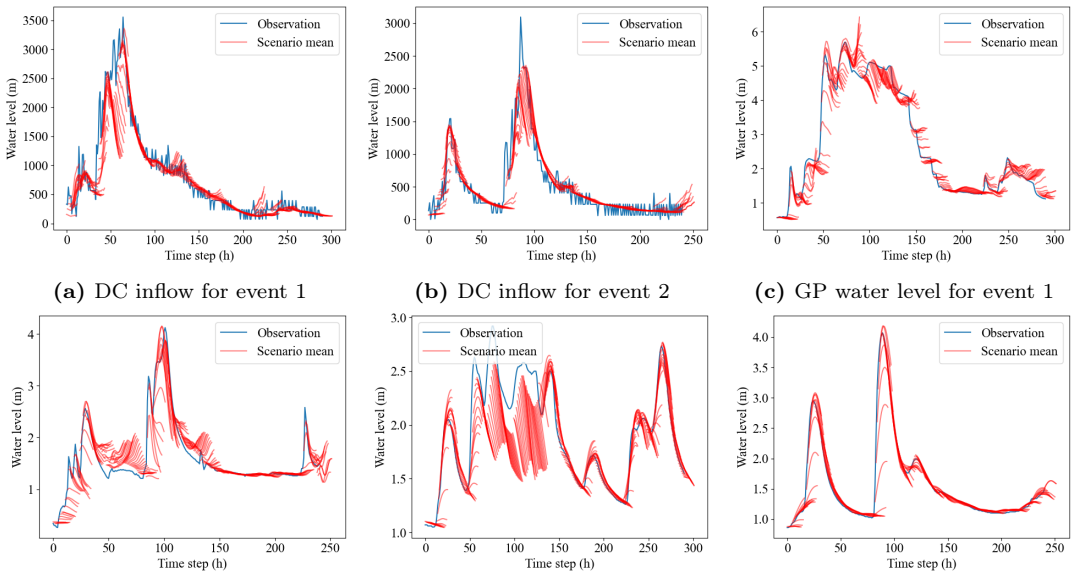
Visually, the BNN model appears to simulate trends and peaks well, as shown in Fig 6.6. One thing to note is that due to limited training data, the models exhibit limitations in peak prediction, as highlighted in Fig 6.6a, Fig 6.6b and Fig 6.6e. For example, as in time steps 50 to 75 of Fig 6.6a, BNN models often predict sharp decreases during periods of rapid increases in inflow and water levels. This is because all period data, including drought and normal seasons, are utilized to train BNN models due to the lack of flood data.

In the previous chapter, we demonstrated that using the energy distance resulted in reduced scenarios that best preserved the statistical metrics of the original scenarios, such as mean, standard deviation, and linear temporal correlation. However, considering the importance of extreme scenarios and computational complexity, the Manhattan (k-median) and Euclidean (k-means) distances could be the most practical choice. Therefore, original scenarios are generated by MC dropout BNN models



(a) DC inflow **(b)** Water level at GP **(c)** Water level at MH

Figure 6.5 | Scenario generation results by MC dropout BNN models.



(a) DC inflow for event 1 **(b)** DC inflow for event 2 **(c)** GP water level for event 1

(d) GP water level for event 2 **(e)** MH water level for event 1 **(f)** MH water level for event 2

Figure 6.6 | Hydrography with the mean of the generated scenario.

for inflow at DC, water level at GP and MH, and then k-means based on the Euclidean distance are utilized for the scenario reduction. Note that, unlike conventional k-means, which uses the average of scenarios in each cluster as a reduced scenario (continuous scenario reduction), the nearest original scenario in each centroid is used as a reduced scenario (discrete scenario reduction), presented in the previous chapter. In addition, we compare its closed-loop performance with that based on the energy distance, which best preserves statistical characteristics.

6.2.3 Risk-aware optimal control

Stochastic Model Predictive Control using CVaR constraints

Let $f(\mathbf{u}, \xi)$ be an expected loss (or a representative value for expected undesired system state) to be minimized, $\mathbf{u} \in \mathbb{R}^{n_u}$ be a control input vector to be optimized, and $\xi \in \mathbb{R}^{n_\xi}$ be variables which represent uncertainty. The probability that the

expected loss $f(\mathbf{u}, \xi)$ does not exceed α is given by

$$\phi(\mathbf{u}, \alpha) = \int_{f(\mathbf{u},\xi)\geq\alpha} p(\xi)d\xi, \tag{6.7}$$

where $\phi(\mathbf{u}, \alpha)$ means the cumulative distribution function for the loss by the control input vector $\mathbf{u}$. The minimum α which makes the cumulative distribution ϕ not less than the given probability threshold β is the Value-at-Risk (VaR), which is defined as

$$\mathrm{VaR}_\beta = \alpha_\beta(\mathbf{u}) = \min\{\alpha \in \mathbb{R} : \phi(\mathbf{u}, \xi) \geq \beta\}. \tag{6.8}$$

Conditional Value-at-Risk (CVaR) means an expected loss for the part where the loss exceeds the VaR value, and it can be given by

$$\mathrm{CVaR}_\beta = (1-\beta)^{-1} \int_{f(\mathbf{u},\xi)\geq VaR_\beta} f(\mathbf{u}, \xi)p(\xi)d\xi. \tag{6.9}$$

To calculate CVaR_β, in (6.9), we have to calculate the VaR_β value first. Rockafellar *et al.* (2000) simplified (6.9) as (6.10) based on the fact that the VaR_β value is $\alpha_\beta(\mathbf{u})$ when $\phi(\mathbf{u}, \xi) = \beta$, and the probability of $f(\mathbf{u}, \xi) \geq \mathrm{VaR}_\beta$ is $1 - \beta$ assuming the probability distribution of ϕ is continuous.

$$\mathrm{CVaR}_\beta \coloneqq \min_\alpha F_\beta(\mathbf{u}, \alpha), \tag{6.10a}$$

$$F_\beta(\mathbf{u}, \alpha) = \alpha + (1-\beta)^{-1} \int_\xi [f(\mathbf{u}, \xi) - \alpha]^+ p(\xi)d\xi, \tag{6.10b}$$

where $[.]^+$ is a ramp function, i.e., $[g]^+ = \frac{g+|g|}{2}$. Assuming the finite number of scenarios with i.i.d. from a continuous distribution of ξ,

$$\mathrm{CVaR}_\beta \coloneqq \min_\alpha \Big(\alpha + \frac{1}{(1-\beta)} \sum_\xi p(\xi)[f(\mathbf{u}, \xi) - \alpha]^+\Big), \tag{6.11}$$

where $p(\xi) \in \Xi$ is the probability of occurrence for a scenario ξ and $\sum_\xi p(\xi) = 1$. Therefore, the objective function which minimizes the CVaR_β value is as follows:

$$\min_{\mathbf{u}} \mathrm{CVaR}_\beta \coloneqq \min_{\mathbf{u},\alpha} \Big(\alpha + \frac{1}{(1-\beta)} \sum_\xi p(\xi)[f(\mathbf{u}, \xi) - \alpha]^+\Big). \tag{6.12}$$

It should be noted that α on the right-hand side in (6.12) is also a decision variable and is dependent on the distribution of $f(\mathbf{u}, \xi)$. CVaR_β is the sum of expected loss whose cumulative probability exceeds a given probability β. Therefore, minimizing CVaR_β leads to minimizing α and the loss $f(.)$ simultaneously by considering trade-offs between α and the loss. This is reasonable in finance, such as portfolio optimization (Rockafellar and Uryasev, 2002), where we want to minimize the expected loss exceeding α as well as α simultaneously because this makes the shape of the loss distribution narrow and values in the tail small.

In general, the loss can be defined as the representative amount of the undesired state. For instance, how much the water level exceeds a threshold can be a risk and a loss $f(.)$ in (6.12). Then, (6.12) can be reformulated as follows:

$$\min_{\mathbf{u}} \mathrm{CVaR}_\beta \coloneqq \min_{\mathbf{u},\alpha} \Big(\alpha_g + \frac{1}{(1-\beta)} \sum_{\xi} p(\xi) [[g(\mathbf{u},\xi) - \mathrm{L}]^+ - \alpha_g]^+ \Big), \tag{6.13}$$

where $g(.)$ denotes a system variable, such as RWL, and L is a given threshold for $g(.)$. Thus, $[g(\mathbf{u},\xi) - \mathrm{L}]^+$ represents the loss $f(.)$. However, this objective makes RWL less than L; if it is not possible, the objective minimizes the tail sum of the exceedance of RWL over α ($\alpha \geq \mathrm{L}$), instead of trying to minimize every exceedance of RWL over L. Otherwise, if we set a state variable like RWL as a loss $f(.)$, minimizing CVaR_β leads to minimizing RWL even though RWL is still under the upper water level. Moreover, if we need to consider other objectives simultaneously, setting CVaR_βs as objectives causes difficulty in choosing weights.

Therefore, it is reasonable to utilize CVaR_β as a constraint which prevents a state variable from exceeding a given threshold, as described in (6.14).

$$\min_{\alpha} \Big(\alpha_f + \frac{1}{(1-\beta)} \sum_{\xi} p(\xi) [f(\mathbf{u},\xi) - \alpha_f]^+ \Big) \leq \mathrm{L}, \tag{6.14}$$

where the loss $f(.)$ is the value of the state variable and L is a given threshold to the state variable as well as threshold to CVaR_β. This means that the sum of α and the expected exceedance over α should be less than the threshold L. This is more robust than (6.13).

Now, the optimization problem can be expressed as follows:

$$\min_{\mathbf{u}} J(\mathbf{u},\xi), \tag{6.15a}$$

$$\text{s.t.} \quad \min_{\alpha} \Big(\alpha + \frac{1}{(1-\beta)} \sum_{\xi} p(\xi) [f(\mathbf{u},\xi) - \alpha]^+ \Big) \leq \mathrm{L}. \tag{6.15b}$$

This CVaR_β constraint is only activated when a CVaR_β value closes to L, and the CVaR_β value can change (or be minimized) to make it less than L. Intuitionally, when the equation (6.15b) is activated, it can be minimized CVaR_β without minimisation (Rockafellar and Uryasev, 2002). This means that the left-hand side of (6.15b) becomes equivalent to CVaR_β only when this constraint is activated. Therefore, we can call it an 'CVaR_β approximation around the point of interest' (Hong *et al.*, 2014). For simplicity, we refer to this approximation as CVaR_β throughout this chapter.

$$\alpha + \frac{1}{(1-\beta)} \sum_{\xi} p(\xi) z_\xi \leq \mathrm{L}, \tag{6.16a}$$

$$z_\xi = [f(\mathbf{u},\xi) - \alpha]^+. \tag{6.16b}$$

How to determine L needs to be considered in order to apply a CVaR_β constraint in a reservoir flood control problem under uncertainty. A physical/operational constraint or upper/lower boundary in a predefined safety range can be a candidate (Hakobyan *et al.*, 2019). For reservoir flood control, physical constraints, e.g., the Flood Water Level (FWL) or maximum discharge capacity, are difficult to utilize as L for the CVaR_β constraint because these are designed against extreme conditions such as a Probable Maximum Flood. For example, for Daecheong reservoir, the highest water level in the last 20 years has been the Elevation Level (EL) 79.21 m, which has never reached the FWL, EL 80.00 m. The maximum discharge is also 3508 m^3/s, far less than a spillway capacity of 11680 m^3/s. In other words, if the FWL or spillway capacity is set as the upper bound of CVaR_β, this CVaR_β constraint has never been activated under flood conditions over the past 20 years.

The operator's loss tolerance seems appropriate to be set as a threshold L. However, there is a disadvantage in that it is highly unlikely to be feasible in extreme events that have not been experienced. Relaxing a hard constraint to make the optimization problem feasible even in extreme situations can be easily achieved by changing it to a soft constraint (Micheli *et al.*, 2022; Meseguer *et al.*, 2006). By setting the weight of this soft constraint very high, this soft constraint works as a hard constraint when there is a feasible solution. Therefore, we set L based on the operator's choice - we select L from historical data in this chapter — and apply CVaR_β as soft constructs.

6.2.4 Numerical experiments

In Chapter 3, we presented eight objectives for practical reservoir flood control. In this chapter, only two objective functions, for simplicity, the minimization of the maximum outflow and the changes between outflow schedules, are applied to simplify the problem. In particular, it is essential to minimize the change between the previous outflow schedule and the current schedule because a sudden change between outflow schedules is physically challenging, and it is disadvantageous to flood control from the perspective of watershed management through cooperation with related organizations. The objectives relating to the RWL are applied by the CVaR_β constraints described in the previous section.

An important constraint is that outflow via spillway gates should start after using all the outflow capacity via turbines. To avoid a nonlinear complementarity equation (Powell *et al.*, 2016), we convert it into a linear soft constraint. In addition, for the CVaR_β constraints, operational thresholds, i.e., the highest (S_U) and lowest RWL (S_L) during the flood event and the attention water level at the downstream monitoring point (L_G) are set as the thresholds.

The objectives and constraints of this research are as follows;

$$\min_{\mathbf{u},\alpha} J := \min_{\mathbf{u},\alpha} J_1 + J_2 + J_3 + J_4,$$

$$J_1 := \max_t O_{spill,t} \times N_p \times w_1, \tag{6.17a}$$

$$J_2 := \sum_t (\Delta O_t^+ + \Delta O_t^-) \times w_2, \tag{6.17b}$$

$$J_3 := \sum_t \Delta O_t^{com} \times w_3, \tag{6.17c}$$

$$J_4 := \sum_t (\Delta C_t^U + \Delta C_t^L + \Delta C_t^G) \times w_4, \tag{6.17d}$$

and subject to the following constraints:

$$O_{spill,t} + O_{turb,t} - O_{total,t} = 0 \quad \forall\, t \in T, \tag{6.18a}$$

$$O_{total,t}^k = O_{total,t}^{k-1} \quad \text{when } t = k, \tag{6.18b}$$

$$\Delta O_t^+ - \Delta O_t^- + (O_{total,t}^k - O_{total,t}^{k-1}) \times \sigma_t = 0 \quad \forall\, t \in T, \tag{6.18c}$$

$$O_{total,t} - \mathrm{OC}_{turb} + \delta O_t^{com} - O_{spill,t} = 0 \quad \forall\, t \in T, \tag{6.18d}$$

$$\mathrm{CVaR}_{\beta,t}^U - S_U - \Delta C_t^U \le 0 \quad \forall\, t \in T, \tag{6.18e}$$

$$\mathrm{CVaR}_{\beta,t}^L - S_L + \Delta C_t^L \ge 0 \quad \forall\, t \in T, \tag{6.18f}$$

$$\mathrm{CVaR}_{\beta,t}^G - L_G - \Delta C_t^G \le 0 \quad \forall\, t \in T, \tag{6.18g}$$

$$\Delta O_t^+,\, \Delta O_t^-,\, \Delta C_t^U,\, \Delta C_t^L,\, \Delta C_t^G,\, \Delta O_t^{com} \ge 0 \quad \forall\, t \in T, \tag{6.18h}$$

and $CVaR_\beta$ constraints as follows;

$$S_{t,\xi_S} - \alpha_t^U - z_{t,\xi_S}^U \le 0 \quad \forall\, t \in T \text{ and } \forall\, \xi_S \in \Xi_S, \tag{6.19a}$$

$$\alpha_t^L - S_{t,\xi_S} - z_{t,\xi_S}^L \le 0 \quad \forall\, t \in T \text{ and } \forall\, \xi_S \in \Xi_S, \tag{6.19b}$$

$$G_{t,\xi_P,\xi_M} - \alpha_t^G - z_{t,\xi_P,\xi_M}^3 \le 0 \quad \forall\, t \in T \text{ and } \forall\, \xi_P \in \Xi_P \text{ and } \forall\, \xi_M \in \Xi_M, \tag{6.19c}$$

$$\mathrm{CVaR}_{\beta,t}^U = \alpha_t^U + \frac{1}{(1-\beta)} \sum_{\xi_S} p(\xi_S) z_{t,\xi_S}^U \quad \forall\, t \in T, \tag{6.19d}$$

$$\mathrm{CVaR}_{\beta,t}^L = \alpha_t^L - \frac{1}{(1-\beta)} \sum_{\xi_S} p(\xi_S) z_{t,\xi_S}^L \quad \forall\, t \in T, \tag{6.19e}$$

$$\mathrm{CVaR}_{\beta,t}^G = \alpha_t^G + \frac{1}{(1-\beta)} \sum_{\xi_P,\xi_M} p(\xi_P \cap \xi_M) z_{t,\xi_P,\xi_M}^G \quad \forall\, t \in T, \tag{6.19f}$$

$$\alpha_t^U,\, \alpha_t^L,\, \alpha_t^G,\, z_{t,\xi}^U,\, z_{t,\xi}^L,\, z_{t,\xi}^G \ge 0 \quad \forall\, t \in T. \tag{6.19g}$$

The first objective (6.17a) minimizes the peak outflow in a prediction horizon with a length N_p. By (6.17b), the difference between the outflow schedules determined at the previous time step $k-1$ and the outflow schedule at the current time step k is minimized, where ΔO^+ means the increase in outflows and ΔO^- means the decrease. Additionally, because the outflow decided at time step k cannot be implemented at time k, a constraint (6.18b) is added.

In (6.17c) and (6.18d), ΔO^{com} is a slack variable which is activated when the total outflow is less than the turbine outflow capacity OC_{turb}. Minimizing $J3$ opens the spillway gates only when the total outflow is larger than OC_{turb}. In (6.18e), (6.18f), and (6.18g), CVaR^U, CVaR^L, and CVaR^G are the CVaR_β approximations for the peak RWL, lowest RWL, and peak water level at the downstream station, GN, respectively. ΔC^U, ΔC^L, and ΔC^G are the exceedances of corresponding operational threshold. Therefore, J_4 minimizes the penalty (exceedances) of CVaR_β soft constraints.

In (6.19), S_{t,ξ_S} means the storage amount under uncertain scenarios ξ_S and G_{t,ξ_P,ξ_M} refers to the water level at GN. Note that the GN water level is affected by both the GP and MH water level; therefore, we need to consider the joint probability $p(\xi_P \cap \xi_M)$ in (6.19f). Because we used two different models built independently to generate scenarios, we can say that the two probability distributions for $p(\xi_P)$ and $p(\xi_M)$ are independent. Therefore, we assume as follows:

$$p(\xi_P \cap \xi_M) = p(\xi_P) \times p(\xi_M). \tag{6.20}$$

In addition, ramp functions $[.]^+$ for z_ξ are linearized as (6.19a), (6.19b), and (6.19c), respectively.

The soft constraint (6.18g) is activated when $\text{CVaR}^G_{\beta,t} \geq L^G$. Therefore, to minimize ΔC^G_t, $\text{CVaR}^G_{\beta,t}$ should be minimized. Now, we can say that (6.18g) leads to

$$\min \text{CVaR}^G_{\beta,t}, \tag{6.21}$$

when $\Delta C^G_t > 0$. For this, α^G_t and z^G_{t,ξ_P,ξ_M} in (6.19f) should be minimized simultaneously. Finally, due to (6.19c), G_{t,ξ_P,ξ_M} is minimized.

6.3 Results and Discussion

6.3.1 Results

We evaluate the closed-loop performance of the Stochastic Model Predictive Control (SMPC) with the Deterministic Model Predictive Control (DMPC) using the average of the original scenarios. The deterministic MPC is formulated like the stochastic MPC except for considering stochastic scenarios. The values of state variables constrained by CVaR_β, i.e., highest RWL, lowest RWL, and highest water level at GN, will be compared with those of DMPC. Note that the inflow and downstream hydrological conditions of the test events used in this chapter are not extreme due to the limitations of flood data, resulting in only minor differences in the comparison results.

Given that the threshold for the peak RWL at DC is EL 78.400 m, the minimum RWL at DC is EL 75.500 m, and the peak water level at GN (hereafter referred to as GWL) is 4.500 m, DMPC cannot preserve the thresholds for both the peak RWL and minimum RWL for Event 1 as shown in Fig 6.5. In this table, negative values

Table 6.5 | DMPC results (exceedance over corresponding thresholds)

Event	peak RWL (EL m)	minimum RWL (EL m)	peak GWL (m)
1	-0.111	-0.012	0.037
2	0.054	-0.012	0.982

Table 6.6 | SMPC results when $\beta = 0.99$ (exceedance over corresponding thresholds)

Event	m	peak RWL (EL m)	minimum RWL (EL m)	peak GWL (m)
1	10	-0.070	0.007	0.026
1	20	-0.051	0.196	0.026
1	30	-0.046	0.211	0.024
1	40	-0.044	0.214	0.023
1	50	-0.044	0.213	0.026
2	10	0.164	0.001	0.992
2	20	0.180	0.942	0.993
2	30	0.153	0.942	0.993
2	40	0.184	0.006	0.993
2	50	0.185	0.942	0.993

represent excess amounts, while positive values indicate margins from the corresponding thresholds. This is mainly because the inflow scenarios for the peak inflow periods cannot mimic the real ones when inflow increases abruptly, as explained in Section 6.2.2.

In the case of the operator's highest risk aversion, i.e., $\beta = 0.99$, the SMPC outperforms the DMPC for all m, the number of reduced scenarios. In all cases, the SMPC preserves the thresholds for minimum RWL and peak GWL, as shown in Table 6.6. The peak RWL still does not satisfy the threshold, but the largest excess is 0.07m when the number of reduced scenarios m is 10, which is smaller than 0.111 m for the DMPC.

In addition, as demonstrated in Tables 6.6 and 6.7, the closed-loop performance becomes more conservative as the number of scenarios m in SMPC increases. Specifically, the excess over the threshold for the peak RWL decreases, providing more margin for the other two indicators. While it is challenging to observe significant differences in the results under high risk-averseness $\beta = 0.99$, as shown in Table 6.6. We can see a more noticeable trend when considering slightly lower risk-averseness. Table 6.7 shows that incorporating more scenarios enhances the robustness of the results. As m increases, the peak RWL decreases, and the minimum RWL increases. Notably, for Event 2, while both DMPC and SMPC with $\beta = 0.99$ satisfy all three thresholds, SMPC with $\beta = 0.7$ only meets all thresholds when $m = 50$.

The SMPC results dependent on the operator's risk averseness β are straightforward. As β increases, the excesses (negative) become larger, or the margins (positive) become smaller, as demonstrated in Table 6.8.

For the closed-loop experiment results of RWL and GWL, as illustrated in Fig 6.7 and 6.8, RWL, where constraints are frequently activated, shows significantly different patterns from historical operational data. Both actual operational records and SMPC results, as well as DMPC results, exhibit distinct distributions. While operational records are concentrated in the middle range of EL 73.5 m to EL 75.0

Table 6.7 | SMPC results when $\beta = 0.7$

Event	m	peak RWL (EL m)	minimum RWL (EL m)	peak GWL (m)
1	10	-0.089	-0.009	0.032
1	20	-0.089	-0.009	0.031
1	30	-0.088	0.003	-0.027
1	40	-0.088	0.595	-0.025
1	50	-0.088	1.004	-0.024
2	10	0.101	-0.010	0.987
2	20	0.100	-0.010	0.987
2	30	0.101	-0.010	0.987
2	40	0.101	-0.002	0.987
2	50	0.101	0.942	0.987

Table 6.8 | SMPC results when $m = 10$

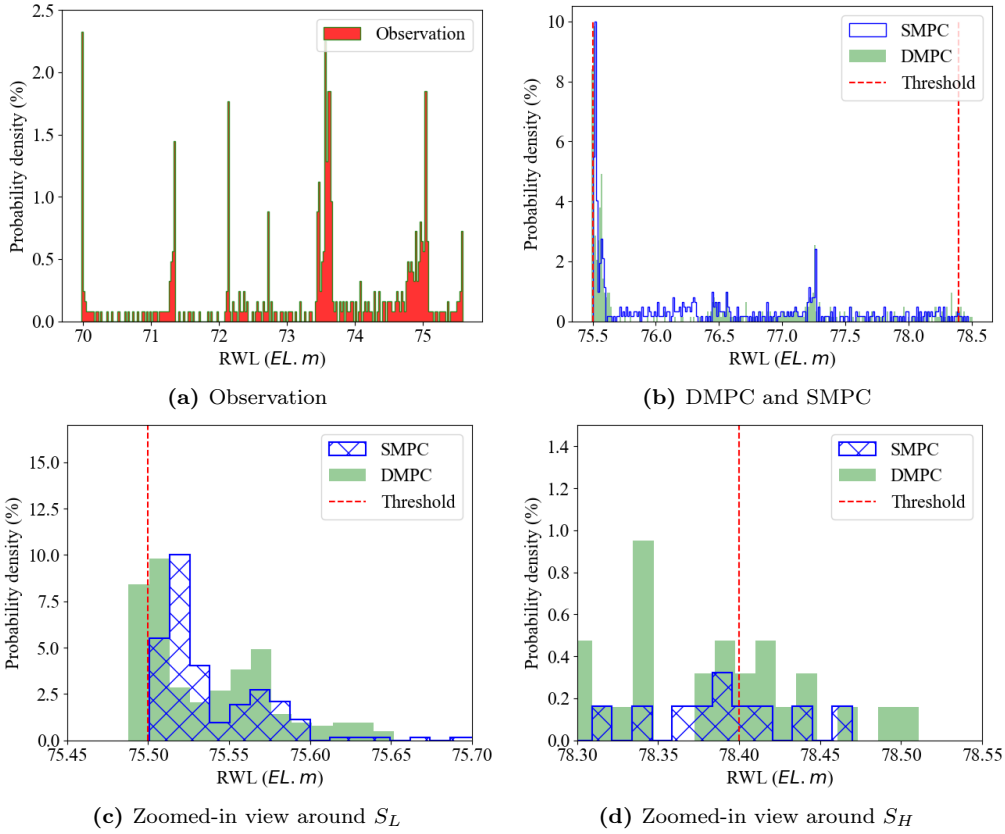
Event	β	peak RWL (EL m)	minimum RWL (EL m)	peak GWL (m)
1	0.99	-0.070	0.007	0.026
1	0.90	-0.070	-0.005	0.030
1	0.80	-0.080	0.000	0.032
1	0.70	-0.089	-0.009	0.032
2	0.99	0.164	0.001	0.992
2	0.90	0.129	-0.008	0.989
2	0.80	0.110	-0.010	0.988
2	0.70	0.101	-0.010	0.987

m, DMPC shows high frequency near the threshold S_L, EL 75.5 m. This tendency is more clearly shown in SMPC than in DMPC. These distributional differences are related to constraint activation. Specifically, both DMPC and SMPC show more frequent activation of the S_L constraint compared to the S_H constraint, even though SMPC does not exceed S_L while DMPC does. Note that the difference in water level ranges between operational records and DMPC/SMPC results stems from the discrepancy between constraints and objective functions of the numerical experiment versus actual operational conditions.

In contrast, GWL exhibits patterns similar to historical operational data since the CVaR_β constraint is rarely activated, and GWL is predominantly influenced by the previous GWL and tributary inflows, rather than reservoir outflows.

In contrast, GWL exhibits patterns similar to historical operational data since the CVaR_β constraint is rarely activated, and GWL is predominantly influenced by the previous GWL and tributary inflows, rather than reservoir outflows. This fact significantly limits us from showing the effectiveness of the stochastic MPC approach in simultaneously addressing downstream flood conditions.

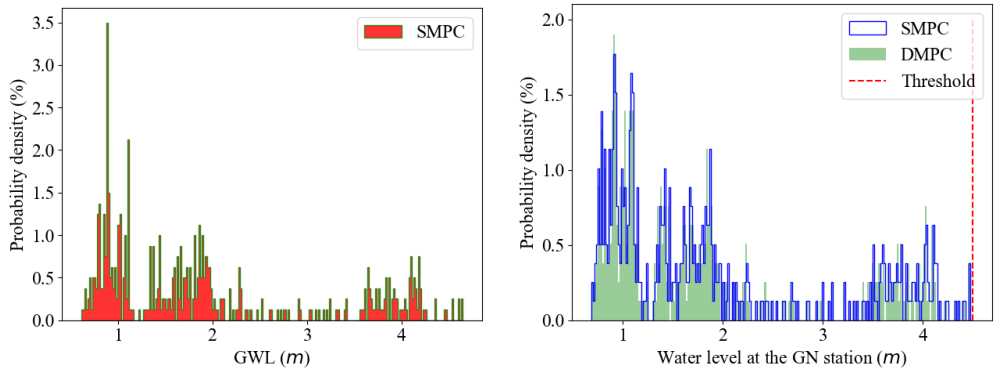
Although there is little difference between recorded peak GWLs for SMPC and DMPC, the maximum values of expected GWLs calculated by the downstream model during MPC iterations under SMPC consistently exceed the recorded peak GWLs in our study events, as presented in Table 6.9. Conversely, for DMPC, the expected peak GWLs are lower than the actual peak GWLs in the simulation results. This indicates that SMPC proactively addresses unfavorable scenarios, providing insight into its capacity to generate robust results when confronting uncertainty.



(a) Observation **(b)** DMPC and SMPC

(c) Zoomed-in view around S_L **(d)** Zoomed-in view around S_H

Figure 6.7 | Distribution of RWLs of observation, DMPC, and SMPC ($\beta = 0.99$ & $m = 10$).



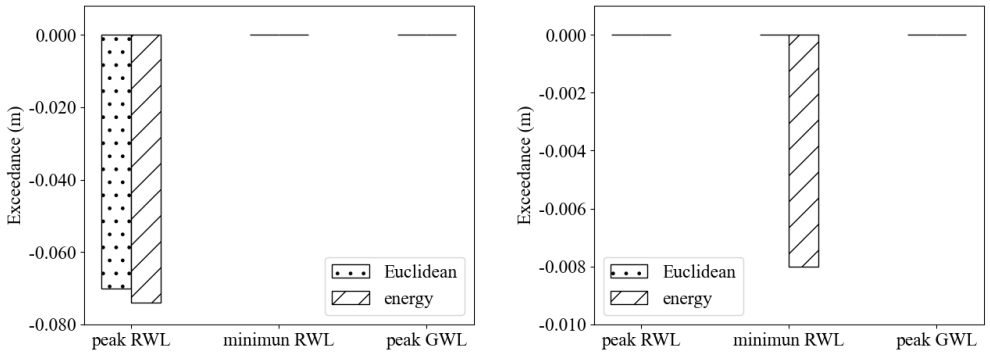
(a) Observation **(b)** DMPC and SMPC

Figure 6.8 | Distribution of GWLs of observation, DMPC, and SMPC ($\beta = 0.99$ & $m = 10$).

Table 6.9 | Expected and recorded peak GWLs(m) when $\beta = 0.99$ & $m = 10$

Event	Expected, SMPC	Recorded, SMPC	Expected, DMPC	Recorded, DMPC
1	4.474	4.483	4.463	4.429
2	3.508	3.848	3.518	3.601

In scenario reduction, we compare the Euclidean distance using the k-means clustering algorithm with the energy distance. In the previous chapter, we demonstrated that the energy distance better preserved the mean, standard deviation, and temporal Pearson correlation of the original scenarios compared to the Euclidean distance. However, the Euclidean distance-based approach performed superiorly in terms of including scenarios close to the upper and lower bounds. Considering that these differences diminish as the number of reduced scenarios m increases, we compare the excess of these three metrics for $m = 10$.



(a) Event 1 **(b)** Event 2

Figure 6.9 | Results by two different scenario reduction measures.

As illustrated in Fig 6.9, in terms of exceedance, the cases applying the Euclidean distance exhibit less exceedance over the thresholds. This is obvious because reduced scenarios based on the Euclidean distance include higher and lower scenarios than those based on the energy distance. Considering both this fact and the significant computational complexity in applying the energy distance, as presented in the previous chapter, we can conclude that the Euclidean distance is a more suitable measure for the scenario reduction of hydrological variables in the context of closed-loop MPC performance.

Finally, no significant differences are observed between SMPC and DMPC regarding the original objectives, i.e., minimizing spillway outflows and changes in outflow schedules. For instance, with SMPC ($\beta = 0.99$ and $m = 10$), the peak outflows are 3636 m^3/s for Event 1 and 1870 m^3/s for Event 2, comparable to DMPC's 3577 m^3/s and 1861 m^3/s. Similarly, the number of changes in consecutive outflow schedules for SMPC (80 and 98 times) shows minimal difference from DMPC (91 and 92 times). However, considering that SMPC's peak RWL is lower than DMPC's, it is understandable that SMPC's maximum discharge typically exceeds that of DMPC in most cases.

6.3.2 Discussion

Stochastic MPC using CVaR_β can readily incorporate operator risk-averseness through adjusting β, demonstrating more robust performance than deterministic MPC with high β. The impact of varying β on performance is straightforward,

indicating that CVaR_β constraints effectively and intuitively reflect operator risk-averseness in closed-loop performance.

In reservoirs designed for extreme events, physical thresholds have limited utility during typical flood control situations, while the operator's loss tolerances as thresholds may make MPC infeasible. This limitation can be addressed by designing CVaR_β as soft constraints rather than hard constraints or objectives. Indeed, the numerical experiment demonstrates that thresholds based on operator loss tolerance produce feasible results while minimizing excesses, even in extreme scenarios.

Compared to the observed distribution of RWLs, SMPC exhibits less conservative results. By considering the range of distribution as a measure of risk, the incorporation of CVaR_β constraints in SMPC enables the maintenance of higher water levels while achieving lower risk (distribution), which implies more efficiency in flood control.

From a methodological perspective of scenario reduction, the Euclidean distance, which is computationally efficient, demonstrates comparable performance to the energy distance. It proves particularly effective in terms of excess over the threshold. Moreover, considering the number of reduced scenarios shows minimal impact when using the Euclidean distance, $m = 10$ appears sufficient for practical implementation.

However, our numerical experiments had certain limitations. In downstream simulations, reservoir outflows have limited influence on the downstream target water level due to the dominant effects of the previous GN water level and tributary inflows. The strong autocorrelation observed at GN can be attributed to the backwater effect caused by a large downstream weir, as well as the fact that only a small amount of water is discharged from DC during our study flood events to reduce the downstream water level. Consequently, this modeling approach restricts our ability to clearly demonstrate the effectiveness of stochastic MPC for the simultaneous management of the downstream water level. Addressing this limitation would require a more sophisticated hydrodynamic river routing model. To the best of our knowledge, hydrodynamic models have rarely been employed in receding horizon MPC approaches due to computational complexity, except in simplified reservoir systems involving short river reaches (Breckpot *et al.*, 2013b; Montero *et al.*, 2013; Ficchi *et al.*, 2014).

Additionally, our case study had limitations in clearly demonstrating the closed-loop performance of different scenario reduction measures - floods with higher peak inflows and longer durations would be necessary for a comprehensive evaluation.

6.4 Conclusions

In this chapter, we present a reservoir flood control methodology from a risk management perspective that incorporates both quantitative risk assessment and operator preferences. First, we demonstrate the applicability of Conditional Value-at-Risk (CVaR) constraints in stochastic Model Predictive Control (MPC) for reservoir flood control. In particular, considering the limitations of physical constraints, such as

the Flood Water Level (FWL), that are rarely activated in practice, we introduce $CVaR_\beta$ as a soft constraint to reflect operator-defined risk preferences. This approach is found to effectively capture the operator's risk tolerance.

We also show that a K-means-based scenario reduction method using the Euclidean distance outperforms the computationally intensive energy distance-based method in terms of peak and minimum Reservoir Water Levels (RWLs) in a stochastic MPC approach for reservoir flood control.

While stochastic MPC using scenario generation and reduction has been extensively studied in various fields such as finance and electricity markets, its application to hydrological facilities has received relatively little attention. In particular, few studies have explored how the $CVaR_\beta$, which allows for the explicit consideration of quantitative risk, can be applied to reservoir operation. Thus, this chapter aims to contribute to the advancement and broader adoption of stochastic receding horizon MPC approaches in water resource management.

Although this approach is applied to a single reservoir in this chapter, the stochastic MPC approach with $CVaR_\beta$ as soft constraints is expected to be readily extensible to a variety of hydrological contexts. This is attributed to both the flexibility of the $CVaR_\beta$ formulation in specifying risk tolerance through the choice of the confidence level (β), and the structural flexibility achieved by implementing $CVaR_\beta$ as a soft constraint.

Data Availability for this chapter

The data and code of this chapter are available at https://data.4tu.nl/datasets/dc6b16b1-aab9-4ddb-a96b-8289ae3c8d5b. All data can be used under the CC-BY-4.0 licence. The hydrological and operational data of the Daecheong reservoir were obtained from K-water's website (http://kwater.or.kr).

7 Conclusions

Research outcomes

This chapter summarizes the research outcomes corresponding to the objectives outlined in Section 1.4. The overall outcomes matched to the research objectives are illustrated in Fig 7.1, which we also outline below by topic.

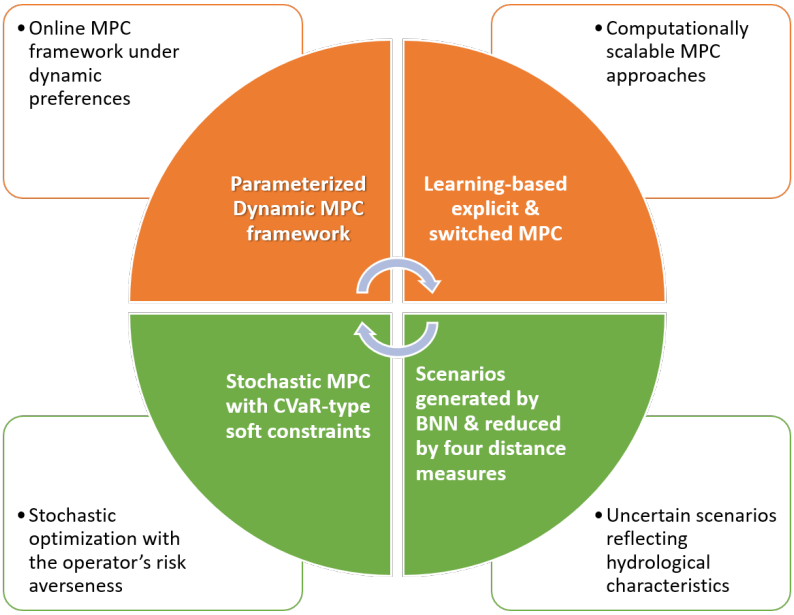


Figure 7.1 | A depiction of the overall contributions of this thesis research by topic.

Development of an Model Predictive Control (MPC) framework that incorporates dynamic preferences, integrating practical objectives

In Chapter 3, we develop a parameterized MPC approach, where a nonlinear multi-objective optimal control problem with dynamic operator preferences is transformed into more tractable linear MPC formulations with dynamic parameters. Our methodology aligns with multi-objective MPC techniques for linear systems that incor-

porate time-varying, state-dependent decision criteria through parametric optimization. This novel formulation allows the simultaneous optimization of both control inputs and parameters that represent operator preferences. We identify and formalize practical flood control objectives, emphasizing some critical yet previously underexamined factors such as minimizing the frequency of outflow schedule adjustments. To make the flood control problem tractable despite the inclusion of many nonlinear formulas and the dynamic preference assumption, we employ Genetic Algorithms to determine optimal weights and parameters for the parameterized linear MPC formulation, which produces the optimal control inputs with the optimal weights and parameters coming from the GA. We designate this integrated framework as Parameterised Dynamic Model Predictive Control (PD-MPC). Our numerical experiments demonstrate PD-MPC's resilience to inflow uncertainties, leveraging the inherent robustness of receding horizon approaches rather than directly addressing uncertainty. The results establish PD-MPC's superior performance compared to fixed-parameter MPC formulations, even when those fixed parameters were specifically tailored for individual objectives. This confirms our hypothesis that MPC weights and parameters must dynamically adapt to evolving hydrological conditions to meet changing preferences. Continuous parameter updates within the PD-MPC framework facilitate real-time optimal decision-making, thereby improving performance across essential metrics such as peak outflows, reservoir water levels, and stability of the outflow schedules. This adaptive capability proves essential for generating control inputs that effectively reflect operators' dynamically changing preferences.

Evaluating the potential of data-driven models to enhance the practicality the MPC approach under dynamic preference assumptions.

In Chapter 4, we propose data-driven surrogate models in the form of explicit and switched MPC approaches for a flood control problem with dynamically changing weights of multi-objectives. These approaches have the advantage of low computational complexity for online implementation compared to a PD-MPC framework proposed in Chapter 3. Although PD-MPC presents the optimal control inputs, it requires solving computationally complex nonlinear control problems online, which may make it impractical to find solutions in real-time scenarios, for example. The switched MPC approach, i.e., data-driven surrogate models that map states to optimal weights and parameters, replaces the gradient-free randomized search algorithm, which was time-consuming for solving PD-MPC online. It can thus significantly reduce complexity. In addition, since linear MPC optimization is conducted every time step using updated information, results are robust to errors and avoid constraint violations. The explicit MPC approach, i.e., a data-driven model that maps states directly to optimal control inputs, optimal outflow schedules here, has the lowest complexity for implementing online. However, a DNN model was found to sometimes result in infeasible control inputs and needs more training data to cover every possible situation.

Evaluating the effectiveness of a Bayesian Neural Network (BNN) model for hydrological scenario generation and analysing various distance measures in scenario reduction methodologies

In Chapter 5, we evaluate the effectiveness of a BNN-based scenario generation and compare the scenario reduction approaches based on different distance measures. The results suggest that the MC dropout BNN model effectively generates uncertain inflow scenarios. Without explicitly defining temporal correlations between elements in a scenario, the BNN model demonstrates reasonable performance (in terms of the RMSE and NSC) and reproduces the characteristic behavior of hydrological prediction, namely the degradation of performance with increasing prediction horizon. Regarding the distance measures for scenario reduction, using the energy distance demonstrates superior performance in preserving the statistical properties of the original scenario set generated by the BNN model. The original temporal correlation structure is best maintained by the reduced scenarios based on the energy distance. The span between the upper and lower envelopes of original scenarios represents the flood risk space resulting from uncertain inflows. To take into account this span, we propose the size of the envelope as a metric for assessing the reduced scenarios using the l_1-norm of scenarios. For this, the Manhattan and Euclidean distance metrics demonstrate superior performance. Moreover, in terms of preservation of temporal correlation and standard deviation of original sets, both the Manhattan and Euclidean distances exhibit comparable performance, following the energy distance. In using the Manhattan and Euclidean distances, the results show that no substantial differences are observed in both statistical properties and extreme scenario preservation. From the computational efficiency perspective, the use of the energy distance demonstrates significant limitations compared to other distance measures. Given that the time horizon is usually less than an hour for flood control, a scenario reduction process requiring more than 10 minutes proves impractical. Consequently, considering the importance of extreme scenarios for optimal flood operation, the Manhattan and Euclidean distances can be practical choices for scenario reduction.

Validation of Conditional Value-at-Risk (CVaR)-type soft constraints within a stochastic MPC framework to operationalise operators' risk tolerance in control

In Chapter 6, we demonstrate that stochastic MPC using CVaR_β can readily incorporate operator risk-averseness through adjusting β, demonstrating more robust performance than deterministic MPC with high β. The impact of varying β on performance is straightforward, indicating that CVaR_β constraints effectively and intuitively reflect operator risk-averseness in closed-loop performance. Compared to the observed distribution of RWLs, stochastic MPC exhibits less conservative results. By considering the range of distribution as a measure of risk, the incorporation of CVaR_β constraints in stochastic MPC enables the maintenance of higher water levels while achieving lower risk (distribution), which implies more efficiency

in flood control. Additionally, from a methodological perspective of scenario reduction, the Euclidean distance, which is computationally efficient, demonstrates comparable performance to the energy distance. It proves particularly effective in terms of excesses over the threshold. Moreover, considering the number of reduced scenarios shows minimal impact when using the Euclidean distance, $m = 10$ appears sufficient for practical implementation.

Limitations and future works

Nevertheless, this research has several limitations.

First, the MPC formulation applied here does not consider the final states of the system. MPC with policy search algorithms (Song and Scaramuzza, 2022b) or a value function produced by a Reinforcement Learning (RL) model (Arroyo *et al.*, 2022a) can give a chance to find the approximation of the terminal cost such that our MPC framework can consider the whole period of a flood event. To the best of our knowledge, this approach has never been applied to a reservoir system for the purpose of flood control.

In terms of scenario reduction methodologies, the widely used clustering algorithms are employed for the Manhattan and Euclidean distances, whereas a simple 1-step forward selection method is utilized for the Wasserstein and energy distances. While this study aims to compare popular distance measures using their established and efficient algorithms, the distinction between the effects of distance measures and reduction algorithms remains unclear.

For training data-driven models in this research, i.e., the explicit MPC, switched MPC, a BNN model to generate uncertain scenarios, and a ARX model for downstream water level prediction, a sufficient amount of data covering a wide range of possible scenarios is crucial. For example, our models may not adequately capture how reservoir outflows affect downstream target water levels when the training data contains only instances of small outflows from the upper reservoir.

The most important limitation of this research, however, is that the Parameterised Dynamic Model Predictive Control (PD-MPC) framework with dynamic preferences proposed in Chapter 3 and 4 needs to be integrated with the stochastic MPC approach with CVaR-type constraints presented in Chapter 6. This integration denotes the connection between the yellow part and green part in Fig 7.1. For this, a new evaluator that can quantify the stochastic optimization results using CVaR-type objectives or soft constraints is necessary in the proposed PD-MPC framework. However, this approach can cause another critical research question: how to formulate operators' preferences under stochastic optimization and whether the parameters and weights of objectives are sufficient to define the preferences in this context. We plan to address this challenge in subsequent studies.

Reflections on relevance to practice

Optimization-based control approaches are not generally adopted in practical reservoir flood control despite their proven effectiveness in various control problems and

the fact that the fundamentals of reservoir flood control coincide with the receding horizon control concept. This research focuses on addressing the limitations of applying a receding horizon MPC approach in practice. The dynamic preferences, practical objectives, computational feasibility, and risk management under uncertainty are explored.

This research investigates the dynamically changing preferences depending on varying hydrological conditions. Although there may be debate on how to define operator preferences, PD-MPC can reflect changes in weights without having to calculate the Pareto set. Moreover, explicit and switch MPC approaches by data-driven surrogate models can lessen the computational burden significantly while performing similarly to PD-MPC. Through this approach, optimal control based on receding horizon MPC, whose process closely resembles the flood control process using simulation models, has taken a step closer to practical online implementation.

Regarding the importance of reflecting the operator's risk tolerance quantitatively under uncertainty, a stochastic MPC approach with CVaR constraints demonstrates exceedance risk quantitatively and responds straightforwardly depending on risk tolerance. With the Euclidean distance-based K-means algorithm, the stochastic MPC has low computational complexity and better performance compared to the deterministic counterpart. Additionally, the effectiveness of the BNN model to produce uncertain inflow scenarios has been presented. This implies that a BNN model, or similar probabilistic data-driven models, can be utilized to predict inflows without forecasts, and it can replace the existing simulation models, which are being utilized to predict inflows with uncertain forecasts.

For the countries or reservoirs operated by predefined rules strictly regulating all details of the outflow amount, the frameworks in this research can be applicable alternatives for the automation. However, in general, this research does not aim to replace the entire simulation-based decision-making process with the automation framework of flood control. For the flood control of large multipurpose reservoirs in South Korea, where operators have a wide range of freedom in their decisions and need to take most responsibility for flood damages, there are many things that need to be considered, e.g., the operational opinions of related organizations and political pressures when hydrological conditions are extremely risky. Internalizing all those things is impossible due to the difficulty in formulating those constraints and objectives, as well as the complex relationships among them.

Therefore, the research outputs can be provided to the operators as references or supporting information. The operators can consult the outflow schedules generated by the PD-MPC framework and/or stochastic MPC approach. In principle, based on the author's experience, no operator willingly implements the optimal outflow schedules directly and promptly; instead, operators want to check, verify, and review them in detail. In this process, it would be a good opportunity to rethink his/her own decision based on simulation models and a limited number of forecasts. The operator can find some hints to improve the decisions or to make a final decision in the middle of an extreme flood, in which it is very difficult to receive accurate information, e.g., rainfall forecasts and soil moisture conditions, in advance. In

addition, the operator can utilize quantitative information, e.g., uncertain scenarios and flood risk, which are not generally provided quantitatively; in fact, he/she has considered them implicitly based on their experiences. Through this research, the operator can have a chance to consider them explicitly.

For the practical application, the frameworks should be tested in practical operational circumstances. Operators can evaluate the optimal outflow schedules from the frameworks in real-time operation, while implementing their own decisions. Through this evaluation, the feasibility of the research outputs can be verified and, more importantly, operators will realize the effectiveness and applicability of the receding horizon MPC based flood control proposed in this thesis.

Overall, the methods and frameworks developed in this research hold significant potential for practical implementation in reservoir flood control operations. By integrating dynamic operator preferences into real-time decision-making, reservoir operators can effectively respond to the increasing unpredictability and severity of flood events exacerbated by climate change. Consequently, the adoption of the proposed approaches can enhance flood management efficiency, mitigate downstream flood risks, and ultimately reduce both economic losses and adverse impacts on affected communities. Furthermore, the scalability and adaptability of these data-driven and control-based methodologies suggest broader applicability across various hydrological systems and regions, underscoring their value for future policy-making and operational practices.

Appendix A

Formulation of the Evaluator

The evaluator estimates how much the control inputs from the MPC design with a particular weight/parameter set satisfy all objectives, including nonlinear objectives. The evaluator returns the (weighted) sum of penalties, and GA finds the weights/parameters which minimise the objective value, $E_{\mathbf{z},\mathbf{u}}$.

Generally, the evaluator is formulated based on the objectives presented in Section 3.2.1. The objectives from $\tilde{J}1$ to $\tilde{J}5$, which are applied in the MPC formulation in linearised forms, are reformulated into non-linear equations, leveraging exponential functions to normalise all penalties. The evaluator also incorporates $\tilde{J}6$, $\tilde{J}7$, and $\tilde{J}8$, which aim to maintain the continuity of spillway gate conditions, to ensure that the outflow remains below the peak inflow up to the current time step, and to confirm full utilisation of turbine capacity before opening spillway gates, respectively (see Section 3.2.1). This is motivated by the difficulty of linearising $\tilde{J}6$ and $\tilde{J}8$, and the similarity of $\tilde{J}7$ to $\tilde{J}1$, which focuses on minimizing the peak outflow. In addition, for $\tilde{J}3$ and $\tilde{J}4$, the evaluator penalises only increases in outflow, in contrast to the linear MPC formulation of the objectives, which penalises both increases and decreases. Regarding $\tilde{J}5$, direct application of S_H is not feasible in the evaluator due to its potential variability. Instead, two alternatives are introduced: penalising instances where RWL increases within a prediction horizon and/or approaches FWL closely. It is worth noting that the evaluator uses a nonlinear system simulation (i.e., nonlinear height-volume curve) to calculate water levels, as mentioned in Section 3.2.3.

Some weights/parameters in the evaluator, such as weights of each penalty value in Equation (A.2) to (A.8), need to be set by a discussion with operators. In this article, weights in Equation (A.1), which are from $w_{e,1}$ to $w_{e,8}$, are assigned to one, except for $w_{e,1}$, $w_{e,2}$, and $w_{e,3}$. The weights for $P_{\tilde{J}1}$ and $P_{\tilde{J}2}$ are set as 0.9 and 0.1, respectively. This is because both objectives work similarly, except for when the outflow changes abruptly. The weight for $P_{\tilde{J}3}$ is set as 0.2 because $P_{\tilde{J}1}$ works similarly to $P_{\tilde{J}3}$ in general. Therefore, although it is necessary to explicitly include $\tilde{J}3$ in MPC formulation with the same search range, assigning a small weight for this objective is reasonable in Evaluator. The return of the evaluator is as follows:

$$\begin{aligned} E_z =& P_{\tilde{J}1} \times w_{e,1} + P_{\tilde{J}2} \times w_{e,2} + P_{\tilde{J}3} \times w_{e,3} + P_{\tilde{J}4} \times w_{e,4} + \\ & P_{\tilde{J}5} \times w_{e,5} + P_{\tilde{J}6} \times w_{e,6} + P_{\tilde{J}7} \times w_{e,7} + P_{\tilde{J}8} \times w_{e,8}. \end{aligned} \tag{A.1}$$

Peak outflow: $\tilde{J}1$ and $\tilde{J}2$

To penalize for the largest outflow, an exponential function is introduced for the scaling:

$$\begin{aligned} P_{\tilde{J}1} &= A^{\frac{\max O^k_{spill,t}}{B}}, \\ P_{\tilde{J}2} &= A^{\frac{\sum O^k_{spill,t}}{B \times H}}, \end{aligned} \tag{A.2}$$

where the exponent, A, and the other constant, i.e., denominator B, are selected to prevent $P_{\tilde{J}1}$ and $P_{\tilde{J}2}$ from being too small or large. Here, we set A to 2 and B to 1000. The prediction horizon length is denoted as H. The results are not hugely dependent on these values.

Step-wise outflow changes in a prediction horizon: $\tilde{J}3$

Unlike the objective of Equation (3.4) in Section 3.2.1, the changes in outflows in a prediction horizon are considered binary numbers to penalize equally for any changes in outflows. This approach allows us to overcome a common challenge faced by linear MPC, i.e., the objective values fluctuate according to the state, and there are still tiny differences resulting in a penalty that is too minor to impact the optimisation process significantly.

In addition, only increasing cases for the first four times are considered. First, a decrease in outflows is preferred, even though penalising only increases in outflows can lead to wild fluctuation when a prediction horizon is short. In reservoir flood control, a prediction horizon is generally short, as mentioned in Section 6.1. Therefore, only increases in outflows are considered by the evaluator because, in the linear MPC formulation, both increases and decreases are penalised already furthermore, because a change in outflows for the distant future is not a big problem, even though no change is the best, the increases for the first four times, i.e., from $t = k$ to $k + 3$, are penalised.

$$\begin{aligned} P_{\tilde{J}3} &= \sum_{t=k}^{k+3} C_{3,t} \times \omega_t, \\ C_{3,t} &= \begin{cases} 1 & \text{if } O^k_{spill,t+1} - O^k_{spill,t} > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{A.3}$$

where $C_{3,t}$ is also the binary value that is one when there is a change at time t at time step k from previous time steps $k-1$. The time-dependent weight ω_t decreases by time t from 0.5 to 0.1.

Changes in outflows calculated at consecutive time steps: $\tilde{J}4$

A similar concept to the previous one is applied to changes in outflows calculated at consecutive time steps as follows:

$$\begin{aligned} P_{\tilde{J}4} &= \sum_{t=k+1}^{k+4} C_{4,t} \times \omega_t, , \\ C_{4,t} &= \begin{cases} 1 & \text{if } O^k_{spill,t} - O^{k-1}_{spill,t} > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{A.4}$$

where $C_{4,t}$ is also the binary value that is one when there is a change at time t at time step k from previous time steps $k-1$. The weight ω_t is the same as the previous one in Equation (A.3).

Peak RWL: $\tilde{J}5$

A similar concept to Equation (3.6) in Section 3.2.1 is applied to the RWL, except for the S_H because S_H can change in PD-MPC. Instead, we add two different objectives, which are penalising when RWL increases in a prediction horizon and approaches FWL. The objective is as follows:

$$\begin{aligned} P_{\tilde{J}5} &= A^{[\max(RWL_t) - S_U]} + A^{[RWL_{k+H-1} - RWL_k]} + A^{[S_L - \min RWL_t]} \times C + P_{FWL}, \\ P_{FWL} &= \begin{cases} \text{large value} & \text{if } \max RWL^k_t > FWL - 0.1, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{A.5}$$

where the value in each bracket is zero if it is negative – we do not present a max function for this in Equation (A.5) so as not to look complicated. Because we want the penalties to have similar values, exponential functions are adopted. Here, we set C as two, the same as A, to circumvent RWL falling under S_L. In addition, a large value, 2000 in our research, is added to $P_{\tilde{J}5}$ when RWL is close to FWL. The exact magnitude of this 'large value' is not crucial; it simply needs to be sufficiently large to prevent the selection of undesirable $\mathbf{z}$.

Continuity of spillway gates condition: $\tilde{J}6$

Because the evaluator and GA can optimise this nonlinear objective, we can remove it from the MPC design and formulate linear MPC. A similar formula as Equation (3.7) is adopted, but we express this also using an exponential function as follows:

$$
\begin{aligned}
P_{\tilde{J}6} &= \sum_{t=k}^{k+H-1} (A^{\rho_{t,close}} + A^{\rho_{t,open}})/C, \\
\rho_{t,close} &= \begin{cases} 1 & \text{if } O^k_{spill,t} > 0 \text{ and } O^k_{spill,t+1} = 0, \\ 0 & \text{otherwise.} \end{cases} \\
\rho_{t,open} &= \begin{cases} 1 & \text{if } O^k_{spill,t} = 0 \text{ and } O^k_{spill,t+1} > 0, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{A.6}
$$

where the exponent A and denominator C are selected to scale this objective value to one, so A and C are set as two.

Peak outflow compared to the peak inflow: $\tilde{J}7$

To penalize the largest outflow when it is above the peak inflow up to the current time step, we assign a large value of γ to $\tilde{J}7$ when peak outflow is over peak inflow to the present, as mentioned in Section 3.2:

$$
P_{\tilde{J}7} = \begin{cases} \text{large value} & \text{if } \max_t O^k_{total,t} > \max_t I_t, \\ 0 & \text{otherwise.} \end{cases} \tag{A.7}
$$

where γ is set as 1000 in our research.

Turbine's priority over spillway: $\tilde{J}8$

Outflows by turbines should precede opening the spillway gates. In some cases, such as in South Korea, changing turbine outflows is not considered flood control because the total capacity of all turbines is generally much smaller than the spillway capacity. Therefore, if the optimal turbine outflows are less than the total capacity of turbines, even if the optimal spillway outflow is larger than zero, i.e., spillway gates open, the evaluator will penalize the weight/parameter set to prevent GA from selecting the set. This is formulated as follows:

$$
P_{\tilde{J}8} = \begin{cases} \text{large value} & \text{if } O^k_{spill,t} > 0 \text{ and } O^k_{turb,t} < MO_{turb}, \\ 0 & \text{otherwise,} \end{cases} \tag{A.8}
$$

where $O^k_{turb,t}$ is outflow via turbines at time t decided at time step k, and MO_{turb} is the outflow capacity of turbines. The large value, here, is set to 100000 to rigorously prevent this scenario.

Appendix B

Piecewise Affine approaches

The PWA approximation approach is motivated by the relationship between RWL and other state features in Fig B.1 and Fig B.2. It looks like the weight sets can be clearly clustered.

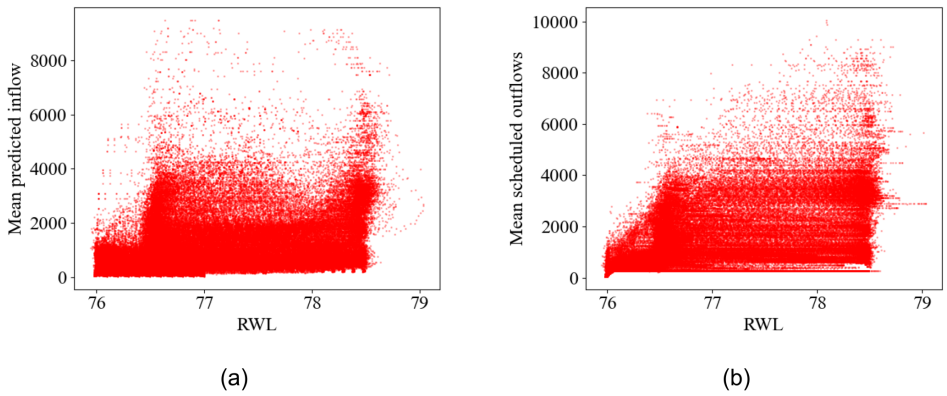


Figure B.1 | The relationship among state features when ZI= 1: (a) between RWLs and means of predicted inflow, (b) between RWLs and means of scheduled outflows.

Various methods for identifying hybrid models using Piecewise Affine (PWA) approximation have been proposed (Moradvandi *et al.*, 2023). These methods involve clustering a dataset, generating affine regressors for clusters, and identifying partitions of the state space. In this study, we do not look to find time-variant models and switching rules but to derive optimal parameters of the model based on states, which here are the optimal weights for the different objectives. Thus, the regressor maps the state space to the weights for the MPC. For this task, machine learning approaches can be used to find the optimal weights and switching rules.

Therefore, we utilize the K-means approach (Pedregosa *et al.*, 2011) to cluster state vectors and define an affine function mapping states to weights for clusters, denoted as 'W-CLR'. The main difference from conventional PWA identification is that explicit (polyhedral) partitioning for the entire state space is unnecessary. Instead, the optimal cluster is estimated at each time step based on the Euclidean distance between cluster centres and the new state. The number of clusters for

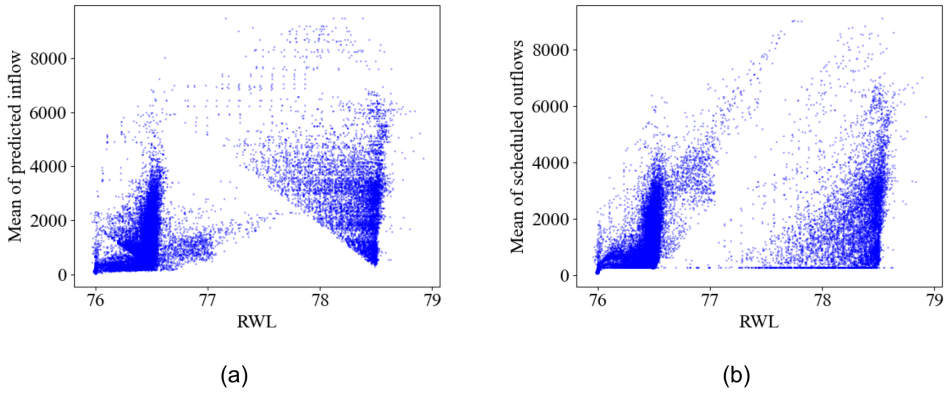


Figure B.2 | The relationship among state features when ZI= 2: (a) between RWLs and means of predicted inflow, (b) between RWLs and means of scheduled outflows.

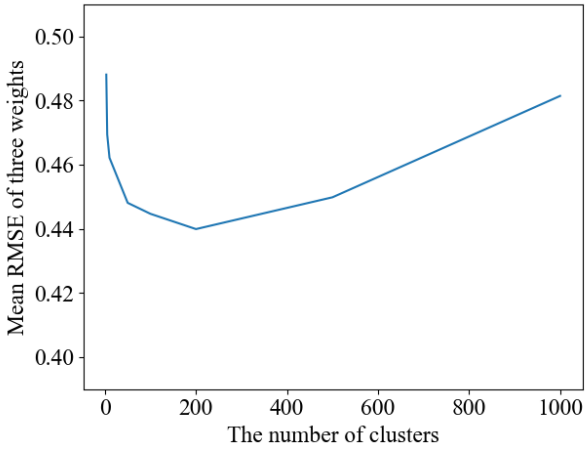


Figure B.3 | Hyperparameter grid search results (to compare three hyperparameters by the same scales, the scale of the y-axis is fixed to [0.4, 0.5]).

K-means is selected by grid search based on RMSE values for the validation events like the Random Forest regressor in Section 4.3.4. As shown in Fig B.3, because performance does not improve beyond 100 clusters and degrades beyond 200 clusters, we choose 100 clusters as sufficient.

In addition to K-means clustering and affine regression, we utilize the Linear Random Forest (LRF) presented in Ao *et al.* (2019) to approximate the traditional PWA approach. LRF applies almost the same methodology as the Random Forest regression, except that the final values are derived from the linear function in each final node. A switched MPC approach using LRF is denoted as 'W-LRF'.

W-CLR tends to increase outflows later than S-MPC, and S-MPC shows higher RWL than W-CLR, as seen in Table B.1 and B.2. This is because model errors accumulate when the number of models used increases. W-CLR utilizes three types of data-driven models, which are a binary cluster, a cluster, and regressors, unlike S-MPC, which has only two models. Therefore, W-CLR is likely to miss the best timing to increase outflow frequently. The fewer models for the best weights show

better results in terms of the peak outflow and peak values of the absolute performance evaluator. W-LRF shows moderate performance between W-CLR and S-MPC in all aspects.

Table B.1 | Average results for all test events and the three switched approaches

	Peak outflow	Peak RWL	Minimum RWL	Total E	Peak E	No. changes
W-CLR	9324.2	79.15	75.93	2214.5	105.0	28.1
W-LRF	7704.7	79.16	75.95	2212.4	68.3	30.7
S-MPC	4650.1	79.61	75.97	2667.7	52.3	23.9

Table B.2 | Average results for each test event and the three switched approaches

	Event	Peak out.	Peak RWL	Min. RWL	Total E	Peak E	No. changes
W-CLR	1	9109.4	78.8	75.9	1412.3	81.5	23.5
W-LRF	1	6987.6	78.8	75.9	1354.2	41.5	24.5
S-MPC	1	5225.7	79.5	76.0	1791.2	50.4	14.5
W-CLR	2	7580.5	79.1	75.9	1958.9	62.7	39.4
W-LRF	2	7839.2	79.0	75.9	1891.0	70.4	45.5
S-MPC	2	4154.6	79.4	76.0	2353.1	41.9	30.1
W-CLR	25	11282.8	79.6	76.0	3272.3	170.7	21.5
W-LRF	25	8287.3	79.7	76.0	3391.8	92.9	22.2
S-MPC	25	4570.1	79.9	76.0	3858.9	64.7	27.1

Bibliography

Ahmad, A., A. El-Shafie, S. F. M. Razali and Z. S. Mohamad, 2014. "Reservoir Optimization in Water Resources: a Review." *Water Resources Management* 28 (11): 3391-3405.

Ahmad, M., M. A. Al Mehedi, M. M. S. Yazdan and R. Kumar, 2022. "Development of machine learning flood model using artificial neural network (ANN) at Var River." *Liquids* 2 (3): 147–160.

Akiba, T., S. Sano, T. Yanase, T. Ohta and M. Koyama, 2019. "Optuna: A Next-generation Hyperparameter Optimization Framework." In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Alarab, I., S. Prakoonwit and M. I. Nacer, 2021. "Illustrative discussion of mc-dropout in general dataset: uncertainty estimation in bitcoin." *Neural Processing Letters* 53 (2): 1001–1011.

Alasali, F., S. Haben, H. Foudeh and W. Holderbaum, 2020. "A comparative study of optimal energy management strategies for energy storage with stochastic loads." *Energies* 13 (10): 2596.

Alessio, A. and A. Bemporad, 2009. "A survey on explicit model predictive control." *Nonlinear Model Predictive Control: Towards New Challenging Applications* pages 345–369.

Allgower, F., R. Findeisen and Z. K. Nagy, 2004. "Nonlinear model predictive control: From theory to application." *Journal-Chinese Institute Of Chemical Engineers* 35 (3): 299–316.

Anitescu, M., 2000. "On solving mathematical programs with complementarity constraints as nonlinear programs." *Preprint ANL/MCS-P864-1200, Argonne National Laboratory, Argonne, IL* 3.

Ao, Y., H. Li, L. Zhu, S. Ali and Z. Yang, 2019. "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling." *Journal of Petroleum Science and Engineering* 174: 776–789.

Arroyo, J., C. Manna, F. Spiessens and L. Helsen, 2022a. "Reinforced model predictive control (RL-MPC) for building energy management." *Applied Energy* 309: 118346.

Arroyo, J., C. Manna, F. Spiessens and L. Helsen, 2022b. "Reinforced model predictive control (RL-MPC) for building energy management." *Applied Energy* 309: 118346.

Arya, V., N. Garg, R. Khandekar, A. Meyerson, K. Munagala and V. Pandit, 2001. "Local search heuristic for k-median and facility location problems." In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 21–29.

Aydin, B. E., G. H. Oude Essink, J. R. Delsman, N. van de Giesen and E. Abraham, 2022. "Nonlinear model predictive control of salinity and water level in polder networks: Case study of Lissertocht catchment." *Agricultural Water Management* 264: 107502. DOI: https://doi.org/10.1016/j.agwat.2022.107502.

Badillo, S., B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert and J. D. Zhang, 2020. "An introduction to machine learning." *Clinical pharmacology & therapeutics* 107 (4): 871–885.

Bao, Y. and J. M. Velni, 2023. "A Hybrid Neural Network Approach for Adaptive Scenario-based Model Predictive Control in the LPV Framework." *IEEE Control Systems Letters* .

Bemporad, A., 2023. "A Piecewise Linear Regression and Classification Algorithm With Application to Learning and Model Predictive Control of Hybrid Systems." *IEEE Transactions on Automatic Control* 68 (6): 3194-3209. DOI: 10.1109/TAC.2022.3183036.

Bemporad, A. and D. M. de la Peña, 2009. "Multiobjective model predictive control." *Automatica* 45 (12): 2823–2830.

Berberich, J., J. Köhler, M. A. Müller and F. Allgöwer, 2022. "Linear tracking MPC for nonlinear systems—Part I: The model-based case." *IEEE Transactions on Automatic Control* 67 (9): 4390–4405.

Bergstra, J., R. Bardenet, Y. Bengio and B. Kégl, 2011. "Algorithms for hyper-parameter optimization." *Advances in neural information processing systems* 24.

Bergstra, J., D. Yamins and D. Cox, 2013. "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures." In *International conference on machine learning*, pages 115–123. PMLR.

Bøhn, E., S. Gros, S. Moe and T. A. Johansen, 2021. "Reinforcement learning of the prediction horizon in model predictive control." *IFAC-PapersOnLine* 54 (6): 314–320.

Breckpot, M., O. M. Agudelo and B. De Moor, 2013a. "Flood Control with Model Predictive Control for River Systems with Water Reservoirs." *Journal of Irrigation and Drainage Engineering* 139 (7): 532-541.

Breckpot, M., O. M. Agudelo and B. De Moor, 2013b. "Flood control with model predictive control for river systems with water reservoirs." *Journal of irrigation and drainage engineering* 139 (7): 532–541.

Breiman, L., 2001. "Random forests." *Machine learning* 45: 5–32.

Cai, W., A. B. Kordabad, H. N. Esfahani, A. M. Lekkas and S. Gros, 2021. "MPC-

based reinforcement learning for a simplified freight mission of autonomous surface vehicles." In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 2990–2995. IEEE.

Castelletti, A., A. Ficchí, A. Cominola, P. Segovia, M. Giuliani, W. Wu, S. Lucia, C. Ocampo-Martinez, B. De Schutter and J. M. Maestre, 2023. "Model Predictive Control of water resources systems: A review and research agenda." *Annual Reviews in Control* .

Chang, C.-N., E. D. M. Singer and A. D. Koussis, 1983. "On the mathematics of storage routing." *Journal of Hydrology* 61 (4): 357–370.

Chawla, N. V., K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, 2002. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research* 16: 321–357.

Che, D. and L. W. Mays, 2015. "Development of an optimization/simulation model for real-time flood-control operation of river-reservoirs systems." *Water resources management* 29: 3987–4005.

Chen, J., P. an Zhong, W. Liu, X.-Y. Wan and W. W.-G. Yeh, 2020. "A multi-objective risk management model for real-time flood control optimal operation of a parallel reservoir system." *Journal of Hydrology* 590: 125264.

Cheng, C. and K. W. Chau, 2001. "Fuzzy iteration methodology for reservoir flood control operation 1." *JAWRA Journal of the American Water Resources Association* 37 (5): 1381–1388.

Cheung, Y.-W. and K. S. Lai, 1995. "Lag order and critical values of the augmented Dickey–Fuller test." *Journal of Business & Economic Statistics* 13 (3): 277–280.

Chiang, P.-K. and P. Willems, 2015. "Combine evolutionary optimization with model predictive control in real-time flood control of a river system." *Water Resources Management* 29: 2527–2542.

Chung, S., M. R. Hipsey and J. Imberger, 2009. "Modelling the propagation of turbid density inflows into a stratified lake: Daecheong Reservoir, Korea." *Environmental Modelling & Software* 24 (12): 1467–1482.

Coxon, G., J. Freer, I. K. Westerberg, T. Wagener, R. Woods and P. Smith, 2015. "A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations." *Water resources research* 51 (7): 5531–5546.

De Nicolao, G., L. Magni and R. Scattolini, 1996. "On the robustness of receding-horizon control with terminal constraints." *IEEE Transactions on Automatic Control* 41 (3): 451–453.

Delgoda, D. K., S. K. Saleem, M. N. Halgamuge and H. Malano, 2013. "Multiple Model Predictive Flood Control in Regulated River Systems with Uncertain Inflows." *Water Resources Management* 27 (3): 765-790.

Di Baldassarre, G. and A. Montanari, 2009. "Uncertainty in river discharge observations: a quantitative analysis." *Hydrology and Earth System Sciences* 13 (6): 913–921.

Ding, D., M. Zhang, X. Pan, M. Yang and X. He, 2019. "Modeling extreme events in time series prediction." In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1114–1122.

Ding, W., C. Xu, M. Arief, H. Lin, B. Li and D. Zhao, 2023. "A survey on safety-critical driving scenario generation—A methodological perspective." *IEEE Transactions on Intelligent Transportation Systems* .

Dires, F. G., M. Amelin and G. Bekele, 2023. "Inflow scenario generation for the ethiopian hydropower system." *Water* 15 (3): 500.

Dupačová, J., N. Gröwe-Kuska and W. Römisch, 2003. "Scenario reduction in stochastic programming." *Mathematical programming* 95: 493–511.

Dvorkin, Y., Y. Wang, H. Pandzic and D. Kirschen, 2014. "Comparison of scenario reduction techniques for the stochastic unit commitment." In *2014 IEEE pes general meeting| conference & exposition*, pages 1–5. IEEE.

Elbashir, S., 2011. "Flood routing in natural channels using Muskingum methods." .

Farina, M., L. Giulioni and R. Scattolini, 2016. "Stochastic linear model predictive control with chance constraints–a review." *Journal of Process Control* 44: 53–67.

Ferreau, H. J., H. G. Bock and M. Diehl, 2008. "An online active set strategy to overcome the limitations of explicit MPC." *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal* 18 (8): 816–830.

Feurer, M. and F. Hutter, 2019. "Hyperparameter optimization." *Automated machine learning: Methods, systems, challenges* pages 3–33.

Feurer, M., J. Springenberg and F. Hutter, 2015. "Initializing bayesian hyperparameter optimization via meta-learning." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29.

Ficchi, A., L. Raso, P.-O. Malaterre, D. Dorchies, M. Jay-Allemand, F. Pianosi, P.-J. van Overloop and G. Thirel, 2014. "Short term reservoirs operation on the Seine River: performance analysis of Tree-Based Model Predictive Control." .

Gad, A. F., 2021. "Pygad: An intuitive genetic algorithm python library." *arXiv preprint arXiv:2106.06158* .

Gal, Y. and Z. Ghahramani, 2016. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning." In *international conference on machine learning*, pages 1050–1059. PMLR.

Germscheid, S. H., A. Mitsos and M. Dahmen, 2024. "Scenario Reduction Methods for Risk-Averse Demand Response Scheduling under Price Uncertainty." In

Computer Aided Chemical Engineering, vol. 53, pages 3223–3228. Elsevier.

Germscheid, S. H., F. T. Röben, H. Sun, A. Bardow, A. Mitsos and M. Dahmen, 2023. "Demand response scheduling of copper production under short-term electricity price uncertainty." *Computers & Chemical Engineering* 178: 108394.

Ghobadi, F. and D. Kang, 2022. "Multi-step ahead probabilistic forecasting of daily streamflow using Bayesian deep learning: A multiple case study." *Water* 14 (22): 3672.

Gill, M. A., 1978. "Flood routing by the Muskingum method." *Journal of hydrology* 36 (3-4): 353–363.

Giuliani, M., J. R. Lamontagne, P. M. Reed and A. Castelletti, 2021. "A State-of-the-Art Review of Optimal Reservoir Control for Managing Conflicting Demands in a Changing World." *Water Resources Research* 57 (12).

Gopikrishnan, P., V. Plerou, Y. Liu, L. N. Amaral, X. Gabaix and H. E. Stanley, 2000. "Scaling and correlation in financial time series." *Physica A: Statistical Mechanics and its Applications* 287 (3-4): 362–373.

Haddad, O. B., M. Farhangi, E. Fallah-Mehdipour and M. A. Marino, 2014. "Effects of inflow uncertainty on the performance of multireservoir systems." *Journal of Irrigation and Drainage Engineering* 140 (11): 04014035.

Hakobyan, A., G. C. Kim and I. Yang, 2019. "Risk-aware motion planning and control using CVaR-constrained optimization." *IEEE Robotics and Automation letters* 4 (4): 3924–3931.

Hart, W. E., J.-P. Watson and D. L. Woodruff, 2011. "Pyomo: modeling and solving mathematical programs in Python." *Mathematical Programming Computation* 3: 219-260.

Havens, K., H. Paerl, E. Phlips, M. Zhu, J. Beaver and A. Srifa, 2016. "Extreme weather events and climate variability provide a lens to how shallow lakes may respond to climate change." *Water* 8 (6): 229.

Heijden, T. van der, M. A. Mendoza-Lugo, P. Palensky, N. van de Giesen and E. Abraham, 2025. "Incorporating risk in operational water resources management: Probabilistic forecasting, scenario generation, and optimal control." *Water Resources Research* 61 (3): e2024WR037115.

Heijden, T. van der, P. Palensky, N. van de Giesen and E. Abraham, 2022. "Day Ahead Market price scenario generation using a Combined Quantile Regression Deep Neural Network and a Non-parametric Bayesian Network." In *2022 IEEE International Conference on Power Systems Technology (POWERCON)*, pages 1-5.

Heirung, T. A. N., J. A. Paulson, J. O'Leary and A. Mesbah, 2018. "Stochastic model predictive control—how does it work?" *Computers & Chemical Engineering* 114: 158–170.

Heitsch, H. and W. Römisch, 2003. "Scenario reduction algorithms in stochastic programming." *Computational optimization and applications* 24: 187–206.

Heitsch, H. and W. Römisch, 2007. "A note on scenario reduction for two-stage stochastic programs." *Operations Research Letters* 35 (6): 731–738.

Hertneck, M., J. Köhler, S. Trimpe and F. Allgöwer, 2018. "Learning an approximate model predictive controller with guarantees." *IEEE Control Systems Letters* 2 (3): 543–548.

Hewing, L., K. P. Wabersich, M. Menner and M. N. Zeilinger, 2020a. "Learning-based model predictive control: Toward safe learning in control." *Annual Review of Control, Robotics, and Autonomous Systems* 3: 269-296.

Hewing, L., K. P. Wabersich, M. Menner and M. N. Zeilinger, 2020b. "Learning-based model predictive control: Toward safe learning in control." *Annual Review of Control, Robotics, and Autonomous Systems* 3 (1): 269–296.

Hirose, N., R. Tajima and K. Sukigara, 2018. "MPC policy learning using DNN for human following control without collision." *Advanced Robotics* 32 (3): 148–159.

Hong, L. J., Z. Hu and L. Zhang, 2014. "Conditional value-at-risk approximation to value-at-risk constrained programs: A remedy via Monte Carlo." *INFORMS Journal on Computing* 26 (2): 385–400.

Horejšová, M., S. Vitali, M. Kopa and V. Moriggia, 2020. "Evaluation of scenario reduction algorithms with nested distance." *Computational Management Science* 17 (2): 241–275.

Hsu, N. S. and C. C. Wei, 2007. "A multipurpose reservoir real-time operation model for flood control during typhoon invasion." *Journal of Hydrology* 336 (3-4): 282-293.

Hu, J. and H. Li, 2019. "A new clustering approach for scenario reduction in multi-stochastic variable programming." *IEEE Transactions on Power Systems* 34 (5): 3813–3825.

Hu, M., G. H. Huang, W. Sun, Y. Li, X. Ding, C. An, X. Zhang and T. Li, 2014. "Multi-objective ecological reservoir operation based on water quality response models and improved genetic algorithm: A case study in Three Gorges Reservoir, China." *Engineering Applications of Artificial Intelligence* 36: 332–346.

Huang, I.-H., M.-J. Chang and G.-F. Lin, 2022a. "An optimal integration of multiple machine learning techniques to real-time reservoir inflow forecasting." *Stochastic Environmental Research and Risk Assessment* 36 (6): 1541–1561.

Huang, J., H. Qin, Y. Zhang, D. Hou, S. Zhu and P. Ren, 2023. "Short-term prediction method of reservoir downstream water level under complicated hydraulic influence." *Water Resources Management* 37 (11): 4475–4490.

Huang, X., B. Xu, P.-a. Zhong, H. Yao, H. Yue, F. Zhu, Q. Lu, Y. Sun, R. Mo, Z. Li

et al., 2022b. "Robust multiobjective reservoir operation and risk decision-making model for real-time flood control coping with forecast uncertainty." *Journal of Hydrology* 605: 127334.

Jain, S. K., L. Shilpa, D. Rani and K. Sudheer, 2023. "State-of-the-art review: operation of multi-purpose reservoirs during flood season." *Journal of Hydrology* page 129165.

Jain, S. K. and K. Sudheer, 2008. "Fitting of hydrologic models: a close look at the Nash–Sutcliffe index." *Journal of hydrologic engineering* 13 (10): 981–986.

Jeong, G., D.-G. Yoo, T.-W. Kim, J.-Y. Lee, J.-W. Noh and D. Kang, 2021. "Integrated quality control process for hydrological database: a case study of Daecheong Dam basin in South Korea." *Water* 13 (20): 2820.

Jospin, L. V., H. Laga, F. Boussaid, W. Buntine and M. Bennamoun, 2022. "Hands-on Bayesian neural networks—A tutorial for deep learning users." *IEEE Computational Intelligence Magazine* 17 (2): 29–48.

Kao, I.-F., Y. Zhou, L.-C. Chang and F.-J. Chang, 2020. "Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting." *Journal of Hydrology* 583: 124631.

Karg, B. and S. Lucia, 2020. "Efficient representation and approximation of model predictive control laws via deep learning." *IEEE Transactions on Cybernetics* 50 (9): 3866–3878.

Kasiviswanathan, K. and K. Sudheer, 2013. "Quantification of the predictive uncertainty of artificial neural network based river flow forecast models." *Stochastic environmental research and risk assessment* 27: 137–146.

Katoch, S., S. S. Chauhan and V. Kumar, 2021. "A review on genetic algorithm: past, present, and future." *Multimedia Tools and Applications* 80 (5): 8091-8126.

Kaut, M. and W. Stein, 2003. *Evaluation of scenario-generation methods for stochastic programming.* Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, Institut für Mathematik.

Khatibi, R., M. A. Ghorbani, M. H. Kashani and O. Kisi, 2011. "Comparison of three artificial intelligence techniques for discharge routing." *Journal of hydrology* 403 (3-4): 201–212.

Khodabakhsh, R. and S. Sirouspour, 2016. "Optimal control of energy storage in a microgrid by minimizing conditional value-at-risk." *IEEE Transactions on Sustainable Energy* 7 (3): 1264–1273.

Khorshidi, M. S., M. R. Nikoo, M. Sadegh and B. Nematollahi, 2019. "A multi-objective risk-based game theoretic approach to reservoir operation policy in potential future drought condition." *Water Resources Management* 33: 1999–2014.

Kim, H.-G., H.-Y. Lee, J. Kim, S. Lee, K. O. Boo and S.-E. Lee, 2021. "Effect of

Model Domain on Summer Precipitation Predictions over the Korean Peninsula in WRF Model." *Atmosphere* 31 (1): 17–28.

Kingma, D. P. and J. Ba, 2014. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* .

Ko, S.-K., D. G. Fontane and J. W. Labadie, 1992. "MULTIOBJECTWE OPTIMIZATION OF RESERVOIR SYSTEMS OPERATION 1." *JAWRA Journal of the American Water Resources Association* 28 (1): 111–127.

Kong, Y., Y. Mei, X. Wang and Y. Ben, 2021. "Solution Selection from a Pareto Optimal Set of Multi-Objective Reservoir Operation via Clustering Operation Processes and Objective Values." *Water* 13 (8). DOI: 10.3390/w13081046.

Labadie, J. W., 2004. "Optimal operation of multireservoir systems: State-of-the-art review." *Journal of Water Resources Planning and Management* 130 (2): 93-111.

Laio, F., G. Di Baldassarre and A. Montanari, 2009. "Model selection techniques for the frequency analysis of hydrological extremes." *Water Resources Research* 45 (7).

Langson, W., I. Chryssochoos, S. Raković and D. Q. Mayne, 2004. "Robust model predictive control using tubes." *Automatica* 40 (1): 125–133.

Le, X.-H., H. V. Ho, G. Lee and S. Jung, 2019. "Application of long short-term memory (LSTM) neural network for flood forecasting." *Water* 11 (7): 1387.

Le Ngo, L., H. Madsen and D. Rosbjerg, 2007. "Simulation and optimisation modelling approach for operation of the Hoa Binh reservoir, Vietnam." *Journal of Hydrology* 336 (3-4): 269-281.

Lee, S., T. Kang and K. Lee, 2017. "An operational model of a reservoir system simulation for real-time flood control in the Han River Basin." *Journal of Flood Risk Management* 10 (4): 499–510.

LemaÃŽtre, G., F. Nogueira and C. K. Aridas, 2017. "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning." *Journal of machine learning research* 18 (17): 1–5.

Li, H., P. Liu, S. Guo, B. Ming, L. Cheng and Y. Zhou, 2018. "Hybrid two-stage stochastic methods using scenario-based forecasts for reservoir refill operations." *Journal of Water Resources Planning and Management* 144 (12): 04018080.

Li, J., J. Zhou and B. Chen, 2020. "Review of wind power scenario generation methods for optimal operation of renewable energy systems." *Applied Energy* 280: 115992.

Li, L., S. Rong, R. Wang and S. Yu, 2021. "Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review." *Chemical Engineering Journal* 405: 126673.

Li, X., S. Guo, P. Liu and G. Chen, 2010. "Dynamic control of flood limited wa-

ter level for reservoir operation by considering inflow uncertainty." *Journal of hydrology* 391 (1-2): 124–132.

Li, Y., 2017. "Deep reinforcement learning: An overview." *arXiv preprint arXiv:1701.07274* .

Liberzon, D., 2003. *Switching in systems and control*, vol. 190. Springer.

Lin, N. M. and M. Rutten, 2016. "Optimal operation of a network of multi-purpose reservoir: A review." *Procedia Engineering* 154: 1376–1384.

Liu, D., T. Bai, M. Deng, Q. Huang, X. Wei and J. Liu, 2023a. "A parallel approximate evaluation-based model for multi-objective operation optimization of reservoir group." *Swarm and Evolutionary Computation* 78: 101288. DOI: https://doi.org/10.1016/j.swevo.2023.101288.

Liu, Y., X. Chen, N. Fan, Z. Zhao and L. Wu, 2023b. "Stochastic Day-ahead operation of cascaded hydropower systems with Bayesian neural network-based scenario generation: A Portland general electric system study." *International Journal of Electrical Power & Energy Systems* 153: 109327.

Löhndorf, N., 2016. "An empirical analysis of scenario generation methods for stochastic optimization." *European Journal of Operational Research* 255 (1): 121–132.

Maddu, R., I. Pradhan, E. Ahmadisharaf, S. K. Singh and R. Shaik, 2022. "Short-range reservoir inflow forecasting using hydrological and large-scale atmospheric circulation information." *Journal of Hydrology* 612: 128153.

Mae, Y., W. Kumagai and T. Kanamori, 2021. "Uncertainty propagation for dropout-based Bayesian neural networks." *Neural Networks* 144: 394–406.

Makhorin, A. "Glpk (gnu linear programming kit), version 5.0."

Malekmohammadi, B., B. Zahraie and R. Kerachian, 2011. "Ranking solutions of multi-objective reservoir operation optimization models using multi-criteria decision analysis." *Expert systems with applications* 38 (6): 7851–7863.

Merkuryeva, G., Y. Merkuryev, B. V. Sokolov, S. Potryasaev, V. A. Zelentsov and A. Lektauers, 2015. "Advanced river flood monitoring, modelling and forecasting." *Journal of computational science* 10: 77–85.

Meseguer, P., F. Rossi and T. Schiex, 2006. "Soft constraints." In *Foundations of Artificial Intelligence*, vol. 2, pages 281–328. Elsevier.

Micheli, F., T. Summers and J. Lygeros, 2022. "Data-driven distributionally robust MPC for systems with uncertain dynamics." In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 4788–4793. IEEE.

Montanari, A. and A. Brath, 2004. "A stochastic approach for assessing the uncertainty of rainfall-runoff simulations." *Water Resources Research* 40 (1).

Montero, R. A., D. Schwanenberg, M. Hatz and M. Brinkmann, 2013. "Simplified hydraulic modelling in model predictive control of flood mitigation measures along rivers." *Journal of Applied Water Engineering and Research* 1 (1): 17–27.

Moradvandi, A., R. E. Lindeboom, E. Abraham and B. De Schutter, 2023. "Models and methods for hybrid system identification: a systematic survey." *IFAC-PapersOnLine* 56 (2): 95–107.

Morari, M. and J. H. Lee, 1999. "Model predictive control: past, present and future." *Computers & chemical engineering* 23 (4-5): 667–682.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel and T. L. Veith, 2007. "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations." *Transactions of the ASABE* 50 (3): 885–900.

Moridi, A. and J. Yazdi, 2017. "Optimal allocation of flood control capacity for multi-reservoir systems using multi-objective optimization approach." *Water Resources Management* 31: 4521–4538.

Myo Lin, N., X. Tian, M. Rutten, E. Abraham, J. M. Maestre and N. van de Giesen, 2020. "Multi-objective model predictive control for real-time operation of a multi-reservoir system." *Water* 12 (7): 1898.

Nash, J. E. and J. V. Sutcliffe, 1970. "River flow forecasting through conceptual models part I—A discussion of principles." *Journal of hydrology* 10 (3): 282–290.

Nourani, V., A. H. Baghanam, J. Adamowski and O. Kisi, 2014. "Applications of hybrid wavelet–artificial intelligence models in hydrology: a review." *Journal of Hydrology* 514: 358–377.

Nubert, J., J. Köhler, V. Berenz, F. Allgöwer and S. Trimpe, 2020. "Safe and fast tracking on a robot manipulator: Robust mpc and neural network control." *IEEE Robotics and Automation Letters* 5 (2): 3050–3057.

Ortiz-Partida, J. P., T. Kahil, T. Ermolieva, Y. Ermoliev, B. Lane, S. Sandoval-Solis and Y. Wada, 2019. "A two-stage stochastic optimization for robust operation of multipurpose reservoirs." *Water Resources Management* 33: 3815–3830.

Pan, I., M. Babaei, A. Korre and S. Durucan, 2014. "Artificial Neural Network based surrogate modelling for multi- objective optimisation of geological CO2 storage operations." *Energy Procedia* 63: 3483-3491. DOI: https://doi.org/10.1016/j.egypro.2014.11.377. 12th International Conference on Greenhouse Gas Control Technologies, GHGT-12.

Panaretos, V. M. and Y. Zemel, 2019. "Statistical aspects of Wasserstein distances." *Annual review of statistics and its application* 6 (1): 405–431.

Park, Y., S. Kim, K. Chon, H. Lee, J. H. Kim and J.-K. Shin, 2021. "Impacts of heavy rain and floodwater on floating debris entering an artificial lake (Daecheong Reservoir, Korea) during the summer." *Desalination and Water Treatment* 219: 399–404.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, 2011. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12: 2825–2830.

Peitz, S. and M. Dellnitz, 2018. "A survey of recent trends in multiobjective optimal control—surrogate models, feedback control and objective reduction." *Mathematical and computational applications* 23 (2): 30.

Peng, Y., K. Chen, H. G. Yan and X. L. Yu, 2017. "Improving Flood-Risk Analysis for Confluence Flooding Control Downstream Using Copula Monte Carlo Method." *Journal of Hydrologic Engineering* 22 (8).

Pflug, G. C. and A. Pichler, 2014. *Multistage stochastic optimization*, vol. 1104. Springer.

Pflug, G. C. and A. Pichler, 2016. "From empirical observations to tree models for stochastic optimization: convergence properties." *SIAM Journal on Optimization* 26 (3): 1715–1740.

Pianosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson and T. Wagener, 2016. "Sensitivity analysis of environmental models: A systematic review with practical workflow." *Environmental Modelling & Software* 79: 214–232.

Pianosi, F., B. Dobson and T. Wagener, 2020. "Use of Reservoir Operation Optimization Methods in Practice: Insights from a Survey of Water Resource Managers." *Journal of Water Resources Planning and Management* 146 (12).

Powell, K. M., A. N. Eaton, J. D. Hedengren and T. F. Edgar, 2016. "A Continuous Formulation for Logical Decisions in Differential Algebraic Systems using Mathematical Programs with Complementarity Constraints." *Processes* 4 (1).

Qi, Y. T., J. S. Yu, X. D. Li, Y. X. Wei and Q. G. Miao, 2017. "Reservoir flood control operation using multi-objective evolutionary algorithm with decomposition and preferences." *Applied Soft Computing* 50: 21-33.

Reddy, M. J. and D. N. Kumar, 2006. "Optimal reservoir operation using multi-objective evolutionary algorithm." *Water resources management* 20: 861–878.

Ritter, J., G. Corzo, D. P. Solomatine and H. Angarita, 2020. "Multiobjective direct policy search using physically based operating rules in multireservoir systems." *Journal of Water Resources Planning and Management* 146 (4): 05020002.

Rockafellar, R. T. and S. Uryasev, 2002. "Conditional value-at-risk for general loss distributions." *Journal of banking & finance* 26 (7): 1443–1471.

Rockafellar, R. T., S. Uryasev *et al.*, 2000. "Optimization of conditional value-at-risk." *Journal of risk* 2: 21–42.

Rujeerapaiboon, N., K. Schindler, D. Kuhn and W. Wiesemann, 2022. "Scenario

reduction revisited: Fundamental limits and guarantees." *Mathematical Programming* 191 (1): 207–242.

Sadr, M. A. M., J. Gante, B. Champagne, G. Falcao and L. Sousa, 2021. "Uncertainty estimation via Monte Carlo dropout in CNN-based mmWave MIMO localization." *IEEE Signal Processing Letters* 29: 269–273.

Saltık, M. B., L. Özkan, J. H. Ludlage, S. Weiland and P. M. Van den Hof, 2018. "An outlook on robust model predictive control algorithms: Reflections on performance and computational aspects." *Journal of Process Control* 61: 77-102.

Schwenzer, M., M. Ay, T. Bergs and D. Abel, 2021. "Review on model predictive control: An engineering perspective." *The International Journal of Advanced Manufacturing Technology* 117 (5): 1327–1349.

Scudder, T. T., 2012. *The future of large dams: Dealing with social, environmental, institutional and political costs*. Routledge.

Séguin, S., C. Audet and P. Côté, 2017a. "Scenario-tree modeling for stochastic short-term hydropower operations planning." *Journal of Water Resources Planning and Management* 143 (12): 04017073.

Séguin, S., S.-E. Fleten, P. Côté, A. Pichler and C. Audet, 2017b. "Stochastic short-term hydropower planning with inflow scenario trees." *European Journal of Operational Research* 259 (3): 1156–1168.

Sharma, K. C., P. Jain and R. Bhakar, 2013. "Wind power scenario generation and reduction in stochastic programming framework." *Electric Power Components and Systems* 41 (3): 271–285.

Shelke, M., S. Londhe, P. Dixit and P. Kolhe, 2023. "Reservoir inflow prediction: a comparison between semi distributed numerical and artificial neural network modelling." *Water Resources Management* 37 (15): 6127–6143.

Shi, J., A. Shirali and G. Narasimhan, 2024. "Boosting Time Series Prediction of Extreme Events by Reweighting and Fine-tuning." In *2024 IEEE International Conference on Big Data (BigData)*, pages 1450–1457. IEEE.

Sideratos, G. and N. D. Hatziargyriou, 2012. "Probabilistic wind power forecasting using radial basis function neural networks." *IEEE Transactions on Power Systems* 27 (4): 1788–1796.

Song, Y. and D. Scaramuzza, 2022a. "Policy search for model predictive control with application to agile drone flight." *IEEE Transactions on Robotics* 38 (4): 2114–2130.

Song, Y. L. and D. Scaramuzza, 2022b. "Policy Search for Model Predictive Control With Application to Agile Drone Flight." *Ieee Transactions on Robotics* 38 (4): 2114-2130.

Székely, G. J. and M. L. Rizzo, 2013. "Energy statistics: A class of statistics based

on distances." *Journal of statistical planning and inference* 143 (8): 1249–1272.

Tang, R., W. Ding, L. Ye, Y. Wang and H. Zhou, 2019. "Tradeoff analysis index for many-objective reservoir optimization." *Water Resources Management* 33: 4637–4651.

Teegavarapu, R. S. and S. P. Simonovic, 2001. "Optimal operation of water resource systems: trade-offs between modelling and practical solutions." *IAHS-AISH PUBL.* (272): 257–262.

Theobald, C., F. Pennerath, B. Conan-Guez, M. Couceiro and A. Napoli, 2021. "A Bayesian neural network based on dropout regulation." *arXiv preprint arXiv:2102.01968* .

Uysal, G., R. Alvarado-Montero, D. Schwanenberg and A. Şensoy, 2018a. "Real-time flood control by tree-based model predictive control including forecast uncertainty: a case study reservoir in turkey." *Water* 10 (3): 340.

Uysal, G., D. Schwanenberg, R. Alvarado-Montero and A. Şensoy, 2018b. "Short term optimal operation of water supply reservoir under flood control stress using model predictive control." *Water Resources Management* 32: 583–597.

Vagropoulos, S. I., E. G. Kardakos, C. K. Simoglou, A. G. Bakirtzis and J. P. Catalao, 2016. "ANN-based scenario generation methodology for stochastic variables of electric power systems." *Electric Power Systems Research* 134: 9–18.

Van Der Heijden, T., P. Palensky, N. Van De Giesen and E. Abraham, 2023. "Closed-loop simulation testing of a probabilistic DR framework for Day Ahead Market participation applied to Battery Energy Storage Systems." In *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*, pages 1–6. IEEE.

Van Overloop, P.-J., 2006. *Model predictive control on open water systems*. IOS Press.

Wang, D., Z. J. Shen, X. Yin, S. Tang, X. Liu, C. Zhang, J. Wang, J. Rodriguez and M. Norambuena, 2021. "Model predictive control using artificial neural network for power converters." *IEEE Transactions on Industrial Electronics* 69 (4): 3689–3699.

Wang, F., O. C. Saavedra Valeriano and X. Sun, 2013. "Near real-time optimization of multi-reservoir during flood season in the Fengman Basin of China." *Water resources management* 27: 4315–4335.

Wang, H., Z. Lei, X. Zhang, B. Zhou and J. Peng, 2016. "Machine learning basics." *Deep learning* pages 98–164.

Wang, H., M. Olhofer and Y. Jin, 2017. "A mini-review on preference modeling and articulation in multi-objective optimization: current status and challenges." *Complex & Intelligent Systems* 3: 233–245.

Wang, L., T. Koike, M. Ikeda, D. N. Tinh, C. T. Nyunt, O. Saavedra, L. C. Nguyen,

T. V. Sap, K. Tamagawa and T. Ohta, 2014. "Optimizing multidam releases in large river basins by combining distributed hydrological inflow predictions with rolling-horizon decision making." *Journal of Water Resources Planning and Management* 140 (10): 05014006.

Wang, M., J. Qiu, H. Yan, Y. Tian and Z. Li, 2023. "Data-driven control for discrete-time piecewise affine systems." *Automatica* 155: 111168.

Watanabe, S., 2023. "Tree-structured Parzen estimator: Understanding its algorithm components and their roles for better empirical performance." *arXiv preprint arXiv:2304.11127* .

Watts, R. J., B. D. Richter, J. J. Opperman and K. H. Bowmer, 2011. "Dam reoperation in an era of climate change." *Marine and Freshwater Research* 62 (3): 321–327.

Wilson, D. L., 1972. "Asymptotic properties of nearest neighbor rules using edited data." *IEEE Transactions on Systems, Man, and Cybernetics* (3): 408–421.

Xu, B., S. E. Boyce, Y. Zhang, Q. Liu, L. Guo and P.-A. Zhong, 2017a. "Stochastic programming with a joint chance constraint model for reservoir refill operation considering flood risk." *Journal of Water Resources Planning and Management* 143 (1): 04016067.

Xu, B., P.-A. Zhong, Q. Huang, J. Wang, Z. Yu and J. Zhang, 2017b. "Optimal hedging rules for water supply reservoir operations under forecast uncertainty and conditional value-at-risk criterion." *Water* 9 (8): 568.

Xu, D., Z. Chen and L. Yang, 2012. "Scenario tree generation approaches using K-means and LP moment matching methods." *Journal of Computational and Applied Mathematics* 236 (17): 4561–4579.

Xu, M., P. J. van Overloop and N. C. van de Giesen, 2011. "On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow." *Advances in Water Resources* 34 (2): 282-290.

Yadav, J. and M. Sharma, 2013. "A Review of K-mean Algorithm." *Int. J. Eng. Trends Technol* 4 (7): 2972–2976.

Yang, S., M. P. Wan, W. Chen, B. F. Ng and S. Dubey, 2020. "Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization." *Applied Energy* 271: 115147.

Yu, X., Y.-P. Xu, H. Gu and Y. Guo, 2023. "Multi-objective robust optimization of reservoir operation for real-time flood control under forecasting uncertainty." *Journal of Hydrology* 620: 129421.

Zhang, J., F. Chen, Z. Cui, Y. Guo and Y. Zhu, 2020. "Deep learning architecture for short-term passenger flow forecasting in urban rail transit." *IEEE Transactions on Intelligent Transportation Systems* 22 (11): 7004–7014.

Zhang, L., S. Zhuang and R. D. Braatz, 2016. "Switched model predictive control of switched linear systems: Feasibility, stability and robustness." *Automatica* 67: 8–21.

Zhang, W., K. Wang, A. Jacquillat and S. Wang, 2023. "Optimized scenario reduction: Solving large-scale stochastic programs with quality guarantees." *INFORMS Journal on Computing* 35 (4): 886–908.

Zhao, T., X. Cai and D. Yang, 2011. "Effect of streamflow forecast uncertainty on real-time reservoir operation." *Advances in Water Resources* 34 (4): 495–504.

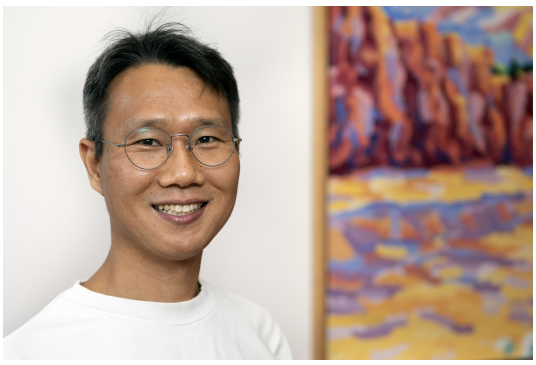
Zhu, F. and P. J. Antsaklis, 2015. "Optimal control of hybrid switched systems: A brief survey." *Discrete Event Dynamic Systems* 25: 345–364.

Zhu, F., P.-a. Zhong and Y. Sun, 2018. "Multi-criteria group decision making under uncertainty: Application in reservoir flood control operation." *Environmental Modelling & Software* 100: 236–251.

Ziel, F., 2021. "The energy distance for ensemble and scenario reduction." *Philosophical Transactions of the Royal Society A* 379 (2202): 20190431.

Zulfiqar, M., K. A. Gamage, M. Kamran and M. B. Rasheed, 2022. "Hyperparameter optimization of bayesian neural network using bayesian optimization and intelligent feature engineering for load forecasting." *Sensors* 22 (12): 4446.

About the author



Ja-Ho Koo was born in Boryeong (Republic of Korea) in October 1978. He graduated from Seoul National University in Seoul, Republic of Korea, in February 2004. He holds two bachelor's degrees, one in civil engineering and the other in business administration. He received an MSc degree in Public Policy from Korea University in 2015. He started his career as a civil engineer in the Korea Water Resources Public Corporation (K-water) in 2004. For more than 20 years, he has had a broad range of working experience relating to water resources management, such as operating multi-purpose reservoirs for water supply and flood control, taking part in developing water-related policies from the perspective of Integrated Water Resource Management (IWRM), integrating and developing a database for water-related data, constructing a tidal power plants, and so on. Currently, Mr. Koo is a PhD candidate at IHE Delft Institute for Water Education, Delft, The Netherlands, as well as a senior manager at K-water in South Korea.

Personal email: koojh78@gmail.com

List of publications

Peer-reviewed publications

Koo, Ja-Ho, Edo Abraham, Andreja Jonoski, and Dimitri P. Solomatine. (2025). Balancing operator's risk averseness in model predictive control for real-time reservoir flood control. *Journal of Hydroinformatics*. 27 (4), 601–624. DOI: https://doi.org/10.2166/hydro.2025.019

Koo, Ja-Ho, Ali Moradvandi, Edo Abraham, Andreja Jonoski, and Dimitri P. Solomatine. (2025). Flood Control of Reservoir Systems: Learning-based Explicit and Switched Model Predictive Control Approaches. *PLOS Water*. 4(5), e0000361. DOI: https://doi.org/10.1371/journal.pwat.0000361

Koo, Ja-Ho, Edo Abraham, Andreja Jonoski, and Dimitri P. Solomatine. (2025). Comparison of scenario reduction approaches for reservoir inflow timeseries generated by a Bayesian Neural Network. *PLOS One*. **(under review)**

List of abbreviations

ARMA	AutoRegressive Moving Average
ARIMA	AutoRegressive Integrated Moving Average
ARX	AutoRegressive models with eXogenous variables
BO	Bayesian Optimization
BNN	Bayesian Neural Network
CVaR	Conditional Value-at-Risk
DC	Daecheong Reservoir
DNN	Deep Neural Network
DMPC	Deterministic Model Predictive Control
DP	Dynamic Programming
EL	Elevation Level
E-MPC	Explicit Model Predictive Control
ENN	Edited Nearest Neighbors
FRWL	Flood season Restricted Water Level
FWL	Flood Water Level
GA	Genetic Algorithm
GN	Geumnam station
GP	Gap station
K-water	Korea Water Resources Public Corporation
LP	Linear Programming
LWL	Low Water Level
MAE	Mean Absolute Error
MAVE	Maximum Allowed Value Estimate
MCDM	Multi-Criteria Decision-Making
MH	Miho station

MPC	Model Predictive Control
MSE	Mean Square Error
NHWL	Normal High Water Level
PD-MPC	Parameterized Dynamic Model Predictive Control
PMF	Probable Maximum Flood
RMSE	Root Mean Square Error
RL	Reinforcement Learning
RWL	Reservoir Water Level
PARMA	Periodic AutoRegressive Moving Average
PWA	Piece-Wise Affine
SMOTE	Synthetic Minority Over-sampling Technique
S-MPC	Switched Model Predictive Control
SMPC	Stochastic Model Predictive Control
TPE	Tree-structured Parzen Estimator
VaR	Value-at-Risk
QRDNN	Quantile Regression Deep Neural Network

List of Figures

List of Tables

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Ja-Ho Koo

born on 18 October 1978, Boryeong, South Korea

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Delft, 3 November 2025

SENSE coordinator PhD education

Dr Ir Peter Vermeulen

The SENSE Director

Dr Jampel Dell'Angelo

The SENSE Research School declares that **Ja-Ho Koo** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 43,2 EC, including the following activities:

SENSE PhD Courses

- Environmental Research in Context (2021)
- Research in context activity: 'Presenting a proposal and organizing Webinars for reservoir operators who will use our research results' (2024)

Other PhD and Advanced MSc Courses

- Computational intelligence and operational water management, IHE Delft (2021)
- Prediction and control with function approximation, COURSERA (2021)
- The Informed Researcher: Information and Data Skills, TU Delft (2021)
- Effective Management of your PhD Research, TU Delft (2021)
- Software Carpentry Workshops, TU Delft (2021)
- Standing up for yourself while keeping a good relation, TU Delft (2021)
- PhD Start-up, TU Delft (2021)
- Creative Tools for Scientific Writing, TU Delft (2021)
- Academic English 2, TU Delft, TU Delft (2022)
- Using Creativity to Maximize Productivity and Innovation in your PhD, TU Delft (2022)
- Research Data Management, TU Delft (2022)
- Career Development – Looking for Work in the Netherlands, TU Delft (2022)
- Project Management for PhD Candidates, TU Delft (2023)
- Brain management, TU Delft (2023)
- Academic English 3, TU Delft (2023)
- Data Visualisations – A practical approach, TU Delft (2023)
- Mental Fitness Intervention Program, TU Delft (2023)
- Decision making under uncertainty, TU Delft (DelftX) (2023)
- Modelling with differential equations, TU Delft (DelftX), (2023)
- Linear algebra 1 & 2, TU Delft (DelftX) (2023)
- Renewable Energy: Small Hydropower Developments, IHE Delft (2024)
- How to formulate successful propositions for your PhD Defence, TU Delft (2024)

Oral Presentations

- Development of a water data model suitable for K-water. EKC: Europe-Korea Conference on Science and Technology. *31 October 2021, Paris, France.*
- Development of MPC formulations for practical reservoir flood control. *IHE Ph.D. Symposium, 19-20 October 2023, Delft, The Netherlands*

IHE DELFT
Institute for Water Education
under the auspices of UNESCO

TU Delft Delft University of Technology

Climate change has made reservoir flood control increasingly critical. This thesis addresses real-time flood control in South Korea, focusing on the limitations of existing operational methods due to uncertain rainfall forecasts. Traditional optimization methods face challenges, including unclear objectives, computational complexity, dynamic operators' preferences, and inflow uncertainty.

To address these issues, the thesis proposes a Parameterized Dynamic Model Predictive Control framework that transforms complex multi-objective optimization into manageable linear problems. Two data-driven approaches — explicit MPC and switched MPC — significantly reduce computation time, enabling faster decision-making during floods.

Bayesian Neural Networks are employed to generate a large set of representative inflow scenarios, which are further refined using scenario reduction. Distance measures are evaluated to prioritize preserving the envelope width of extreme scenarios critical for flood control, ensuring reduced scenario sets maintain the full range of risk conditions while achieving computational efficiency.

Finally, Conditional Value-at-Risk-based stochastic MPC incorporates operators' risk-aversion, improving robustness against inflow uncertainty. Validation on the Daecheong Reservoir demonstrates the effectiveness of these methodologies.

The proposed approaches offer flexibility to adapt to regional constraints, enhancing real-time flood control. This research contributes to improved balancing of operators' objectives in critical situations and facilitates the practical implementation of optimal reservoir flood control.

This book is printed on paper from sustainably managed forests and controlled sources

9 789073 445741