COMPETITION VS COOPERATION: APPLICATION OF GAME THEORY IN THE MULTI-AGENT COORDINATION OF A BC HYDROPOWER SYSTEM

by

Farah Rawas

B.A., Mount Holyoke College, 2017

B.S., University of Massachusetts, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Competition vs Cooperation: Application of Game Theory in the Multi-Agent Coordination of a BC Hydropower System

submitted by Farah Rawas in partial fulfillment of the requirements for

the degree of Master of Applied Science

in Civil Engineering

Examining Committee:

Dr. Ziad Shawwash, Department of Civil Engineering, UBC

Supervisor

Dr. Omar Swei, Department of Civil Engineering, UBC

Supervisory Committee Member

Abstract

Game theory has been gaining popularity as an innovative tool in the coordination of multi-reservoir systems for the optimal release and market policies through finding equilibrium. An equilibrium-based decision-making model (EDM) was developed to coordinate release and market decisions to meet demand and trade electricity from the Peace and Columbia systems in the US and Alberta Markets. The Williston and Kinbasket Reservoirs were taken as the main two agents to represent the Peace and Columbia rivers, respectively. Data on demand, inflows, prices, and release limits were provided by BC Hydro; and the Water Value Function for each reservoir was obtained from the Energy Studies Peace and Columbia Optimizers for the December 2019 study. The policies resulting from the game-theoretic model were compared to these of an existing iterative simulation and coordination model, the Energy Studies Models. The model showed reasonable results for the Peace system with low absolute error and mean absolute deviation for the drawdowns from Williston Reservoir, while the drawdowns from Kinbasket Reservoir showed larger error as compared to energy studies.

Three different solution algorithms were investigated: social optimum, Nash Support Enumeration, and Mixed Integer Linear Programming. In this case study, the Mixed Integer Linear Programming algorithm to find Nash Equilibrium gave the best strategies and rewards. However, the Nash Support Enumeration algorithm is more adaptable to situations with more than two agents. The results suggest that Game Theory is a promising technique that should be further investigated and enhanced to aide Energy Studies in the coordination of reservoir release policies. To further develop the model results, inverse reinforcement learning algorithms in Stochastic Games were investigated and presented. An effective way to validate and compare

this model and the different tools developed by the BC Hydro's system optimization group is by following a model benchmarking framework detailed in this research.

Lay Summary

In British Columbia, over 75% of the installed power generating capacity is at hydroelectric generation stations in the Peace and Columbia river basins. The coordination of these plants is a mathematically complex task that involves solving problems with multiple-decision makers that operate the different generation plants. Game theory has been well-regarded as a tool to solve multi-agent, central coordination systems as opposed to conventional, single-objective optimization techniques. The goal of this research is to develop and test a multi-agent reservoir coordination model based on game-theoretic algorithms to inform operational decisions for the Columbia and Peace river systems, on storage operations, and purchases and sales in the electricity market. The results from this model were compared with existing models developed by BC Hydro. The results indicate that this modelling approach can achieve efficient computational results to optimize the operations and planning of hydro systems under uncertain water inflows scenarios.

Preface

The work presented in this research is carried out by the author Farah Rawas under the supervision of Dr. Ziad Shawwash in collaboration with the Energy Studies team within BC Hydro's Generation Systems Operations department.

The author developed and applied the model and both the Social Optimum and the MILP algorithms. The input data used in the case study were generated from BC Hydro models. Mr. Tim Blair explained the working of the Energy Studies Models and was instrumental in this research. Dr. Ziad Shawwash, and Mr. Tim Blair, specialist engineer at BC Hydro, have provided several valuable suggestions to improve the modelling framework.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

ARD: Arrow Dam

BC: British Columbia

BCH: British Columbia Hydro

cms: Cubic Meter Seconds

CRT: Columbia River Treaty

EDM: Equilibrium Based Decision Making Model

GJ: Giga Joules

GMS: G. M. Shrum Generation Station

GT: Game Theory

KBT: Kinbasket Lake

kWh: Kilowatt Hour

LP: Linear Programming

m: Meter

MCA: Mica Generation Station

MILP: Mixed Integer Linear Programming

MWh: Megawatt Hour

RL: Reinforcement Learning

US: United states

WSR: Williston Reservoir

Acknowledgements

The UBC Vancouver campus is located on the traditional territory of the xʷməθkʷəy̓əm (Musqueam) people. This research project was fully funded by BC Hydro and by a Collaborative Research Development NSERC Grant CRDPJ 543692-19 to Dr. Shawwash.

I owe a debt of thanks to many people since it is their generous help along that has led me here. First and foremost, I present my sincere gratitude to my supervisor, Dr. Ziad Shawwash, for providing me with this opportunity, enlarging my vision of operations research and water resources engineering, and always encouraging me to seek out learning opportunities.

I would like to acknowledge the vast support of the fellow engineers at BC Hydro. In particular, I would like to thank Mr. Tim Blair for his patience and ability to explain complex models in simple terms.

I offer my enduring gratefulness to the faculty, staff, and my fellow students at UBC, who have inspired me to continue my work in this field. I owe special thanks to Dr Omar Swei, who taught me important lessons in the craft of research.

This thesis was written, edited, and finalized during a global pandemic. As the universities around the world closed their doors, and the world went into a lockdown, I fought through the grief and discomfort of the situation to let this work see light.

This humble work is dedicated to

my wonderful nieces, Celine and Celia Abou Afash, who inspire me to push through the struggles and achieve my goals to inspire them to be strong independent women in the field of their passion,

my family for supporting me and tolerating the years and tears of distance while I fulfill my ambitions abroad,

and to every person who was there for me during difficult times, listening, caring, and sharing with me tremendous lessons from their professional and personal experiences.

Chapter 1: Introduction

1.1 Research Question and Motivation

Over 75% of BC Hydro's installed generating capacity is at hydroelectric generation stations in the Peace and Columbia river basins. The coordination of these systems is an essential yet complex task. It was Ganji et al. (2007) argued that the available techniques commonly used in reservoir optimization/operation do not consider interaction, behavior and preferences of water users, reservoir operator and their associated modeling procedures, within the stochastic modeling framework. To address this issue, competitive and cooperative methods are being used. The surveyed literature throughout this thesis demonstrated the use of Game Theory (GT) models in several scenarios: operations research, water resources planning, considering short- and long-term planning and deterministic and stochastic inputs. Previous work, despite the gaps in literature, suggests great utility and opportunity in using both GT and decomposition methods for long term hydropower strategies, coordinating multi agent reservoir optimization problems, and scheduling of hydrothermal systems to minimize costs and optimize bidding in a competitive market. The findings from the literature present an opportunity for research to advance the GT approaches to hydropower optimization and scheduling through analyzing methods that can:

a) account for stochastic inputs to the systems from environmental or economic factors,
b) use decomposition methods to ensure convergence,
c) reduce computation time,
d) coordinate multiple agents in an uncertain environment, and
e) introduce and model communication among agents.

Therefore, the main question that this research will attempt to answer is: can Game Theory help in coordinating system operations?

1.2 Challenges Addressed

In British Columbia, the coordination of hydropower generation is an essential but complex task that involves solving multi-decision-maker problems of multiple generation plants. Competitive and cooperative game theory has been increasingly used to solve multi-agent, central coordination systems as opposed to conventional, single-objective optimization techniques. Usually, those techniques do not account for the much needed "interaction" of the agents and only emphasize the common interest of the system and ignore individual interests (Madani 2010). Optimal coordination of the many facets of reservoir systems requires the assistance of computer modeling tools to provide information for rational management and operational decisions (Labadie 2004). We propose and test game-theoretic algorithms to coordinate and optimize the multi-reservoir system of BC Hydro under several inflow scenarios. The model can be used to inform operational decisions for the Columbia and Peace river systems, on storage operations, and purchases and sales in the electricity market.

1.3 Goals and Objectives

The work in this thesis fulfills three main objectives:

1. Inform operational decisions for Columbia and Peace rivers system, on storage operations, and purchases and sales of market electricity.
2. Build a preliminary coordination model for the Columbia and Peace reservoirs based on Game Theoretic techniques.
3. Propose a benchmarking framework to test and compare the full model to other optimization and coordination stochastic models.

To achieve these goals and objectives, a thorough literature review was conducted on Game Theoretic techniques in operation research, multi-agent coordination problems, and hydropower management. Modelling application were also surveyed, and three solution algorithms for Game Theory were chosen to be used in the models and tested in a case study. Also, practical understanding of the current modelling approaches used within the BC Hydro Generation System Operations group at BCH Hydro was accomplished through several meetings. These meetings also served to discuss the proposed approaches to model the BCH Columbia and Peace rivers systems, and test their coordination based on GT techniques. A proof-of-concept model was built using a variation of GT solution algorithms. Case study data were acquired from BCH as well as results to be tested against the results of the built model. Finally, theory and applications of benchmarking of stochastic optimization were investigated, and a benchmarking scheme was established.

1.4 Overview of Subsequent Sections

The first section of Chapter 2 explains the concept of game theory and presents a literature survey that investigates the different applications of those techniques, with a focus on water resources management literature. The second section of Chapter 2 provides a background on the BC Hydro reservoir system and its management. Chapter 3 explains the research methods, modelling approaches, and game-theoretic algorithms. In Chapter 3.5, the case study is defined, and the results are presented. Chapter 4 discusses the results of the case study and the implications of the model on the operation of the BC Hydro system. In Chapter 5, we present a benchmarking framework to compare this model to other stochastic reservoir simulation and optimization models. Then follows the conclusion in Chapter 6.

Chapter 2: Background

2.1 Game Theory

Game theory (GT) is essentially the mathematical study of competition and cooperation. It was formally introduced in 1944 with the publication of von Neumann and Morgenstern's "Theory of Games and Economic Behavior" and gained ground with the proposal of Nash's equilibrium solution. Traditional GT is an economic theory that models interactions between rational agents as games of two or more players that can choose from a set of strategies and the corresponding preferences (Tuyls and Nowé 2005). It is the mathematical study of interactive decision making in the sense that the agents involved in the decisions consider their own choices and those of others. Choices are determined by stable preferences concerning the outcomes of the agent's possible decisions as well as the relation between their own choices and the decisions of other agents.

In game theory, there are decision makers that play a game to optimize their own objective **(competitiveness)**, knowing that other players' decisions affect their objective value and that their decision affects others' payoffs and decisions **(cooperation)**. Cooperative multitasking is also referred to as ***non-preemption***. The payoffs to players determine the decisions made and the type of the game being played. The concept of ***Pareto efficiency*** *or* ***Pareto optimality*** is a state of allocation of resources from which it is impossible to reallocate resources so as to make any one individual or preference criterion better off without making at least one individual or preference criterion worse off. A ***Nash Equilibrium*** (NE) is a state of the game where no player prefers a different action if the current actions of the other players are fixed (Nash 1953). In other words, if there is a set of strategies for a game with the property that no player can increase its payoff by changing his strategy while the other players keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute a Nash equilibrium. ***Propensity to disrupt*** is another commonly used quantitative method to evaluate the stability of game theoretic allocations with respect to the cooperating agents' powers in a coalition of agents.

Stochastic game theory approaches may have multiple NE states with different values, none of which are strictly optimal (Bab and Brafman 2008). The list of competitive game theory (CGT) solution concepts include: The Core, The Shapley Value; The Nucleolus; The Generalized Shapley; Nash/Nash-Harsany; and others (Dutta 1999; Madani 2010).

Game theoretic approaches to dynamic programming were discussed in (Bertsekas and Tsitsiklis 1996). In a game between a minimizer and a maximizer, the authors introduce a set of randomized strategies for each player. Each player has a probability distribution of selecting their strategy. Those probabilities then are added to the probability of the system state transition and expected cost. The policy for each player consists of functions that choose at a state the probability distribution over a set of strategies. The cost to go function here becomes the equilibrium value of the game. Discounted games are analogous to the Bellman equation but for two players. ***Sequential games*** can be embedded to existing general game framework such that each player selects a choice with full knowledge of the preceding, or the following, choice of the other player and the state transition resulting from that choice. Sequential stochastic games are important since they are the subject of multi-agent inverse reinforcement learning as explained later in GT in Multi-Agent Systems and Reinforcement Learning. A Q-learning algorithm, where agents learn how to act optimally in controlled Markovian domains (Watkins and Dayan 1992), can be derived for sequential games in which we can obtain a Bellman equation that satisfies the Q-factors.

2.1.1 Classification and elements of game models

2.1.1.1 Normal Games

Normal games have three main components: the players, also referred to in this thesis as agents, the actions that the players can do, and their payoffs or utility functions. Normal form games list players' payoffs as a function of their actions in a matrix representation. In normal form games, the notion of time, or the sequence of the actions of the players is not accounted for.

2.1.1.2 Stochastic and Markov Games

A stochastic game is a very broad framework which generalizes both Markov Decision Processes and repeated games. A stochastic game is a collection of normal-form games that the agents play repeatedly. The game played at any time depends probabilistically on the previous game played and the actions of the agents in that game.

2.1.1.3 Example of Stochastic Repeated Games: OPEC Oil Cartel

An application for infinitely repeated games is the Organization for Oil Exporting Countries (OPEC). The oil cartel OPEC seeks to maintain high prices by restraining its members' production levels through explicit quotas. It has had mixed success and the price history of world oil has been fluctuating from being low and stable to high and unstable, as it has been in recent years. This price history can be rationalized by way of some critical ideas: demand uncertainty, international politics, and repeated-game perspective. OPEC exists because its members realize that they are in a repeated game; bound by persistence in high demand for oil. Dutta 1999 modeled OPEC Oil market as a repeated game with demand uncertainty, where 2 exporters (a large exporter country and a small exporter country) have different strategies for how much to produce (high or low production) based on the level of the demand (high or low demand). The profit is represented in a matrix for good demand and bad demand as price per barrel as discussed below.

Suppose the countries as SA and VA with respective output levels Q_L and Q_H for SA and q_L and q_H for VA. The payoffs ($ per barrel) for taking specific decisions (output levels based on demand) are presented in the payoff matrix in Table 1.

Table 1: Payoff matrices of two demand scenarios. Adopted from Dutta 1999

Good Demand		
SA\ VA	**q_L**	**q_H**
Q_L	160, 100	136, 119
Q_H	170, 85	140, 98
Bad Demand		
SA\ VA	**q_L**	**q_H**
Q_L	88, 55	80, 70
Q_H	100, 50	90, 63
Where, SA and VA are the two agents, q_{L}, q_H represent the low and high output levels of agent VA Q_L , Q_H represent the low and high output levels of agent SA		

Here, the game represented in Table 1 is divided into two scenarios representing a good or a bad demand year, and the payoffs inside the table (*eg. 160, 100*) represent the payoffs of agents SA and VA respectively (in that order) for taking the decisions q_x or Q_x under a certain scenario, given the other's decision.

In a repeated game, let's say SA and VA agree to cooperate on a strategy to produce low in a *good* demand year, then with the profit numbers given in Table 1, in a repeated game setting, SA would have a profit for sustaining a low output strategy of $160 + 160\delta + 160\,\delta^2 + \ldots$ (where δ is the discount factor). Whereas if SA decides to overproduce (Q_H) then the payoff would be penalized by VA's decision to defect in the following year and SA's payoff would become:

$170 + 140\,\delta + 140\,\delta^2 + \ldots$

This assumes that any quota violation is observable to the cartel partners.

Now let $\boldsymbol{p}$ denote the probability that a demand is robust in any period. With this dimension, a modified stage-game payoff matrix can be created where SA and VA each have a set of four decisions to make within each stage: Q_L regardless of demand (in all cases), Q_H regardless of demand (in all cases), Q_L only if demand is good (Q_H if demand is bad), and Q_L only if demand is bad (Q_H if demand is good).

If SA was to abide by the low-production policy for a good demand year, the stream of future profit for SA would like the following:

$[160p + (1-p)\,90]\,\delta + [160\,p + (1-p)\,90]\,\delta^2 + \ldots$

Equating this payoff with the payoff stream of a strategy to defect allows us to calculate the discount factor δ condition in terms of *p* which ensures that a country cooperates.

In this case, if SA overproduces, it will get a lifetime profit of:

$170 + [140p + (1- p)\, 90]\, \delta + [140\, p + (1- p)\, 90]\, \delta^2 + \ldots$

It is not profitable to overproduce if:

$$160 + \frac{[160p + (1- p)\, 90]\, \delta}{1- \delta} >= 170 + \frac{[140p + (1- p)\, 90]\, \delta}{1- \delta}$$

Therefore the condition is:

$$\delta \geq \frac{1}{1 - 2p}$$

From this analysis, (Dutta 1999) concludes that in any market it is the smaller producers who have the most to gain from cheating on OPEC. Hence, they are the most likely quota violators. In real life, determining who has the best incentive to cheat proved to be a much more complex topic on which several books were published. A similar analysis can be carried out even if quota violations are unobservable. In that case there will be strategic price uncertainty (in addition to that caused by demand uncertainty) as OPEC triggers occasional price wars on account of low prices.

2.1.2 Cooperative vs Non-cooperative Games

Some of the major distinction is between non-cooperative and cooperative game theory is that in non-cooperative games (Bauso, 2014):

i) every player seeks its best response based on the available information and in order to maximize its own payoff,

ii) there are no binding agreements on optimal joint actions,
iii) pre-play communication is possibly allowed.

In cooperative games:

i) the players seek optimal joint actions, or reasonable cost/reward sharing rules that make the coalitions stable
ii) pre-play communication is allowed, and
iii) side payments are also allowed.

Note that while non-cooperative game theory is by far more widespread than cooperative game theory, there is a large consensus on the idea that cooperative game theory has a broader range of applications and has become a major design tool in engineered systems such as electrical power grid systems and water distribution systems.

2.1.3 Game Equilibrium

A ***Nash Equilibrium*** (NE) is a state of the game where no player prefers a different action if the current actions of the other players are fixed (Nash 1953). In other words, if there is a set of strategies for a game with the property that no player can increase its payoff by changing his strategy while the other players keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute a Nash equilibrium. What makes the notion of equilibrium compelling is that all matrix games have a Nash equilibrium, although there may be more than one.

2.2 Game Theory in Reservoir Management and Operation Research Literature

In dynamic programming, iterative algorithms are used to help estimate the value function. For solving problems with multiple criteria and multiple decision makers, different optimization methods can predict different outcomes. Classic reservoir optimization methods are designed to prevail over the high-dimensional, dynamic, nonlinear, and stochastic characteristics of reservoir systems; however, there continues to be a gap between theoretical developments and real-world implementations (Labadie 2004). Conventional optimization methods usually convert the multi-decision-maker problems of the whole system into a single-decision-maker problem, with a single composite objective (Madani 2010). Consequently, those schemes only emphasize the common interest of the whole organization and ignore individual systems' interests. Real world water management cases involve ongoing relationships that are realized through cooperative strategies (Ristić and Madani 2019). In a scenario of predominant dependence on hydro generation such as in British Columbia, Canada (Shawwash et al. 2000), the coordination of hydropower generation is an essential but complex task that involves coordination of plants that might not be connected hydraulically or that are owned by different agents (Faria et al. 2018). Cooperative game theory has been attracting the attention to solve the multi agent, central coordination hydropower and energy production systems. This is due to the nature of the method that reflects different behaviors performed by different parties, which are otherwise neglected by the other conventional methods (Madani 2010; Parrachino et al. 2006). In what follows is a discussion of the algorithms that have been used in solving large scale stochastic problems and a survey of the different applications of those techniques in recent literature, with a focus on water resources management.

Competitive-Cooperative Coevolutionary Paradigm for Dynamic Multiobjective Optimization

In this paper, (Chi-Keong Goh and Kay Chen Tan 2009) presented a coevolutionary paradigm that incorporates the competitive and cooperative mechanisms observed in nature to facilitate adaptive problem decomposition in coevolution. Empirical studies were conducted by the author for both dynamic and static environments. The main idea of competitive-cooperative coevolution is to allow the decomposition process of the optimization problem to adapt and emerge rather than being hand designed and fixed at the start of the evolutionary optimization process. Existing coevolutionary techniques can be divided into two main classes: competitive coevolution and cooperative coevolution.

According to the authors, an explicit way of implementing cooperative coevolution in optimization techniques is to split a solution vector into different subcomponents and assign multiple evolving subpopulations to optimize the individual subcomponents. Here, each species subpopulation will compete to represent a particular subcomponent of the multi-objective (MO) problem, while the eventual winners will cooperate to evolve for better solutions. Through such an iterative process of competition and cooperation, the various subcomponents are optimized by different species subpopulations based on the optimization requirements of that particular time instant. An early attempt to integrate the cooperative model for MO optimization is to decompose the problem along the decision space, and each subpopulation is optimized by the MO genetic algorithm. One major issue of these MO coevolutionary algorithms is their dependence on appropriate manual decomposition of the problem into various subcomponents. The game-theoretic approach of modeling cooperation attempts to alleviate the issue of parameter dependencies by decomposing the optimization problem into only two subpopulations.

Cooperative game theory and last addition method in the allocation of firm energy rights

(Faria et al. 2018) integrates the concepts of cooperative game theory (CGT) and last addition (LA) allocation method to compute the firm energy rights of each hydro plant in the Brazilian hydropower system. Their model uses deterministic parameters. The last addition allocation method is widely used to allocate energy between hydro plants proportionally to the incremental benefit that exists when the system is simulated with and without this plant. A firm energy allocation satisfies the conditions of fairness of a cooperative game if none of the agents have interest in leaving the coalition, where the sum of firm energy allocated to any agent should be greater than or equal to the energy amount that an agent could generate when operating alone, i.e. maximizing only its own energy. The GT-LA algorithm starts by searching for allocations close to optimal non-cooperative allocations, iteratively, that satisfy a set of constraints. Then, it computes the squared sum of the percentage difference between the new and first allocation. The algorithm stops when the maximum violation found by the cooperative game model is less than an error epsilon. Their decision variables include individual firm energy of each hydro plant, firm energy of a subset of hydro plants, average power generated, turbine outflow, water spillage, available water volume stored at each plant. This paper also investigated the performance of the Benders algorithm and the mixed integer linear programming when dealing with the firm energy allocation problem to compute cooperative game constraints and concluded that there is no significant advantage in solving the problem via Benders decomposition. However, the authors conclude that methods that aim to reduce computational time in computing the cooperative game constraints should be analyzed.

Bayesian and Robust Nash Equilibrium in Hydro-Dominated Systems Under Uncertainty

In this paper (Moiseeva and Hesamzadeh 2018) use modified Benders decomposition algorithm (MBDA) to solve equilibrium problems with equilibrium constraints (EPEC), reformulated as a stochastic mixed-integer linear programming (MILP) model, which is solved for a global optimum. The model uncertainty comes from wind power production, demand levels, inflow uncertainty. The authors model the interactions of price-making hydropower producers: profit-maximizing hydropower producers compete in electricity market by submitting strategic bids. The bids are received by the system operator, which orders them and determines the system price. Two main aspects that differentiate hydropower producers from thermal generators are the water value, accounting for future opportunities for the water in the reservoirs and the waterway coupling of the reservoirs in a congested network.

Solving the EPEC resulted in finding a Nash equilibrium: a stable set of strategies, from which no generator wants to deviate unilaterally, given that the strategies of the competitors are held fixed. EPEC was solved using primal and dual variables and upper level problem. MBDA was applied to the EPEC formulation to decompose the initial MILP into a small coordinating master problem and an LP subproblem, which is easier to solve. The application of the solution algorithm – MBDA – to EPEC problem reformulated as a MILP did not require the optimal tuning of the disjunctive parameter.

Simulation-Optimization Model to Derive Operation Rules of Multiple Cascaded Reservoirs for Nash Equilibrium

(Wu et al. 2019) proposed a game model for hydropower operation to obtain operation rules according to Nash equilibrium. The model assumes that each agent controls one cascaded hydropower reservoir in which the objective of each is to maximize its total profit over all simulation periods. The operation rule makes decision of total cascade power generation according to energy storage, and decides the power allocation among reservoirs. The novelty of the proposed model is that the analysis combines reservoir simulation-optimization and power market games. The operation of multiple cascaded hydropower reservoir system can be formulated as an *N*-player noncooperative game model, in which each player's action is defined by an operation rule and payoff function. When a cascade player believes that the other players will use certain rules, it can get an optimal rule for its own, and the changes of its rule may cause a profit loss for the others. Then other cascades then need to re-optimize their rules. The process can be repeated for several rounds until no agent can improve its operation rule while the other's is fixed. The Nash equilibrium is obtained for the rule sets of cascades and evaluated by simulation horizon profits. The authors use successive low dimensional local search to obtain near optimal solution, in which new values of only one variable are tested to find a better one at one step. A case study of three cascade hydropower systems located in China was used with a hypothetical market. The game model was solved using linear, quadratic, and cubic rule curves. Results showed the effect of different kinds of models on rule curves and the potential impact of market reformation to the operation of the cascaded hydropower reservoirs. For the studied cascaded reservoirs, the profit increasing percentages can be 2.6%–3.9% with 0.5%–2.0% energy loss, comparing the game model to the energy maximization model.

Noncooperative Game Theory Framework for Risk-Based Optimization for River Levee System

Game theory is also applied within risk-based optimization frameworks. For example, (Hui et al. 2016) used risk-based optimization for levee planning to minimize the expected annual total costs of annual damage cost (EAD) and annualized construction cost (ACC) for two arbitrary landholders in the Cosumnes River in California. Acting independently, each self-interested landholder would optimally determine the height of its own levee with the same risk-aversion. The system was solved in five different cases using iterative multiple-shot noncooperative planning and modeled different scenarios of reversible decisions, irreversible decisions, identical riverside conditions, and different riversides conditions.

Stochastic competitive analysis of hydropower and water supplies within an energy–water nexus

Hu et al (2018) used strategic competitive behaviors regarding energy-water linkage, with stochastic market scenarios to provide guidance for water-energy nexus policy.

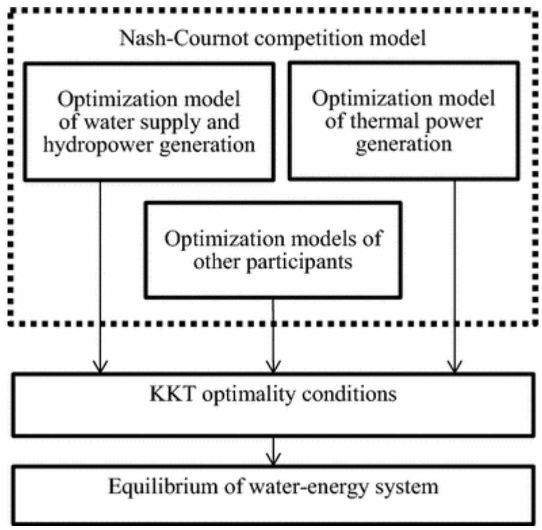


Figure 1: Flowchart of competitive analysis of hydropower and water supplies within an energy–water nexus (Hu et al 2018)

The authors built competitive Nash–Cournot models simulating interactive behaviors among participants in energy and water systems (Figure 1) (Hu et al. 2018). In the model, the profit-maximizing *n*-problem of hydropower generation and water supply is formulated. Next, the optimization problems of coal-fired and gas-fired power companies are established. The Karash–Kuhn–Tucker (KKT) optimality conditions are then derived for the Nash–Cournot competition models in energy and water systems. The results established a relationship between hydropower generation and carbon dioxide emissions and showed that a decrease in hydropower generation raises carbon dioxide emissions by approximately 6–10% for an 80% hydropower scenario, and by 12–19% for a 60% hydropower scenario in Tai-Chung and Chang- Hwa reservoir systems on the Dajia River in Taiwan.

The profit-maximizing problems in energy and water resource management assume a sequential reservoirs system, including at least one upstream and one downstream reservoir. The reservoir system generates hydropower and supplies water demand. Total power is sold by the reservoir system and other thermal power companies. Based on the inverse demand curve for the electricity market and the total power production, the equilibrium price and revenue in the market are obtained.

Stochastic Dynamic Game between Hydropower Plant and Thermal Power Plant in Smart Grid Networks

(Forouzandehmehr et al. 2016) considered a theoretical smart grid network with one pumped-storage hydropower plant and one thermal power plant as two price makers. They studied the competitive interactions between an autonomous pumped-storage hydropower plant and a thermal power plant in order to optimize power generation and storage. A stochastic dynamic game is formulated to characterize this competition and the instantaneous market price is modeled as a Cournot duopoly game. The solutions are derived using the stochastic Hamilton-Jacobi-Bellman[1] (HJB) equations but also show that the proposed game can converge to a

[1] HJB is the partial differential equation solved by the value function in continuous time; the corresponding discrete-time equation is usually referred to as the Bellman equation.

feedback Nash equilibrium. The proposed framework and games can reduce the peak-to-average ratio and total energy generation for the thermal plant, which helps power plant operation and reduces CO2 emission. The system and game model are defined in Figure 2 below:

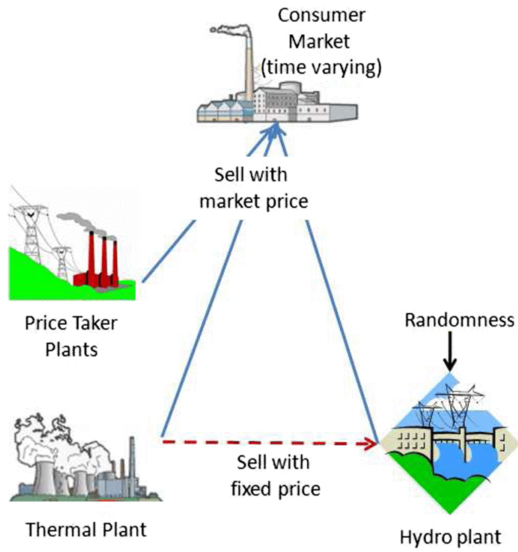


Figure 2: Hydro-thermal system and game model schematic (Forouzandehmehr et al. 2016)

The problem is decomposed into a thermal sub problem and a hydro sub problem that are solved in parallel through a constraint-relaxed iterative algorithm. In their proposed method, the optimal scheduling of the thermal and pumped-storage power plants constitutes a game because the output of each power plant affects the price, and therefore, affects both the payoffs of power plants. In order to reach an NE, each power plant chooses its output to maximize its payoff given the output (strategy) of other power plant (player). The instantaneous market price is considered as a Cournot model. In the proposed stochastic differential game, the dynamics come from the water volume in the reservoir, where the stochastic variable captures the natural inflow and natural loss (leakage) to the reservoir. The hydro plant decides how much power to produce, and the thermal plant decides how much to sell to the market or sell to the hydro plant for the pumped- storage plant. Here, the dynamics comes from the varying water volume in the reservoir, and the stochastic aspect captures the uncertain natural inflow to and losses from the reservoir. The solutions are derived using the stochastic Hamilton–Jacobi–Bellman equations.

The reservoir dynamic model was formulated as a linear differential equation of the reservoir volume as a function of discharge rates, stored thermal power, turbine and generator efficiencies, acceleration of gravity, water discharge rates, and leakage rates. The market prices were modeled as a spot price which is a function of the operating costs to meet the residual demand. Then, the hydro plant revenue is modeled and used in the stochastic game objective of the pumped-storage hydro to maximize the utility over a time interval by controlling the discharged water. Similarly, the stochastic game of the thermal plant player is to control the selling and storing of power to maximize its utility. In summary, the model has two value functions (hydro and thermal) that are in the linear quadratic forms of the controls (water discharge, thermal power sales).

Simulation results: they investigated the performance of the proposed game numerically. The simulations were performed for each hour in one-day duration, using different values of the constant K, which is the constant price for thermal player to sell to hydro. The amounts of generation and storages of both hydro and thermal followed the fluctuation of the demand function throughout the day. The payoff functions variations were correlated to the changes in plant output and market values, as well as a function of K.

Competitive multi-agent scheduling with an iterative selection rule

(Nicosia et al. 2018) studied a theoretical multi-agent (k agents) scheduling problem where the operator of a single machine iteratively selects the next task to be processed from a set of tasks submitted by the agents. The problem is a minimization problem of task completion time that makes use of central coordination mechanism that regulates access of agents' tasks to the machine by employing a decision rule, iteratively. In this paper, the goal is to design system-wide rules which, given the selfish decisions of the users, maximize the total social welfare. Each agent pursues the minimization of a given objective function, such as makespan, sum of completion times or sum of weighted completion times. Additionally, an external **coordination mechanism**, aiming at reaching a high throughput, or number of processed tasks per time unit, regulates access of agents' tasks to the machine. An independent coordinator employs a decision rule to manage the usage of the resource and thus determines the respective schedules on the shared system. Often, such a situation follows a global objective function which a central authority explicitly pursues, as far as possible under the actions of the agents.

However, in (Nicosia et al. 2018), no global objective function is manifestly taken into account. In the algorithm, each agent *i* wants to optimize its own objective function (sum of weighted completion times), which only depends on the completion times of its tasks. They say that an agent or task *wins* a round, if it is selected to be scheduled on the machine, otherwise it *loses*. Their results proved that the algorithm basically assigns blocks of tasks for agents as late as possible and then inserts the tasks of the last agent in shortest processing time (SPT) (first order, moving the reserved blocks to an earlier starting time if non-preemption would cause idle times.

The situation described in this paper could also be viewed in a game-theoretic setting, in which the algorithms induce *strategies* for the agents and each solution of the problem determines the corresponding agents' payoffs. Thus, they could also apply the concepts of ***extensive games*** with a ***decision tree***. In this framework, determining an *optimal strategy* of a single agent, assuming that also the other agents may follow a selfish optimal strategy, can be done by backward induction, but it requires exponential time in general. (Nicosia et al. 2018) noted that when dealing with *k*-agent scheduling, it would be interesting to identify polynomially solvable special cases, either by restricting the scheduling environment or the agents' strategies.

2.2.1 Modelling Advancement: Multi-Agent Inverse Reinforcement Learning

Game theoretic algorithms can perform better results in the reservoir coordination problem when the utility function representing the agent's stochastic payoffs is enhanced. In this case, a utility function can be learnt from previous policies. Inverse Reinforcement Learning (IRL) is a widely cited machine learning algorithm, with application in planning and operations. Inverse reinforcement learning (IRL) aims to determine the underlying reward function from an expert (or optimal agent) from its behavior data and its environment dynamics (Abbeel and Ng 2004; Ng and Russell 2000). As in (Bertsekas & Tsitsiklis, 1996), the goal of standard reinforcement learning is to find a policy such that the associated value function is maximized (reward function), for all states. IRL addresses the fundamental problem of finding the reward function in building a computational model for sequential decision making (Choi and Kim 2015; Lin et al. 2018). In IRL, it is generally assumed that the expert acts in an environment modeled as a Markov decision process (MDP). Under the MDP formalism, the IRL problem is defined as finding the reward function that the expert is optimizing given the behavior data of state-action histories and the environment model of state transition probabilities.

Since the publication of Ng and Russell on IRL, the framework has been extended to multi-agent setting in multi-agent inverse reinforcement learning (MIRL) (Dimitrakakis and Rothkopf 2012; Goldman and Zilberstein 2003; Lin et al. 2018; Reddy et al. 2012). For example, (Reddy et al.

2012) addressed the issue of using inverse reinforcement learning to learn the reward function in a multi agent general sum stochastic game setting, where the agents can either cooperate or be strictly non-cooperative. They derived the necessary conditions for the estimated reward to guarantee the optimality of the agents' policies and an iterative algorithm for estimating the agents reward function using IRL. The application of IRL in game-theoretic reservoir coordination problem is considered a novel approach.

2.2.2 Examples of Game Theory and Multi-Agent Reinforcement Learning

Due to the stochastic nature of inflows, there is a pressing need for a model that is adaptive to stochastic fluctuations in seasonal or daily inflows and does not make over-simplifying assumptions in hydropower planning. In Multi-Agent Reinforcement Learning (MARL), the game theory model of a Stochastic Game (SG) is adapted to model the multi-agent-environment (Bab and Brafman 2008). MARL deals with the problem of learning to behave well through trial and error interaction within a multi-agent dynamics environment when the environmental dynamic and the algorithms employed by the other agents are initially unknown. ***Evolutionary game theory*** (EGT) assumes that the game is played repeatedly by players randomly drawn from large populations, uninformed of the preferences of the opponent players (Tuyls and Nowé 2005). The basic properties of a Multiagent System corresponds exactly with that of EGT (Tuyls and Nowé 2005), this is why GT meets the concept of reinforcement learning (RL) in EGT.

A stochastic game can be a ***repeated game*** that consists of a set of agents, a set of states, a set of available actions, a function of transition probabilities and a function for reward (Figure 1).

In EGT, the question becomes not what strategy to adopt, but how a player can learn to optimize its behavior and maximize its return. It describes how agents can make decisions in complex environments where they interact with other agents. In such complex environments, software agents must be able to learn from their environment and adapt to its non-stationarity (Tuyls and Nowé 2005)

For multi agent reinforcement learning, several solution algorithms exist; however, the challenge remains in finding solutions suited for multi-stage cases. This is especially relevant for multi-agent reservoir systems with coordination mechanism to satisfy system level constraints, or the cooperation of different users and different water utilizations in a same river basin. For multi-agent systems characterized by distributed decision processes at the agent level with a coordination mechanism organizing the interactions among individual decision processes at the system level, (Yang et al. 2009) presented a decentralized (distributed) optimization method known as constraint-based reasoning, which allows individual agents in a multiagent system to optimize their behaviors over various alternatives. The method achieves optimization among agents with a bargaining scheme, in which the i^{th} agent optimizes its objective with a selected priority for collaboration and sends the solution back to all other agents with which it interacts. The method also uses a “central” processor which makes available to all agents in the next round of bargaining information on system cost and degree of constraint violation. (Giuliani and Castelletti 2013) proposed novel decision-analytic framework based on multi-agent systems to model and analyze different levels of cooperation and information exchange among multiple decision makers. The multi-agent operation models focused on multi-agent reservoir systems with coordination mechanism to satisfy system level constraints, or the cooperation of different users and different water utilizations in a same river basin.

(Tuyls and Nowé 2005) discussed the mathematical connection of evolutionary game theory with multi-agent reinforcement learning; a relationship that has been attracting attention of researchers from different fields such as economics, computer science, and artificial intelligence.

Optimizing Information Exchange in Cooperative Multi-Agent Systems

Cooperative agents are able to share information at the offline planning stage as if they were centrally controlled (Goldman and Zilberstein 2003). Decentralized cooperative MAS is the approach introduced by (Goldman and Zilberstein 2003) to solve the problem of information sharing in MAS such as computational processes distributed in an information space. Their paper focuses on agents that may need some of this information to get synchronized from time to time, but they cannot assume that communication is free and information can be exchanged at each moment. In their theoretical framework, they focus on a decentralized partially-observable Markov Decision Process with communication. Here, the two-agent MDP is defined with sets for: states, control actions, functions for reward, observation, transition probability, as well as the alphabet of messages and the cost of transmitting atomic message. In their model, "observability" or "joint synchronization" is only achieved by communication. They describe the interaction among the agents as a process in which agents perform an action, observe their environment, and then send a message that is instantaneously received by the other agent. In this framework, there exists a local policy for each agent to map the histories of observations and histories of messages received since the last synchronization and the last action. The resulting joint policy is defined to be a pair of local policies, one for each agent.

In this paper, the authors use a two-way communication with $\sum$ as the communication language (noting that the choice is up to the agent designer) and focus on informative messages. Other types of messages can define: commitments, reward/punishment, and world information. They experimented with their approach in a problem of two agents that have to meet at some location as early as possible, in a 2D grid environment. The case where agents can communicate their observations (i.e., their actual locations), incurring a cost , and compared the utility attained by both agents in the following four different scenarios: optimizing with no communication, ideal communication, sub-goal communication (a heuristic solution which assumes that the agents have a notion of sub-goals), and a greedy approach where agents optimize myopically at each communication. The greedy meta-level approach was able to produce near-optimal solutions in their example. In this approach, for each possible distance between the agents, a policy of communication is computed such that it stipulates when it is the best time to send that message. By iterating on this policy agents are able to communicate more than once and thus approximate the optimal solution to the decentralized control with communication problem.

A game theory–reinforcement learning (GT–RL) method to develop optimal operation policies for multi-operator reservoir systems

In this paper, (Madani and Hooshyar 2014) presented a numerical example of 3 single purpose reservoir systems in a cooperative game theory allocation paradigm and uses reinforcement learning (RL) algorithm. In simple reservoir operations, the RL agent can be considered as the operator who makes release decisions; the set of discrete reservoir storage levels can be considered as the environment. The objective of learning in this case would be finding the best release strategy for a given level of storage with respect to the operation objective(s) (e.g., maximizing the hydropower generation profits, minimizing expected flood costs, minimizing water shortage costs, maximizing recreational benefits) and constraints (e.g., upper and lower storage/release levels, maximum ramping rate, maximum temperature).

The Madani and Hooshyar 2014 GT-RL method has three major steps. First, the obtained benefits are extensively determined under each possible partial coalition and total (grand) coalitions. Here, the conventional social planner approach (grand coalition) using RL was used where they calculated immediate rewards, then updated RL- Q factors, and updated the policy in a 12 month, 100 years, 100 episodes loop. Then, they evaluated policy performance and plotted the number of learning episodes versus the average annual revenue (learning curve). For the partial coalitions obtainable benefits, only 1 coalition at a time is possible (2 and 1) when solved in multi-agent mode (an agent for the coalition and an agent for the single reservoir). The second step is to apply different cooperative game theory solution methods (Nash-Harsanyi, Shapley, and Nucleolus) to find fair and efficient allocations of the incremental benefits of cooperation among the agents based on different defined notions of fairness. Finally**,** the stability of each cooperative allocation solution is examined to find the allocation solution with the highest acceptance potential.

The GT approach to reservoir management in this case demonstrated that the social-planner approach, or the grand coalition, were the operator maximizes the total benefit of the system, is not a stable one. This is because the individual payoffs of the reservoirs can be maximized in other ways, such as partial coalitions, or by non-cooperative behavior, and would be higher that there individual benefit in the coalition, bearing in mind that this comes at the expense of other agents' benefit. Therefore, in a scenario like that, individual hydropower agents do not have an economic incentive to partake and stay in a grand coalition. The authors concluded that Nash-Harsanyi is most suitable solution method for the allocation in their game.

Another observation that was not mentioned in the paper, was that the non-cooperative behavior of a single reservoir, sometimes, but not all the times, grants the agent better benefits than the grand coalition or partial coalition. The framework presented in this paper was applied to reservoirs with identical objectives and deterministic inflows. It would be interesting to see how adding stochasticity to this framework changes the outcomes.

2.3 Managing the BC Hydro System

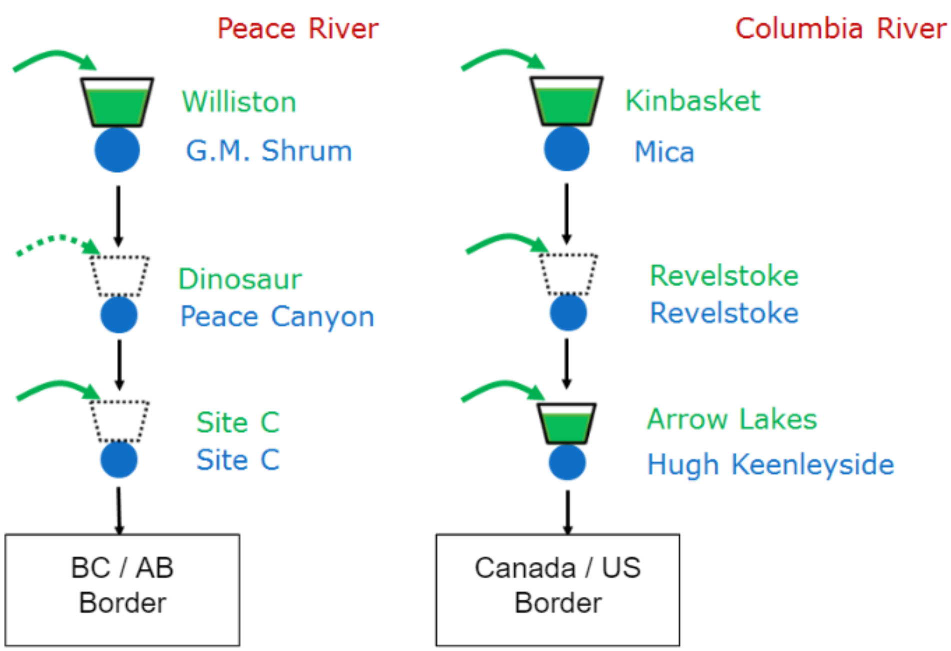


Figure 3: Schematic of the Peace and Columbia River Reservoirs and Generation Stations

2.3.1 BC Hydro System Overview

BC Hydro generate over 43,000 gigawatt hours of electricity annually, through 100+ generating units, to supply more than 1.9 million residential, commercial and industrial customers. Over 75% of BC Hydro's installed generating capacity is at hydroelectric generation stations in the Peace and Columbia river basins (“Energy in BC” 2020). This research is concerned with modelling these two systems.

2.3.1.1 Peace River

The Peace system on the Peace River in northeastern B.C. The biggest facility is the GM Shrum at the Williston Reservoir, with a capacity of 2,730 MW. The other facilities on the Peace system are Peace Canyon at the Dinosaur Reservoir, and Site C Dam, which is still under construction.

2.3.1.2 Columbia River

The total capacity of the Columbia system is 58% of BC Hydro's total capacity. Mica Generation Station at the Kinbasket Reservoir alone, has a capacity of 2,746 MW. Revelstoke Generation Station and Hugh Keenleyside Dam at the Arrow lake are also part of the Columbia system.

These reservoir release policies need to comply with the Columbia River Treaty (CRT) and subsequent agreements.

2.3.1.2.1 Columbia River Treaty

The Columbia river treaty is an international treaty between the United States (US) and Canada with the explicit intent to coordinate flood control and optimize power generation within the Columbia basin, which spans both countries. The treaty subsequent agreements manages environmental concerns such as desirable river flows and fish migration.

Objectives for Supplemental Operating Agreements (SOAs)

Power objectives include minimizing spill and optimizing energy production and power value in Canada and the US.

Operations for power objectives may be combined with non-power objectives (Columbia River Treaty Operating Committee 2019). When appropriate, the Operating Committee will make suitable arrangements for delivery of power relating to sharing of benefits from operational agreements.

Table 2: Columbia River Treaty Minimum and Maximum Flow Limits at the Arrow Reservoir (Columbia River Treaty Operating Committee 2019)

	Minimum Outflow in cfs	Min outflow in cms	Max outflow Limit in cfs	Max outflow in cms
January	10000	283.16847	70000	1982.17929
Feb	10000	283.16847	60000	1699.01082
March	10000	283.16847	-	-
15-Apr	10000	283.16847	-	-
30-Apr	10000	283.16847	-	-
May	5000	141.584235	-	-
Jun	5000	141.584235	-	-
July	10000	283.16847	-	-
Aug 15-Dec	10000	283.16847		

Table 3: Columbia River Treaty Canadian Storage Accounts (Columbia River Treaty Operating Committee 2019)

Canadian Storage Accounts	Million Acre Feet	Million Cubic Meter
Duncan Reservoir	1.4	1726.9
Arrow Reservoir	7.1	8757.7
Mica Reservoir	7	8634.4
Total	15.5	19118.9

2.3.2 Modelling Framework: Energy Studies

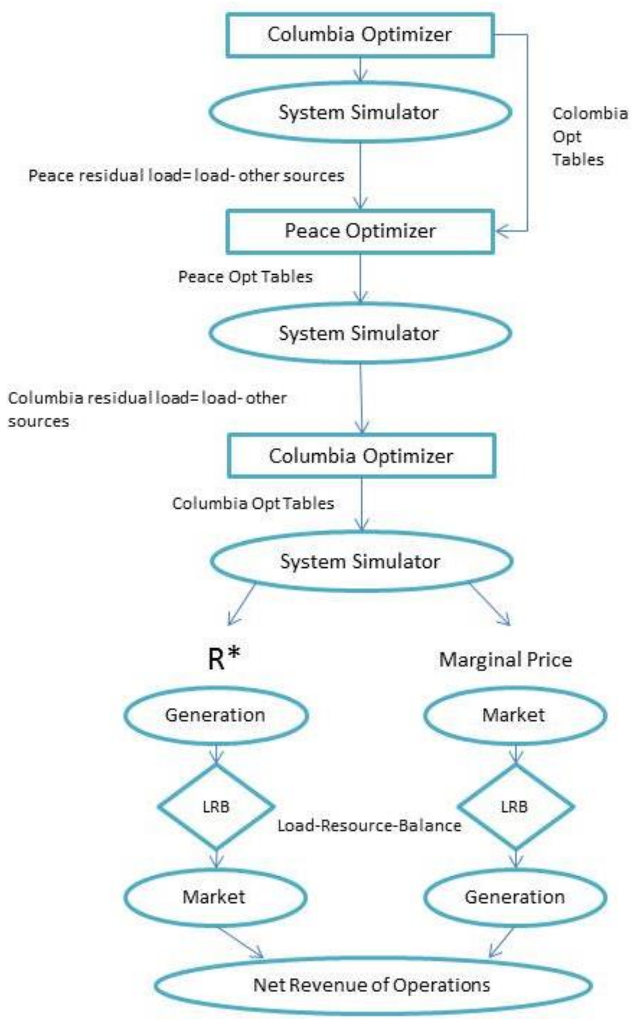
In operations planning, BC Hydro makes a decision regarding reservoir releases to (1) meet forecasted domestic electricity demand; and (2) maximize total benefits from energy trade in electricity markets and the value of water stored in reservoirs for a multiyear planning horizon. Reservoir release policies need to hedge against stochastic inflows in a snowmelt-dominated environment to achieve the best benefit, in particular, reducing spill in wet years and saving water for power generation in dry years (Guan et al. 2018).

BC Hydro uses an Energy Study to optimize the dispatch of BC Hydro's resources. It also identifies water inflows, market prices and loads as the primary variable inputs to the Energy Study influencing the Cost of Energy (*BC Hydro Rate Application 2016*). The Energy Study models (ESM) are used by BC Hydro to inform operational decisions on system storage operations, thermal dispatch, and purchases and sales of market electricity. The same models are used for BC Hydro's ongoing financial forecasting of the Cost of Energy.

Peace and Columbia River systems are each optimized using scenario SDP where historic inflow ensembles are used and reservoir levels are discretized into stages to find optimal decisions. The optimization model is called the Marginal Cost Model. The planning period for Williston goes from the beginning of every month in the current calendar year and 4 years into the future, while the planning period for Columbia system has additional months and goes to an extra year due to treaty considerations.

The system simulator for ESM currently plays the role of coordinating between the optimizations of Peace and Columbia. This happens through a series of iterations where in each iteration, the optimizer for each system at once fulfills the demand to its capacity based on the release decisions (R*) and marginal price, then results in a residual load (Figure 4). The residual load is then fulfilled by the optimization model of the other system through the same process and outputs. To reconcile the difference in planning period months between the two systems, the simulator only run until December 2024. The system simulator double checks the outputs of the optimizers and using the load-resource balance equation, calculates the breakdown between system generation and market net purchases. Subsequently, the system simulator determines the net revenue of operations (NRO). It is important to note that the Bellman equation that is being solved is bound by the gas state variable that controls the market price for electricity (in GJ). The prices derived from the gas price-duration curves (that are load-based) inform the decisions to buy (when prices are low) and sell (when prices are high). This process is also built into each system optimizer to maximize the revenue while serving load. The optimizers work at a monthly time-step and provide an optimal “average” value in cubic meters per second (cms). The simulator takes average monthly discharge and adjusts for daily decisions and market activity.

The model converges at a tolerance value (difference between iterations) of almost 3 million CAD.



In summary, the system simulator uses an "optimal policy approach" where at a given elevation, the optimal policy is obtained from the optimizers, then, the energy generation is calculated, then the residual load, then decision about market prices are made (buy, sell, do nothing). A second approach "marginal cost approach" is being contemplated by the ESM team where at a given elevation, the marginal price is derived from the value function, then market activity is decided, then the necessary reservoir generation for the residual load is calculated, leading to the final reservoir operations forecast (GSO, 2020).

Figure 4: Schematic of Energy Studies Modelling

2.3.2.1 Trading Decisions

Today, a large share of electricity is sold in wholesale in electricity markets. Like generating companies with pumped hydro storage capacities, BC Hydro use their reservoir storage capacity and market trade to buy electricity at the market when the price is low and sell electricity when the price is high, while trying to mitigate the risk of positive and negative imbalances (Löhndorf et al. 2013).

Hydro storage plants face some challenges associated with trading due to:

- Day ahead-prices uncertainty and the uncertainty of the development of electricity prices over time, and uncertainty of water inflow into the reservoir
- A system of hydro storage plants with multiple reservoirs requires a coordinated water release policy, since upstream releases influence reservoir levels. In addition to the day ahead bidding decisions, a generating company must decide about water releases from multiple reservoirs at a time. In the decision-making process, they consider future decisions and states of the system as well as their probabilities.

Those challenges to hydro generating companies in general, and BC Hydro in particular, present a need to solve a multi-stage stochastic programming problem to optimize the market bidding and storage decisions with multiple storage units and several price scenarios.

2.3.2.2 Simulation and Optimization Models – SSDP

BC Hydro strives to develop tools using stochastic optimization techniques to generate reservoir release policies that take into account stochastic inflows, prices, and load with proper modeling of the CRT and subsequent agreements for operations planning. The models in current use are either deterministic, e.g., the hydro simulation model (HYSIM), generalized optimization model (GOM) (Fane 2003), CRT optimization model (CRTOM), and short-term optimization model (STOM) (Shawwash et al. 2000) or they are stochastic only for the GMS facilities, e.g., models that evolved from the stochastic dynamic programming (SDP) model by Druce (1990). More research work on applying stochastic optimization techniques to BC Hydro's system includes studies by Abdalla (2007), who used reinforcement learning (RL) (Barto and Sutton 1998); Shabani (2009), who used RL; Guan et al. (2013), who used stochastic dual dynamic programming (SDDP) (Pereira and Pinto 1991; Goor et al. 2011; Poorsepahy-Samian et al. 2016); and Mamun et al. (2015), who used goal programming (Charnes et al. 1955).

2.4 Uncertainty

Uncertainty in simulating water resource systems makes it difficult to assess how effective different water management decisions will be (Dobson et al. 2019). Different types of uncertainties can undermine the credibility of simulation and optimization models of reservoir operations.

A common conceptual classification of uncertainties affecting simulation models distinguishes between *aleatory* and *epistemic* uncertainty (Beven et al. 2015). One of the most read Water Research papers to date is (Cosgrove and Loucks 2015), which discussed non-stationarity in water supply and demand. Practically unavoidable in WRS is the variability in hydrological forcing, such as inflows into reservoirs. A common practice in the field is to assume that inflows are aleatory and stationary processes (Dobson et al. 2019), although the validity of the stationarity assumption is highly debated. Reservoir operation is then stochastically optimized under this statistical model, for example, via Stochastic Dynamic Programming. Similar considerations apply to other system variables that can be regarded as aleatory uncertainties, such as water demand and evaporation from reservoir surfaces, which are often modeled using similar statistical models to inflows. According to (Cosgrove and Loucks 2015), evidence to date suggests we will be observing more variability, resulting in more frequent floods and droughts of greater intensity and duration. At the same time, demand for water for agriculture and energy production in particular will be influenced by climate change, technological development and urbanization and human responses.

Another source of *epistemic* uncertainty is the choice and formulation of model outcomes such as, in the case of WRS optimization, the metrics of system performance (or objectives hereafter). For example, for the inflow, demand, evaporation, and fisheries releases, the binary choice is between a more sophisticated or less sophisticated representation of the process. Another important modeling choice is how to represent cooperation between the two companies when making release or pumping decisions. Previous literature demonstrated that model assumptions about coordination between connected reservoirs can dramatically impact the performance of optimized reservoir operations (Giuliani and Castelletti 2013; Tilmant and Kinzelbach 2012; Wu et al. 2019). In (Dobson et al. 2019), it was found that the assumption about the level of cooperation between water agents has a greater impact on estimating objective values than any other modeling choice.

2.5 Inflows and Climate Change in BC

Water inflows are very important inputs to the optimization problem and drivers of uncertainty in the model. Several studies have been concerned with analyzing trends and studying the variations in streamflow for the Pacific Northwest, in which changes in annual total, and summer mean stream flows were attributed to changing climate and variability.

(Barnett et al. 2005) predicted that by 2050, the Columbia River system would not be able to sustain historical levels of both spring and autumn water releases for hydropower generation and releases for spring and summer salmon runs. (Forbes et al. 2019) conducted trend, detection, and attribution analyses using naturalized streamflow observations and routed land surface model runoff for 10 sub-basins in the Columbia River Basin (CRB) during water years 1951–2008. In their study, all sub-basins showed significant declines in the observed amount of annual total streamflow, except for the Middle and Upper Snake and Upper Columbia Sub-basins.

Furthermore, when analyzing the distribution of the monthly flows at The Dalles, (Forbes et al. 2019) found that even though June had the greatest median flow at The Dalles, it had the greatest significant decreasing trend for the period and was the start of significant decreases for the entire June–October season. Previous studies have also analyzed the trends in streamflow for the Pacific Northwest including the CRB. (Luce and Holden 2009) showed decreases in annual mean flow over the period 1948–2006 and (Stewart et al. 2005) showed similar changes in the fraction of annual total flow and consistent decline in the fraction of annual total flow in June. On average, the annual total streamflow for the CRB decreased by approximately 15% between 1951 and 2008 (i.e., the percentage of change in volume over time). Of that 15%, roughly 77% was during the June–October months with 40% solely in June (peak flow) and 31% in July–September (summer mean). More specifically, flow in June has declined by 28% on average. The fact that these declines are in five consecutive months during the year is particularly worrisome for all the inhabitants and the natural ecosystem of this region. On average, these five months provided 49% of the annual total flow with June providing 22% itself.

Reservoir release policies need to hedge against stochastic inflows in a snowmelt-dominated environment to achieve the best benefit, in particular, reducing spill in wet years and saving water for power generation in dry years. Inflows come mainly from snowmelt during the freshet period extending from mid-April to mid-August, with very large variations, and the inflow volume is the primary contributor to the annual inflow volume (Guan et al. 2018). The forecast of this inflow volume (seasonal volume forecast) is heavily used in operations planning.

Therefore, the treatment of inflows variation in short- and long-term optimization and planning can be very critical to the model performance.

2.6 Summary

In the literature review, we have seen models that cover topics related to:

- incorporating the competitive and cooperative mechanisms in theoretical optimization problems through iteration (Chi-Keong Goh and Kay Chen Tan 2009)
- using modified benders decomposition algorithm (MBDA) to solve equilibrium problems with equilibrium constraints in hydro-dominated systems (Moiseeva and Hesamzadeh 2018)
- theoretical competitive multi-agent scheduling through the minimization problem of task completion time by a central coordination mechanism (Nicosia et al. 2018)
- game theory–reinforcement learning (GT–RL) methods to develop optimal operation policies for multi-operator reservoir systems but with deterministic inflows (Madani and Hooshyar 2014)

However, as explained earlier, in a scenario of predominant dependence on hydro generation such as in British Columbia, Canada, the coordination of hydropower generation involves the coordination of plants that might not be connected hydraulically or that are owned by different agents (Faria et al. 2018).

Managing the hydro dominated system in BC requires coordinating the operations of the Peace and Columbia systems that represent 29% and 58% of BC Hydro's total capacity respectively. This is not only the case in British Columbia, in other regions with hydro-dominated systems of energy it is common to face the issue of coordinating multiple systems of hydraulically independent reservoirs. For example, Hydro-Québec Production owns and operates 61 generations plants over 26 major reservoirs, located on 13 watersheds (Hydro-Québec (Montréal) 2009). Norway has 1660 hydropower plants with the 30 largest reservoirs providing about half the storage capacity. This allows for the split of electricity production into two categories, flexible and intermittent. Flexible hydropower plants can produce electricity even in periods when there is little precipitation and inflow is low while the large available reservoir storage capacity makes it possible to even out production over the years, seasons, weeks and days, within the constraints set by the license and the watercourse itself (Energifakta Norge, 2015).

In BC, the coordination of the multi-agent system of reservoirs operations informs the operations decisions for each system on: reservoir releases, electricity generation, spills, and storage levels, while taking into account the operations decisions for the other system simultaneously. In other words, it is solving the Bellman equation for both systems simultaneously, finding the optimal releases for both systems while taking into account the releases of the other. At BC Hydro, the coordination process at the planning stage happens through simulating the systems while assuming a specific capacity of the other the system and iterating to convergence. However, this process is time-consuming and can be improved for better results.

The literature shows a promising role for Game Theory to help coordinate systems where cooperation and competition are happening among its agents. However, in the literature, there are no practical and realistic applications that use game theory to develop optimal operation policies for multi-operator reservoir systems but with stochastic inflows and market prices. To the understanding of the author, up to the publication of this thesis, there were no models that test the effectiveness of GT algorithms in solving multi-agent reservoir problems.

Therefore, this thesis addresses the gaps of:

A. Coordinating multi-agent reservoir systems with stochastic inflows in a Game environment
B. Verification and testing of Game-Theoretic algorithms in finding equilibrium

The work in this thesis fulfills three main objectives:

a. Inform operational decisions for Columbia and Peace rivers system, on storage operations, and purchases and sales of market electricity.
b. Build a preliminary coordination model for the Columbia and Peace reservoirs based on Game Theoretic techniques.
c. Propose a benchmarking framework to test and compare the full model to other optimization and coordination stochastic models.

Chapter 3: Methods for Equilibrium-Based Decision-Making Model (EDM)

3.1 Modelling Concept and Approach

Multi-agent systems, such as BC Hydro's reservoir system, are conventionally optimized for the maximization to the entire system's benefit. This is referred to as "social planning". Game-theoretic allocations, on the other hand, can help us find alternative solutions to maximize individual agent's (river systems or hydro generation stations) in more realistic scenarios accounting for market and demand variables, or maintenance periods. In the literature review in Chapter 2, models such as (Madani and Hooshyar 2014) demonstrated the optimization of a multi-agent system in several ways: cooperation, partial cooperation, and non-cooperation. A typical game model can span various solution strategies, such as to cooperate or be greedy. Different strategies define different objective functions formulations, because each strategy defines a different class of games and solution.

This section explains the methods used in a novel model to coordinate the decisions for a multi-agent hydropower system using Game Theory. The Equilibrium-based Decision-Making Model (EDM), uses a modified General Sum Stochastic Game approach that shifts modeling from a single agent simulation-optimization to two-agent system characterized by a co-dependent decision process at the agent level with a coordination mechanism organizing the interactions among the individual decision processes at the system level. Figure 5 shows the overarching modelling concept that guided the design of the model. The model (EDM) acts as a coordination step after the optimal releases and value functions for each agent are determined by the Marginal Cost Model.

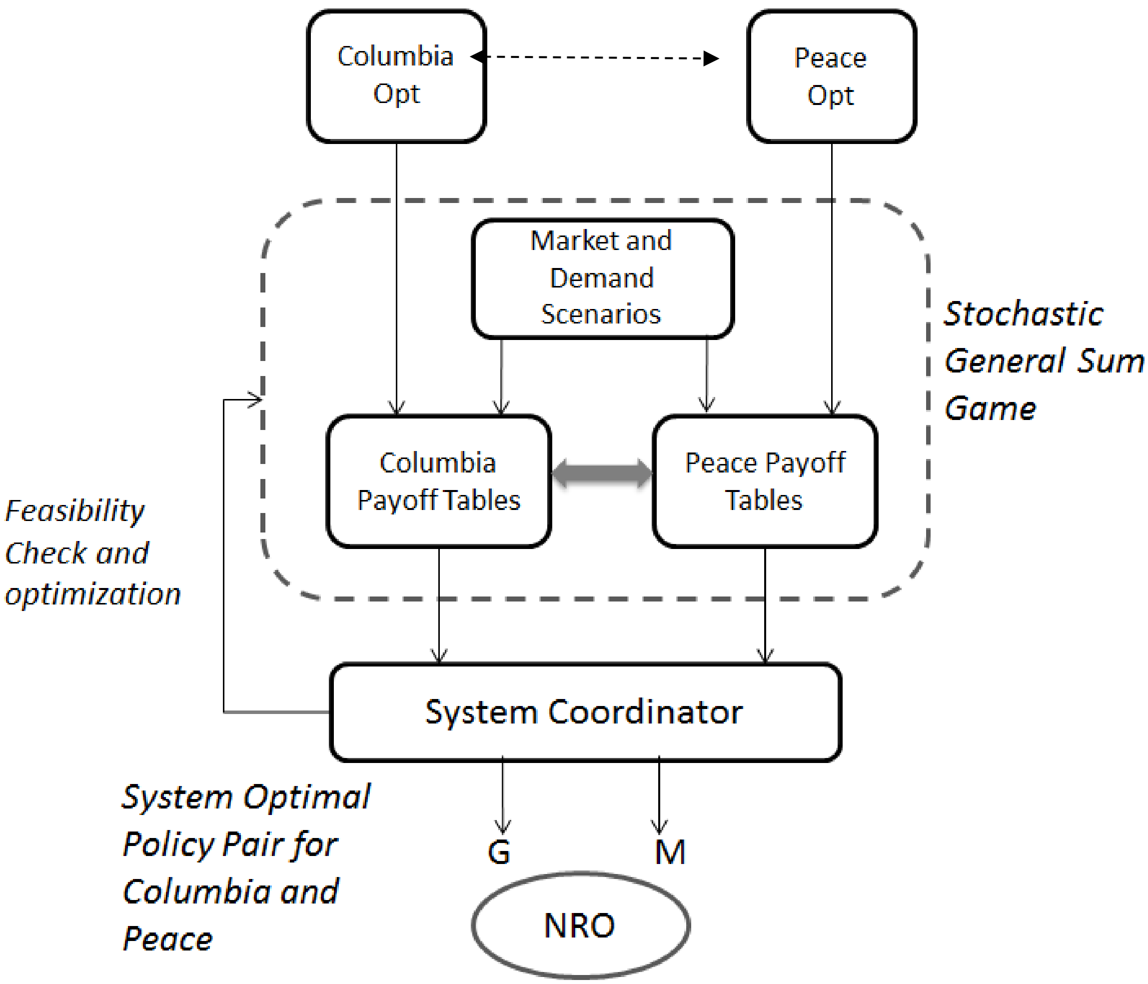


Figure 5: Modelling Concept for the Coordination for Peace and Columbia Reservoir System. Columbia Opt and Peace Opt represent the individual agent's optimizations that result from the Marginal Cost Model.

3.2 Methods for Finding Equilibrium

Nash equilibrium (Nash 1950) is the most central solution concept for games. It defines how rational agents should act in settings where an agent's best strategy may depend on what another agent does, and vice versa. For the concept to be operational, it needs to be accompanied by an algorithm for finding an equilibrium (Sandholm et al. 2005). There are many computational complexities associated with computing a Nash equilibrium; these complexities and a handful of algorithms to overcome them are widely cited in the computer science literature. In this model, the game equilibrium is found by a social optimum method, and a Nash equilibrium method. The social optimum **(SO)** equilibrium is obtained by maximizing the rewards for the system through a single reward function. The Nash equilibrium solution is found by two algorithms: support enumeration algorithm (Avis et al. 2010; Roughgarden 2010) and a mixed-integer linear programming algorithm (Sandholm et al. 2005), these algorithms were chosen based on a ease of application, accessible solvers, and speed benchmarks cited in the literature.

3.2.1 Definition of Nash Equilibrium

A Multi-Reservoir General-sum discounted stochastic game (GSSG) is defined as a tuple $\langle S,N,A,P,R,\gamma\rangle$, where:

- N is a finite set of n agents (reservoirs).
- S is a finite set of M states (elevations)
- A_i is a set of actions (generation, buying, selling) available to agent i, and a vector $(a_1,a_2,\dots,a_n)^T \in \times_i A_i$ represents the joint action of all the agents.
- $P: S\times A\times S'\rightarrow R$ is the state transition probability function. $P(s'|s,a)$ denotes the probability of transitioning from state s to state s′ by taking joint action a.
- $R= r_1, \dots, r_n$ where $r_i: S\times A\rightarrow R$ is the reward function or payoff value for agent i for taking a specific action.
- $\gamma \in [0,1]$ is the discount factor.

In this definition, iand A_i are assumed to be finite. A policy is defined as $\pi: S\times A\rightarrow R$, where A is the joint action set of all agents. In GSSGs the agents have full observability of the state of the world. In game theory, a *strategy* is generally defined as the action that an agent takes in a given state. In case of a *pure strategy*, there is one action a_i that the agent always prefers taking in a state *s*, but, in the case of a *mixed strategy* an agent has a probability distribution over the set of actions he can take in a given state.

A strategy profile $ST=(ST_1,\ldots,ST_n)$ is a Nash equilibrium if, for all agents *i*, ST_i is a best response to ST_{-i}, where, ST_{-i} are the strategies of all agents not including *i*.

3.2.2 Nash Equilibrium method: Support Enumeration (Nash)

The packages QuantEcon.jl and Games.jl (Sargent and Stachurski 2019)were used to compute all mixed Nash equilibria of a 2-player (non-degenerate) normal form game using a brute force support enumeration algorithm. The algorithm checks all the equal-size support pairs; if the players have the same number *n* of actions, there are *2n* choose *n-1* such pairs. This should only be used for small games.

3.2.3 Mixed Integer Linear Programming (MILP)

The mixed integer program formulation from (Sandholm et al. 2005) was used as another way to find the Nash equilibrium of the game at each time step. The formulation is said to outperform both the Lemke-Howson and the Support Enumeration algorithm in solving complex games.

A mixed integer program is a linear program in which some of the variables are constrained to be integers. This algorithm is based on penalizing regret, which is defined as follows:

The regret of pure strategy *si* is the difference in utility for player *i* between playing an optimal strategy (given the other player's mixed strategy) and playing *si*.

For every pure strategy s_i, there is a binary variable b_{si}. If this variable is set to 1, the probability placed on the strategy must be 0. If it is set to 0, the strategy is allowed to be in the support, but the regret of the strategy must be 0. The formulation has the following variables other than the b_{si}. For each player, there is a variable u_i indicating the highest possible expected utility that that player can obtain given the other player's mixed strategy. For every pure strategy s_i, there is a variable p_{si} indicating the probability placed on that strategy, a variable u_{si} indicating the expected utility of playing that strategy (given the other player's mixed strategy), and a variable r_{si} indicating the regret of playing s=. The constant U_i indicates the maximum difference between two utilities in the game for player i. The formulation follows below:

$$
\begin{array}{llcll}
\multicolumn{5}{l}{\textbf{find } p_{s_i}, u_i, u_{s_i}, r_{s_i}, b_{s_i} \textbf{ such that}} \\
(\forall i) & \displaystyle\sum_{s_i \in S_i} p_{s_i} & = & 1 & (1) \\
(\forall i)(\forall s_i \in S_i) & u_{s_i} & = & \displaystyle\sum_{s_{1-i} \in S_{1-i}} p_{s_{1-i}} u_i(s_i, s_{1-i}) & (2) \\
(\forall i)(\forall s_i \in S_i) & u_i & \geq & u_{s_i} & (3) \\
(\forall i)(\forall s_i \in S_i) & r_{s_i} & = & u_i - u_{s_i} & (4) \\
(\forall i)(\forall s_i \in S_i) & p_{s_i} & \leq & 1 - b_{s_i} & (5) \\
(\forall i)(\forall s_i \in S_i) & r_{s_i} & \leq & U_i b_{s_i} & (6) \\
\multicolumn{5}{l}{\textbf{domains: } p_{s_i} \geq 0, u_i \geq 0, u_{s_i} \geq 0, r_{s_i} \geq 0, b_{s_i} \in \{0,1\}.}
\end{array}
$$

Figure 6: Formulation of the mixed integer program for finding Nash Equilibrium

The first four constraints ensure that the p_{si} values constitute a valid probability distribution and define the regret of a strategy. Constraint (5) ensures that b_{si} can be set to 1 only when no probability is placed on s_i. On the other hand, Constraint 6 ensures that the regret of a strategy equals 0, unless $b_{si} = 1$, in which case the constraint is vacuous because the regret can never exceed *Ui*. Technically, Constraint 3 is redundant as it follows from Constraint 4 and r_{si}¸ 0.

This algorithm was formulated in Julia using the optimization package JuMP, and Cbc solver, all of which are open-source software.

3.3 Reward Calculation

A key step in the model algorithm is creating payoff tables that define the payoffs (or payoff functions) for strategy pairs of the two agents, that lead to the determination of a Nash Equilibrium. In this model, two reward equations were formed to capture the costs and benefits of a strategy, with respect to the other agent's strategy. For the social optimum (SO) algorithm, the reward is calculated using one equation for the entire system (Equation 1). For the Nash equilibrium (Nash and MILP), the reward was calculated separately for each agent, given the other agent's policy (Equation 2).

Equation 1: Social Reward Equation

$$\sum\left(-\boldsymbol{Val}(\boldsymbol{elevation}_{\boldsymbol{t+1}})_{\boldsymbol{i}} + \boldsymbol{Market\ Price}_{\boldsymbol{t}} * (\boldsymbol{Sales} - \boldsymbol{Purchases})_{\boldsymbol{i,t}} - \boldsymbol{Price} * (\boldsymbol{other\ sources})\right) + \boldsymbol{FR}$$

Equation 2: Competitive Reward Equation

$$-\boldsymbol{Val}(\boldsymbol{elevation}_{\boldsymbol{t+1}})_{\boldsymbol{i}} + \boldsymbol{Market\ Price}_{\boldsymbol{t}} * (\boldsymbol{Sales} - \boldsymbol{Purchases})_{\boldsymbol{i,t}} - \boldsymbol{Price} * (\boldsymbol{other\ sources}) + \boldsymbol{FR}$$

The term *Val* in each equation represents the value of storage reached by each policy for each agent, based on the inflows on the scenario year in consideration. The value function is obtained from the Marginal Cost Model. The term *FR* represents the future rewards, which is the model's way of incorporating the occurrence of all other water inflow scenarios when calculating the discounted expected future reward (Equation 3).

Equation 3: Expected value of the future reward of each agent, where ∂^t is the discount factor, and $p_{i|j}$ is the transition probability from scenario *j* to *i*, multiplied by the value function f_{yi}, *y is the ending elevation and w is the water year* .

$$\partial^t * \sum_{j=1}^{j=W} p_{i|j} * f_{yi}$$

The calculation of the future reward is represented in the schematic of Figure 7. The value of the reward is maximized at each time step independent of the following time steps, but the possibilities of occurring weather scenarios for step *t+1* are captured through the future reward function.

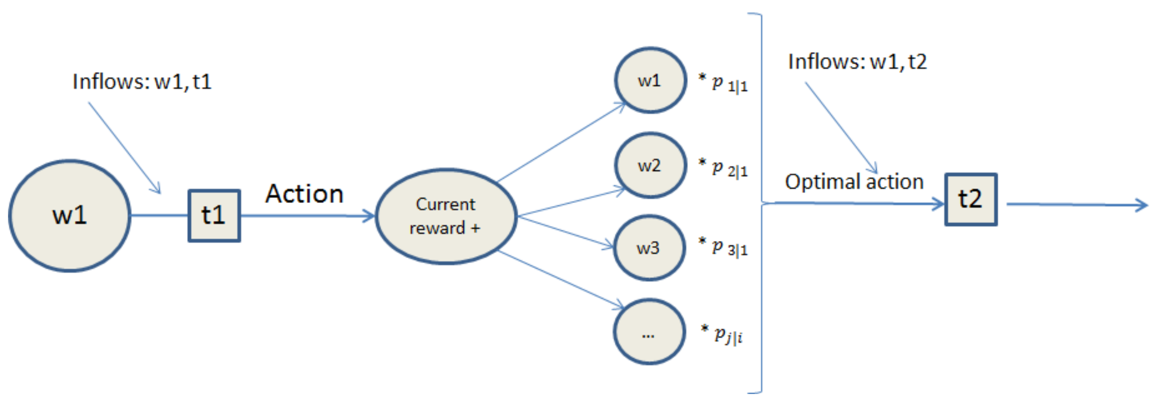


Figure 7: Calculating the expected discounted future reward at each time step in the model, where *w1* represent scenario year 1 for water inflows.

3.4 Final Model Algorithm and Schematic

The model, with its three solution branches (S.O, Nash, and MILP), is divided into three main subroutines. The first subroutine generates the game environment based on the given market state and constraints. The game environment allocates each agent's policies. Then, the reward matrix is generated through the reward calculation subroutine. Lastly, the Nash equilibrium is determined, using the appropriate reward functions, and the optimal policy pair is obtained. The model uses a sequential decision making structure of the reservoir system optimization problem by decomposing the original problem into subproblems that are solved sequentially over each stage (i.e time period).

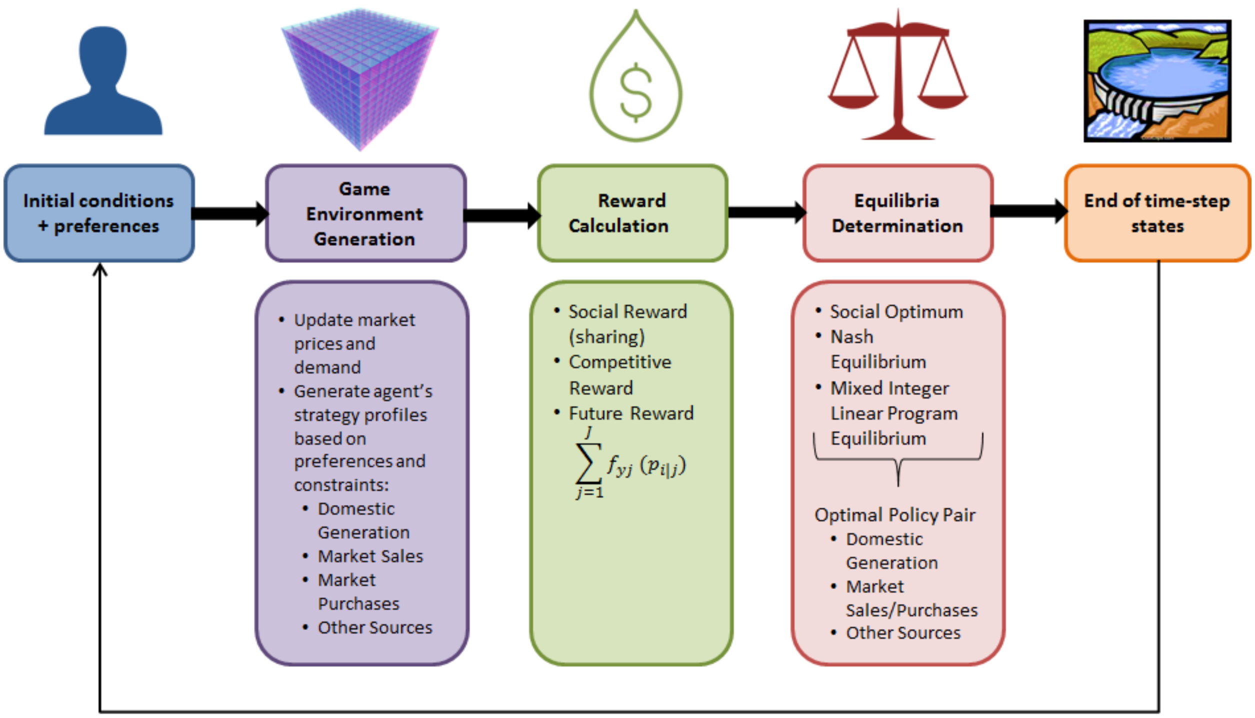


Figure 8: Flowchart of the equilibrium model subroutines

3.4.1 Model Inputs

The model input constitutes of the preferences and data, the preferences are as follows:

- Game size: length of policy vectors for each agent
- Exploration bounds around the optimal release strategy

- Market price upper and lower threshold

The model input data are as follows:

- Discount factor
- Domestic demand
- Inflows
- Mid C electricity prices
- Market price elasticity curve
- Output of the MCM: Value function and optimal or target releases for discretized elevation levels.
- Transition probabilities for each agent
- Upper and lower bounds on elevations and releases for each agent
- Storage to elevation curves
- Releases of dependent agents (such as Arrow for Columbia System)

3.4.2 Model Outputs

After the simulation of the entire planning period is done, and equilibrium at each step is determined, the model outputs the following:

- Agent's policy for domestic generation at each time step
- Agent's policy for market sales and purchases at each time step
- Agent's elevation at each time step
- Agent's rewards
- Other sources used to meet the demand

3.4.3 Game Generation Step

The game generation step or subroutine of the model utilizes logical decision-making for trading that was decided on with the BC Hydro energy Studies group. The model starts by comparing the market price to a given threshold, if the market price is above the threshold, then the model constructs the agent's policies to generate hydropower to meet the demand and sell to the electricity market. If the market price is below the threshold, then the model constructs the agent's policies to purchase electricity from the market to meet domestic demand. If the market price is between these thresholds, then the model does not construct any buying or selling policies.

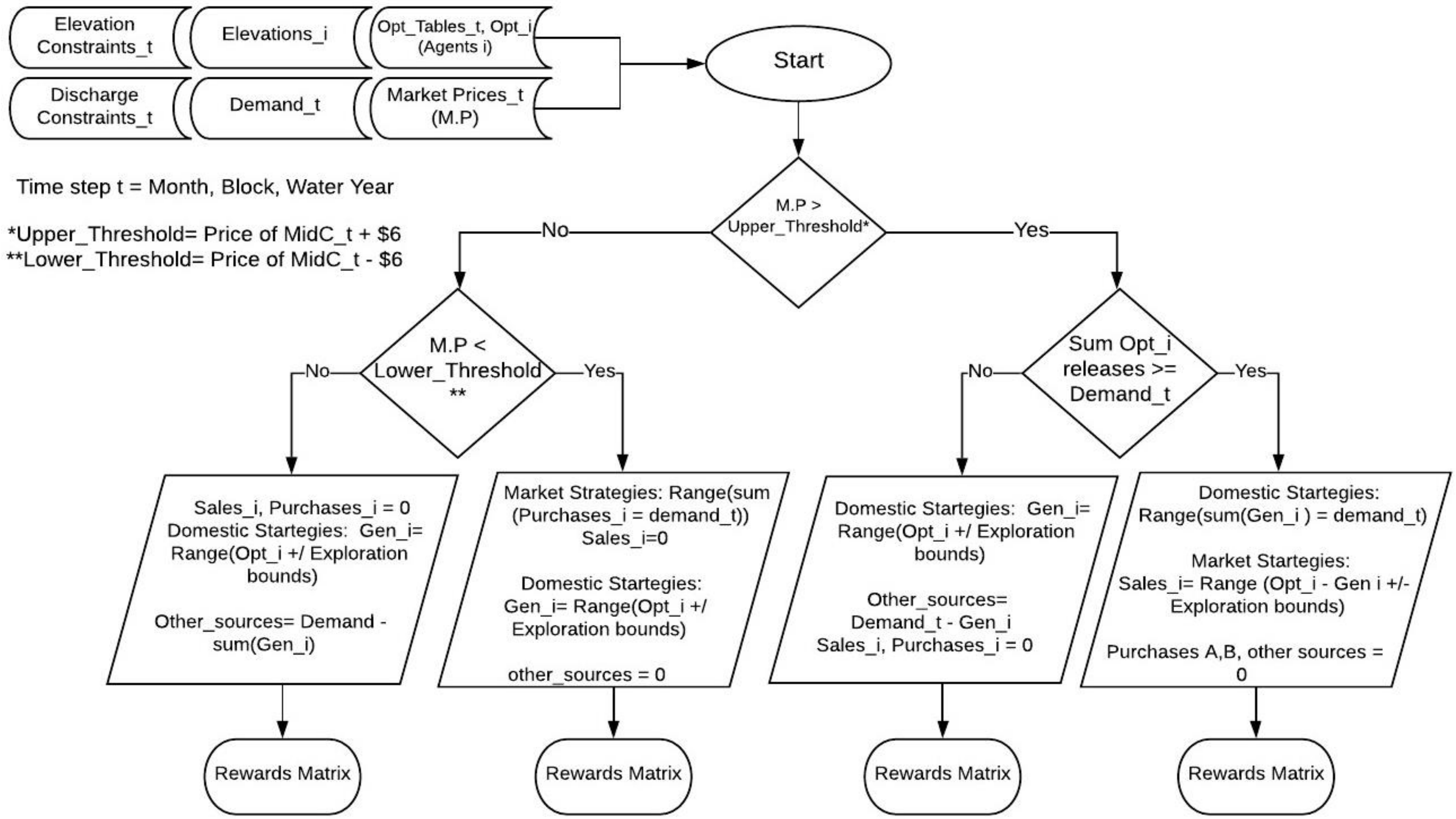


Figure 9: Flowchart of the game generation algorithm at time *t*. *Opt_i* = optimal release from MCM for agent I, *Gen_i* = hydro generation for agent I for domestic demand, *Sales_i*: hydro generation for market sales.

3.4.3.1 Optimization of the Columbia River

The optimal releases from the Kinbasket reservoir in the game generation step are obtained through a linear program step where the constraints on the Arrow reservoir are applied in order to not violate the Columbia River Treaty.

Table 4: Linear Programming problem formulation of the Columbia River

Objective Function	$\textit{maximize } z = c_1x_1 + c_2x_2$
Subject to	$a_L < x_1 < a_U$ $b_L < x_2 < b_U$ $d_L < s_1 - x_1 + I < d_U$ $F_L < s_1 + x_1 + I - x_2 < F_U$
Where,	
	z: value to be maximized x_1, x_2: decision variables representing the releases from Mica and Arrow $c1, c2$: coefficients for the decision variables a_L, a_U: lower and upper bounds on Mica releases b_L, b_U: lower and upper bounds on Arrow releases d_L, d_U: lower and upper bounds on Mica Storage F_L, F_U: lower and upper bound on Arrow storage I: local inflows

Chapter 4: Case Study and Results

4.1 Case Study: December 2019 Study

4.1.1 Study Time Frame and Outcomes

The model was used to conduct a study based on data from the December 2019 Energy Study for the Peace and Columbia systems. The EDM model was run on a monthly time step from January 2020 to December 2020, where each month was divided into 5 blocks (3 light load hour blocks and 2 heavy load hour blocks), coordinating the discharges, generation, and market activity for the Kinbasket (KBT) and Williston (WSR) reservoirs and taking into account the predetermined discharges of the Arrow reservoir at every time step. The results were then compared to the policy derived from the Energy Study of December 2019 for the period of January to December 2020.

4.1.2 Data

4.1.2.1 Water Years Inflow Scenarios

The availability of water has a major influence on BC Hydro's import or export decisions with Alberta and the US. Local inflows for the Kinbasket, Williston, Revelstoke, and Arrow reservoirs were obtained from historical streamflow data. There are 46 ensembles, referred to as "water years", derived from the inflows of the years 1973 to 2018. The cumulative annual inflows, shown for illustrative purposes, are shown in below. As shown, 1973 is a dry year, 1995 is a wet year and 1984 is a relatively average year. Climate change could potentially impact the viability of these inflows to be representative of future conditions; however, that is beyond the scope of this research.

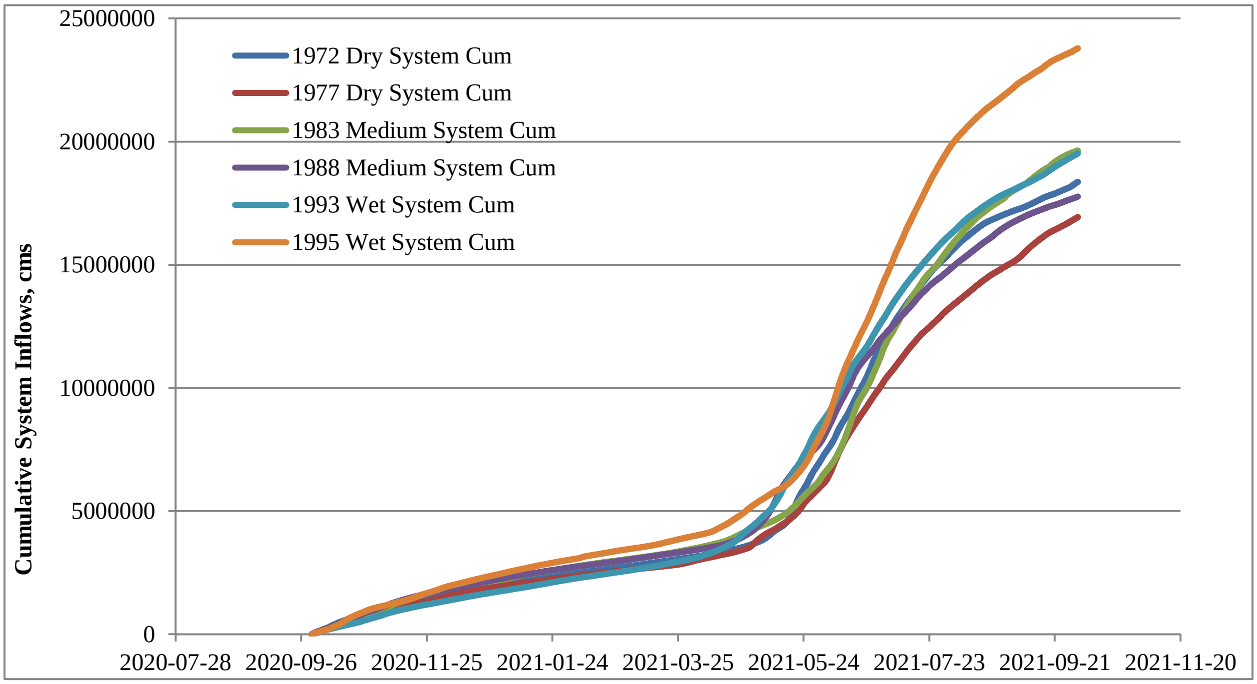


Figure 10: Cumulative system inflows (cms) for selected water years

4.1.2.2 Transition Probabilities

The transition probabilities represent the conditional probability of occurrence of a scenario of inflow in the next month, given the inflow scenario at the month we are in. They are calculated based on the history of inflow and forecasted inflow and are used to calculate the expected value of being at a specific storage state in the reservoir game. The transition probabilities used in this case study are considered to be Markovian and were determined by the BC Hydro Energy Studies group.

4.1.2.3 Market Elasticity Curve

In the BCH ESM, the forecasted market prices derived from the gas price-duration curves (that are load-based) inform the decisions to buy (when prices are low) and sell (when prices are high). This is done by including the gas state variable that controls the market price for electricity (in GJ) in the Bellman equation being solved to maximize the revenue while serving load. In the EDM model, price duration curves () were used to represent the price paid per MWh for the electricity sold or purchased. The price duration curves were developed by the BCH Generation System Operations group as part of the Water Value Project (Guan et al., 2018).

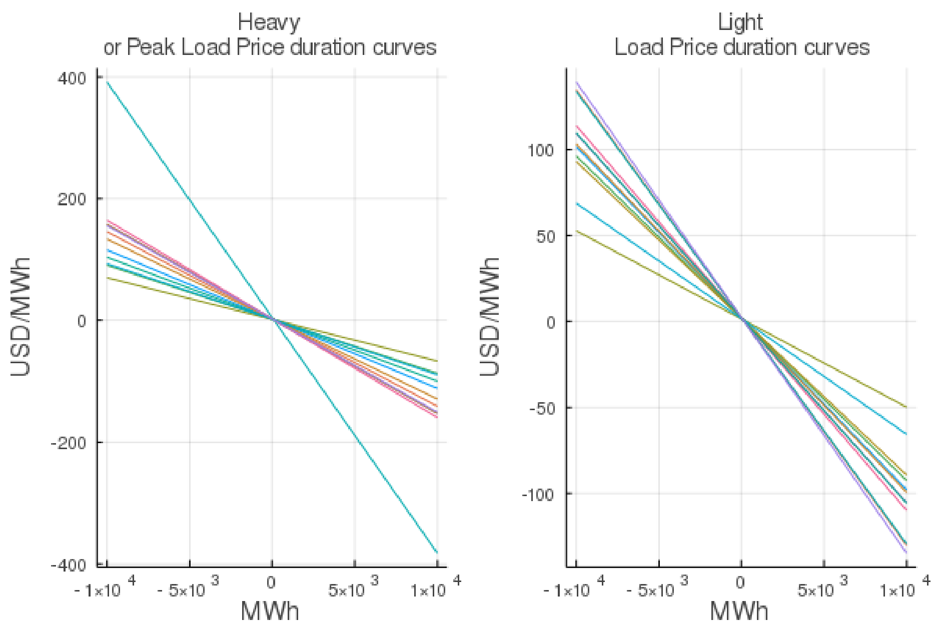


Figure 11: Heavy and Light Load hour monthly price duration curve

4.1.2.4 Storage-Elevation Curves

A relationship between the amount of water in storage and the vertical level of water in the reservoir had to be assumed for Kinbasket, Arrow, and Williston Reservoirs.

4.1.3 Constraints

4.1.3.1 Mass Balance:

The reservoir storage at any time step is the sum of the storage at the previous time step, current period inflows, and releases through turbines and spillways.

4.1.3.2 Storage Constraints

Reservoir storage must be operated within the storage limits which are equivalent to the minimum and maximum physical storage requirements for the reservoir.

For the Peace system, the storage is bound by the requirements for the Williston Reservoir, while for the Columbia system, the Kinbasket reservoir in addition to non-treaty storage is taken into consideration. Treaty storage was not taken into account. In a separate linear program, the upper and lower bounds on the Kinbasket reservoir storage and releases are considered as constraints, as well as the associated upper and lower bounds on the downstream, Arrow reservoir storage.

4.1.3.3 Power Generation

Power generation in a reservoir is a function of the reservoir's elevation. Typically, a coefficient describing the average power generated per unit release at each starting storage state is calculated. The HK values for Mica generation station (Kinbasket Reservoir) and the GMS generation station (Williston Reservoir) vary with elevation level, and range between 1.17 to 1.5.

4.1.3.4 Generation Limit:

Generation for each reservoir in each time step must be within the maximum and minimum generation limits.

4.1.3.5 Load-Resource Balance:

The system load must be equal to the sum of energy generated or traded to outside markets. Energy can be bought or sold during each time step at current market prices.

4.1.3.6 Demand

In ESM, demand is a discrete variable that must be met from generation or market purchases. Forecasted monthly demand for each water year ensemble was used in the model and divided to represent the 5 load blocks.

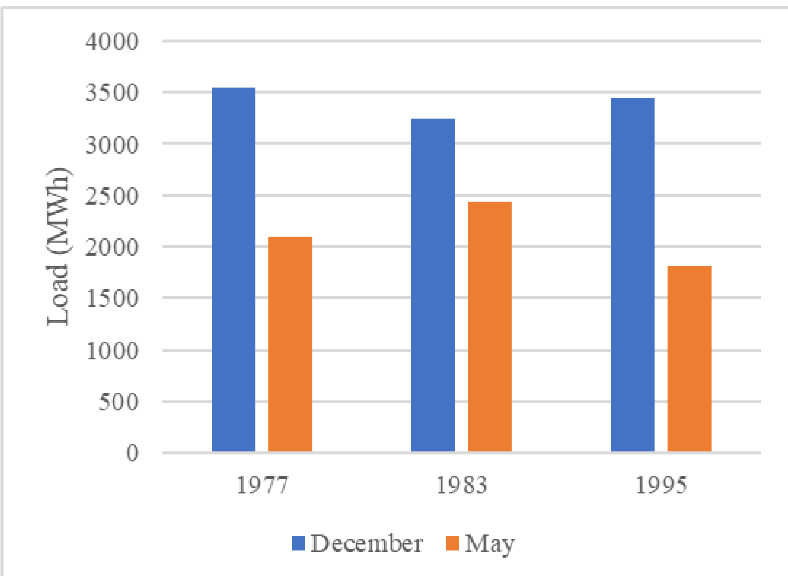


Figure 12: Domestic electricity load on the BCH system, forecasted for May and December 2020, from the 1977, 1983, and 1995 water years.

4.1.3.7 Flow Constraints and Peace Ice Cover

River ice is managed in North America by maintaining high minimum flows in low-temperature months and draw down this level gradually, in order to prevent the melting ice from causing floods. The Peace Canyon flows, downstream of G.M.S were taken as constraints for the Peace River system.

4.1.4 Error Calculation

The error in the policy, with respect to the policy determine by ES, was determined in two ways: mean absolute deviation (MAD), as in , and mean absolute percent error (MAPE), as in .

Equation 4: Mean absolute deviation formula.

$$MAD = \frac{\sum |x_i - \bar{x}|}{n}$$

Equation 5: Mean absolute percent error formula.

$$MPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|x_t - A_t|}{A_t}$$

4.2 Results

4.2.1 Error and Deviation

The output of the three runs of the game model with the three different algorithms (S.O = social Optimum, Nash = support Enumeration Nash Equilibrium, and MILP = Mixed Integer LP Nash Equilibrium) was compared to the monthly input of the Energy Studies (ES) for the period January 2020 to December 2020. The output was consolidated into one light Load Hour (LLH) block and one Heavy Load Hour (HLH) block for each month to match the output of the ES.

The total discharge policies can be seen plotted side-to-side in Figures 18-29 for the scenario years 1977, 1983, 1995. Although the values differ, there is consistency among the three algorithms when it comes to the monthly total discharge model policy deviation from the Energy Studies policy (Table 6). When the game model recommends a higher or lower discharge than the Energy Studies, it is usually all three algorithms agreeing on the direction of the deviation.

Table 5: Total discharges and standard deviation for the year (Jan 2020- Dec 2020) for the water scenarios: 1977, 1983, and 1995

		MCA LLH			MCA HLH			GMS LLH			GMS HLH		
		1977	**1983**	**1995**	**1977**	**1983**	**1995**	**1977**	**1983**	**1995**	**1977**	**1983**	**1995**
Total Year Discharge	E.S	6617	5135	5997	8247	6626	8606	9197	11260	13114	12846	14426	15123
	MILP	6912	7390	6044	6978	7390	6676	12520	13674	15117	10830	13370	11568
	Nash	6912	7390	6044	9294	9657	9247	12955	14080	15403	12654	13337	15502
	S.O	3871	7390	6044	6978	7390	6676	12955	14080	15403	12654	13337	15502
St Dev	E.S	528	513	460	461	470	440	599	480	426	566	565	450
	MILP	353	396	362	344	396	327	522	426	616	342	478	284
	Nash	353	396	362	260	278	162	512	435	591	413	431	580
	S.O	279	396	362	344	396	327	512	435	591	413	431	580

The error, summarized in Table 6 represents the mean absolute deviation (MAD) for the model's different algorithms monthly release policy, values highlighted as a heat map, with green representing the lower values and red representing the higher values. The lowest deviation ranged from 218 cms to 350 cms, with the highest deviation being 514 cms. The MCA LLH policy had larger deviation and error than the GMS LLH and HLH policy. The MCA release policies are governed by both the Mica and Arrow stations. Any deviations in the existing Arrow release, will increase the deviation of the releases of Mica. However, GMS releases do not depend on the PCN downstream releases in this model, and is not subject to further deviation.

Table 6: Mean absolute deviation (MAD) for monthly release policies from the different algorithms, values highlighted as a heat map, with green representing the lower values and red representing the higher values

		MCA LLH			MCA HLH			GMS HLH			GMS LLH		
		1977	**1983**	**1995**	**1977**	**1983**	**1995**	**1977**	**1983**	**1995**	**1977**	**1983**	**1995**
Mean Absolute Deviation (cms)	Nash	614.8	577.7	624.4	546.5	521.8	481.2	349.2	442.8	444.2	413.9	345.7	535.5
	S.O	614.8	577.7	624.4	543.3	517.7	395.8	218.8	274.9	454.1	455.4	385.9	562.2
	MILP	420.1	577.7	624.4	546.5	521.8	481.2	218.8	274.9	454.1	455.4	385.9	562.2

The error in Table 7 represents the percent error in the total discharge policy for that year. The total discharge error trends are similar to the mean monthly policy deviation trends in which the MCA LLH policy had the larger deviation. The magnitude of the error did not differ between the wet, dry, or medium year. The three algorithms performed equally well in the three scenario years.

Table 7: Percent error for the total release policies from the different algorithms, values highlighted as a heat map, with green representing the lower values and red representing the higher values

		MCA LLH			MCA HLH			GMS LLH			GMS HLH		
		1977	**1983**	**1995**	**1977**	**1983**	**1995**	**1977**	**1983**	**1995**	**1977**	**1983**	**1995**
Mean Percent Error	MILP	4.5	43.9	0.8	15.4	11.5	22.4	2.5	5.2	0.0	17.8	18.7	11.8
	Nash	4.5	43.9	0.8	12.7	45.8	7.4	0.8	2.4	1.9	37.6	18.4	18.2
	S.O	41.5	43.9	0.8	15.4	11.5	22.4	0.8	2.4	1.9	37.6	18.4	18.2

Looking at the two error calculations, the 1977, 1983, and 1995 MAD performance was fairly consistent, but the MAPE is clearly much higher in 1983. Since the MAD was calculated for the monthly releases, while the MAPE was calculated for the total releases of the policy year, it is expected that the year-to-year differences in MAPE be higher than those of MAD, which is average over the months. In other words, the month to month differences of the error in release policy between the wet, dry, and medium year, hence the MAD for MCA LLH is consistently high among 1977, 1983, and 1995, and consistently moderate for GMs LLH for example. However, the cumulative releases for a policy year is impacted greatly by the wetness or dryness of the year, hence, the MAPE for MCA LLH and HLL was higher in 1983 than 1977 and 1995.

Although errors exists, the discharge plots (figures 18 to 29) show that the constraints for upper and lower discharge were met by the model for the exception of GMS LLH and HLH of May 1995 and July 1983 for the MILP algorithm for GMS HLH. This suggests that a possible model adjustment is needed for WSR reservoir to an LP formulation sub-step, similar to the way that KBT reservoir is modeled (Columbia LP step). The use of the Columbia LP step reduced the variation between the results of the three game algorithms but did not help mitigate the error in the policy.

The market policy for the years were chosen because the model prioritizes fulfilling the demand from the reservoirs and does not consider the rest of the BC Hydro system as resources.

The model can perform 5520 LPs in under 4 minutes, more will be discussed in the benchmarking section in Chapter 5. In the discussion, it can be argued that the results of these models prove that the concept of game theory valid for multi-0-reservoir modelling, despite the magnitude of the error. By considering the results valid, and looking at the optimized reward, one can deduce that the game model, or the equilibrium-based model that takes each agent’s benefits into consideration simultaneously, can achieve better rewards when optimizing the multi-agent hydro system, and therefore, this method is worth investigating further. Section 5.5 in Chapter 5 presents possible model advancement using AI techniques.

4.2.2 Rewards

The reward for these years can be seen in Figure 13 and Figure 14. The reward from the ES results was recalculated according to equations 1 and 2. It can be seen that the game equilibrium algorithm yielded higher rewards for both agents and the system. This result was also consistent when comparing the policy rewards for all 46 inflow scenarios testes, represented in Figure 15 and Figure 16.

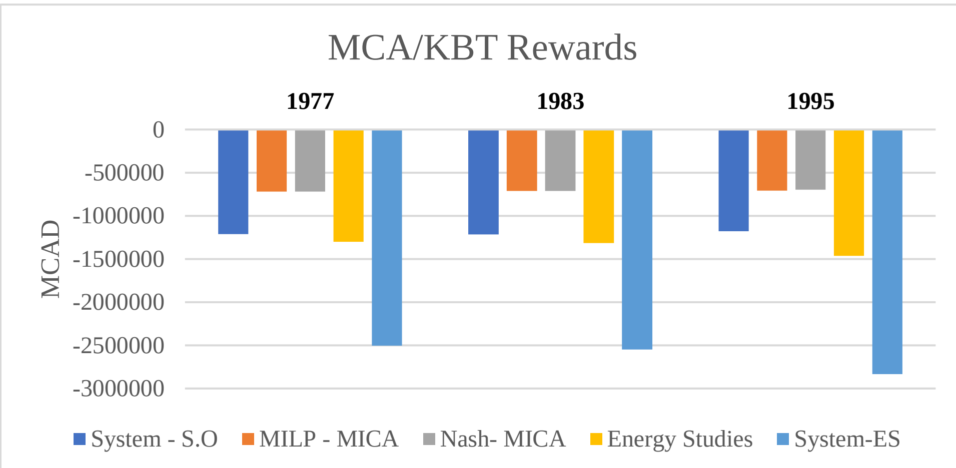


Figure 13: Total year agent and system reward for the scenario years for Kinbasket Reservoir

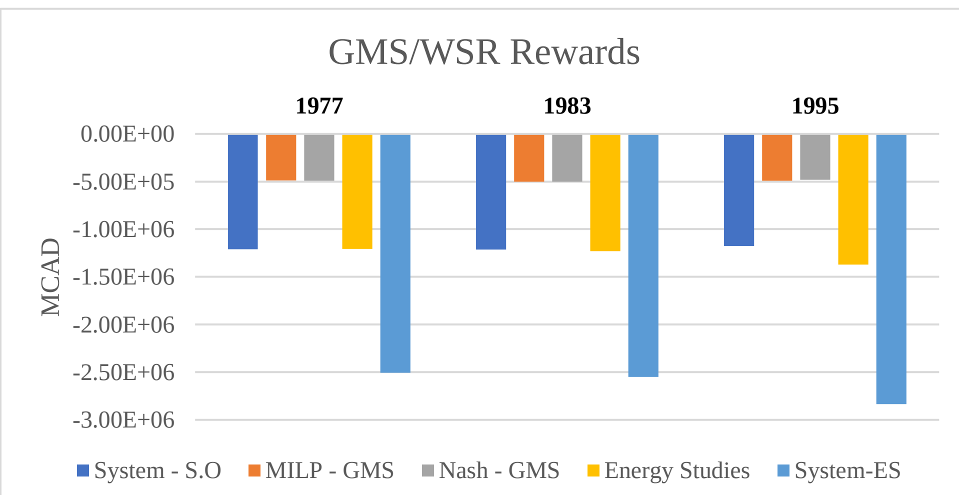
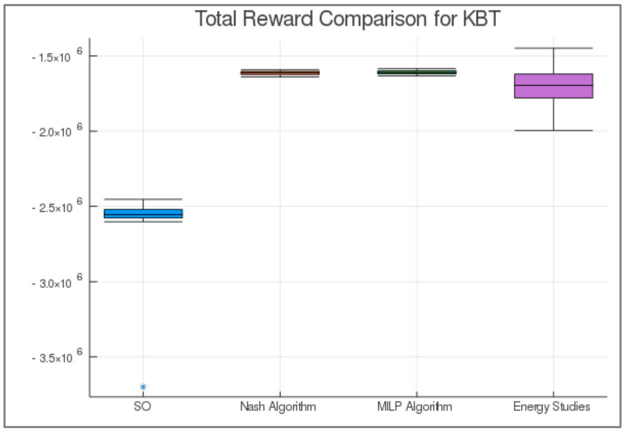


Figure 14: Total year agent and system reward for the scenario years for Williston Reservoir



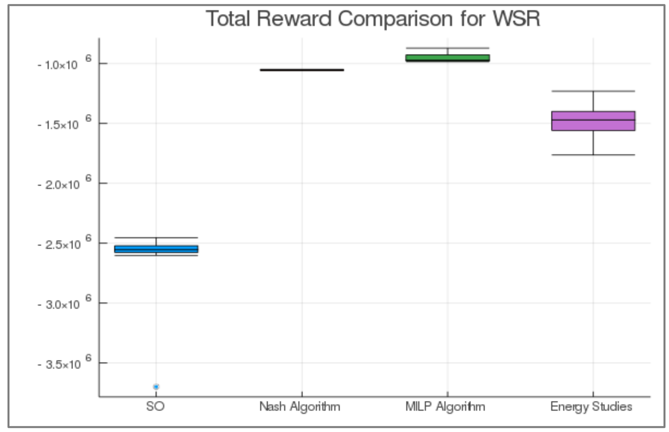


Figure 15: Total policy year reward boxplot (MCAD) for Kinbasket Reservoir

Figure 16:Total policy year reward boxplot (MCAD) for Williston Reservoir

4.2.3 LLH Discharge Policy

The Light Load Hour total discharge policies can be seen plotted side-to-side in figures 18-23 for the scenario years 1977, 1983, 1995. The figures show the instances when any of the game model algorithms recommends a higher or lower discharge than the Energy Studies and by how much. It can be seen that, usually, all three algorithms agree on the direction of the deviation. However, not all three algorithms perform similarly all the time. While the MAD and MAPE can show us that the MILP performs slightly better than the other algorithms, the bars in the figures below confirm this observation. The dotted lines show the lower and upper bounds on releases, which also an important factor in determining the model fidelity to realizing these constraints. It was only in three instances where the upper limit on releases was exceeded (Figures 22, 26, and 28). In the three instances it was for the GMS station, which is expected since these bounds were not applied to GMS as hard constraints.

4.2.3.1 Dry Year

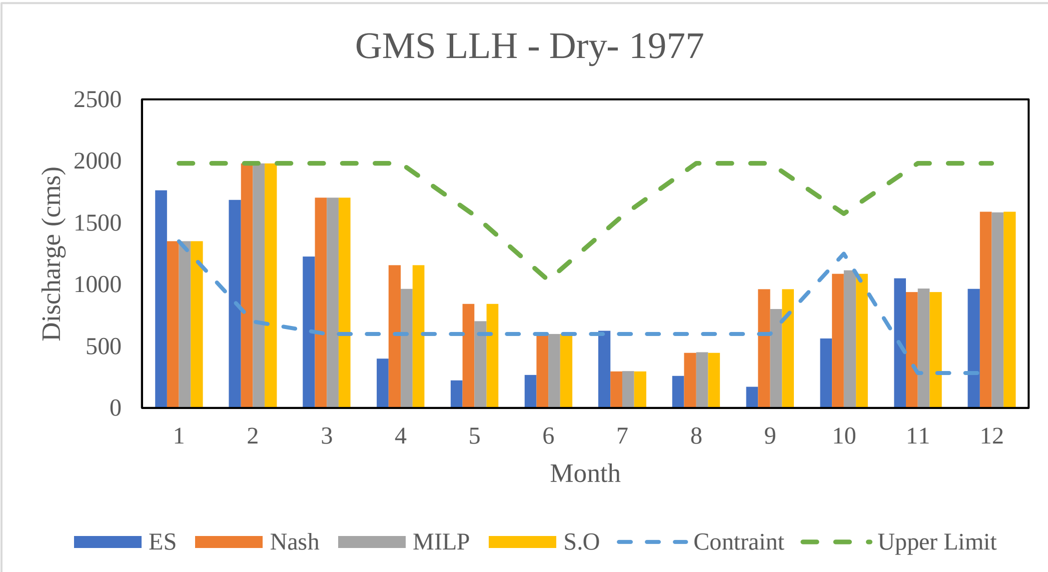


Figure 17: GMS LLH monthly discharges (cms) of the different algorithms for the dry year inflow scenario (1977)

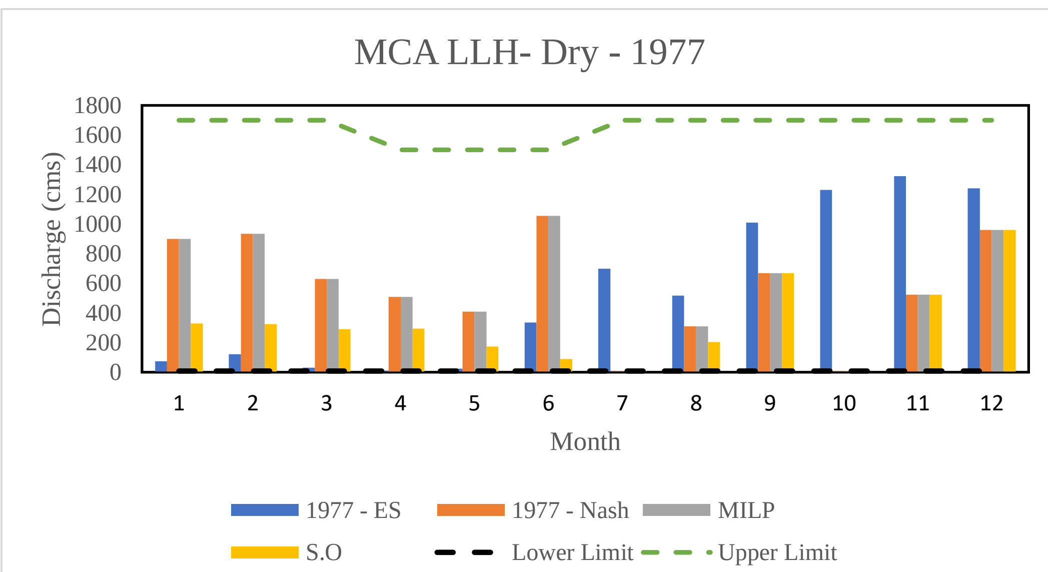


Figure 18: MCA LLH monthly discharges (cms) of the different algorithms for the dry year inflow scenario (1977)

4.2.3.2 Medium Years

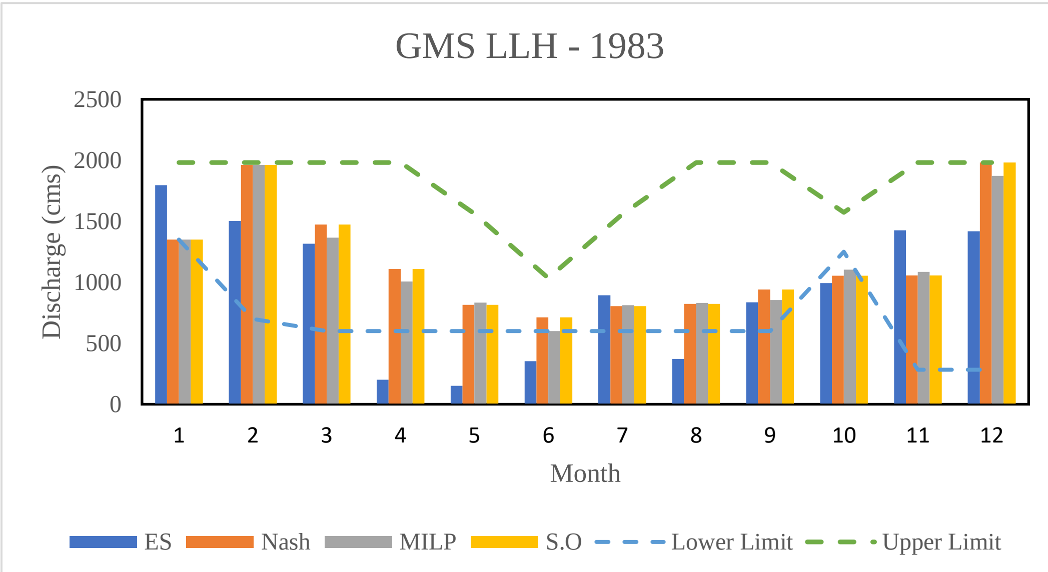


Figure 19: GMS LLH monthly discharges (cms) of the different algorithms for the medium inflow year scenario (1983)

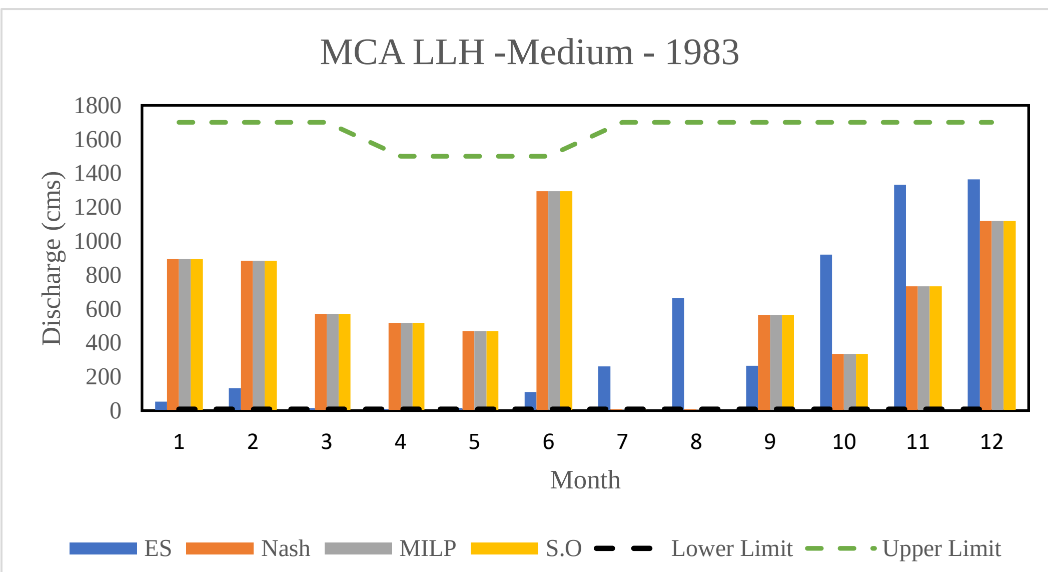


Figure 20: MCA LLH monthly discharges (cms) of the different algorithms for the medium inflow year scenario (1983)

4.2.3.3 Wet Years

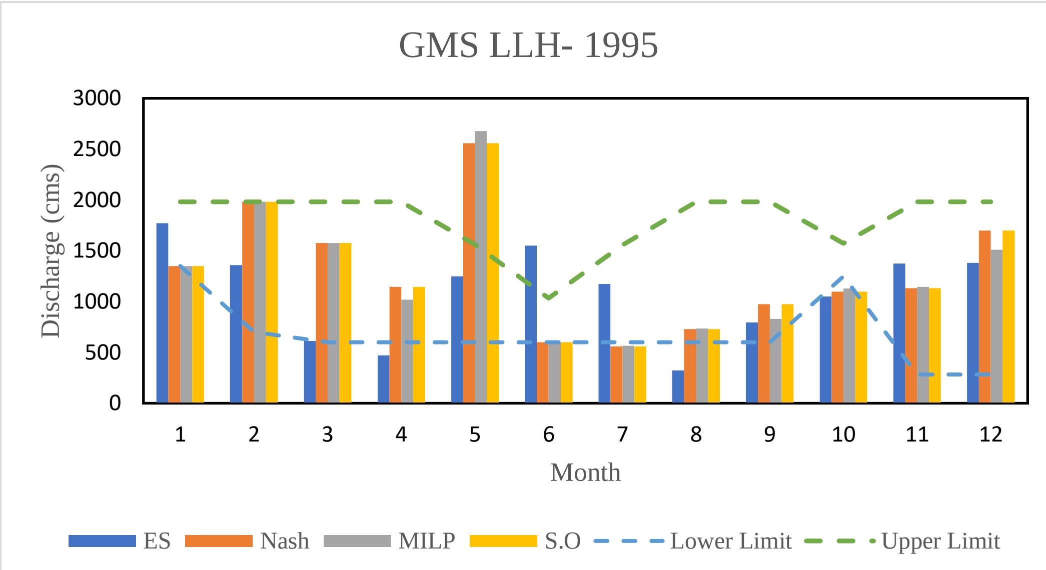


Figure 21: GMS LLH monthly discharges (cms) of the different algorithms for the wet year inflow scenario (1995)

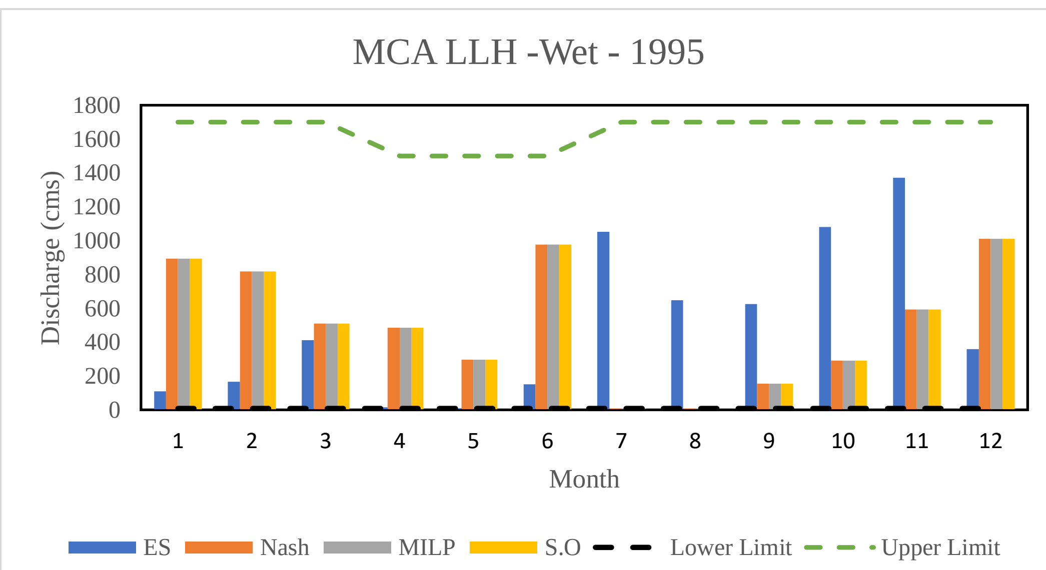


Figure 22: MCA LLH monthly discharges (cms) of the different algorithms for the wet year inflow scenario (1995)

4.2.4 HLH Discharge Policy

The Heavy Load Hour total discharge policies can be seen plotted side-to-side in figures 24-29 for the scenario years 1977, 1983, 1995.

4.2.4.1 Dry Years

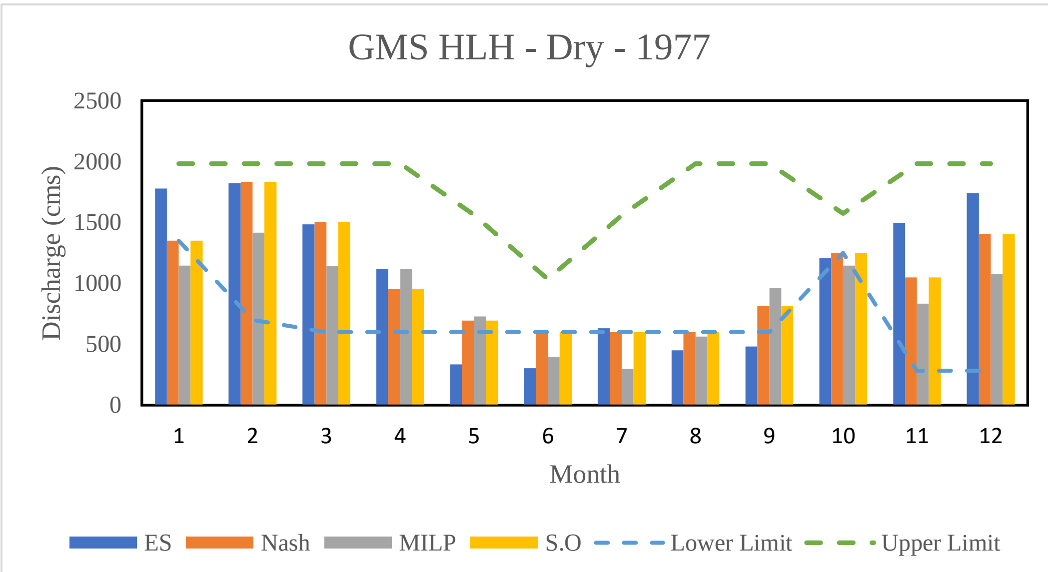


Figure 23: GMS HLH monthly discharges (cms) of the different algorithms for the dry year inflow scenario (1977)

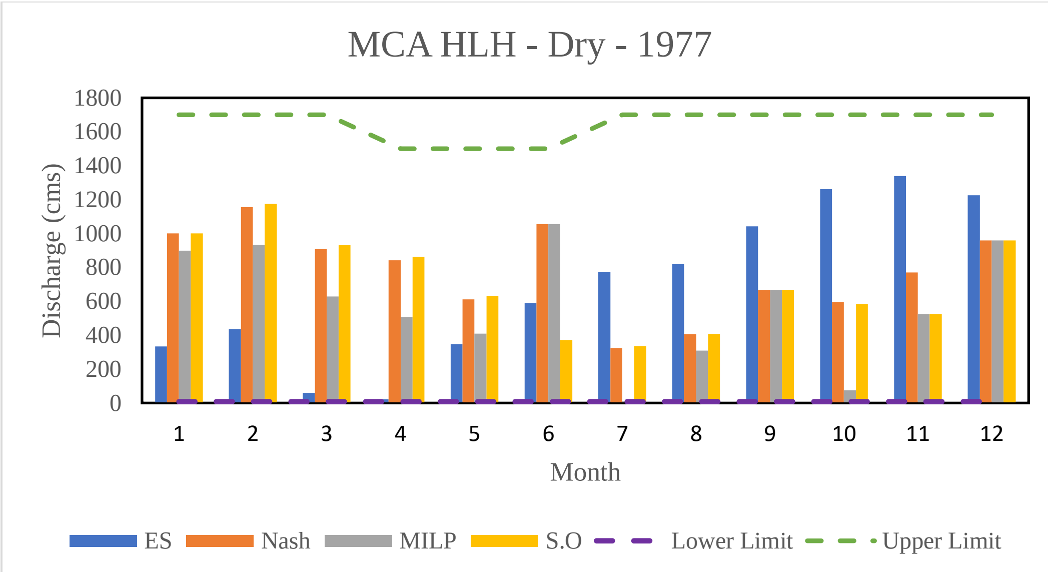


Figure 24: MCA HLH monthly discharges (cms) of the different algorithms for the dry year inflow scenario (1977)

4.2.4.2 Medium Years

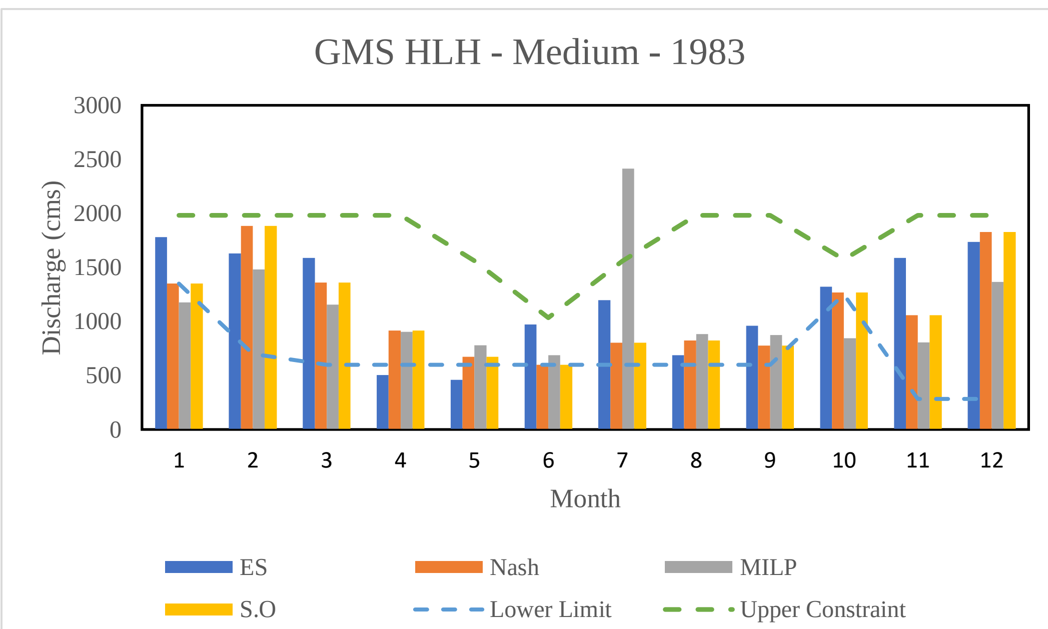


Figure 25: GMS HLH monthly discharges (cms) of the different algorithms for the medium inflow year scenario (1983)

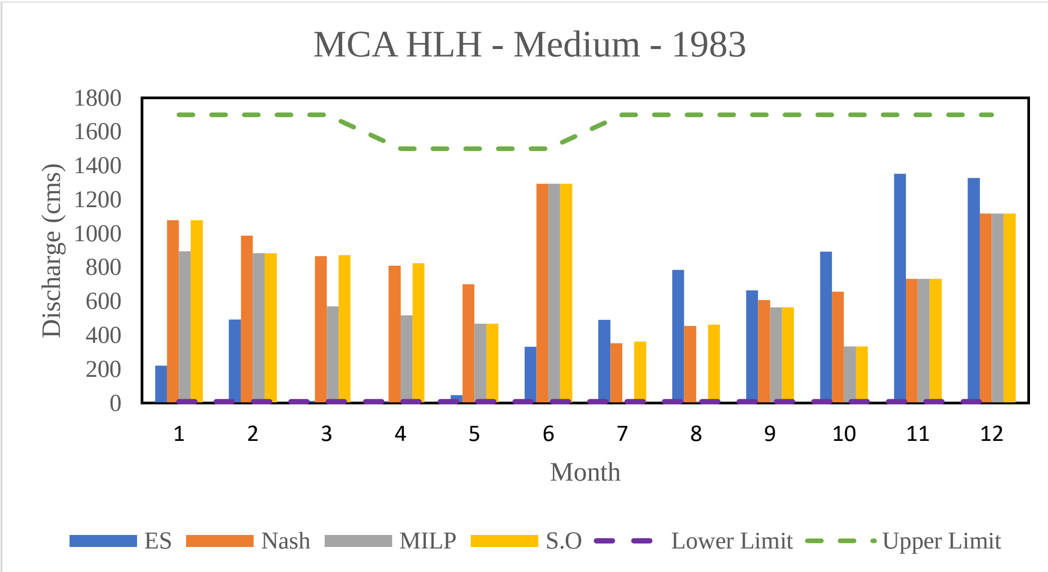


Figure 26: MCA HLH monthly discharges (cms) of the different algorithms for the medium inflow year scenario (1983)

4.2.4.3 Wet Years

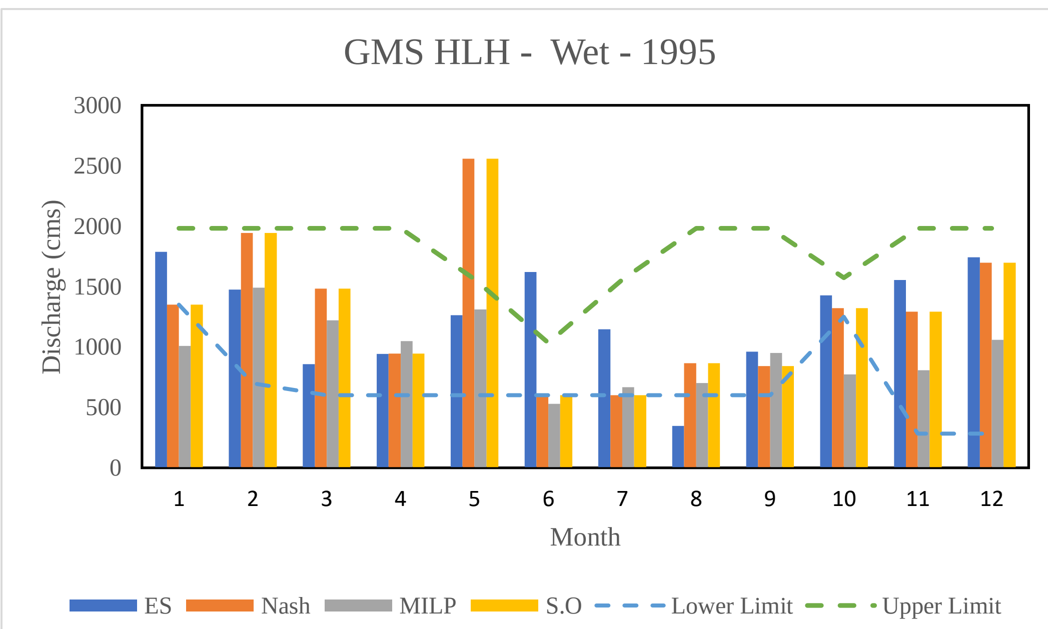


Figure 27: GMS HLH monthly discharges (cms) of the different algorithms for the wet year inflow scenario (1995)

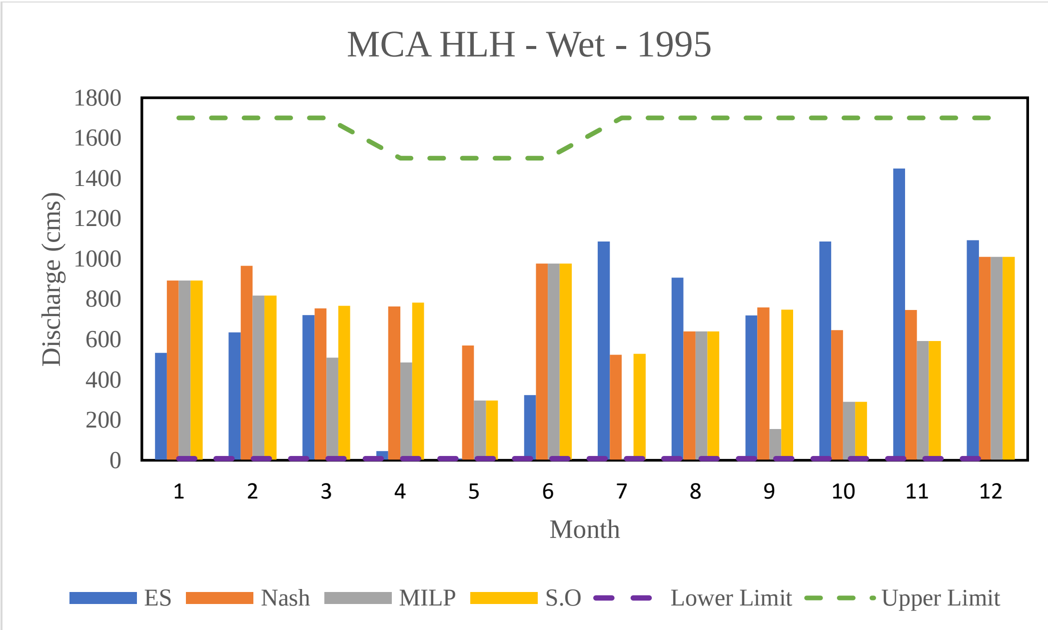


Figure 28: MCA HLH monthly discharges (cms) of the different algorithms for the wet year inflow scenario (1995)

4.2.5 Market Policy

There was no consistency in the market policy recommendations among the three algorithms. For LLH and HLH purchases, the Nash algorithm recommended significantly higher GWh per policy year than the other algorithms. For LLH sales, the three algorithms recommended higher sales than ES, while for HLH, the sales for all 4 models were within 2000 to 4000 GWh.

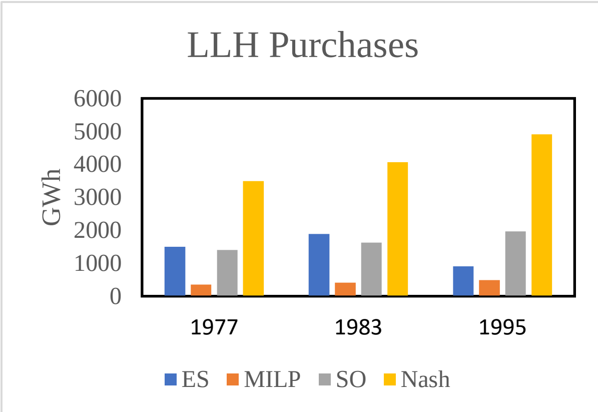


Figure 29: System total LLH purchases for the policy year (Jan 2020 to Dec 2020)

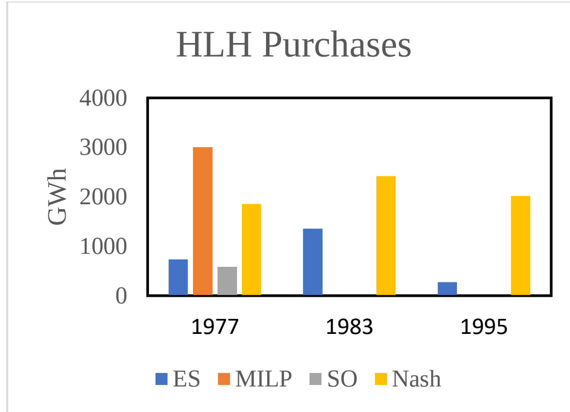


Figure 30: System total HLH purchases for the policy year (Jan 2020 to Dec 2020)

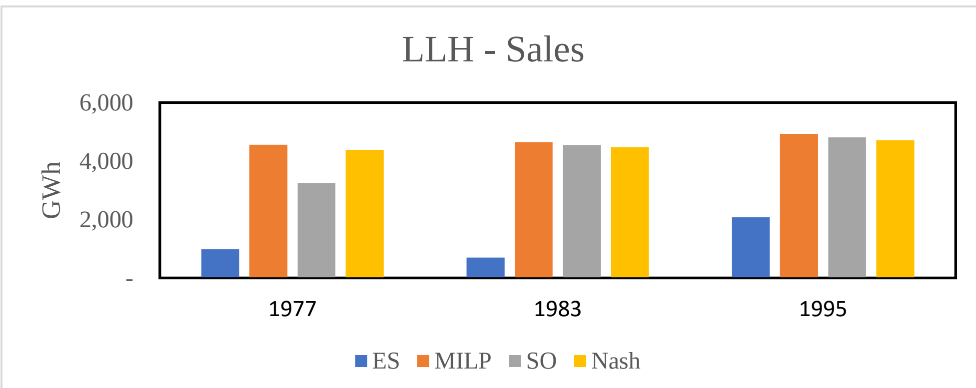


Figure 31: System total LLH sales for the policy year (Jan 2020 to Dec 2020)

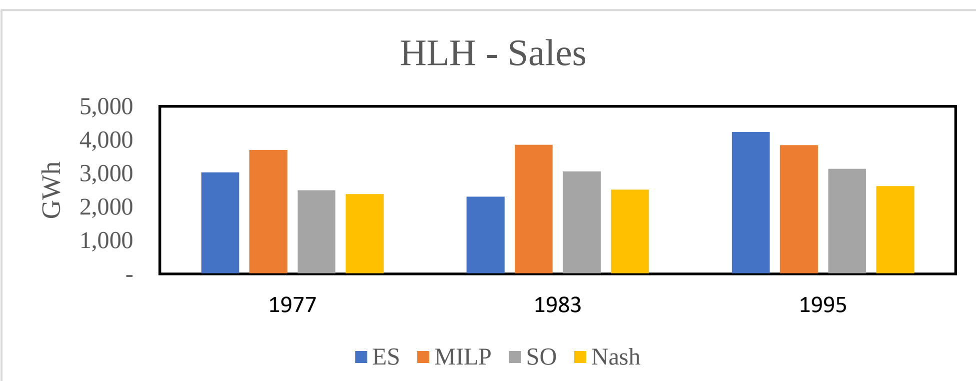


Figure 32: System total HLH sales for the policy year (Jan 2020 to Dec 2020)

4.2.5.1 Other sources

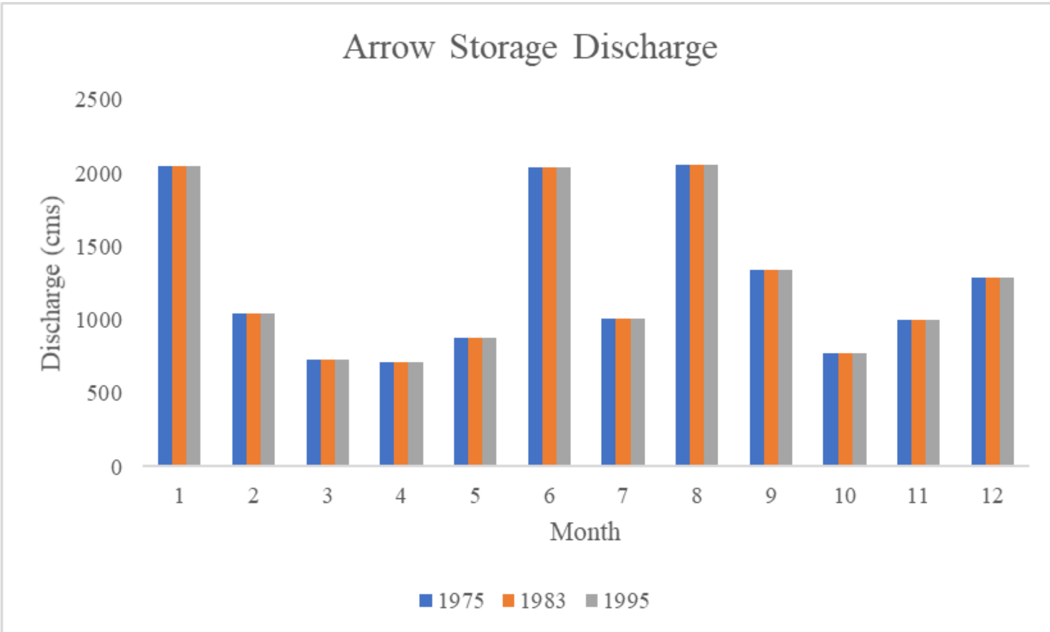


Figure 33: Arrow dam plant monthly discharges (cms)

Chapter 5: Discussion

5.1.1 Distinguishing Characteristics of the Equilibrium-Based Decision-Making Model

In multi objective optimization problems (MOOP), the decision for each of the objectives is assumed to be made by the same agent. If a MOOP problem contains individual rational agents (or groups of agents) for each objective function, then it is a game problem, not a MOOP problem. What distinguishes Game Theory (GT) from MOOP is the capacity of GT to analyze situations where we have interdependence among the players. In an equilibrium case, the individual rational agents are assumed to act in their own self-interest, while taking into account the other agent's possible policies. Equilibrium is reached when no agent can improve on their reward anymore. In a pareto optimal case, the agents can improve on their own rewards, as demonstrated by the results of the case study, where the social reward for the system shared by the agent was less than the reward achieved competitively by each agent. Pareto optimality ensures the agents are rewarded for the system's best interest and the overall objective, such as meeting the demand, while the cost of getting to equilibrium is shared by the agents or tackled by the system. Nash equilibrium may not always be Pareto optimal, hence the rewards are achieved.

On the other hand, there are similarities in the overall approach of optimization problems and equilibrium problems. Optimization problems are often approached using mathematical programming techniques. These problems are generally approached in the following sequence (Fourer et al. 2003):

1) Formulating a model by identifying the variables, objectives and constraints that represent the problem at hand

2) Collecting data defining a problem instance

3) Generating a specific objective function and constraint equations from the model and data

4) Solving the problem using a solver that applies an optimization algorithm to find the optimal values of the variables

5) Analyzing the results

In this thesis, it was demonstrated by the methods investigated that these five steps were preserved.

1) The variables were identified, and the system objectives and constraints were identified
2) Data was collected on the inflows, agents constraints, and properties such as HK values
3) The objective was substituted by the agent's utility functions, and the constraints were coded into the game generation step
4) Game solution algorithms were used to solve for equilibrium
5) Results were analyzed

According to (Labadie 2004), optimal coordination of the many facets of reservoir systems requires the assistance of computer modeling tools to provide information for rational management and operational decisions. This GT-based model developed in this thesis, with the aide of the Julia programming language, was able to achieve computationally promising results by cutting the time required to solve the problems by half.

5.1.2 Limitations and simplifications

There are several simplifications that prevent the model from generating a policy similar to the ES model. The first and most pronounce simplification was not accounting for daily variations. ES model is optimized using a daily and hourly time step which helps account for the demand and price variations of these hours. This model, instead, lumps these variations into monthly LLH and HLH blocks. This might have resulted in variation of input data used, such as monthly demand, which was adapted to fit into the 5-block scheme.

Market prices were also a differentiating factor in this model and ES. While in ES, market prices are generated via a stochastic model, the prices used here were applied deterministically, not accounting for future variability in circumstances or inflow conditions. The discount factor also did not change with time. When modelling the market decisions, there was no separation of the Alberta and US trades, which could have impacted the value of buying and selling through the transmission cost difference.

Finally, the BC Hydro system is a large one, consisting of 31 generating stations sending electricity along 75,000 kilometres of power lines, is not limited by the Columbia and Peace rivers. The vast resources were simply modeled in this thesis as an emergency reserve by imposing a "penalty" value on their use by the modelled agents. This way, the agents in the game were driven to complementary strategies to fully meet the demand. A true game of the BC Hydro would have thousands of strategies to meet the demand and achieve market gain. Previous applications show that it is best to use a variant of SDP (such as SDDP) to obtain the optimal allocations of the resources for the game model.

5.1.3 Functionality as a decision-making tool

BC Hydro strives to develop tools using stochastic optimization techniques to generate reservoir release policies that consider stochastic inflows, prices, and load with proper modeling of the CRT and subsequent agreements for operations planning. Results from the case study show that the game theoretic approach presents a promising tool for BC Hydro for solving multi-objective problems, since it produces policies that fall within the set constraints and maximize the reward. If the errors in the results and the deviation from the ES policy were correctly attributed to the model's simplified representation of other system variables, then the development and enhancement of the model which are beyond the scope of this research can lead to a product better suited for energy studies. The expansion of the model to a daily time step should be considered as a first step in the model enhancement. Also, several variants of multi agent solution concepts can be investigated to maximize the total payoff of the game involving multi-agent reservoir systems with a coordination mechanism to satisfy system level constraints. After enhancements to the existing formulations have been implemented, advanced solution methods should be investigated and used in the model. In the following sections, I highlight some findings from literature on Multi-Agent Inverse Reinforcement Learning.

Chapter 6: Benchmarking Framework for Stochastic Models

6.1 Background

Benchmarking has been used widely to compare models in the fields of hydrology, water resources management, operations research, and optimization. The goal from conducting model benchmarking is to compare and to provide insight into the performance of a variety of models used in stochastic optimization and optimal control. Statistical methods are widely used to asses model performances; however, as argued by (Seibert 2001), it is important to compare the model results with results obtained in some other way, which highlights the importance of choosing appropriate benchmark problems.

Benchmarking is performed in several different ways based on the discipline, problems addressed, and the techniques that are being used in the comparison. Principles of benchmarking have been borrowed from economics and especially from the manufacturing industry as they were used as tools to enhance performance. As outlined in the book Benchmarking Theory and Practice (Rolstadås and International Federation for Information Processing 1995), there are four phases to the benchmarking process are as follows.

Table 8: Phases of the Benchmarking Process, (Rolstadås and International Federation for Information Processing 1995)

First Phase	Preparation for Benchmarking: This phase addresses the questions of “what is the subject to be benchmarked and who are the best “competitors”, and what is the best data collection method?
Second Phase	identifying the strengths of each competitor or model
Third Phase	defining functional goals and action plans for the analysis
Fourth Phase	planning and implementation

While the management of water reservoirs has been a popular area of research in the stochastic programming literature (Salas and Powell 2018), there is a paucity in examples on benchmarking stochastic optimization models in this field. Therefore, in this Chapter, we review the different benchmarking methods that have been used in peer reviewed literature, highlighting novel methods in stochastic algorithms optimization from operation research, and conclude with recommendations for benchmarking plan for iterative models of SDDP and existing Energy Study Models.

6.1.1 Overview of Widely Used Methods:

Statistical Significance Tests: a nonparametric, complete-block, two-way layout Friedman test is commonly used to detect if there was a difference between the algorithms (nuisance factor).

Convergence Speed: in an operational context, such as continuous reservoir management for hydropower production, calibration speed is of the essence. In this respect, the convergence speed of each algorithm can be tracked and compared to the overall best objective function value. This is done to prevent an algorithm that converges fast to a poor optimum to be ranked higher than it should.

Dispersion: the dispersion metric of Lunacek and Whitley (2006) was computed to further investigate the causes of algorithm performance or lack thereof. It uses iterative random sampling of the search space to measure the average pair-wise Euclidian distance between m-best parameter sets from a population of *n* parameter sets, where the size of m is fixed and *n* is variable. A decrease in the average Euclidian distance when *n* is increased means that the fitness landscape has a converging global structure. Further information on the dispersion metric is available in Lunacek and Whitley (2006).

Benchmark Series: Used in hydrology, benchmark series are used to compute the goodness-of-fit of a model with respect to the benchmark using Equation 6.

Equation 6: Goodness-of-fit equation

$$X^2 = \sum \frac{(\text{observed - expected})^2}{\text{expected}}$$

Performance Profiles: The performance profile for an algorithm is the (cumulative) distribution function for a performance metric. For example, the ratio of the computing time of a solver versus the best time of all of the solvers can be used as the performance metric. Performance ratio is the computing time required to solve problem p by solver s divided by the minimum time of all the solvers to solve the same benchmark problem.

Then define the function p_s as the cumulative distribution function for the performance ratio, such that $p_s(T)$ is the probability of solver s that a performance ratio $r_{p,s}$, is within a factor T of the best possible ratio. Then plot $P[0,1]$ against different values of T. $(1-p_s(T))$ is the fraction of problems that the solver cannot solve within a factor τ of the best solver, including problems for which the solver in question fails.

6.1.2 Examples from Literature

Comparison of Stochastic Optimization Algorithms for Hydropower Reservoir Operation with Ensemble Streamflow Prediction

Ensemble streamflow prediction (ESP) is meant to produce multiple and reliable scenarios of possible river flows to represent forecast uncertainty. (Côté and Leconte 2016) present a comparison between four optimization algorithms in a test bed in which ensemble streamflow predictions (ESPs) are updated each time a decision is taken, which is an operating procedure that is closer to real world practice. The comparison was performed on the Rio Tinto Alcan (RTA) hydropower system in Québec, Canada, which consists of six generating stations in series and three major reservoirs. The tested optimization algorithms are the deterministic optimization approach (DA) currently used by RTA and three explicit stochastic optimization approaches, i.e., stochastic dynamic programming (SDP), sampling stochastic dynamic programming (SSDP), and a scenario tree approach. The objective of the problem is to maximize, on a weekly time step (T = 52), the function describing the expected value of the energy export and imports under several constraints (water value function). The first algorithm deterministically solves the objective function as the sum of payoffs for each ESP scenario independently. The second algorithm (SDP) decomposes the problem into small optimizations in a backward recursive equation starting from the end of the planning horizon to the first period. The third algorithm (SSDP) uses the transition probabilities of streamflow scenarios

The fourth approach models the stochastic process by a scenario tree, in which each node has one predecessor. In this approach, an original set of scenarios is transformed into a tree-node structure. The tree is then reduced by successively removing nodes while preserving the probabilities associated with the original inflow sequences.

They computed the average time required for one release decision (or one period) in the test bed for each method, with the number of scenarios, points for state variables, and number of decision variables and constraints where applicable. The SSDP turned out to be the slowest algorithms since the water value function has to be evaluated over each scenario. They then compared the water storage in target reservoirs and the water spillage in system for the planning horizon based on different operating policies given by the different algorithms. The authors also compare the effect of dispersion in the ESP scenarios and historical sequences on the different methods. As expected, all the stochastic methods outperformed the deterministic approach. Also, it was concluded that scenario based methods are superior to probability distribution methods, while noting that the SDP probability distribution method has several advantages in computation time and discretization points. The study employed an empirical approach to artificially increase flow volumes in order to account for the impact of climate change on watersheds in northern environments. Since in a climate change context, it is likely that dispersion in the historical data will not be representative of future dispersion. The authors propose that using stochastic optimization approaches as a decision-making tool to properly manage hydropower systems and generate maximum gains will be even more important for hydrological regimes characterized by increased variability, such as that caused by climate change. A similar approach was implemented by (Cote et al. 2011) to compare the stochastic optimization of Hyro-Quebec hydropower installations by SDP and SSDP.

Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration

(Arsenault et al. 2014) benchmarked ten stochastic optimization methods that were used to calibrate parameter sets for three hydrological models on 10 different basins. Optimization algorithm performance was compared for each of the available basin-model combinations. The 10 optimization methods were then programmed for each of the model-basin pairs using their default parameter values. Each optimization algorithm was used to complete 40 different model calibrations, with each optimization run being limited to 25,000 evaluations. The objective function value was saved for every model evaluation, thereby generating a trace for each calibration run.

The aim of this study was to try and find the best optimization algorithm for a given hydrologic model calibration problem based on the problem characteristics. Therefore, the authors conducted an overall comparison based on several parameters of interest: convergence speed and ability of attain low objective function values. A general performance assessment is made first. Then, method performance with respect to model complexity, basin type, convergence speed, and computing power were addressed.

Integrating Logistical and Technical Criteria into a Multi-team, Competitive Watershed Model Ranking Procedure

This study investigated hydrological model comparison techniques and illustrated a systematic watershed model comparison and selection process, integrating a full range of relevant criteria. The process by (Cunderlik et al. 2013) is believed by the authors to be broadly useful. It consists of:

- Screening for a set of candidate models on the basis of prerequisite model attributes; assessing hydrologic simulation performance using various conventional statistical metrics
- Assessing operational logistics performance, reflecting somewhat subjective but centrally important issues around relative feasibility and suitability of candidate models in the intended context of use
- Integrating the hydrologic and operational performance results, which are evaluated using a weighted-matrix approach, into a single coherent and comprehensive ranking system.

The process was applied to evaluation of watershed models for operational hydroelectric inflow forecasting in British Columbia, Canada. An important feature of the study was its horse-race project management approach, involving a supervised competition between expert teams using different models but the same data sets. The horse-race format involved a direct competition between the four candidate models with the winner declared on the basis of both hydrologic simulation success and a variety of practical criteria. The compared models that passed the screening were:

- UBCE Watershed Model (UBCWM),
- the Environment Canada modification of the Hydrologiska Byråns Vattenbalansavdelning model (HBV-EC),
- the National Weather Service River Forecasting System (NWSRFS), and
- the University of Waterloo WATFLOOD.

The models were evaluated for operational logistics to reflect the practical issues associated with the forecasting operations. In this part of the process, the following features and sub-features were compared:

1. Modeling time, with subfeatures of calibration time/effort required, central processing unit (CPU) time for model execution, and data pre/post-processing time/effort intrinsically required by the model
2. Sensitivity to input data availability, with sub-features of meteorological data requirements, and other data requirements
3. Model robustness and diagnostics, with sub-features of code stability, diagnostic capabilities, and code stewardship
4. Ease/speed of state- and driving-variable adjustment and updating, with sub-features of ease with which input meteorological data may be manually altered, and ease with which internal states may be adjusted or updated.

For the hydrologic model performance, ?? study watersheds from British Columbia were used to represent 3 different hydroclimatic and physiographic settings: Alouette, Finlay, and Mica. A wide variety of statistical performance measures were reviewed to determine a set of measures suitable for the evaluation of the watershed models. Ultimately, the main objectives of the calibration in all study watersheds were to (1) maximize the Nash–Sutcliffe coefficient of efficiency of daily flows (CE), (2) maximize the coefficient of variation (R^2) of annual flow volumes, (3) minimize bias in monthly flow volumes, (4) minimize bias in annual flow volumes, and (5) minimize bias in March 1, April 1, and May 1 snow water equivalent (SWE).

With the availability of different metrics to measure the performance, the authors were presented with an interesting challenge to rank the models. The decision was made to rely upon weighting matrices, wherein individual scores were assigned within a potential range of 0–1, to provide relative scorings for each of the models. This, of-course, could be used to have a bias towards one of the models. Two different matrices were employed (Table 9, Table 10), one with respect to the operational logistics as applied to BC Hydro's routine forecasting needs, and the other one with respect to hydrologic performance of the individual models. The two matrices were combined into an overall model suitability matrix Table 11.

Table 9: Operational Logistics Matrix (Cunderlik et al 2013)

Feature	Subfeature	UBCWM	HBV-EC	NWSRFS	WATFLOOD
Modeling time	Calibration time/effort required	1.00	0.80	0.70	0.60
	CPU time for model execution	1.00	1.00	1.00	0.60
	Data preprocessing /postprocessing time/effort intrinsically required by model	1.00	0.80	0.50	0.50
Sensitivity to input data availability	Meteorological data requirements	1.00	0.70	0.70	0.50
	Other data requirements	1.00	0.90	0.80	0.70
Model robustness and diagnostics	Code stability	0.70	1.00	1.00	0.70
	Diagnostic capabilities	0.70	0.70	1.00	0.70
	Code stewardship	0.50	0.80	1.00	0.80
Ease/speed of state- and driving-variable adjustment and updating	Ease with which input meteorological data may be manually altered	1.00	1.00	0.90	0.80
	Ease with which internal states may be adjusted or updated	1.00	1.00	0.90	0.70
Total score		0.91	0.88	0.85	0.66

Table 10: Hydrologic Performance Matrix (Cunderlik et al 2013)

Measure/score	UBCWM	HBV-EC	NWSRFS	WATFLOOD
Calibration	0.79	0.76	0.81	0.73
Validation	0.64	0.65	0.60	0.61
Overall calibration–validation	0.72	0.71	0.70	0.67
Calibration–validation consistency	0.85	0.89	0.79	0.88
Interwatershed consistency	0.83	0.83	0.70	0.89
Overall consistency	0.84	0.86	0.75	0.89
Overall hydrologic performance	0.75	0.75	0.71	0.72

Table 11: Combined Performance Matrix Hydrologic Performance Matrix (Cunderlik et al 2013)

Measure/score	UBCWM	HBV-EC	NWSRFS	WATFLOOD
Hydrologic performance	0.75	0.75	0.71	0.72
Operational logistics performance	0.91	0.88	0.85	0.66
Combined performance	0.83	0.81	0.78	0.69

Using Confidence Intervals for benchmarking stochastic models:

(Liu et al. 2017) proposed a way to add **confidence intervals** (CI) to the performance profile method (PPM) and data profiles method (DPM) benchmarking methods to make it a framework for benchmarking stochastic methods. In benchmarking stochastic optimization models, it is important to have a unique and similar starting search point among all considered models. Numerical example, models were required to complete at least 10,000 evaluations as termination criteria.

Steps of the algorithm of the benchmarking:

1. Generate matrices for sample mean, and confidence bounds
2. Compare sample mean, determine the winner (the one with the higher profile)

3. Compare the bound for the “worse” performance, if the winner is still a winner it wins (one with higher profile still is higher), if it loses, there is no difference between the two models.

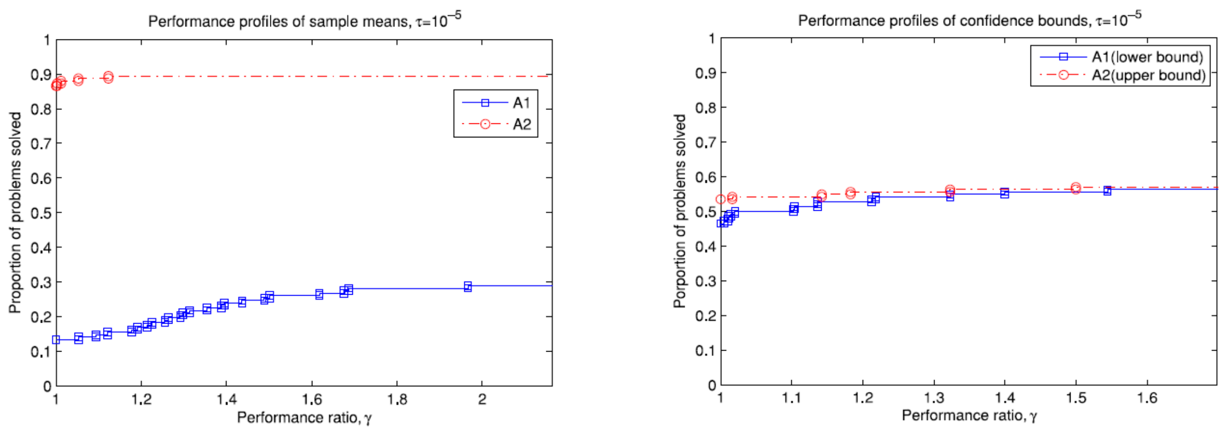


Figure 34: Results of the benchmarking problem using stochastic performance profiles and stochastic performance profiles method

The numerical example compared genetic algorithms (GA), differential evolution (DE), and stochastic particle swarm optimization (SPSO). They benchmarked DE8, GA, and SPSO using PPM. In this step, SPSO performed better than DE8 and GA, while GA performed worst amongst the three. They conduct another PPM comparison for the upper bound CI matrix of SPSO together with the lower bound CI matrices of DE8 and SPSO. Figure 34 is an example of comparing two stochastic optimization algorithms with PPM. The results, shown in Figure 34, show that algorithm A2 performs much better than A1 (right side). However, when the sample variances are taken into account (left side), the confidence upper bound for A2 and confidence lower bound for A1 are displayed, and it is clear that A2 performs significantly better than A1.

A comparison of approximate dynamic programming techniques on benchmark energy storage problems: Does anything work?

(Jiang et al. 2014) investigated the effectiveness of several techniques that fall under the realm of approximate dynamic programming (ADP) on a simple energy storage and allocation problem. The problem seeks to optimally control (through profit maximization) a storage device that interacts with both the grid and an uncertain energy supply. The energy storage and allocation problem is formulated as a Markov decision process. In their benchmark, they consider a stochastic wind supply, stochastic electricity prices, and a deterministic demand. They consider two algorithms that are based on value function approximation (cost-to-go function): approximate policy iteration and approximate value iteration. Several parameters where chosen to represent a common set of problems resulting in 17 stochastic benchmark problem formulations. Five popular algorithms were then tested on the 17 problems, and the percent optimality for each algorithm determined to be the value of the approximate policy divided by the value of the optimal (backward dynamic programming) policy. In their comparison, they compare the performance of each algorithm based on the set of problem architecture they solve, but since none of the algorithms consistently performed better, they concluded that none of the current techniques work reliably in a way that can scale up to more complex problems.

Benchmarking a Scalable Approximate Dynamic Programming Algorithm for Stochastic Control of Grid-Level Energy Storage

(Salas and Powell 2018) enhanced the approach used in (Jiang et al. 2014) of designing test problems to benchmark an approximate dynamic programming algorithm for a portfolio of heterogeneous storage devices in a time-dependent environment, where wind supply, demand, and electricity prices may evolve stochastically. They assessed the optimality of the back-propagation (BPTT) algorithm by comparing the performance of the resulting approximate policy to optimal for time-dependent deterministic problems with 2000 time periods, and optimal policies for a variety of stochastic problems. The deterministic comparison was done against benchmark problems that can be solved exactly using the LP formulation. The stochastic benchmarks consist of discretized problems for which the exact solution can be found. First, they benchmark BPTT on deterministic solutions, then, they benchmark on stochastic problems and compare the numerical performance of BPTT against approximate value iteration. It is important to mention that to conduct the benchmarking; they had to tune the process of state aggregation and choice of step size algorithm for smoothing new estimates into. To compare the value of the approximate policy generated by these two algorithms over iterations, they define a metric F_n which is the objective value given after n iterations, divided by the true optimal value, and a corresponding standard error. They found that the model they benchmarked was able to design time-dependent control policies that are within 0.08% of the optimal deterministic problems and within 1.34% of the stochastic ones.

Assisted History Matching Benchmarking in Petroleum Reservoir Engineering

In petroleum engineering, history matching involves the conditioning of numerical reservoir flow models to field performance through the quantitative integration of historical dynamic data (Bhark and Dehghani 2017). A study has been done to benchmark four techniques for assisted history matching that have been applied in the oil and gas industry for asset management applications. The techniques benchmarked were: Design of Experiments (DoE), genetic algorithm (GA), Ensemble Kalman Filter and Ensemble Kalman Smoother (EnKF/ES), and streamline-based generalized travel time inversion (GTTI).

The foundational capabilities of the techniques that this industry is interested in are as follows:

1. To parameterize and calibrate any reservoir model attribute
2. To integrate with any type of numerical simulation method and grid structure, for any time duration
3. To quantify data misfit
4. To output an ensemble of historically matched and calibrated flow models (posterior probability distributions) suitable for both deterministic analysis (conservative, best technical estimate) and probabilistic analysis (uncertainty range)

This study focuses primarily on the method of DoE. DoE workflows involve parameter identification for reservoir simulation, application of parameter samples in simulation of historical period, construction of proxy models for individual history matching errors, and exhaustive Monte Carlo sampling of input parameter combinations that lead to discrete model selection.

The benchmarking analysis relied on two methods. The first one is a common one-variable-at-a-time OVAT 2 point sensitivity analysis, and the other is a Standardized Pareto Analysis method t-test to represent the dependencies of parameters. The OVAT method involves perturbing a parameter at its high and low extreme values and recording its response behavior. This tool is useful for broad understanding of the positive or negative change in a modeled response due to the positive or negative change in a parameter. The authors claim that this is useful for the initial quality control of the simulation workflow definition.

A standardized Pareto analysis uses t-test for statistical significance of parameter impact on history matching error. It captures the main effects of single parameter and first order interaction effects of two parameters, and whether the magnitude of this effect is statistically significant. The t-test computes the error at min and max values of individual parameters as a ratio of the difference in the mean relative to the variance of several experiments. The calculations are repeated at different random selections of the remaining parameters to distribute potential influence of all other controlled effects on current experiment.

Benchmarking model-free and model-based optimal control

In this paper, (Koryakovskiy et al. 2017) apply benchmarking problem to optimal control Reinforcement Learning (RL) model and Model Predictive Control (NMPC) in robotics subject to uncertainties. Based on the quantitative comparison, they are able to identify the strong and weak points of both algorithms, and explore potential benefits of their combination. The study conducted in this article is set up as follows:

- **In the first step (I),** they establish optimal control ("OC") solutions for the ideal benchmark problem. Then they consider the NMPC formulation and derive the corresponding RL formulation from it. They highlight the changes introduced in both formulations and discuss their effects. Subsequently, they address the strengths and weaknesses of NMPC and RL in terms of their ability to adapt to *structural* and *parametric* uncertainties.

- **In the second step (II)**, they investigate NMPC and RL methods that are explicitly unable to adapt to uncertainties. They introduce the term *frozen* to refer to this inability.
- **In the third step (III),** the effect of uncertainties and the ability to adapt to them is analyzed for NMPC methods that have explicitly been equipped with the knowledge about the uncertainties and for RL that is allowed to interact with the real system for an additional 5% of the learning time. They introduce the term *adaptive* to distinguish these from the *frozen* methods.

The two-dimensional benchmark example studied in this article is a pendulum attached to a cart in its center of mass through a massless rod. In an ideal scenario, both the cart and the pendulum can move without friction along their respective degrees of freedom (x axis for the cart and rotary motion of the pendulum). In a second scenario, uncertainty is employed in the form of viscous friction at the rotary joint, producing an internal torque. Depending on whether or not this friction is included in the model, uncertainty in friction can be considered as a parametric or as a structural uncertainty. This problem was used to investigate control scenarios for swing-up motions of the cart-pendulum system from a given initial state. To make the two algorithms comparable, a discount rate, which is inevitable for solving a continuing task in RL, was applied to the objective function of NMPC.

For experimental measures of the “frozen” methods, they employ the R^2 coefficient of determination as a similarity measure of trajectories. It quantifies the deviation of the trajectory obtained by the algorithms from an optimal trajectory. Second, they employ regret as a measure to evaluate the performance of the methods against uncertainty. Regret, which is commonly used in RL, quantifies the amount of additional cost which is incurred due to suboptimal actions taken by a controller with respect to the optimal control actions. Lower values of regret indicate a controller, whose behavior is closer to the optimal one. Due to the stochastic nature of RL, they plot a mean value of the regret averaged over 50 runs. They summarize the results of the pendulum-cart problem in the following table.

Table 12: Summary of results of Koryakovskiy et al. 2017 benchmark problem

Table 3. Summary of results.

Category	Findings
Achieved similarity of "iNMPC" and "RL" methods on the ideal system	**More than** 90.3%
Break-even point: the difference in energy consumed by ideal and noisy systems after which "RL-adapt" performance becomes better then "iNMPC"	6.1%
Best performing algorithm under parametric uncertainties	"iNMPC-adapt"
Best performing algorithm under structural uncertainties	"iNMPC" before the break-even point and "RL-adapt" after the break-even point

In benchmarking different stochastic optimization methods, it is important to test for the algorithm strength and vulnerabilities (ex. Computation time) as well as the algorithm's policy decision outcomes (ex. Storage at target reservoirs). Table 13 represents all the techniques found to benchmark models. The first step in any benchmarking exercise is to identify the key features that the models are expected to perform, usually referred to as benchmarking standards. The second step is to identify key performance statistics. The last step is to identify a fair problem of interest to challenge the different mod

Table 13: Summary of papers reviewed, in order of publication date

Paper Title	**Reference**	**Number of Models Benchmarked**	**Type of Models**	**Methods used for Benchmarking**
Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales	Carpenter and Georgakakos 2004	2	Specially distributed and specially lumped streamflow forecast models.	KS- test statistics
Stochastic optimisation of Hydro-Quebec hydropower installations: a statistical comparison between SDP and SSDP methods	Cote et al. 2011	2	SDP and SSDP	Statistical analysis with synthetic flow scenarios, comparing storage and spill policy decisions
Optimizing Trading Decisions for Hydro Storage Systems Using Approximate Dual Dynamic Programming	Löhndorf et al. 2013	2	SDDP integrated with ADP and a deterministic LP counterpart	Case study analysis, convergence behavior
Integrating Logistical and Technical Criteria into a Multiteam, Competitive Watershed Model Ranking Procedure	Cunderlik et al. 2013	2013	UBC Watershed Model, HBV-EC, NWSRFS, WATFLOOD	Horse-race competition, application to case study. Comparison of operational logistics, comparison of hydrologic performance, and combined performance, decision matrices

Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration	Arsenault et al. 2014	10	ASA, CMAES, CS, DDS, PS, HS, PSO, DE, GA, SCE-UA	Statistical significance, convergence speed, dispersion
A comparison of approximate dynamic programming techniques on benchmark energy storage problems: Does anything work?	Jiang et al. 2014	5	Support vector regression (SVR), Gaussian process regression (GPR), local parametric methods (LPR), clustering method called Dirichlet cloud with radial basis functions (DCR). And Monotone-ADP	Percent optimality of benchmark problem
Comparison of Stochastic Optimization Algorithms for Hydropower Reservoir Operation with Ensemble Streamflow Prediction	Cote and Leconte 2015	4	Deterministic optimization, SDP, SSDP, scenario tree approach	Test bed, algorithm speed, policy comparison for storage and spillage,
Assisted History Matching Benchmarking in Petroleum Reservoir Engineering	Bhark and Dehghani 2017	4	Design of Experiments (DoE), genetic algorithm (GA), Ensemble Kalman Filter and Ensemble Kalman Smoother (EnKF/ES), and streamline-based generalized travel time inversion (GTTI)	One-variable-at-a-time OVAT and t-test standardized Pareto Analysis

Benchmarking model-free and model-based optimal control	Koryakovskiy et al. 2017	3	Reinforcement Learning (RL) model and Model Predictive Control (NMPC)	R^2 coefficient of determination and Measure of Regret
Using Confidence Intervals for benchmarking stochastic models	Liu et al. 2017	3	GA, DE, and SPSO	Confidence Intervals and Performance Profiles
Benchmarking a Scalable Approximate Dynamic Programming Algorithm for Stochastic Control of Grid-Level Energy Storage	Salas and Powell 2018	2	Back-propagation (BPTT) algorithm and LP	Comparison of value for approximate policy generated and percent-difference from optimal deterministic problem

6.2 Proposed Framework

The schematic below illustrates the steps that are proposed to test and compare the stochastic models of the BC Hydro Generation System Operations (GSO). The following table recommends the benchmarking metrics to measure the performance of stochastic optimization models.

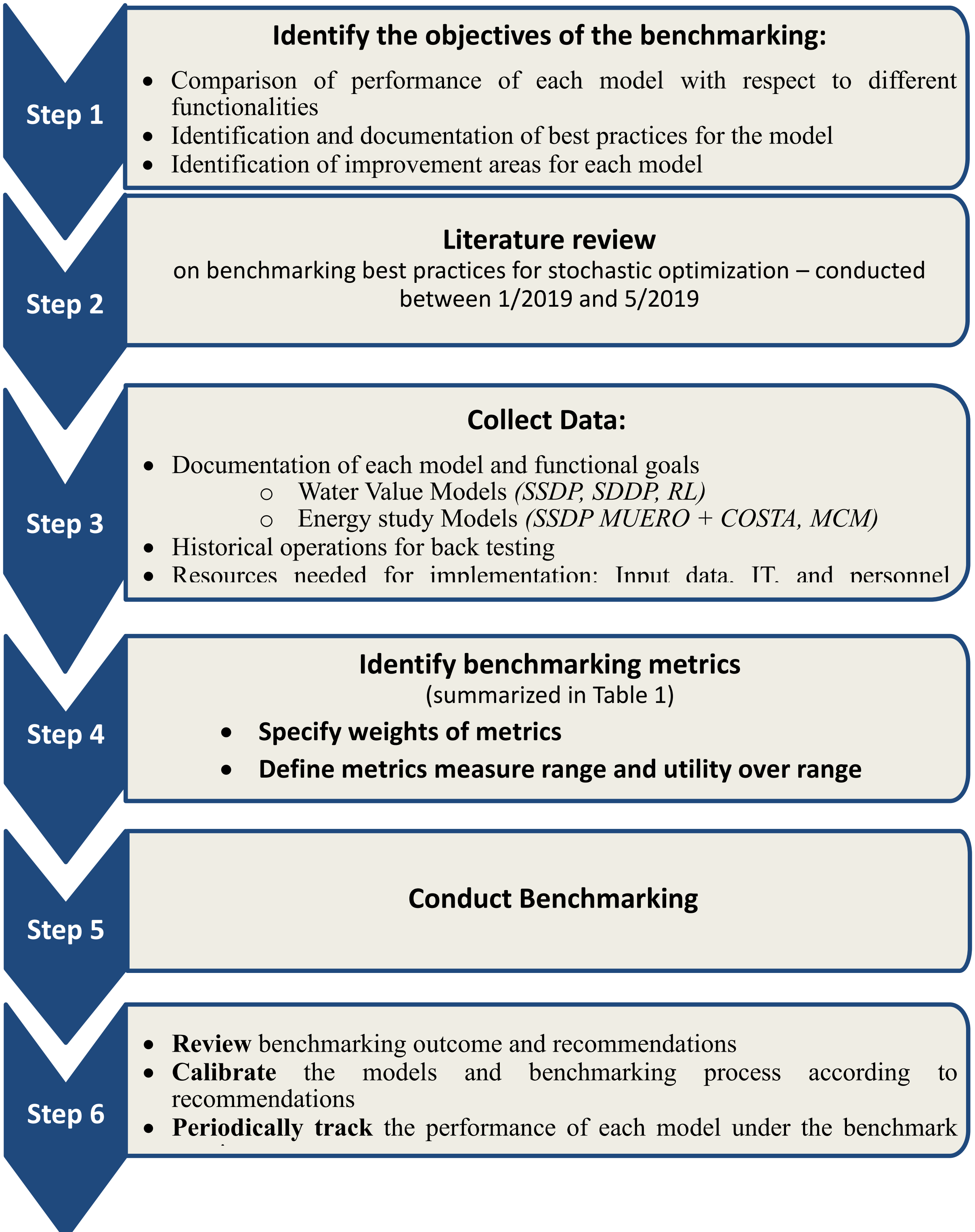


Figure 35: Schematic of the benchmarking process

Table 14: Benchmarking metrics to measure the performance of stochastic optimization models

Description of Performance Measure	Rationale/ BM Activity	Measurement
1. Model efficiency, user friendliness, and resource cost/requirements.	• To evaluate the indirect cost that each model has on the time of GSO engineers and IT resources such as personnel training time, manual labor time, and time required to run the models.	**1-a)** Surveying GSO models end-users and collecting data on the full time required to complete a study, the fraction of time for manual labor vs. automated process.
		1-b) Comparing the models for sample mean of performance profiles by diving the time by the minimum achieved time required to finish a study. Separate data collection survey for model maintenance and training costs should be conducted.
2. Model forecast performance	• To back-test and identify vulnerabilities in the models and determine the methods that best represent real conditions. • To assess the impact of different transition probabilities calculation method on the accuracy of water values.	**2-a)** Simulating a historical period using observed inputs (i.e temperature, inflows, loads, external market prices) and the forecasts that was available at each point in time, and comparing each model's outputs with observed values.
		2-b) Using statistical metrics such as Mean absolute percentage error (MAPE) to compare predicted with actual state variables, and the Goodness-of fit measure to compare the observed and simulated variables.
3. Model outcome for expected profit and operation rules	• To evaluate the model alignment with BC Hydro's objective to maximize risk	**3-a)** Measured by running each of the models using a unified inflow scenario tree and comparing the resultant policies with operations for a chosen period of time.

	neutral long-term net revenue from operations. • To back-test against short term historical operations, and identify models' vulnerabilities and shortfalls. • To identify the approaches that best support operations and planning needs. • Measuring the water value sensitivity to extreme weather (prolonged wet or dry sequences).	**3-b)** Comparing the resultant net profit of operations for operating under the policies of benchmark models, and comparing to actual historical data.
4. CRT- Evaluation of the Columbia River Treaty constraints modelling method	• The treaty coordinates flood risk management across the Columbia Basin. This impacts the Arrow reservoir by providing a maximum storage that is based on the forecast of the seasonal inflow volume the The Dalles dam.	**4-a)** Measured by tracking storage at Arrow reservoir resulting from the optimization under each method, from the runs implemented in performance measure n. 3-a.

Chapter 7: Conclusions and Recommendations

Game theory has been proven in literature and the models developed in this research to be a helpful tool in the coordination of multiple reservoirs for the optimal release and market policy. An equilibrium-based decision-making model (EDM) was developed to coordinate release and market decisions to meet demand and trade electricity from the Peace and Columbia systems in BC. The policies resulting from the game-theoretic model were compared to these of an existing iterative coordination model, Energy Studies Model. The model showed reasonable results for the Peace system with drawdowns from WSR with low MSE, while the drawdowns from KBT had larger error as compared to ES. This was explained by the over-simplification of the constraints at the KBT, REV, and ARW reservoirs and Columbia River Treaty. The results also demonstrated that the Nash equilibrium may not always be Pareto optimal, hence the rewards achieved by the system (Social Optimum) were less than the reward achieved competitively by each agent.

Game theory has four key components that need to be enhanced and adapted to help solve this problem: players, strategies, payoffs, and equilibrium. In this work, the best representations for the agents, their actions, and payoffs where modeled, and several algorithms for equilibrium were explored. It was found that the Mixed Integer Linear Programming algorithm to find Nash Equilibrium gave the best strategies. However, the support enumeration Nash algorithm is more adaptable to situations with more than two agents. Other key game components can be addressed to better represent the BC Hydro reservoir systems. For example, using inverse reinforcement techniques, the agent's utility function can be learnt from previous policies.

Although the errors do not validate the equilibrium-based decision-making model as a functional tool for modelling the BC Hydro system, the results suggest that Game Theory is a promising technique that should be further investigated to aide Energy Studies in the coordination of reservoir release policies. An effective way to validate and compare this technique and the different tools developed by the BC Hydro's system optimization group is by following a model benchmarking framework.

It is therefore recommended to further develop the proposed equilibrium-based decision-making model to incorporate the Inverse Reinforcement Learning tools presented in section 1.1. It is also recommended to follow the proposed benchmarking framework to comprehensively investigate and verify the effectiveness of game theory algorithms as tools in the multi-agent coordination of reservoir systems, compare the results with other models, and identify the gaps in each model.

Bibliography

Abbeel, P., and Ng, A. (2004). "Apprenticeship learning via inverse reinforcement learning." *Proceedings of the twenty-first international conference on machine learning*, ACM, 1.

Arsenault, R., Poulin, A., Côté, P., and Brissette, F. (2014). "Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration." *Journal of Hydrologic Engineering*, 19(7), 1374–1384.

Avis, D., Rosenberg, G. D., Savani, R., and von Stengel, B. (2010). "Enumeration of Nash Equilibria for Two-Player Games." *Economic Theory*, Springer, 42(1), 9–37.

Bab, A., and Brafman, R. I. (2008). "Multi-Agent Reinforcement Learning in Common Interest and Fixed Sum Stochastic Games: An Experimental Study." *Journal of Machine Learning Research*, 9, 41.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P. (2005). "Potential impacts of a warming climate on water availability in snow-dominated regions." *Nature*, 438(7066), 303–309.

Bauso, D. (2014). "Types of Games and Representations." *Game Theory: Models, Numerical Methods and Applications*, Foundations and Trends® in Systems and Control, Now Publishers, 379.

Bertsekas, D. P., and Tsitsiklis, J. N. (1996). "Extensions." *Neuro-dynamic Programming*, Athena Scientific, Belmont, MA.

Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., Page, T., Phillips, J. C., Rougier, J. T., Simpson, M., Stephenson, D. B., Smith, P. J., Wagener, T., and Watson, M. (2015). "Epistemic uncertainties and natural hazard risk assessment – Part 1: A review of the issues." *Natural Hazards and Earth System Sciences Discussions*, 3(12), 7333–7377.

Bhark, E., and Dehghani, K. (2017). "Assisted History Matching Benchmarking: Design of Experiments-based Techniques." Society of Petroleum Engineers, Amsterdam, The Netherlands, 35.

Chi-Keong Goh, and Kay Chen Tan. (2009). "A Competitive-Cooperative Coevolutionary Paradigm for Dynamic Multiobjective Optimization." *IEEE Transactions on Evolutionary Computation*, 13(1), 103–127.

Choi, J., and Kim, K.-E. (2015). "Hierarchical Bayesian Inverse Reinforcement Learning." *IEEE Transactions on Cybernetics*, 45(4), 793–805.

Columbia River Treaty Operating Committee. (2019). "Columbia River Treaty Detailed Operating Plan For Canadian Storage: 1 August 2019 Through 31 July 2020."

Cosgrove, W. J., and Loucks, D. P. (2015). "Water management: Current and future challenges and research directions." *Water Resources Research*, 51(6), 4823–4839.

Cote, P., Haguma, D., Leconte, R., and Krau, S. (2011). “Stochastic optimisation of Hydro-Quebec hydropower installations: a statistical comparison between SDP and SSDP methods.” *Canadian Journal of Civil Engineering*, 38(12), 1427–1434.

Côté, P., and Leconte, R. (2016). “Comparison of Stochastic Optimization Algorithms for Hydropower Reservoir Operation with Ensemble Streamflow Prediction.” *Journal of Water Resources Planning and Management*, 142(2), 04015046.

Cunderlik, J. M., Fleming, S. W., Jenkinson, R. W., Thiemann, M., Kouwen, N., and Quick, M. (2013). “Integrating Logistical and Technical Criteria into a Multiteam, Competitive Watershed Model Ranking Procedure.” *Journal of Hydrologic Engineering*, 18(6), 641–654.

Dimitrakakis, C., and Rothkopf, C. A. (2012). “Bayesian Multitask Inverse Reinforcement Learning.” *Recent Advances in Reinforcement Learning*, S. Sanner and M. Hutter, eds., Springer Berlin Heidelberg, 273–284.

Dobson, B., Wagener, T., and Pianosi, F. (2019). “How Important Are Model Structural and Contextual Uncertainties when Estimating the Optimized Performance of Water Resource Systems?” *Water Resources Research*, 55(3), 2170–2193.

Dutta, P. K. (1999). *Strategies and games: theory and practice*. MIT Press, Cambridge, Mass.

“Energy in BC.” (2020). <https://www.bchydro.com/energy-in-bc/operations/our-facilities.html> (Apr. 17, 2020).

Faria, Victor. A. D., de Queiroz, A. R., Lima, L. M. M., and Lima, J. W. M. (2018). “Cooperative game theory and last addition method in the allocation of firm energy rights.” *Applied Energy*, 226, 905–915.

Forbes, W. L., Mao, J., Ricciuto, D. M., Kao, S.-C., Shi, X., Tavakoly, A. A., Jin, M., Guo, W., Zhao, T., Wang, Y., Thornton, P. E., and Hoffman, F. M. (2019). “Streamflow in the Columbia River Basin: Quantifying Changes Over the Period 1951-2008 and Determining the Drivers of Those Changes.” *Water Resources Research*, 0(0).

Forouzandehmehr, N., Han, Z., and Zheng, R. (2016). “Stochastic Dynamic Game between Hydropower Plant and Thermal Power Plant in Smart Grid Networks.” *IEEE Systems Journal*, 10(1), 88–96.

Giuliani, M., and Castelletti, A. (2013). “Assessing the value of cooperation and information exchange in large water resources systems by agent-based optimization: MAS Framework for Large Water Resources Systems.” *Water Resources Research*, 49(7), 3912–3926.

Goldman, C. V., and Zilberstein, S. (2003). “Optimizing Information Exchange in Cooperative Multi-agent Systems.” 8.

Guan, Z., Shawwash, Z., and Abdalla, A. (2018). “Using SDDP to Develop Water-Value Functions for a Multireservoir System with International Treaties.” *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, 144(2), 05017021.

Hu, M.-C., Huang, T., Yu, H.-L., and Tung, C.-P. (2018). “Stochastic competitive analysis of hydropower and water supplies within an energy–water nexus.” *Stochastic Environmental Research and Risk Assessment*, 32(9), 2761–2769.

Hui, R., Lund, J. R., and Madani, K. (2016). “Game theory and risk-based leveed river system planning with noncooperation: GAME THEORY AND RISK-BASED LEVEED RIVER SYSTEM PLANNING.” *Water Resources Research*, 52(1), 119–134.

Hydro-Québec (Montréal). (2009). “Annual report.” *Annual report*, Montréal.

Jiang, D. R., Pham, T. V., Powell, W. B., Salas, D. F., and Scott, W. R. (2014). “A comparison of approximate dynamic programming techniques on benchmark energy storage problems: Does anything work?” *2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, IEEE, Orlando, FL, USA, 1–8.

Koryakovskiy, I., Kudruss, M., Babuška, R., Caarls, W., Kirches, C., Mombaur, K., Schlöder, J. P., and Vallery, H. (2017). “Benchmarking model-free and model-based optimal control.” *Robotics and Autonomous Systems*, 92, 81–90.

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.

Lin, X., Adams, S. C., and Beling, P. A. (2018). “Multi-agent Inverse Reinforcement Learning for General-sum Stochastic Games.” (Journal Article).

Liu, Q., Chen, W.-N., Deng, J. D., Gu, T., Zhang, H., Yu, Z., and Zhang, J. (2017). “Benchmarking Stochastic Algorithms for Global Optimization Problems by Visualizing Confidence Intervals.” *IEEE Transactions on Cybernetics*, 47(9), 2924–2937.

Löhndorf, N., Wozabal, D., and Minner, S. (2013). “Optimizing Trading Decisions for Hydro Storage Systems Using Approximate Dual Dynamic Programming.” *Operations Research*, 61(4), 810–823.

Luce, C. H., and Holden, Z. A. (2009). “Declining annual streamflow distributions in the Pacific Northwest United States, 1948–2006.” *Geophysical Research Letters*, 36(16).

Madani, K. (2010). “Game theory and water resources.” *Journal of Hydrology*, 381(3), 225–238.

Madani, K., and Hooshyar, M. (2014). “A game theory–reinforcement learning (GT–RL) method to develop optimal operation policies for multi-operator reservoir systems.” *Journal of Hydrology*, 519, 732–742.

Ministry of Petroleum and Energy. (2015). “Energifakta Norge.” *Energifakta Norge*, <https://energifaktanorge.no/en/norsk-energiforsyning/kraftproduksjon/> (Aug. 20, 2020).

Moiseeva, E., and Hesamzadeh, M. R. (2018). “Bayesian and Robust Nash Equilibria in Hydrodominated Systems Under Uncertainty.” *IEEE Transactions on Sustainable Energy*, 9(2), 818–830.

Nash, J. (1953). “Two-person cooperative games.” *Econometrica*, 21(1), 128–140.

Ng, A. Y., and Russell, S. (2000). “Algorithms for Inverse Reinforcement Learning.” *Proc. Intl. Conf. Mach. Learning (ICML’00)*, 663–670.

Nicosia, G., Pacifici, A., and Pferschy, U. (2018). “Competitive multi-agent scheduling with an iterative selection rule.” *4OR*, 16(1), 15–29.

Parrachino, I., Dinar, A., and Patrone, F. (2006). “Cooperative Game Theory And Its Application To Natural, Environmental, And Water Resource Issues : 3. Application To Water Resources.” Policy, Research working paper, Washington, DC: World Bank.

Reddy, T. S., Gopikrishna, V., Zaruba, G., and Huber, M. (2012). “Inverse reinforcement learning for decentralized non-cooperative multiagent systems.” *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1930–1935.

Ristić, B., and Madani, K. (2019). “A Game Theory Warning to Blind Drivers Playing Chicken With Public Goods.” *Water Resources Research*, 55(3), 2000–2013.

Rolstadås, A., and International Federation for Information Processing. (1995). *Benchmarking: theory and practice*. Springer, Boston, MA.

Roughgarden, T. (2010). “Computing equilibria: a computational complexity perspective.” *Economic Theory*, 42(1), 193–236.

Salas, D. F., and Powell, W. B. (2018). “Benchmarking a Scalable Approximate Dynamic Programming Algorithm for Stochastic Control of Grid-Level Energy Storage.” *INFORMS Journal on Computing*, 30(1), 106–123.

Sandholm, T., Gilpin, A., and Conitzer, V. (2005). “Mixed-Integer Programming Methods for Finding Nash Equilibria.” *Proceedings of the National Conference on Artificial Intelligence*.

Sargent, T. J., and Stachurski, J. (2019). “Quantitative Economics with Julia.” <https://julia.quantecon.org/>.

Seibert, J. (2001). “On the need for benchmarks in hydrological modelling.” *Hydrological Processes*, 15(6), 1063–1064.

Shawwash, Z. K., Siu, T. K., and Russell, S. O. D. (2000). “The B.C. Hydro short term hydro scheduling optimization model.” *IEEE Transactions on Power Systems*, 15(3), 1125–1131.

Stewart, I. T., Cayan, D. R., and Dettinger, M. D. (2005). “Changes toward Earlier Streamflow Timing across Western North America.” *Journal of Climate*, 18(8), 1136–1155.

Tilmant, A., and Kinzelbach, W. (2012). “The cost of noncooperation in international river basins: THE COST OF NONCOOPERATION.” *Water Resources Research*, 48(1).

Tuyls, K., and Nowé, A. (2005). “Evolutionary game theory and multi-agent reinforcement learning.” *The Knowledge Engineering Review*, 20(01), 63.

Watkins, C. J. C. H., and Dayan, P. (1992). “Q-learning.” *Machine Learning*, 8(3), 279–292.

Wu, X., Li, S., Cheng, C., Miao, S., and Ying, Q. (2019). "Simulation-Optimization Model to Derive Operation Rules of Multiple Cascaded Reservoirs for Nash Equilibrium." *Journal of Water Resources Planning and Management*, 145(5), 04019013.

Yang, Y.-C. E., Cai, X., and Stipanović, D. M. (2009). "A decentralized optimization algorithm for multiagent system–based watershed management." *Water Resources Research*, 45(8), W08430.