OPTIMAL RESERVOIR OPERATION SYSTEM BASED ON ARTIFICIAL INTELLIGENCE AND METAHEURISTICS ALGORITHMS

KARIM SHERIF MOSTAFA HASSAN IBRAHIM

DOCTOR OF PHILOSOPHY (ENGINEERING)

LEE KONG CHIAN
FACULTY OF ENGINEERING AND SCIENCE
UNIVERSITI TUNKU ABDUL RAHMAN
SEPTEMBER 2023

OPTIMAL RESERVOIR OPERATION SYSTEM BASED ON ARTIFICIAL INTELLIGENCE AND METAHEURISTICS ALGORITHMS

By

KARIM SHERIF MOSTAFA HASSAN IBRAHIM

A thesis submitted to the Department of Civil Engineering,
Lee Kong Chian Faculty of Engineering and Science,
Universiti Tunku Abdul Rahman,
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy (Engineering)
September 2023

ABSTRACT

OPTIMAL RESERVOIR OPERATION SYSTEM BASED ON ARTIFICIAL INTELLIGENCE AND METAHEURISTICS ALGORITHMS

Karim Sherif Mostafa Hassan

Over the immediate past decades, global warming across the world and in Malaysia has caused extreme changes to the climate by disturbing its hydrological cycle, causing drought to some areas and floods to another. As a result of this, it has become more important to meet the rising need for water in order to facilitate increased agricultural, residential, and industrial production. Therefore, the process of developing a prediction for water inflow and optimising the release operation in a reservoir has become an increasingly crucial undertaking for the management of a reservoir. Conventionally, the simulation and forecasting scenarios were calculated using traditional linear models but however, traditional linear models lack the capability to grasp the dynamic and non-linear aspects inherent in hydrological applications. Therefore, if inflow projections were more accurate, there would be a greater need for monitoring of the water quality, and the management of the reservoir would be more effective. Following this, the difficulties caused by flash flooding and the water crisis in Malaysia and the rest of the world may be mitigated with the aid of machine learning. In this study, four different machine learning models were proposed to forecast the reservoir inflow namely, Support Vector regression (SVR), Multi-layer perceptron neural

network (MLPNN), Adaptive Neuro Fuzzy Inference System (ANFIS) and Extreme Gradient Boosting (XG-Boost). All the four models were given historical data for training and testing that were collected over 19 years (2000-2019) at Klang Gate Dam which is located in the Gombak District. The data was divided into monthly and daily timeframe while the daily is further time lagged and subdivided into 7 main scenarios starting with scenario-1 to scenario-7 with one, three, and five-days lag for water level and inflow. Four optimisation algorithms were used to simulate reservoir operation over a 12-month period and generate a release curve. For the sake of validation four statistical analysis namely coefficient of determination (R^2), Mean Square Error (MSE), Median Absolute Error (Mead) and root mean squared error were adapted to validate and test the machine learning models. On the other hand, risk analysis test was carried out to test the optimisation algorithm output. The risk analysis consisted of reliability, resiliency to recover from failure, vulnerability degree and finally sustainability. Results revealed that: (1) For the monthly period, all four models were capable of predicting effective monthly reservoir inflows by achieving at least an R^2 of 0.5; the XG-Boost model was rated as the best model, followed by the MLPNN, SVR, and finally ANFIS. (2) For forecasting daily inflow, the XG-Boost still surpasses all other models, but with diminished efficiency. The models were still placed in the same order, with the ANFIS doing very poorly in scenarios 2, 3, and 4. (3) The best scenarios for daily inflows are scenarios 5, 6, and 7, since the models were developed using 1, 3, and 5 days of anticipated inflow, and XG-Boost consistently beats all other models. Moreover, for the optimisation phase all four models were able to produce release curves with a certain degree of validity. Firefly algorithm (FA) was ranked as the best optimisation algorithm after obtaining a vulnerability percentage of 39% as the finest optimisation algorithm model; followed by the particle swarm (PSO) at 59.99%, then the genetic algorithm (GA) with a vulnerability of 68%, and lastly, the nuclear reaction (NRO) with 72%. The present study proposes a novel approach for developing a one-size-fits-all system for forecasting water reservoir inflow and optimizing reservoir operation. The proposed system is based on a combination of machine learning and optimisation techniques, and it has the potential to eliminate the need for tedious and time-consuming manual data collection and interpretation. This could lead to significant improvements in the efficiency and effectiveness of water resources management, and it could also facilitate the development of more informed and timely water resources policies at the national and international levels.

ACKNOWLEDGEMENT

LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 18/09/2023

SUBMISSION OF THESIS

It is hereby certified that **Karim Sherif Mostafa Hassan Ibrahim** (ID No: **20UEM01058**) has completed this thesis entitled "OPTIMAL RESERVOIR OPERATION SYSTEM BASED ON ARTIFICIAL INTELLIGENCE AND METAHEURISTICS ALGORITHMS" under the supervision of Ir. Prof. Dr. Huang Yuk Feng (Supervisor), and Ir. Dr. Koo Chai Hoon (Co-Supervisor) from the Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, and Dr. Al Mahfoodh Ali Najah Ahmed (Co-Supervisor) from the Department of Civil Engineering, Universiti Tenaga Nasional (UNITEN).

I understand that University will upload a softcopy of my thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

(Karim Sherif Mostafa Hassan Ibrahim)

APPROVAL SHEET

This dissertation/thesis entitled "**OPTIMAL RESERVOIR OPERATION SYSTEM BASED ON ARTIFICIAL INTELLIGENCE AND METAHEURISTICS ALGORITHMS**" was prepared by KARIM SHERIF MOSTAFA HASSAN IBRAHIM and submitted as partial fulfilment of the requirements for the degree of Doctor of Philosophy (Engineering) at Universiti Tunku Abdul Rahman.

Approved by:

(Prof. Ir. Dr. Huang Yuk Feng)
Date:…18/9/2023…………….
Professor/Supervisor
Department of Civil Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

(Ir. Dr. Koo Chai Hoon)
Date:…18/9/2023……
Associate Professor/Co-supervisor
Department of Civil Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

Dr. Al Mahfoodh Ali Najah Ahmed
Senior Lecturer
Dept. of Civil Engineering
College of Engineering
Universiti Tenaga Nasional

(Dr. Al Mahfoodh Ali Najah Ahmed)
Date:…18/9/2023………………...
External Co-supervisor
Department of Civil Engineering
College of Engineering, Universiti Tenaga Nasional (UNITEN)

DECLARATION

I hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

Name Karim Sherif Mostafa Hassan Ibrahim

Date :18/09/2023___________________

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACO	Ant Colony Optimisation
KGD	Klang Gate Dam
RCGA	Real Coded Genetic Algorithm
ANFIS	Adaptive Network-Based Fuzzy Inference System
ANN	Artificial Neural Network
MLPNN	Multi-layer Perceptron Neural Network
MCM	Million Cubic Meter
DP	Dynamic Programming
EC	Evolutionary Computation
PSO	Elitist-Mutated Particle Swarm Optimisation
NRO	Nuclear reaction Optimisation
FA	Firefly Optimisation Algorithm
FT	Fourier Transform
GA	Genetic Algorithm
GP	Genetic Programming
IDP	Incremental Dynamic Programming
SVR	Support Vector Regression
SVM	Support Vector Machine
XG-Boost	Extreme Gradient Boosting
CS	Conventional Reservoir Operation System
IS	Integrated Losses Reservoir Operation System

CHAPTER 1

INTRODUCTION

1.1 General Introduction

It is well acknowledged that the global and regional communities are now grappling with one of the most significant environmental and social challenges: the water crisis. It is projected that climate change brought on by human activity, sometimes known as so-called global warming, will make the current global water issue even more severe. Global warming has caused numerous changes in the large-scale hydrological cycle, including an increase in atmospheric water vapor content, altered precipitation patterns, intensities, and extremes, decreased snow cover and widespread ice melting, and adjustments to soil moisture and run-off, have been linked to the observed warming over the past several decades.

Considering that there is a shortage of fresh water, water resource management is crucial for efficiently managing the limited supply of water that is now accessible and for sustaining it. It is possible to describe water resource management as the process of planning, generating, providing, and overseeing the best use of water and its resources. Additionally, managing water resources is crucial in averting natural calamities like drought and flood. Dams/water-reservoirs are one kind of infrastructure that will directly contribute to regulating the quantity of fresh water and mitigating natural disasters. The reservoir inflow parameter is crucial to the system's proper management and functioning. In order to minimise the effects of water surpluses and water deficits, decision-making in the water reservoir sector requires good forecasting for the inflow parameter. Prior to the development of technology, reservoir level forecasting was done via the use of the available hydrologic data and a traditional linear water balance Equation. Due to the stochastic nature of meteorological processes like evaporation, rainfall, and temperature, the findings, however, were not precise and were questionable.

Recently, artificial intelligence (AI)-based data-driven approaches have developed as key tools to overcome the disadvantages of traditional modelling methodology. These techniques are driven by the data. These data-driven techniques are based on collecting and reusing information implicitly present in hydrological time series. This is done without explicitly taking into consideration the physical rules that underpin the process.

For planners and decision-makers, operating dam reservoirs is one of the greatest difficulties in effectively using the available water supplies. Finding the right operational guidelines for the water stored behind the dam is their primary goal to ensure that it is used as efficiently as possible to satisfy various water usage needs.

Due to the current availability of water resources, there is a significant lot of interest in developing a computer-aided optimisation model to address the challenge of finding the best operating rule for a dam reservoir (Zor et al., 2017). The need to master a suitable advanced optimisation method to effectively conduct and deal with difficult applications has received substantial attention since a dam reservoir system is very nonlinear and complex with various limitations and purposes (Damian, 2019). Furthermore, when it comes to the design, development, management, and operation of water resources systems, there is always an element of uncertainty. It is caused by the fact that many of the elements that influence the performance of water resources systems are not known with certainty and cannot be understood with certainty while a system is being planned, constructed, built, managed, or operated (Loucks and van Beek, 2017) perhaps one of the effective tools available to overcome these uncertainties would be a metaheuristic algorithm.

The metaheuristics are derived from a variety of natural principles, drawn from fields such as biology, ethology, and physics. Modern metaheuristics are significantly more flexible than traditional optimisers (Sörensen et al., 2018).

Unlike traditional optimisers, modern metaheuristics do not require the objective and constraints to be continuous, differentiable, linear, or convex. Additionally, modern metaheuristics can typically solve large-scale problems in an effective manner. In spite of the fact that a metaheuristic may not identify the best answer, it can often find one that is almost optimal.

The research that is presented in this thesis intends to construct a reliable and robust machine learning model for the purpose of predicting reservoir inflow. It is also aimed at deriving an optimum operational strategy for reservoirs through metaheuristic algorithms and, finally examining the reservoir operation performance. This will allow the problem of uncertainty to be resolved. In the next parts, the problem statement, the aim and objectives, and the possible contribution of the study are provided.

1.2 Problem Statement

The effects of climate change extend to every living thing on the planet, putting their continued existence in jeopardy. Because of the major changes in water dynamics that have resulted from this, immediate action must be taken to avert any negative impacts and the catastrophes that may follow them.

These problems may be solved by regulating the amount of water that is already accessible. The planning and forecasting of reservoirs are implemented in order to tackle these problems. Even while reservoir forecasting is an extremely helpful tool for both civil and environmental engineering, it does not, on its own, offer an optimal answer for the process of formulating a reservoir operation strategy. As a result, it needs to be supported with optimisation algorithms so that engineers may make credentialed judgments on acceptable water resources, whether short-term or long-term, in the manner that is both the most accurate and the most efficient.

In the past, forecasting reservoir levels were accomplished by the use of mathematical water volume balance Equation. The basic water Equation, also known as the water balance Equation, is no longer suitable and accurate as a result of an increase in the amount of data as well as the inclusion of numerous hydrological variables as inputs (Jothiprakash and Magar, 2012). This is in addition to the lack of other data for various reasons. Because of this, it became necessary to use models based on artificial intelligence as a replacement tool for reservoir forecasting.

Despite the satisfactory performance exhibited by certain conventional AI algorithms like ANFIS, SVM, and ANN in reservoir storage or release estimation, these methods still possess certain limitations. For instance, they have restricted feature extraction capabilities and tend to require more time when utilized as stand-alone models (Zhang et al., 2018). Hence, this current research aims to overcome such limitations by developing an artificial intelligence model coupled with optimisation algorithms to enhance the model performance.

The management plan for a reservoir depends on stochastic factors as well as the values of deterministic factors and preset behavior, where even little changes may have a big impact on the efficacy of water release and the functionality of the reservoir (He et al., 2018). The formulation of an adequate strategy for the operation of reservoirs is complicated and difficult as a result of all of these factors. The quantity of water released from a reservoir is governed by operating rule curves, which establish relationships between parametric and non-parametric variables such the volume of water flowing into the reservoir, the amount of water evaporating, and the reservoir's level of water.

In order to operate a reservoir system in real-time, A forecasting model and rule curves may be coupled together to approximate the inflow of water as a stochastic vector, despite the historical usage of linear programming in Hydrology field it still requires further validation.

The algorithmic dexterity of classical optimisation techniques, such as nonlinear programming, linear programming, and dynamic programming, is limited in several ways. For example, linear programming is incapable of resolving issues with nonlinear or non-convex objective functions. Additionally, the computational time tends to increase when addressing problems with a large number of decision variables (Chitsaz and Banihabib, 2015). While most traditional optimisation techniques take into account several different goals, they typically do so in a limited manner by examining each aim separately and assigning weights. Hence, such methodologies have shown an imbalance between exploitation and exploration at different phases of the search (Reddy and Kumar, 2006; Črepinšek et al., 2013).

In addition to all the above, in many published studies, projected data are typically treated as separate entities and not integrated into reservoir operations. Moreover, the absence of comparisons between optimisation methods based on forecasted inflow by machine learning models attempting to build an ideal reservoir operating rule curve is a foregone conclusion and there is currently a lack of a comprehensive optimisation system that utilizes artificial intelligence model outputs to generate an optimal release strategy.

In this situation, it is necessary to look into innovative techniques to help in creating an ideal water release strategy for dams and reservoirs that could minimise the gaps between the water demand and release.

1.3 Aim and Objectives

The main goal of this research project is to develop a powerful hybrid system for inflow forecasting using a machine learning model on a daily and monthly time frame basis. Following this, a simulation of the reservoir operations followed by optimising the operation is carried out to seek optimal release curves. The specific objectives of this study are as stated in the four points below:

i) To develop four AI models, namely Support Vector Regression (SVR), Multilayer perceptron Neural Network (MLPNN), Extreme Gradient Boosting (XG-Boost), and Adaptive Neuro-Fuzzy Inference System (ANFIS) for reservoir inflow forecasting.

ii) To optimise reservoir operation by adopting four novel optimisation algorithms, the (Genetic Algorithms (GA), the Particle swarm (PSO), the Nuclear reaction Optimisation Algorithm (NRO) and the Firefly Optimisation Algorithm (FA).

iii) To propose smart optimising release guidelines through simulating reservoir water balance, based on the observed historical data, using the AI techniques as stated in objective one, along with the loses; into the four proposed optimisation algorithms through developing a semi-autonomic optimisation system.

iv) To assess the reservoir performances operation under each model and to identify the most effective optimisation algorithms, in terms of reliability, resilience, vulnerability, and sustainability indicators.

1.4 Scope of study

This research places a significant emphasis on the use of artificial intelligence models and metaheuristics algorithms in the optimisation of a reservoir system's release strategy. Peninsular Malaysia is the location of the researched study region with the utilisation of data for 20 years. The performance of the model, which is based on AI approaches, was examined, and compared to the findings acquired from simulations, which were created by feeding the model historical data to reduce the prediction's uncertainty.

In detail, a one-of-a-kind optimisation model with two stages was developed, with the optimisation process being divided into "actual" and "predicted" inflows on the basis of three separate quantitative hydrological periods, which are designated as "high," "medium," and "low" inflows, respectively.

In addition, one of the scopes of this study is to perform a simulation of the reservoir condition by integrating evaporation to account for reservoir losses. A detailed description of the scope of the study and the data utilised are outlined in Chapter 3 of this thesis.

1.5 Contribution of the study

The uniqueness and innovation of this research is that it undertakes a second analysis on the combined application of inflow projections with various lead periods at KLANG Gate Dam. Developing an optimum reservoir operating strategy using projections for the inflow in the form of ranges (High, Medium, and Low inflow) rather than fixed values, hence, overcoming the inflow uncertainties and expanding the search space which resulted in generating a wider range of outputs hence increasing the overall operational dependability by fusing both instrumental and reconstructed oscillation data to optimise reservoir operation.

Finally, the contribution of this study towards the scientific community is that it closes one of the gaps found in the field which is the estimation of reservoir losses and the effect of these losses on the reservoir operation as evaporation losses are significantly important to the reservoir and in most of the cases the data are not available and need to be estimated (Allawi, Jaafar, Mohamad Hamzah, Abdullah, et al., 2018; Dashti Latif et al., 2021; Hamdan and Zaki, 2016).

This study is novel as it developed a closed loop system that forecast, simulate, and optimise the reservoir operation with a minimal human interference as once the data are being fed to the system, the system will iterate and dynamically Configure itself to find the optimum needed output by solving the non-linear relationship between the water demand and release to the inflow aiming to address the environmental uncertainties, such closed loop system in information technology (IT) field is defined as an autonomic computing system (Obienu, 2018).

From the standpoint of the national economy, developing a reservoir operating system is a significant decision-support tool. Although the established system's accuracy relies on the data's accessibility, it may be utilised for any case study in Peninsular Malaysia. Guideline makers can precisely utilise this information when creating suitable plans for supplying water, timing irrigation, and other activities in order to encourage the recovery and robustness of the national economy, which, regrettably, had been declining in recent years.

1.6 Outline of thesis

This thesis is divided into five chapters, including this one. The context, purpose, and goals of this research, as well as its scope and contributions, are briefly described in Chapter 1.

The discussion in Chapter 2 centers on the conclusions drawn from earlier research investigations that were carried out in the same area; the literature analysis will include optimisation techniques and machine learning models. This highlights current research trends, along with their benefits and drawbacks. The purpose and uniqueness of this study activity (research gap) are described at the conclusion of Chapter 2. The whole research technique, including the development of base and hybridised models to identify their ideal inputs and structures as well as the testing methods (various training and testing cases) to validate the reliability of the models, is described in Chapter 3. In Chapter 4, the performance of the basic models under various input combinations is discussed, along with the release curves that each model produced. This chapter offers a succinct overview of the findings.

In Chapter 5, a short and concluding comment is provided along with suggestions for future research projects linked to the foundation/basis established by this research study.

CHAPTER 2

LITERATURE REVIEW

2.1 Forecasting-Based AI Models

2.1.1 Artificial Neural Network

Neural networks are a subset of artificial intelligence. Since the 1980s, artificial neural networks (ANN) have gained a lot of attention in the field of artificial intelligence (Specht, 1988; Jang, 1991; Zhang et al., 1998). It develops a simplified model which abstracts the neural network of the human brain from the viewpoint of information processing and composes various networks based on multiple connections. Even if each neuron does its task exceedingly slowly and incorrectly, an ANN may learn from examples and then generalise to situations that have never been observed. A network can collaborate effectively, such as with a data collection function (Wu and Feng, 2018).

The number of different neural networks continues to mushroom at an alarming rate. When there are so many new architectural styles and methods to choose from, one needs a map to guide them. Despite the massive growth of neural networks, some well-known models were tested for forecasting the future and improving one's ability to make choices based on that prediction.

The following list summarises the most reliable models (Leijnen and Veen, 2020; Hill et al., 1994; Hounmenou, 2021):

- Multilayer Perceptron Neural Network (MLPNN)
- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)
- Deep Feed Forward (DFF)
- Radial Basis function Network (RBNN)
- Feed Forward Neural Network (FFNN)
- Convolutional Neural Network (CNN)

The multilayer perceptron does not make any previous assumptions about the distribution of the data, in contrast to conventional statistical approaches. It can be programmed to properly generalise when given raw, untested data and is capable of modelling extremely non-linear functions.

These characteristics make the multilayer perceptron an appealing choice for creating mathematical simulations, as well as for evaluating between statistical methods. MLPNN general structure can be seen in Figure 2.1.

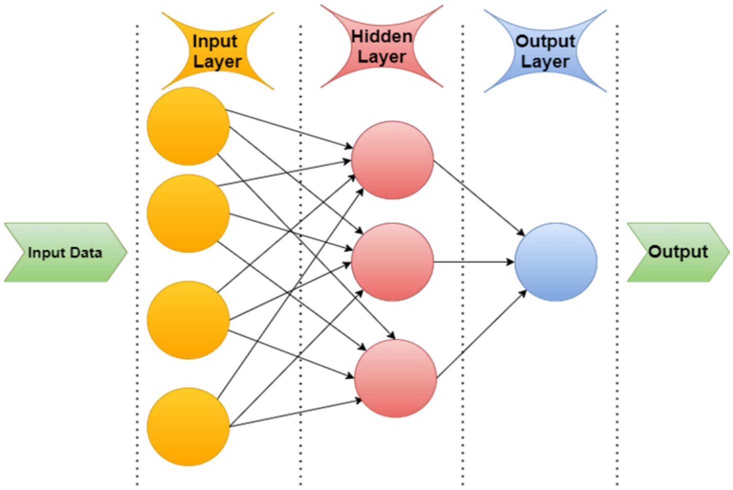


Figure 2.1: MLPNN Model Structure

(Gardner and Dorling, 1998; Taud and Mas, 2018; Ramchoun et al., 2016) With all being said about the multilayer perceptron neural network, it is convincing enough to be selected in this study; however, it will be unfair if a full literature review were not conducted. Therefore, the upcoming few paragraphs will cover literature reviews, including, but not limited to MLPNN, LSTM, and BPNN.

For the purpose of examining data-driven approaches, particularly the ANN, for the prediction of monthly river flow, research located in Turkey was carried out (Terzi and Ergin, 2014). The (RBFNN) was used with a particular dataset of 480 monthly flows since there was an abundance of historical data that was easily available. The Feed-Forward Neural Network (FFNN) was used in contrast to the autoregressive model and ANFIS, and both of these models were evaluated (AR). The output was generated using a kernel transferring function of a Gaussian function in the hidden layer, and the number of hidden neurons was chosen by trial and error. Although it was demonstrated in the study that the FFNN and the RBFNN could forecast streamflow effectively and generate outstanding results, it did not achieve the optimal result, thus it is required to test more models.

(Ababaei et al., 2013) proposed a modified RBFNN, FFNN, and GRNN as reservoir inflow forecasting models by applying K-Nearest Neighbour (KNN) as a data preprocessing technique to improve the model performance to carry a fair comparison between the modified and stand-alone model. When results from stand-alone models are compared to those from the data fusion approach, it becomes clear that the data fusion method may significantly enhance the outcomes of individual models and RBFNN yield better results than the other models.

In a follow-up study on the effectiveness of KNN as a data pre-processing technique (Akbari and Afshar, 2014) implemented KNN to estimate error prediction of real-time inflow forecasting models in two catchments. The updated inflows are compared to those of persistence-based approaches such as autoregressive (AR) and artificial neural network (ANN) models.

The findings indicate that similarity-based error prediction models have the potential to be recognised as an effective option for real-time inflow forecasting, particularly in situations where the persistence in the error series generated by the flow forecasting model is relatively low; however, it will not necessarily yield better results and it is computationally expensive as it complicates the model and adds more iteration time.

In light of the fact that hydrological time series is made up of several frequency components and has nonlinear interactions, data pre-processing techniques need to be utilised to enhance the performance of models used for forecasting. Beside s KNN, one of the primary data pre-processing techniques would be wavelet transform or wavelet decomposition (Nourani et al., 2014; Sifuzzaman et al., 2009; Minu et al., 2010).

Wavelet analysis is a multi-resolution analysis that may be performed in both the time and frequency domains. Controlling the scaling and shifting of a signal representing a time series allows the wavelet transform to breakdown the signal into many resolutions. In the time domain as well as the frequency domain, it has excellent localization characteristics. It also has the benefit of being flexible in terms of picking the mother wavelet, which is the transform function, according to the features of the time series. This is another one of its advantages (Shensa, 1992).

Additionally, a variety of wavelets, such as the Daubechies, Symmlet, Coiflet, Meyer, Gaussian, Mexican caps, Morlet, and Shannon wavelets, may be used in conjunction with machine learning to forecast non-linear time series data. (Minu et al., 2010).

Owing to the good performance achieved by Wavelet decomposition and the fact that reservoir inflow is hard to be predicted and requires a reliable model, (Seo et al., 2015) decided to fuse Wavelet with WA-MLPNN and compare it with the ordinary MLP model. The results showed that the WA-

MLPNN is superior to the ANN with an R value of 0.979 to 0.967, respectively.

Data pre-processing is important to build a reliable and efficient model; however, what is significant is identifying the optimum combination of data for the model to forecast and produce the desired output. A study conducted in 2016 by (Sauhats et al., 2016) put the ANN model through its paces using four distinct input scenarios, such is predicting inflow based on the required day temperature and previous data on temperature and precipitation (Historical). According to the findings of the research, each of the factors that went into the model produced positive predicting outcomes. However, to validate forecasts, further research is required on different network designs and model properties, among other things.

The scientific community has acknowledged the need for more effectively chosen inputs in artificial neural network models (ANNs); hence (Panagoulia et al., 2017) have performed a trial-and-error multi-stage method for input selection variables for the MLP model to forecast river flow. A validation process was carried out by comparing MLP to auto-regressive model. Few combinations of input were used, and the suggested technique identified the last four days' river flow, the past three days' worth of precipitation, and the seasonality as reliable input factors. The author concluded that further study should be carried to identify the optimum input combination for ANN models.

Scientists and researchers have been in a dilemma since artificial neural work was developed as the challenge of selecting the appropriate network size and topology is one of the most significant obstacles that researchers who use neural networks must overcome. The issue is made much more difficult by the fact that when the neural network is taught to make extremely few mistakes, there is a chance that it will not react correctly to patterns that were not used during the training process (Sheela and Deepa, 2013).

Generally, four questions need an answer when using any neural network which are:

- What kind of neural network architecture need to be utilised?
- What is the optimal number of neurons to use?
- How many different patterns ought to be employed during the training process?
- Which algorithm for training should be implemented?

To our regret, there is no simple response to any of these issues. Practically speaking, it is exceedingly challenging to establish a decent network architecture based just on the of inputs and outputs. Practically speaking, it is particularly challenging to develop a decent network architecture based just on the quantity of inputs and outputs. It greatly relies on the quantity of training instances and the difficulty of the classifications being learned.

Millions of hidden layers are needed to solve certain issues with a single input and output, whereas just one hidden layer, or none at all, is needed to solve challenges with a million inputs and outputs. However, it is true that better results may be produced with less training time by selecting the appropriate combination of hidden layers and neurons, Thus, prioritising the search for the optimal ratio of neurons and hidden layers is essential. (Panchal et al., 2011; Karsoliya, 2012; Sheela and Deepa, 2013; Çolak, 2021).

In a comprehensive investigation carried out by (Zhang et al., 2018), the neural network model was trained using twenty-two years' worth of historical inflow data and then validated using eight years' worth of inflow data. The discovery demonstrates that the maximum number of parameters and the number of hidden nodes, both of which have an impact on the accuracy of the simulation, are two important factors for any neural network model.

Iterations and simulation accuracy have a relationship that is directly proportional to one another, meaning that an increase in the maximum number of iterations will also result in an increase in accuracy. A comparison was made between the LSTM and the SVR models and MLPNN trained with back propagation. The results of the comparison show that the LSTM is more accurate than the BP-MLPNN and the SVR, respectively with corresponding R values of 0.9996, 09799, and 0.9404. The author (Zhang et al., 2018) concluded that determining the optimal number of iterations for the BP-MLPNN model should be a top priority to increase the accuracy of the model.

For the objective of forecasting river discharge and comparing various network architectures, a research report published in 2018 used back-propagation neural network models (BPNN). This choice was taken after considering the suggestion to test alternative ANN designs. The model with the choice to have 3-5-1 layers was shown to perform the best after the research examined three separate models with three distinct sets of layers (2-4-1, 3-6-1, and 3-5-1). However, the research underlined that there is no conventional method to identify the optimal number of layers and that this optimal number can only be established by trial and error (Ghose, 2018).

Despite the fact that earlier research explored a variety of activation functions and artificial neural network models using a variety of topologies, the climatic indicators have not yet been subjected to any more scrutiny. When calculating reservoir inflow for effective dam operating procedures, the climatic variability should be considered because it has a significant impact on the climatic cycle. This is especially important when the weather conditions are difficult (Maity and Kumar, 2009).

For the purpose of optimising reservoir operations and maximising hydropower production, a method that is both general and scalable has been presented for predicting reservoir inflow and investigating the optimum weather factors to be considered. A three-layered artificial neural network (ANN) that was hydrologically relevant was fed short-term weather predictions and antecedent hydrological data to forecast reservoir inflow for a lead time of seven days. The scheme's applicability was shown across 23 dams in the United States, each of which had unique hydrological features and climatic regimes. An artificial neural network was fed ensembles of weather prediction fields to investigate probabilistic forecasting. According to the results, it seems that one's ability to predict accurately increases along with a decrease in the coefficient of variation in the reservoir inflow and an increase in drainage area, moreover the results shown that MLPNN is a robust forecasting model due to its reliable performance when projecting water inflow across the 23 dams used in the study (Ahmad and Hossain, 2019).

Researchers in 2019 employed an ANN model that depended on regression parameters and a combination of exogenous components to predict reservoir inflow(Kim et al., 2019). There was a total of twenty-four distinct climatic variables considered. The autocorrelation function, also known as the ACF, was used so that the optimal elements that had the most impact on reservoir inflow could be determined. In light of this, the NINO12 and AMO indices were chosen as the best options. The MLP-ANN model was put through a rigorous test against the other seven AI models, and the results were compared to one another. Despite this, it has shown the greatest performance out of the eight models that were selected, which implies that the ANN is a model that is accurate and has a good capacity for predicting (Kim et al., 2019).

The researcher's curiosity has been sparked by recent events relating to environmental factors. Recent studies have contrasted the NAR and FFNN with two artificial neural network (ANN) models, known as static and dynamic models, for predicting reservoir input. The models were updated with a small number of additional inputs, which included precipitation data and time lags ranging from zero to twelve days. In the course of the research, the question of how many hidden neurons should be used to create the best possible ANN structure model and the environmental factors that influence the model performance was investigated. The NAR dynamic neural network triumphed over both static and dynamic models previously developed. It is strongly advised that in order to increase the accuracy of weather forecasting, precipitation data as well as the date and time index (t) must be considered with at least a three-day lag (Q(t-3)). It was demonstrated that while forecasting high and peak inflows, vibrant neural networks—which operate on records that are being updated automatically and immediately once recorded better than other forms of neural networks (Hadiyan et al., 2020).

Recent research released in 2022 (Ibrahim et al., 2022) enhanced MLPNN and used it to make predictions about the amount of water entering the reservoir on a daily and monthly basis. One-day, three-day, and five-day delays are offered for the daily data. Grid-search and trial-and-error methods were used to fine-tune the MLPNN's parameters. MLPNN was compared to three other forecasting models which were SVR, XG-BOOST, and ANFIS. The overall results revealed that XG-BOOST obtained a better score than MLPNN despite the minor difference in performance between the top two models. Due to this reason, the author recommended that XG-BOOST is an effective model for forecasting purposes.

A summarisation extracted from all the previous literature review conducted generally on artificial neural network models and specifically on MLPNN is shown in the bullet points below:

- At least 60% of the available data is required for training the model.
- The choice of input variables is crucial since it affects the effectiveness of the model.
- Identifying the optimum number of neurons and hidden layers is extremely important to seek outstanding results.
- There is no standard procedure to identify the model parameters, however the most common methods are grid-search and trial-and-error approach.
- MLPNN is a robust model that can be applied to various case studies.
- The main performance test carried out to test any neural network model would be coefficient of correlation (R^2).

2.1.2 Support Vector Regression/Machine (SVR/SVM)

Support vector machines, also known as SVMs, that were developed by (Drucker et al., 1997) are a type of advanced machine learning technique that is based on structural risk minimization (SRM). SRM's most fundamental principle is to concentrate on lowering an upper limit on the simplification error rather than the training error. The SVM produces the optimum network structure using this theory as its foundation. Additionally, the SVM solution is always distinct and globally optimum since it is the same as solving a linear restricted quadratic programming problem. This technique minimises the expected error of a learning model and reduces the issue of over-fitting.

The (SVR) approach, on the other hand, is a regressive or forecasting strategy that maintains all the crucial components of the optimum margin methodology. SVR had proven its regression ability way back in time when it was tested by (Drucker et al., 1997; Müller et al., 1997) to predict time series data.

Grasp these three fundamental ideas, which are given below, is necessary in order to have a comprehensive understanding of the SVM and SVR models. These ideas form the basis of the SVM and a typical SVR model is presented in Figure 2.2.

- Decomposition of the hyperplane into higher dimensional regions.
- Kernel functions and their effects on transferring the nonlinear data while generating a linear output.
- Soft margins and the separation of data.

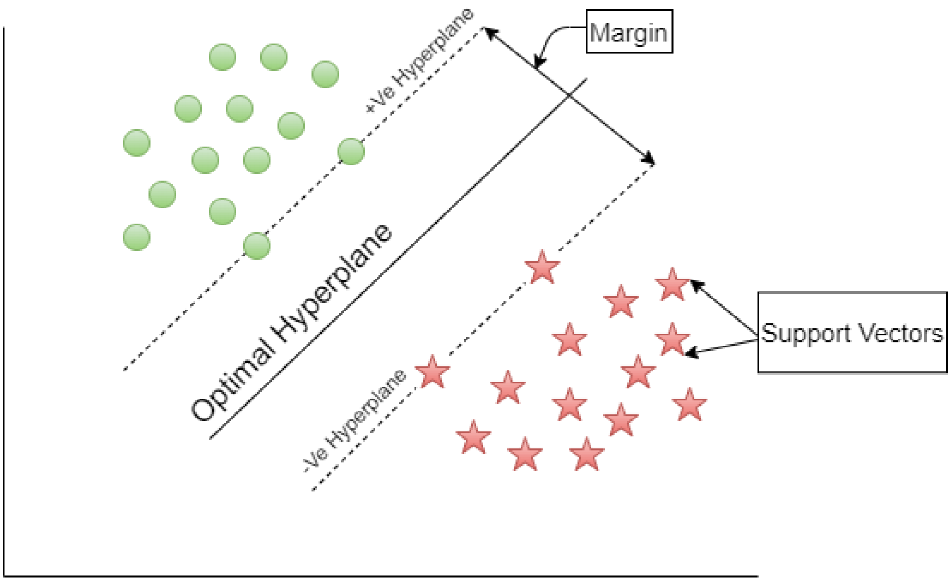


Figure 2.1: SVR Hyperplane

Finding the most effective activation kernel in the SVR/SVM that will lead to an excellent output is one of the most significant challenges that researchers face when they adopted SVR model (LIN et al., 2006; She and Basketfield, 2005; Liong and Sivapragasam, 2002).

The creation of reliable models that can predict inflows during environmental catastrophic such as typhoons is an absolute need for the efficient construction and development of reservoirs. A study conducted by (Lin et al., 2009) included typhoon characteristics as a new input for SVR besides the inflow and rainfall to achieve an even higher level of improvement in the long-term forecasts.

The SVR was compared to MLPNN-based models and the results have shown that the suggested SVM-based models are more accurate, resilient, and efficient than the current BPN-based models, and the features of typhoons should be utilised as input to the reservoir in-flow forecasting models for long lead-time forecasting. It is anticipated that the suggested modelling approach would be helpful in enhancing the reservoir inflow forecasts, and the SVM-based models that have been provided are recommended as an alternative to the models that are currently in use.

In order to predict the stream flow for the Alavian reservoir, a further state-of-the-art study was conducted in 2011. (Noori et al., 2011). In the research, a support vector machine (SVM) that had been trained using (RBF) kernel. The study was designed to focus on maximising the amount of data points that could be fed into the SVM model in order to get the best potential outcomes. In this particular scenario, there were 18 different variables that were used as inputs. To find the variables that provided the best results, techniques such as principal component analysis (PCA), the Gamma test (GT), and forward selection (FS) were utilized to reduce the total number of variables used as inputs and the results have shown that principal component analysis (PCA) was found to be the most effective input analysis selector to be combined with the support vector machine (SVM). This was due to the fact that the PCA was able to reduce the total number of input variables while still maintaining their presence.

The principal component analysis (PCA) was utilised with the aim to reduce the degree of ambiguity in the input variables. Their descriptive article has further material that goes into greater depth about the PCA theory (Gower and Blasius, 2005). The gamma test GT computes the minimal mean square error (MSE). The GA test has some additional remarks that are mentioned here (Stefansson et al., 1997). The FS is a straightforward strategy that is used when there are only a few possibilities to consider for each covariate (N).

Streamflow and Inflow time series are always nonlinear, time-varying, and indeterminate, and the underlying processes of streamflow production are likely to alter significantly throughout low, medium, and high flow periods, particularly when severe events occur (Tripathi et al., 2006). It is quite difficult to anticipate the streamflow precisely. Therefore, a data preprocessing technique is needed. Fortunately, the presence of a powerful de-noise technique known as the wavelet decomposition and reconstruction theory. A strong tool for processing and evaluating time series signals is the wavelet. The wavelet transform is a mathematical technique that was developed from Fourier analysis and is especially designed to allow the identification of non-stationary signals with distinct frequency properties. It may concurrently offer information about time and frequency (Shensa, 1992).

Despite the fact that wavelet transform is an effective tool to de-noise any form of data, it has a major disadvantage which is it necessitates a considerable amount of data and computing time. To overcome this drawback, (Kalteh, 2013) decided to use a new wavelet transform known as the discrete wavelet transform along with the SVR that uses radial basis function (RBF) kernel. The comparison was made in two steps: first, the ANN and SVR stand-alone models were compared to the hybridize wavelet models. The comparison results showed that the wavelet models performed better than the stand-alone models.

Researchers have taken interest of another study that employs the SVR with intermittent wavelet transform for lake level forecasting and compares it to three models from varied backgrounds, including the ARMA, ANFIS, and SVR (Shafaei and Kisi, 2015). Like (Kalteh, 2013), as evidenced by the SVR model, which used the (RBF) as its kernel approach, integrated wavelet models have produced better outcomes than stand-alone models.

As pointed out by (Babur et al., 2016; MULUYE and COULIBALY, 2007; Dettinger, 2011; Gutiérrez and Dracup, 2001), identifying the climate phenomena that dominate the local hydrology and incorporating climate phenomenon indicators into a given modelling framework may help improve reservoir inflow sub-seasonal and seasonal projections. Relying on that fact (Yang et al., 2017) incorporated Indicators of 17 different climatic phenomena into SVR to forecast reservoir inflow and compared it to Random forest and ANN. It was concluded that using climate indicators has improved the model performance significantly despite SVR has shown slightly lower results by being ranked in the second position. Furthermore, a point worth noting made was that Because of the high degree of uncertainty associated with SVR predictions, a careful choice of kernel functions and parameterisation is necessary before their use in actual settings.

In 2018, groundbreaking research was released that models and simulates reservoir flow using support vector regression, artificial neural networks, and deep learning algorithm. This study was carried out by the authors of the study. The process of building the four distinct SVR models required the employment of all four SVR transfer kernels, hence It has come to everyone's attention that the selection of the kernel function has a significant impact on the efficiency of the algorithm.

A time series regression model was employed in 2019 to anticipate daily reservoir levels using the support vector machine (Ahmed and Amr, 2019). The SVM used five cross validation folds in reference to study published by (Hipni et al., 2013) and four distinct scenarios. The capacity of predictive models to generalise was evaluated using cross-validation, which is a data resampling approach (Berrar, 2018). This method also helps avoid over-fitting. The four input parameters used consisted of the following: the first input was daily precipitation, while the second was a combination of daily precipitation and a one-day dam water level measurement as inputs. The third and fourth scenarios, which included daily precipitation and daily dam water levels for seven days. The author recommended that a wavelet or an optimisation method should be coupled with the SVM in future research to improve performance and the effect of using different kernel should be investigated.

A summary obtained from the literature review on SVR shows:

- The choice of kernels is extremely important when dealing with non-linear data as it acts as a hyperplane that creates a relation among the non-linear data to produce a linear output.
- The top two recommended kernels to be used as the transfer function in SVR would be radial basis function (RBF) and sigmoid.
- Wavelet decomposition improves SVR performances; however, the improvement is subjected to the data used and does not guarantee an optimum performance.
- Cross-validation is an effective way to validate the best range of data; however, the number of cross-validation folds need to be investigated.

2.1.3 Adaptive-Neuro-Fuzzy-Inference System (ANFIS)

Understanding that traditional system modelling techniques relying on numerical tools perform poorly when tackling nonlinear and frequently changing problems, such as those found in the hydrology was the main motivation for (Jang, 1993) to develop a new machine learning model aiming to address the non-linearity found in such problems.

First, two hypotheses served as the foundation for the development of the ANFIS. To begin, there is the theory of neural networks, which includes both the fuzzy system (a kind of reasoning system) and Neural Network (a kind of multilayer feed-forward neural network with nodes and neurons that connect the output and input over numerous layers) (Jang, 1991).

The if-then rules serve as the main framework for the second theory, which is known as the Fuzzy Inference System. Each fuzzy rule specifies a certain network behavior locally, and this makes up the basis of the Fuzzy Inference System. The ANFIS model had a total of 5 levels, the first of which was an input layer, and the second of which was a layer that represented the firing intensity of a rule. Layer 3 is the one that calculates the overall firing strength rule, Layer 4 is the one that is called a subsequent layer, and Layer 5 is the one that computes the overall and generates the output that is needed.

In 2009, (Wang et al., 2009) conducted research to anticipate monthly water consumption utilising the ANFIS, in addition to the ARMA (benchmark model), SVR, ANN, and GP. All of these models were compared with each other. According to the findings that were acquired from this research, the ANFIS model was able to get the highest peak flows out of all of the models that were built in various research areas correspondingly to a similar study conducted by (Zounemat-Kermani and Teshnehlab, 2008). As a consequence, the findings of the research are quite positive, and they imply that the ANFIS technique may be effective in modelling monthly discharge time series.

A study to predict the water level at was conducted in 2011(Valizadeh, 2011). The ANFIS used the same hybrid learning technique with some minor modifications. These modifications involved testing the ANFIS using a

Generalized bell function, which is a kind of function that is classified as nonlinear.

Three models were built, each of which contained a distinct type of input, in order to investigate the various quantities and types of input into the ANFIS. Model one had a daily rainfall as its input, while model two had both a daily rainfall and the water level from the previous day as its input. Model three used a combination of the inputs from models one and two. After analysing the data, the author found that model-3 had shown greater performance. Unfortunately, because the length separating the rainfall station and the amount of precipitation that landed directly on the reservoir was ignored the findings that were produced for the scenario that lasted for less than one day were not very impressive. As a result, it was suggested that such specifications be considered as a potential option for the future.

In an investigation to compare between the performance of ANFIS and ANN when using historical inflow and rainfall lumped data vs distributed data, (Jothiprakash and Magar, 2012) tested a few membership functions in ANFIS aiming to obtain the optimal number of membership functions. The outcome of the study shows that a bell-shaped membership function is optimum on top of Gaussian membership, and this is due to the primary difference that bell-shaped membership functions include a smaller number of parameters. As a result, fuzzy set theory may be tackled by tuning the free parameter. In conclusion, due to the clustering effect and the increased cost of fuzzifying a large number of inputs, dispersed input data models fared marginally worse than those using lumped data. The ANFIS models outperform the best ANN models in every category, but notably in peak and longer lead period prediction.

Testing various membership functions in an attempt to find the best combination has attracted all researchers' attention who are adopting ANFIS for forecasting purposes. In order to arrive at the best possible model, a process of trial and error was used by (Kisi et al., 2012) to forecast lake reservoir levels, during which a variety of membership function types and the number of layers were utilised. All ANFIS membership activation functions were tested. The approach with a triangle basis functions and a 3-3-3 structure ended up being the best choice.

(He et al., 2014) studied various membership functions in ANFIS to forecast river inflow in semiarid mountain land and compared it to ANN and SVR. Overall, the findings showed that the triangle membership function, a straightforward straight-line function, was determined to provide the best results among the other potential forms of membership functions, and the performance of ANN, ANFIS, and SVM models in river flow forecasting is good, and there is little to no difference between the training and validation periods in terms of the performance gained using various assessment criteria, however SVR was slightly superior to ANFIS.

The goal of this study by (Seo et al., 2015) was to investigate how the wavelet transform would affect ANFIS in order to estimate the daily reservoir level as precisely and quickly as feasible. ANFIS was trained using a hybrid learning algorithm that combines the least-squares method and the back-propagation descent technique.

Additionally, in order to determine the parameters, the Gaussian membership function was implemented in the nodes of layer 1. One of the ANFIS models is known as the stand-alone model, while the other is known as the WANFIS. The WANFIS model is a merged model that consists of the ANFIS, and a discrete wavelet transform. It has shown that the performance of the model is reliant on the input sets and the mother wavelets. Furthermore, it has been discovered that wavelet decomposition using the mother wavelet, Daubechies, will further increase the effectiveness of the ANFIS model.

In a further investigation on the performance and robustness of ANFIS model for the purpose of river inflow forecasting was carried out by (Nguyen et al., 2018). Both the ANFIS and dynamic evolving neural-fuzzy inference system (DENFIS) models were used for the investigation, the main comparison between both models is that in the global model (ANFIS), a subtractive clustering approach was utilized, but in the DENFIS model, an evolving clustering method was used.

Every piece of data is considered to be an independent candidate for the cluster Centre when using the subtractive clustering method (Priyono et al., 2005; Benmouiza and Cheknane, 2019). Evolving clustering, nevertheless, is comparable to the K-Nearest Neighbor cluster in that it depends on the separation between the selected center and the points (Škrjanc et al., 2019). Both the ANFIS and DENFIS models' training was accomplished by applying either global or local learning methodologies. During the period of testing, ANFIS was executed in the offline mode, whereas the DENFIS model was executed both offline (DENFIS 1) and online (DENFIS 2) modes. The study concluded that when used for real-time inflow forecasting, a DENFIS model with online learning that is started using fuzzy rules discovered after offline learning is an effective strategy that uses learning from both historical data and the most current data.

As a key solution to overcome natural disasters such as floods, (Chang et al., 2016) forecasted inflow aiming to estimate the water shortage rate. Data on reservoir inflows for a 44-year period was collected and supplied to ANFIS as input. The back-propagation neural network, also known as the BPNN, has been selected to serve as a model for comparison to the ANFIS. When the model is being constructed, future water needs, expected monthly inflows (or seasonal inflow) of the reservoir in the next coming quarter, and historical starting reservoir storages are used to Configure the input patterns. On the other hand, the ideal seasonal water shortage rates produced from the NSGA-II serve as output objectives (training targets) for both neural networks. According to the findings, both the ANFIS and the BPNN models provide almost equally excellent performance when projecting water scarcity rates; however, the ANFIS model offers much greater stability.

Despite the good performance that was exhibited by ANFIS model in the previous studies, it is yet to be reliable when dealing with patterns with a high degree of stochasticity and a diverse set of inflow characteristics may be seen in the data. Owing to this fact, (Allawi et al., 2018) proposed effective modifications to the conventional coactive neuro-fuzzy inference system (CANFIS) method. This upgrade contains an update to the back propagation technique, which in turn led to an update of the membership criteria and the induction of the Centre weighted set rather than the global weighted set that was previously utilized in feature extraction and when compared to the conventional ANFIS model used in reservoir inflow forecasting for a semi-arid area, the modified CANFIS model seems to be more useful and reliable.

According to an examination of various ANFIS articles that were studied, and some were cited previously, the following statements can be made:

- The choice of the size and type of membership functions determines how well ANFIS performs since these two variables have the most effects on the computational cost and precision of the ANFIS-based model that is created (Talpur et al., 2017).
- There are no established methods for choosing the kind of MF that will provide the best possible results. MF plays a significant part in the decision-making process.
- ANFIS model could be enhanced by using wavelet decomposition or other clustering variations to cluster the fuzzy rules and deliver an optimal fuzzy rule.

2.1.4 Extreme Gradient Boosting (XG-BOOST)

The Extreme Gradient Boosting (XG-BOOST) technique that was presented by (Chen and Guestrin, 2016) is an innovative implementation approach for the Gradient Boosting Machine, and in particular, K Classification and Regression Trees. The method is founded on the concept of "boosting," which involves combining all of the predictions made by a group of "weak" learners in order to build a "strong" learner via the use of additive training procedures.

XG-BOOST's primary goals are to avoid over-fitting and maximise the use of computational resources. This is accomplished by reducing the complexity of the goal functions, which enables the combination of predictive and regularization terms while still ensuring that the computing performance is optimised (Santhanam et al., 2017). XG-BOOST performance when dealing with highly non-linear data-series such as these found in hydrology is yet to be explored. On the other hand, The XG-Boost algorithm has received a lot of attention for hydrological prediction owing to its great learning performance and efficient training time (Fan et al., 2018). The XG-BOOST algorithm is shown to be a robust predictor, producing more impressive prediction accuracy and generalisation capabilities.

In a recent study, (Yu et al., 2020) adopted XG-BOOST to forecast streamflow at Three Gorge Dam in China. The study used time-series data which have been divided into 7 components, with a Suitable small number of consecutive frequencies in each component to be used by XG-BOOST as an input data. The XG-BOOST was compared to SVR, and the study highlighted that XG-BOOST were more efficient and robust compared to SVR however further testing is needed.

In a short conference paper published November 2021, XG-BOOST performance was investigated to forecast reservoir inflow (Dornpunya et al., 2021). Despite the lack of comparison between other models, the author concluded that when compared to its ability to estimate the monthly inflow, the XG-BOOST model's capacity to accurately forecast the daily influx of water into the reservoir is much superior. If you anticipate the average values of the daily and monthly inflows, your prediction results will be far closer to the actual inflows observed. However, none of these two newly established models has a very strong capacity to define or forecast the dynamics of values that are really severe. In light of this, the author suggests that the model parameters need to be modified in order to increase the quality of the machine learning algorithm that is used for hydrological prediction.

2.1.5 Strengths And Shortcomings of Machine Learning Models

Table 2.1 provides an overview of the benefits and drawbacks associated with machine learning models and the application of these models to predict reservoir inflow.

Table 2.1: Model Strengths and Weaknesses

Model	Strengths	Weaknesses
Multi-layer perceptron Neural Network (MLPNN). (Svozil et al., 1997; Yaseen et al., 2015; Oyebode and Stretch, 2019; Sherif et al., 2021; Dashti Latif et al., 2021)	• The capacity to simulate dynamic and nonlinear hydrological processes without pre-setting any assumptions about the relationships between the input and output variables. • Dynamic as it is able to continually adapt their behaviour in response to the changing conditions of their surroundings. • MLPNN is able to provide accurate models or consistent replies based on the learning process even when an issue does not have a closed-form mathematical model to describe it.	• It is difficult to construct a suitable model beforehand since there is neither a standard nor a set of rules that regulate the right design and development of acceptable models. • MLPNNs occasionally suffer from over-fitting problems. • If ANN presents a test answer, it provides no hint of why and how.
Support Vector Regression (SVR) (Yang et al., 2017; Pour et al., 2020; Drucker et al., 1997; Hipni et al., 2013; Bao et al., 2014)	• Ability to be rapidly trained and has a wide optimised solution. • Efficiencies in high dimensional spaces. • Flexibility for various problems as various kernel functions is available.	• Deficient performance when features are more than samples. • cross-validation (K-fold) is used instead of probability estimation whereby benchmark for K-value identification.

Table 2.1 (Continued): Model Strengths and Weaknesses

Model	Strengths	Weaknesses
Adaptive Neuro Fuzzy Inference System (ANFIS) (Khan and Chai, 2017; Karaboga and Kaya, 2019)	• Potential in a single system to utilise the neural network and fuzzy logic. • Fast and accurate training process. • ANFIS is ideal for providing the hydrological information needed for feasibility assessments.	• Sensitive to the number of clusters. • software becomes more sophisticated as there are more ambiguous rules.
Extreme Gradient Boosting (XG-BOOST) (Cinar, 2019; Santhanam et al., 2017)	• Gradient Boosted Machines (GBM) and Random Forest (RF) vary primarily in that GBM adds a new tree to complement existing ones while RF builds trees independently of one another. • Both lambda and gamma are adjustable in XG-BOOST. These variables are controlling the penalty score. Because of this, the model may be safeguarded against over-fitting. • The error rate decreases with each successive model that is used in gradient boosting, up to the very last (final) model.	• The operation of XG-Boost might be highly time intensive. • Using gradient boosting to construct precise models also calls for a great deal of tuning of variables.

2.2 Metaheuristic Optimisation Algorithm

There are two main categories of stochastic algorithms: heuristic and metaheuristic. While heuristic algorithms are able to identify answers in a fair period of time, they do not guarantee finding the best possible solution. In contrast, metaheuristic algorithms are designed to be well-suited for global optimisation and make use of a set of trade-offs including randomisation and local search (Sharma and Kaur, 2021). This is because randomisation offers a decent technique to transition from local search to the pursuit on a global scale. The majority of probabilistic algorithms may be categorised as metaheuristics; some examples of these are GA, PSO, and FIREFLY, amongst others.

2.2.1 Genetic Algorithm (GA)

The genetic algorithm (GA) is a well-known algorithm that gets its name from its inspiration, the process of biological evolution (Katoch et al., 2021). The Darwinian idea that the strongest and healthiest will survive and thrive in nature was modelled after J.H. Holland made the first suggestion for GA in 1992 (Holland, 1992). The depiction of chromosomes, the selection of individuals based on their fitness, and the use of biologically-inspired operators are the fundamental components of GA (Storn and Price, 1997).

Inversion, a unique component that is often used in GA deployments, was another innovation that Holland brought to the Table. In most cases, the format of the chromosomes is that of a binary string. Each locus (a particular place on a chromosome) in chromosomes may have either the allele 0 or the allele 1. These are different variants of the same gene. In the "solution space," chromosomes may be considered individual points. In order to process these, genetic operators are used, and their population is replaced in an iterative fashion.

The fitness function is utilised to give each chromosome in the population a score out of the possible options. Selection, mutation, and genetic crossover make up the biologically inspired operators. During the selection process, chromosomes are chosen for further processing based on how well they will contribute to the organism's overall fitness. To produce offspring, the crossover operator selects a random locus and then modifies the subsequences of the chromosomes that are next to each other. During the process of mutation, some segments of chromosomes will, on the basis of likelihood, be switched around.

It is a well-established fact that the phenomenon that lies beneath the seemingly random nature of the inflow is actually quite complicated. In a more tangible sense, the modelling of reservoir operations.

To produce an ideal reservoir operation strategy to aid decision-making, a technique that combines the constrained genetic algorithm (CGA) where the ecological base flow needs are treated as reservoir operation water release constraint while optimising the 10-day reservoir storage was developed by (Chang et al., 2010). The performance of GA was measured using generalised shortage Index (GSI) and the results revealed that the CGA technique has the potential to considerably increase the efficiency and efficacy of water delivery capabilities to both human and ecological base flow needs, hence optimising reservoir operations for various water users.

(Li et al., 2010) gave a helpful example of how the genetic algorithm may help with precise streamflow forecasts, despite the presence of a large number of climatic precursors. The SVM and MLR models were used in the experiments, and comparisons were made between those models and a hybrid SVM model that included a genetic algorithm optimised model and bagged.

The bootstrap statistical resampling approach serves as the foundation for the Bagging method, which results in a variety of training sets that are used to construct the SVM model that is part of an ensemble (Breiman, 1996). To produce an optimal model of the SVM, a variety of kernels and parameters must be adjusted. This parameter tuning is frequently done using an approach that requires trial and error. The genetic algorithm uses a combination of three real value variables to determine which chromosome belongs to which person. The genetic algorithm (GA) starts with the generation of a chromosomal population at random. By achieving a higher frequency on the cross-validated correlation coefficient, the findings demonstrate that the hybrid SVM and MLR models that were bagged and GA-optimized outperformed the traditional SVM and MLR models.

(Fallah-Mehdipour et al., 2012) optimised reservoir systems via the use of genetic programming (GP) and genetic algorithms (GP). The operating rule curves are guidelines that each period synchronises the reservoir system parameters by taking into consideration the observations made of the system in the past. In order to consider, the physical and hydrological conditions, these Guidelines make use of deterministic and stochastic variables, which are often stated via the use of a mathematical Equation.

The linear decision rule (LDR), which was first presented by (Revelle et al., 1969), is one of the most often utilised rules in reservoir operation. It is also recognised as being one of the earliest rules to be applied to reservoir operation. The GA with the linear decision rule (LDR) and two different forms of (GP) were tested for the purpose of predicting reservoir inflow. One type of genetic programming utilises stochastic variables, while the other utilizes deterministic variables. According to the findings, both the GA and the GP were able to accurately forecast the amount of water entering the reservoir. Nevertheless, the GP that used deterministic data had shown marginally superior performance in comparison to the GA.

(Hossain and El-shafie, 2014) carried out a performance study to optimise the stream flow by accurately anticipating it since decision-makers in a reservoir structure still need guidance in order to manage the reservoir properly. The GA is widely recognised and has been put into use effectively in the field of reservoir operational development. As a result, it has been used as a benchmark for the ABC (Artificial Bee Colony).

A description of the many rules that may be used as operators in a genetic algorithm is found in (Goldberg and Deb, 1991). In the study, the crossover technique was used to drive the GA. By comparison to the GA, the ABC was superior, as shown by the findings. However, the GA did manage to obtain a high dependability percentage of 91.6%, proving both its excellent performance and its dependability as an optimisation approach.

Most approaches integrating reservoir management have the maximisation of social and economic goals as their primary focus (e.g., power generation, flood control, etc). In spite of this, objectives and limits concerning eco-environmental issues (such as fish habitat, environmental flow, etc.) are gradually being integrated with reservoir operation in response to growing concerns over river ecosystems.

(Chen et al., 2016) replaced the conventional genetic algorithm with Non-dominated Sorting Genetic Algorithm-II (NSGA-II) aiming to optimise the operation of the reservoir by meeting eco-environmental objectives. The Non-Dominated Sorting Genetic Algorithm-II which was developed by (Deb et al., 2002), is an algorithm that shows potential for solving multi-objective problems. NSGA-II is superior to NSGA in a number of respects, including the fact that its sorting algorithm is superior, the inclusion of elitism, and the fact that its sharing parameter does not need to be determined a priori. Based on the case study of Qingshitan reservoir, the performance of the NSGA-II was significantly improved in comparison to GA. The case study demonstrated that the parallel approach, which uses several parallel groupings of populations, was responsible for the improvement. This progress was made possible as a result of an increase in variety brought about by the parallel groupings. The outcome of this study was similar to that found in (Tsai et al., 2015).

In 2016, the genetic algorithm was implemented as an optimisation tool in the Cameron Highland hydropower reservoir system by (Tayebiyan et al., 2016). This is just one example of how the scope of genetic algorithm applications continues to broaden with each passing year. The primary aim of the construction of any hydroelectric reservoir system should be to facilitate the production of renewable energy from the area's native resources. As a result, to improve the system's operational performance, optimising the total power output during working hours has been set as a goal. This will ensure that all of the various physical and organisational restrictions are satisfied. When developing the goal function, there were two primary limitations, which were hydro-plant discharge restrictions and reservoir storage volume requirements. Both of these constraints were considered. The three distinct Guidelines optimised using GA are the BSOPHP, the CSOPHP, and the standard hedging Guideline (SHPHP). Genetic algorithm comes in a few different versions, but the real coded genetic algorithm (RCGA) is the most effective one when dealing with a constrained algorithm as proven by (Abdul-Rahman et al., 2013). In terms of optimisation using RCG, the SHPHP proposal was the best one out of the three offered since it obtained a high level of efficiency, resilience, vulnerability, and lifespan.

In the operation of a reservoir, there are several characteristics that might have an effect on the operation; nevertheless, the reservoir inflow is the one that is considered to be the most crucial. A genetic algorithm has been integrated with an artificial neural network (ANN) to build a hybrid model that is capable of achieving the optimal reservoir inflow by attaining the optimal ANN parameters (Moeeni et al., 2017). This was done to guarantee that the reservoir inflow forecasting was as accurate as possible. The genetic algorithm combined generator with an artificial neural network was the most generally used elite population, and the hybrid model ANN-GA was compared to the stand-alone ANN model. The outcomes demonstrated that the hybrid model performed better than the stand-alone model. In addition to this, when estimating the monthly inflow, it was found that the ANN-GA model had done much better than the other two models, including the SARIMA, which is an integrated autoregressive model. This was determined by comparing the SARIMA to the ANN and the ANN-GA models. Despite this, the SARIMA model performed better at forecasting the short-term, such as cycles lasting a few days or a few weeks.

The motivation behind integrating machine learning models with metaheuristics optimisation algorithms is to overcome the limitation. Owing to the fact that Non-Linear Programming (NLP) approaches suffer from a sluggish rate of convergence and demand substantial computing storage and time, (Krishna et al., 2018) have chosen GA to improve NLP model performance when optimising reservoir operation forming a hybrid model known as GA-NLP. The author highlighted that GA was chosen as it uses coded copies of the system variables rather than the actual variables themselves. Secondly, although almost all conventional optimisation methods start their search at a single point, GAs instead use a population of points to get the best solution. The model is used for one of 2 potential inflow situations and one of three possible priority levels. The overall results demonstrated that GA-NLP provides more accurate and reliable results as the annual irrigation deficit has seen an improvement of 2% compared to the stand-alone model while the annual electricity generation has increased by 18% compared to the operation when stand-alone NLP was used.

In a recent study published in 2021 by (El Harraki et al., 2021), a new developed objective function that combines both the largest possible water deficit and the average frequency of shortages was given to GA. The aim of this study was to optimise the reservoir operation and compared it to the conventional objective function which are calculated based on sum of squared deviations (SSD). The comparison of the pattern and performance indicators maximum deficit and shortage frequency reveals that the proposed objective function produces better results for public water supplies in regard to dependability and vulnerability, and slightly better results for irrigation in terms of shortage frequency. Hence, the author recommends replacing the standard objective function with the newly proposed function.

2.2.2 Particle Swarm Optimisation (PSO)

Particle Swarm Optimisation (PSO) was developed in the middle of the 1990s by (Kennedy and Eberhart, 1995) as part of a socio-cognitive study investigating the concept of "collective intelligence" in biological populations. At the time, Kennedy and Eberhart tried replicating the orchestrated, elegant motion of swarms of birds as part of their research.

In PSO, a set of randomly generated solutions is referred to as the "initial swarm." These solutions then begin to move through the design space in the direction of the optimal solution throughout a series of iterations, also known as "moves." These moves are based on a significant amount of data regarding the design space that is assimilated and shared by all members of the swarm. PSO is inspired by the capacity of groups of animals, such as flocks of birds, schools of fish, and herds of animals, to adapt to their environment, locate rich sources of food, and escape predators by employing an "information sharing" method, and so establishing an evolutionary advantage.

To make the operating system simpler, it is essential to increase the efficacy of inflow forecasting models. To get a more precise assessment of the yearly inflow into the reservoir, PSO was used to enhance the Support Vector Machine by (Wang et al., 2010). The SVM-PSO is a hybrid model that was optimised using the PSO. The radial basis function was chosen to serve as the kernel for the support vector machine that was used in this study. The PSO had an inertia weight that was somewhere between 0.4 and 0.9. For the purpose of making a comparison, BPNN and the Elman neural network model were contrasted with the hybrid SVM model. The findings demonstrated that the SVM that was optimised using PSO could identify the best parameter values;

hence, the SVM that was optimised with PSO performed better than the other two models.

In the work that was released, a further investigation of the PSO impact on (SVM) was studied by developing a better SVM model to forecast monthly streamflow(Guo et al., 2011),. This model was studied in the context of the hydrology field, This published study was a novel in its own because it consisted of three main phases: In order to limit the amount of unreliable data, wavelet de-noising was initially applied to the raw data; after that, chaotic feature assessment and parse-space reconstruction were employed to define the forecasting model's structure; thereafter, PSO was used to the model parameters to identify the ideal parameters.

In a manner that is analogous to the research conducted by (Wang et al., 2010), the value w, which refers to inertia, was determined to linearly change from 0.4 to 0.9. This was done because it was discovered that larger inertia engages in global exploration, whereas smaller inertia engages in the local exploration, as was demonstrated by Lin et al. in 2008. The standard SVM, the linear adaptive insensitive factor (LAIF-SVM), and the non-linear adaptive insensitive factor (NAIF-SVM) were the three SVM models that were constructed and the PSO was used to optimise all of the models' parameters. In all, three SVM models were produced. The ANN model was included in the comparison with the other three models so that the overall dependability of the research could be increased. According to the findings, the PSO was able to significantly enhance the performance of the models, and the NAIF-SVM model emerged as the most successful one after being optimised by the PSO. On the other hand, the number of support vectors grew, which slowed down the training process and had a detrimental influence on accuracy. As a result, it was said in passing that more As a consequence, it was said in passing that more study is required to determine how to speed up training while maintaining a suitable number of vectors in the optimised SVM.

Although PSO is most often affiliated with the SVM, it is nevertheless possible to use it with machine learning models and it will still produce a positive outcome. When there is a high number of decision variables or parameters to be optimised, the PSO algorithm has a sever shortcoming is that it does not ensure that the solution will converge globally Liu et al., 2020). In order to circumvent this shortcoming, research was carried out by (Cheng et al., 2015) to anticipate daily reservoir run-off utilizing a quantum-behaved particle swarm optimisation with artificial neural networks (ANN), a technique known as ANN-QSPO for daily reservoir run-off forecasting. This technique is founded on (ANN) and (QPSO), based on the developer of QPSO (Jun et al., 2004), they claimed that it is theoretically possible to ensure that the global optimum solution will be found in the whole searching space in QPSO. In addition, the outcomes of several simulations of complicated benchmark functions demonstrated that QPSO had a superior capacity for global searching in comparison to PSO. In the ANN-QSPO approach that was suggested, QPSO was used to pick the ideal parameters for the ANN, and when the training process was complete, the ANN was utilised to make predictions. For the purpose of predicting the daily run-off of the Hongjiadu reservoir in Southeast

China, the suggested method was evaluated alongside a stand-alone ANN model. The experiment results demonstrated that the suggested strategy produces much higher levels of forecast accuracy than the basic ANN model.

In a study conducted by (SaberChenari et al., 2016) on one of the largest dams in IRAN Mahabad reservoir dam, the PSO was employed as a tool for the aim of optimising the functioning of the dam. In four cases under ordinary and dry circumstances, streams to the reservoir were considered while lowering the probability of monthly average inflows. This was needed to enable the reservoir scheme for catastrophic situations. In overall, the results showed that the quantities of the discharged water derived by the PSO algorithm had an excellent correlation with the downstream water requirements, and the storage reservoir had an adequate capacity to satisfy the water demands of the following months.

A flash flood is an Inevitable occurrence that may result in significant fatalities. It is most common in rural regions and occurs when there is an excessive amount of rainwater that falls into drainage areas and then gathers into the main flow. The majority of water flows into the river at some point. This causes a significant quantity of water to be released into the river system that is farther downstream. Consequently, the purpose of the research was to make projections about water levels to forestall a significant increase in the volume of the river. The research established the model in stages, each building upon the previous one. The first step was initialising the PSO, which involved setting the settings for the initial swarm particles and the PSO coefficient, the two main parameters. Each step was finished in order.

When the data preparation stage is over, which creates the optimal data set for the model, the third stage is training the PSO procedure. The findings shown that the PSO is extremely capable of lowering the number of inputs and creating an ideal model. The model that used an input of 10 upstream water level measurements and a trip time from upstream to downstream of 5 hours was determined to be the best one. The PSO excels in reducing the number of inputs and generating an ideal model. This demonstrated how the PSO might be used to determine the ideal input parameters for massive amounts of data.

Owing to the fact that PSO is a robust algorithm, it has been adopted by (Karami et al., 2019) to optimise two different reservoirs located in two different countries with totally different objective functions and was compared to whale optimisation algorithm (WOA) and Genetic algorithm (GA). The first case study's objective was to reduce the amount of water that was supplied for agriculture at the Aydoughmoush dam in Iran. According to the findings, the operational rule of the PSO was good; however, it was not as good as WOA, which had an average solution that is quite near to the global solution. The second case study was determined to find ways to reduce the amount of unused electrical power in the Karun-4 reservoir. In reality, water must be discharged to optimise energy production or reduce electricity shortages. According to the findings, the operating guideline for the water that was accomplished by the PSO generates more electrical power than both the GA and the WOA.

Further investigation on the robustness and effectiveness of PSO in comparison to newly developed models such as Shark machine learning Algorithm (SMLA) and conventional model GA was carried out by (Allawi et al., 2019) to validate the newly developed algorithm. The goal function of this research was to minimise the gap between water demand and release. The results that were published have ranked PSO in second placed right after SMLA with a minor difference in terms of reliability, resilience, and vulnerability. The main findings of the study showed that the incorporation of reservoir inflow predictions into the operational period of a reservoir system is advantageous for getting closer to the real behaviour of optimisation models used in the management of reservoir systems and optimising PSO is a critical stage that needs to be carried in a proper manner to obtain an optimum performance.

Although the PSO method searches for the best solution more quickly than other optimisation techniques do, it is plagued by the drawbacks of early completion and limited fine-tuning capabilities. Because of their low inertia weight, particles tend to settle into their respective local optimums during the early stages of development. In contrast, the fine-tuning impact of particles is relatively weak in the latter stage due to the enormous inertia weight, and it is possible that the ideal solution will not be reached at this point.

Hence, a study conducted by (Chen et al., 2020) proposed an adaptive random inertia weight (ARIW) Equation, to enhance the inertia of particle adaptivity when optimising Panjiakou reservoir operation in China. An ARIW technique is suggested as a means of making the PSO algorithm more effective in its operation. Randomized inertia weights are produced via a triangle probabilistic density function and these weights are modified as a result of particles swarm evolution. It has been shown via the application that the

ARIW-PSO method is more effective than both the traditional PSO and the GA algorithms. When the ARIW-PSO algorithm is used, the water deficit for the water supply tends to be levelled out, which helps to alleviate the loss that is produced by a lack of water. The results reflected the similar outcome obtained by (Al-Aqeeli and Mahmood Agha, 2020) when optimising multi-objective reservoir operation.

2.2.3 Firefly Optimisation Algorithm (FA)

One of the most unique and interesting species that can be found in the natural world are fireflies, sometimes known as lightning bugs. They are able to produce light as a result of specialised photometric organs that are located extremely near to the body surface and are covered by a window of cuticle that is transparent. Bioluminescent signals are known to perform a variety of functions, including as components of courting rituals, techniques of attracting prey, methods of social orientation, and as a warning system to attackers (Lewis et al., 2020). The phenomena of fireflies illuminating is a subject of ongoing study, taking into consideration both the biological and sociological components of the phenomenon.

Inspired by the magnificent behaviour of fireflies when they are in groups has motivated (Yang, 2009a) to develop an optimisation algorithm that mimic the firefly behaviour in 2009. The following are the three fundamental principles used by FA:

1. Fireflies are mutually attractive since they are all of the same gender.

2. As distance increases, both the appeal and the brightness diminish. Therefore, when there are two flashing fireflies, the one that is less brilliant will travel closer to the one that is brighter. If there is not a firefly that is brighter than the one that is now moving, then it will travel at random.
3. The light level is specified by an objective function that either maximises or minimises the optimisation space.

Despite the fact that it has been shown to be extremely effective at optimising typical benchmark functions (Adil Hashmi, 2013; Łukasik and Żak, 2009; Yang, 2013), the Firefly Algorithm is only hardly ever employed to solve issues that are generally encountered in hydrology or specifically in reservoir optimisation problems (Kumar and Kumar, 2021).

In contrast of its rarity, (Garousi-Nejad et al., 2016a) used FA to tackle two long-term operations that both included the same reservoir but served distinct goals. These activities were the provision of irrigation water and the generation of hydropower. In the first step of the process, the performance of the FA was evaluated using five distinct mathematical test functions. In these cases, the FA converged to near-global solutions more quickly than the GA did. In addition to this, in comparison to the GA, the FA was able to acquire the value of the objective function that was closer to the global solution. After that, the results of the reservoir operating systems fully exposed the tremendous capability of FA and stressed its ability in addressing complicated restricted optimisation issues.

Furthermore, it can be determined that the FA attained a closer average value of an objective function (3.6078) to non-linear programming (NLP) (3.3727) than the average value of the objective function acquired by GA (6.6754). In addition, the results of five runs of the FA demonstrated a smaller standard deviation (0.06), whereas the results of five runs of the GA exhibited a standard deviation of 0.31, which is approximately five times worse. The author concluded that the findings indicated that the FA was capable of arriving at more optimum solutions than the GA and NLP. The author has found that even with all these advantages, FA has some limitations such as the optimal parameters are hard to be defined and it requires relatively large amount of data for training purposes.

In an attempt to overcome the limitations of firefly mentioned previously, (Garousi-Nejad et al., 2016b) developed a modified version from the original firefly and named it as MFA. The MFA was used to find solutions for three benchmark reservoir operating issues, which were known as discrete four reservoir (DFP), continuous four reservoir (CFP), and lastly continuous 10-reservoir (CTP). The conclusion that can be drawn from the findings of the first issue, the DFP, is that the MFA is capable of calculating the optimal value of the objective function, which is equivalent to the global optimum that can be attained via LP. Applying the MFA to solve the DFP led to the discovery of four distinct alternative optimum solutions, none of which had been reported before by other scientists utilising various methods. This is an additional significant success brought about by the use of the MFA. Concerning the CFP, the optimum solution achieved with the MFA is 99.99% of the LP solution.

This little difference demonstrates that the MFA is trustworthy when applied to issues involving continuous operations involving many reservoirs. When it comes to the results of the MFA for the third problem, the CTP, the benefits and superior performance of the MFA were brought to light by solving the complex CTP and reaching the best value of the objective function which was 99.21% of the LP solution. This highlighted the advantages and superior performance of the MFA. With all being said, the author has concluded that the FA is a strong non-linear algorithm capable of solving non-linear problems by expanding its search space.

2.2.4 Nuclear Reaction Optimisation Algorithm (NRO)

In the field of artificial intelligence (AI), optimisation strategies are becoming more relevant as optimisation challenges get more complicated and demands for operative accuracy and iteration rate increase. In 2019, (Wei et al., 2019) developed a novel and versatile approach to produce electricity that is based on the way nuclei react with one another (NRO). This technique is completely novel and has been experimented out on 28 modern benchmark and 23 traditional benchmark problems. Ten alternative optimisation methods were evaluated with the generated model (NRO) for the purpose of validation, namely atom search optimisation algorithm (ASO) (Zhao et al., 2019), Biogeography based optimisation (BBO) (Simon, 2008), butterfly optimisation algorithm (BOA) (Arora and Singh, 2019), Cuckoo search (CS) (Yang and Deb, 2010), Multi-Verse Optimiser (MVO) (Mirjalili et al., 2016), lighting search algorithm (LSA) (Shareef et al., 2015), Salp Swarm Algorithm (SSA)

(Mirjalili et al., 2017), Harmony Search Algorithm (HS) (Yang, 2009b) and finally Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Igel et al., 2007).

Based on the developer of the model (Wei et al., 2019), the quality of the solutions presented by NRO is superior, with a more desirable mean answer and a lower standard deviation. According to the preceding investigation, the performance of NRO's exploitation in terms of multimodal low-dimensional functions compares well to that of the algorithms that were evaluated for comparison as shown in the spider chart by Figure 2.3. On the other hand, NRO is yet to be tested for real case study optimisation problem and therefore the application of developing a reservoir operation scenario using NRO is among the novelties of the current research.

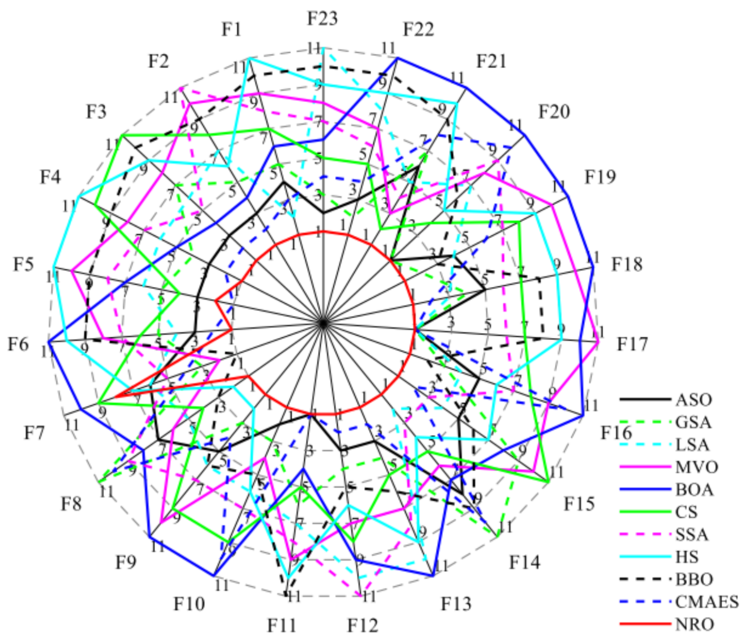


Figure 2.2: Spider Chart for Comparison between NRO vs 10 Models

Most reservoir operating models include historical inflow projections (Niu et al., 2019). The inflow prediction based on historical data, however, cannot match the actual inflow throughout the whole forecast horizon owing to the limitations of present forecasting tools. This means that inflow prediction will always be susceptible to forecast uncertainty, which might hinder optimum reservoir management (Liu et al., 2020; Allawi et al., 2019). Although some studies tried to address the inflow forecast uncertainties (Zhang et al., 2019; Liu et al., 2019), it remains completely unexplored.

As a result, it is also necessary to assess the inflow uncertainty associated with the reservoir operating rules by replacing the historical inflow data with more reliable data obtained through projecting inflow based on machine learning models to reduce the errors that might have occurred due to faulty equipment readings at the site. A novel two-stage optimisation model will be developed for this study, dividing the optimisation process into two stages (instrumental and forecast inflow) based on three different hydrological periods (high, medium, and low inflow).

To lessen the uncertainty of the prediction, the current study utilised this model to substitute the historical inflow. Firstly, historical inflow was utilised as an input for machine learning models; as a result, the model will be depending on a pattern thus decreasing the inflow uncertainty.

Besides addressing the inflow uncertainty gap that was found in previous studies, the novelty of this study could be further seen in the choice of the algorithms as in the current study for the first time for solving optimisation problems, four hybrid models (XG-Boost-based) using four optimisation algorithm Nuclear Reaction Optimisation (NPO), Real coded genetic algorithm RCGA, Firefly (FA) and particle swarm optimisation (PSO) are used. These models' accuracy is examined to solve the inflow uncertainty forecasting problem, and meanwhile developing a release scenario and hence developing a semi-integrated-real-time system.

Despite the popularity of optimisation algorithms, the core questions on performance issues are still partially answered due to limited insightful analyses. Mere investigation and comparison of results may not reveal the reasons behind what is poor or better performance. Therefore, tuning and analysing exploration and exploitation of swarm and physics-based metaheuristic algorithms is crucial to obtain an optimum model (Hussain et al., 2019). With all being said, this study was intended to overcome this gap by tuning the four models to obtain the best exploration and exploitation performance.

1

CHAPTER 3

METHODOLOGY

3.1 Introduction

The operating strategy of a reservoir is a set of guidelines that the decision maker of the reservoir system should adhere to in an attempt to meet their objectives, which may include the supply of water, the prevention of flooding, or the production of hydroelectric power. This chapter covers an in-depth discussion on the approach taken to forecast, simulate and optimise the reservoir operation as well as the concerns that go along with it. The data preprocessing methods carried out in this study are reviewed in detail along with the method of handling and replacing missing data.

This research is divided into three main phases; phase (1) is the development of machine learning forecasting base models which are able to project volume of water flowing into the reservoir catchment using historical data available. The output and the performance of these models were verified using a few statistical tests and the models were ranked based on the results obtained from the statistical metrics tests. In phase (2), the objective function was formulated without violating the reservoir constraints which were the storage and release levels.

In phase (3), the reservoir water loss condition was simulated using the best performing machine learning model from phase (1). The simulation was integrated with the metaheuristic optimisation algorithm and subsequently developing a closed-loop system capable of forecasting, simulating, and optimising with minimal human interference. Finally, the risk analysis to assess the metaheuristic models and reservoir performance were computed using four main indicators known as vulnerability, resilience, reliability, and sustainability.

The storage capacity of the reservoir as well as the volume of water inflow value for each step must be obtained to do this after the development of the operational rules that were achieved by utilising optimisation approaches. This may be accomplished by obtaining the value of the inflow. The amount of water that will be stored in the reservoir at the commencement of the step is calculated using the anticipated amount of water that will flow into the reservoir at the beginning of the phase. With knowledge of both the beginning storage and the inflow category (i.e., the value that is expected to be received), it is possible to derive from the operating rule graph for this specific month, the determination of the ideal amount of water to be released. Keeping in mind that the volume of the inflow is uncertain at the beginning of each time step, which is why the anticipated inflow value is employed. On the other hand, since the actual inflow was received after each stage, it would be utilised to calculate the reservoir capacity after each time step in the following formulation.

3.2 Study Area and Data Collection

The Klang Gate Dam (KGD) is located in the Ulu Klang region, Gombak District, Selangor, Malaysia. Data collection of the KGD was undertaken in this project. The Klang Gate Dam was the very first dam to be built in Malaysia, and it did not become fully operational until the year 1958. The dam was constructed with an average height of 37.0 m and a maximum length of 139 m. It can be found in the Gombak district in the state of Selangor. There is a capacity of 25,104 million litres of water that may be stored in the catchment reservoir of the dam. A full diagram of the KGD is shown in Table 3.1 while the location of the dam is shown in Figure 3.1.

The nation's capital city of Kuala Lumpur then and now, has the Bukit Nanas water treatment facility that receives water supply from KGD and treats it. Of course in modern times, there are many more sources of water for the capital. Each day, the Bukit Nanas facility generates 144 million gallons of filtered water. The Klang Gate Ridge, the biggest quartz dyke in the world, is located close to the dam. The Klang Gate Ridge is 200 meters wide and extends over 22 kilometres.

Table 3.1: Klang Gate Dam Properties

Item	Description
River Name	Klang River
Structure Type	Arched Concrete Gravity Dam
Date Of Completion	June 1980
Elevation (M)	97.87
Length (M)	138.72
Height (M)	36.89
Location From City Centre (Km)	19.9
Catchment Area (Km^2)	77
Flood Control Volume (M^3)	6.2 X10
Total Storage Capacity (M^3)	35.4 X10
Spillway Type	Overflow Gated Spillway

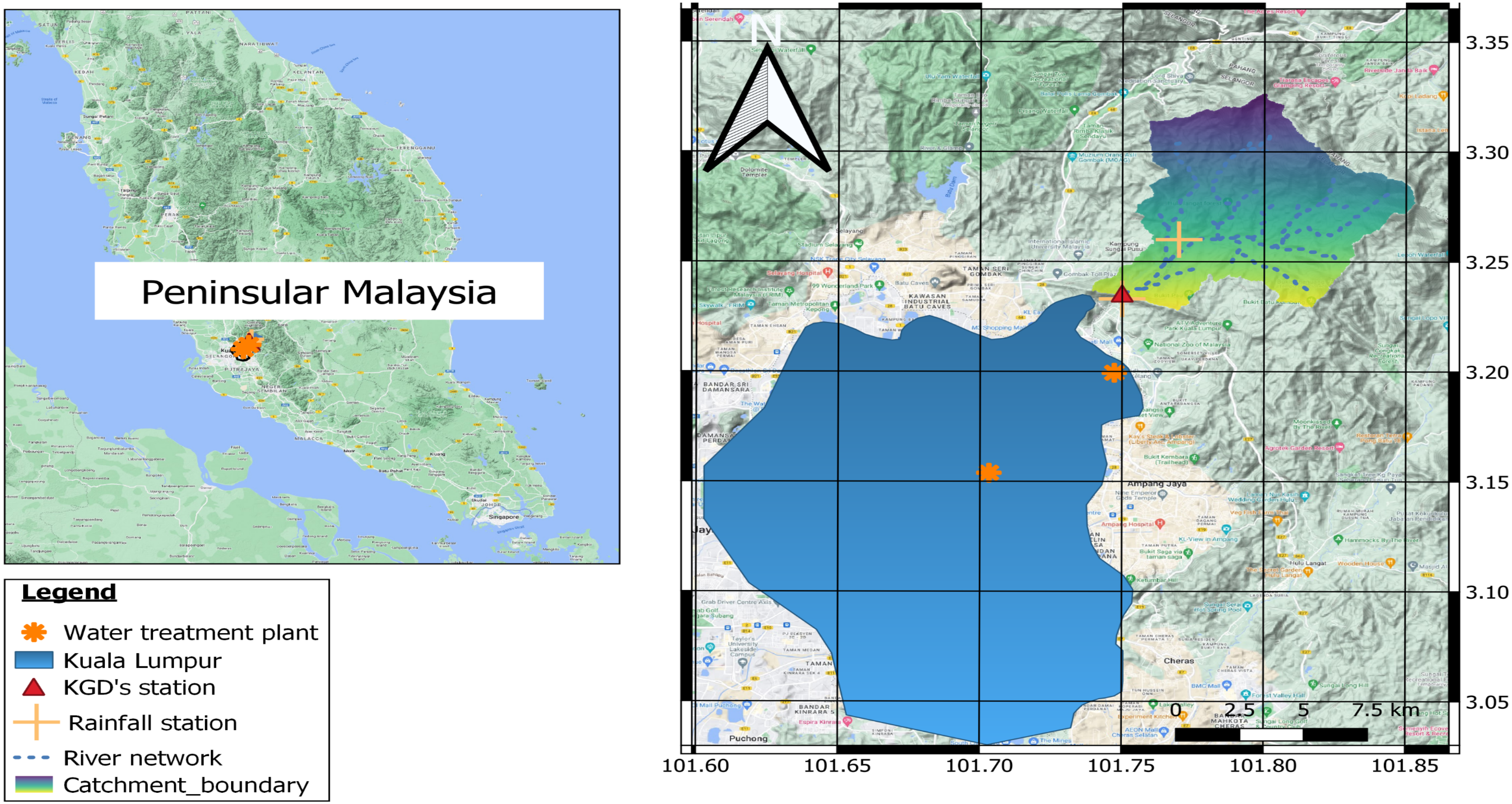
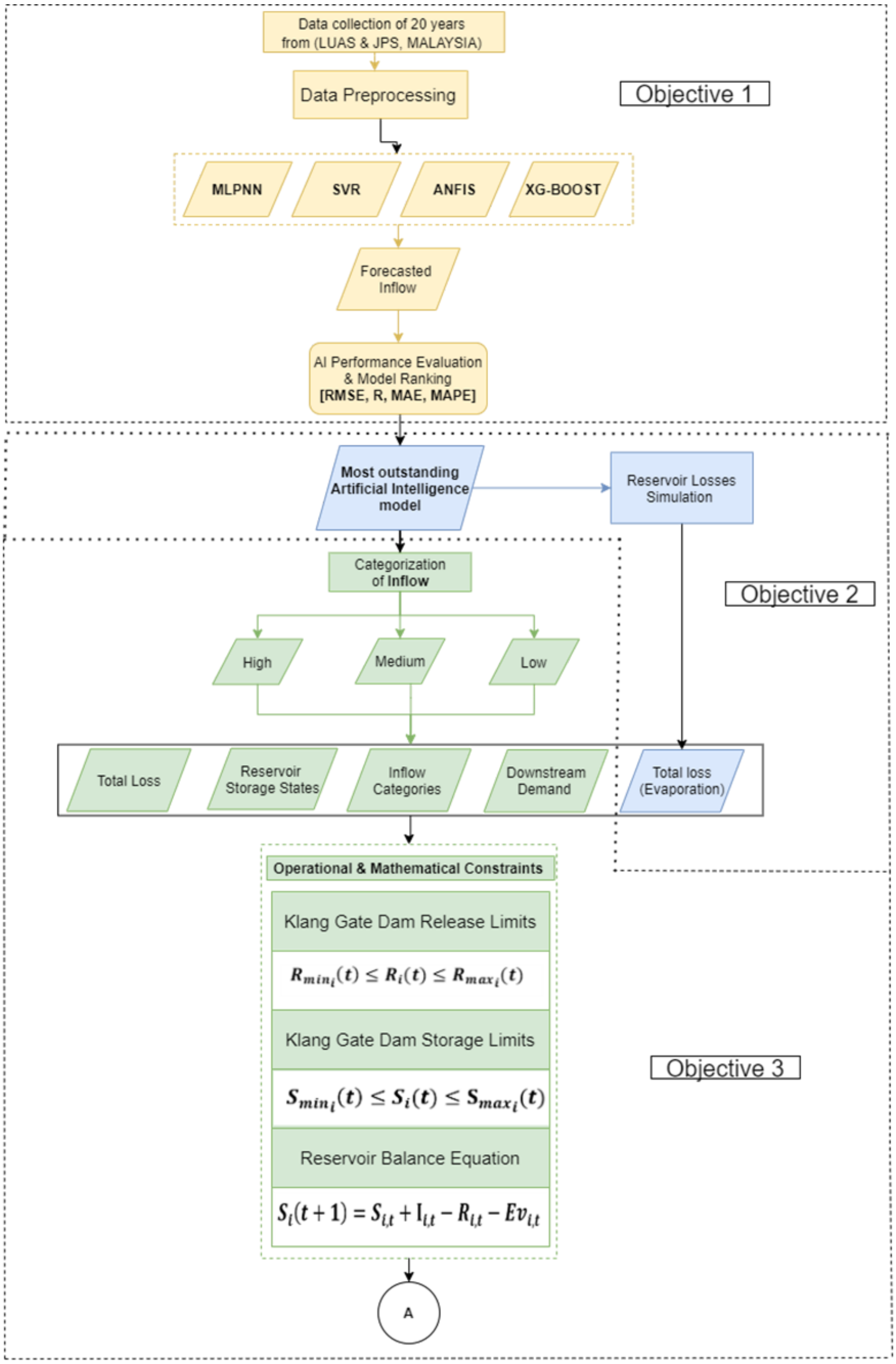


Figure 3.1: Klang Gate Dam Location

3.3 Study Flowchart

Figure 3.2 shows the overall study flowchart in this present study. The full explanation on the flowchart is covered in the next few sections under the methodology chapter. The first phase in this research was the data collection and processing hence the next section will cover all the data pre-processing approaches to refine the data. All the machine learning models, and the optimisation algorithms specified in the flowchart is explained briefly in this chapter.

(a)


Data collection of 20 years from (LUAS & JPS, MALAYSIA)
Data Preprocessing
Objective 1
MLPNN
SVR
ANFIS
XG-BOOST
Forecasted Inflow
AI Performance Evaluation & Model Ranking [RMSE, R, MAE, MAPE]
Most outstanding Artificial Intelligence model
Reservoir Losses Simulation
Objective 2
Categorization of Inflow
High
Medium
Low
Total Loss
Reservoir Storage States
Inflow Categories
Downstream Demand
Total loss (Evaporation)
Operational & Mathematical Constraints
Klang Gate Dam Release Limits
$R_{min_i}(t) \leq R_i(t) \leq R_{max_i}(t)$
Klang Gate Dam Storage Limits
$S_{min_i}(t) \leq S_i(t) \leq S_{max_i}(t)$
Reservoir Balance Equation
$S_i(t+1) = S_{i,t} + I_{i,t} - R_{i,t} - Ev_{i,t}$
Objective 3
A

(b)

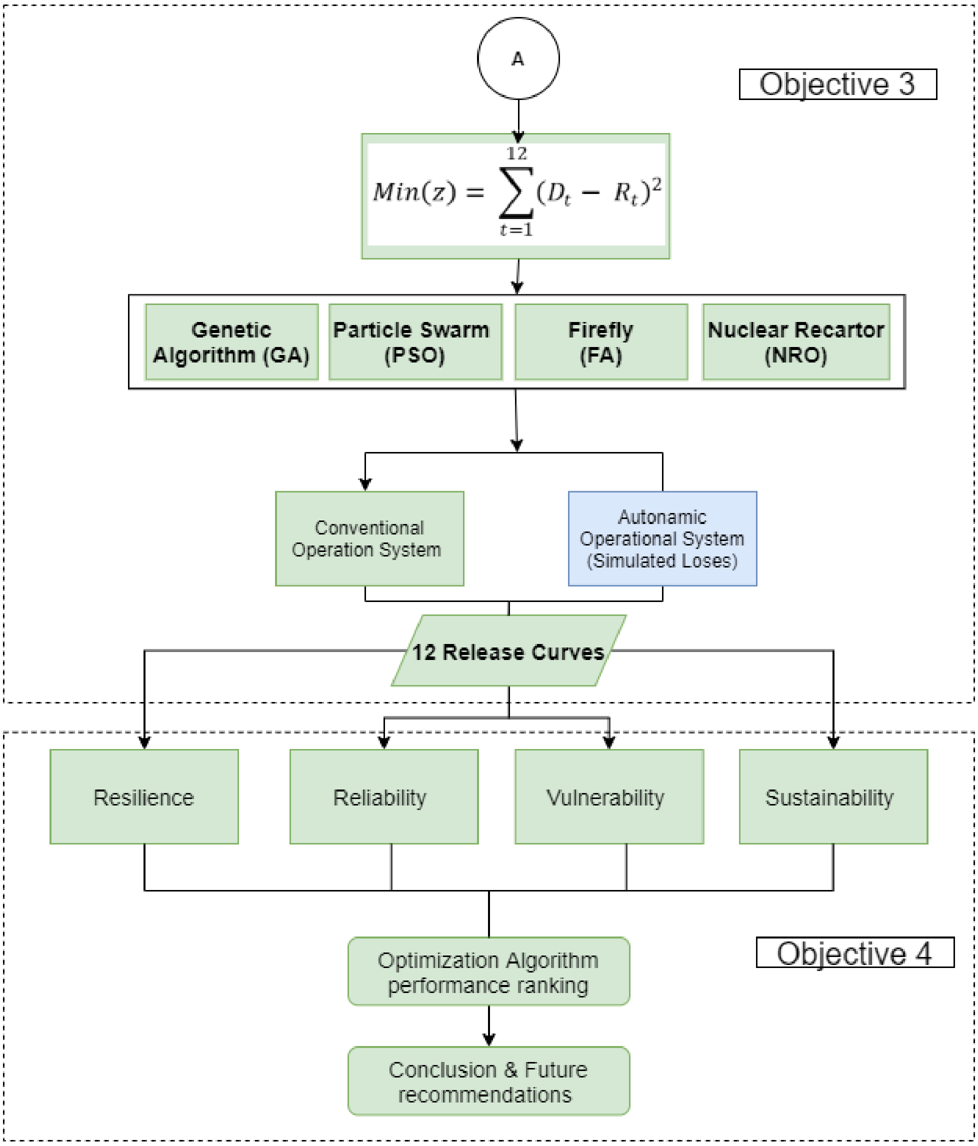


Figure 3.2: Overall Study Flowchart (a) – (b)

3.4 Data Preprocessing

Data analysis is challenging when there are missing values (MVs). For researchers, the existence of Missing Values might cause significant issues. Discarding the instances that include MVs is the easiest method to handle them. However, this strategy is only useful when the data includes only a limited number of cases of MVs and when the inference process will not be seriously biased by the examination of the entire examples. When such faults are present, data preprocessing is typically necessary and readily available to prepare and tidy up the data.

Data preprocessing encompasses data preparation, which is further enhanced by the incorporation, cleaning, normalisation, and transformation of data; as well as data reduction activities, such as feature extraction, instances selection, discretisation, and so on. A final dataset that may be regarded as valid and helpful for future data mining algorithms is the outcome that is anticipated to come about as the consequence of reliable chaining of data preparation operations.

The data collected in this study was supplied by the Lembaga Urus Air Selangor (LUAS) and Jabatan Pengairan dan Saliran Malaysia (JPS). However, unfortunately, despite the many years of operation of the dam, only 20 years of almost complete sets of daily natural inflow at KGD along with water level, rainfall, and evaporation were available during the period between 2000 and 2019, at the KGD reservoir and river. It is important to mention that some of the data were missing and some appear to be faulty which might have occurred due to various reasons, for instance, manual data entry techniques, equipment malfunctions, and inaccurate measurements.

In the current study, the data preparation procedure consisted of these few steps mentioned below:

i) Data Cleansing

ii) Data Integration

iii) Data Transformation

iv) Data normalization

3.4.1 Data Cleansing

Analysing the data to find mistakes and irregularities in the database is the first stage in data cleansing. In other words, this stage is known as data auditing, and it is at this stage that all kinds of irregularities inside the database will be discovered (García et al., 2015). Data cleansing becomes increasingly important when numerous data sources need to be combined, as in reservoir databases or international web-based information systems. This is due to the sources frequently including duplicate data in various forms. In addition, data analysis will be used to uncover issues with data quality by obtaining information about the data attributes. Data mining and data profiling are two methods used as data analysis in the current study. Data segmentation places a focus on the examination of specific instances of particular attributes.

Data mining, on the other hand, focuses on identifying a specific data trend in a vast dataset. Hence, in the current study all the missing data were replaced by averaging the current data for each category and replacing the missing data with the simple moving average approach (SMA). The SMA approach works when a value is absent at a certain time point, the average of the neighbor data points over a given window size and a predetermined number of prior data points are computed to fill the gap. The SMA has been proven to be effective and robust in the replacement of missing data (Hansun, 2013).

3.4.2 Data Integration

It included the process of combining data from two different sources (LUAS & JPS). This procedure was carried out with extreme caution to prevent redundancy and inconsistencies in the data set that was ultimately produced.

A data integration strategy had to meet a number of criteria. First of all, both when integrating several data sources and when using a single data source, it should identify and correct all significant mistakes and inconsistencies. The method involved human examination, and clean data from many sources was combined into a centralised database or data warehouse. Additionally, data integration was done in conjunction with schema-related data transformations based on extensive information, never alone.

The identification and unification of variables and domains, the study of attribute correlation, the duplication of tuples, and the discovery of conflicts in the data values of diverse sources were all carried out within the context of data integration.

3.4.3 Data Transformation

During this stage of preprocessing, the data were either transformed or consolidated in order to make the results of the mining process more applicable or maybe more effective. The smoothing of data, the generation of features, the aggregation or summarisation of data, the normalisation, discretisation, and generalisation of data are all subtasks that fall under the umbrella of data transformation. A major sensitive transformation method that was used in the current study is known as column transformer.

Column transformation in machine learning is the process of preprocessing or applying certain transformations to specific columns (features) in a dataset. This estimator enables the transformation of individual columns or column subsets of the input independently. The attributes produced by each transformer are then combined to create a single feature matrix. The objective is to modify the data in a way that improves machine learning model performance or renders the data better suited to analysis (Ahsan Md et al., 2021).

3.4.4 Data Normalisation

In the current study, various parameters were considered and each parameters have its own unit and reference which impacts in the analysis of the data by each measuring unit that was utilised. Every single one of the characteristics have to make use of a standard scale or range, in addition to being represented in the same units of measurement. The process of normalising the data that were used revolved around giving all of the input variables an equal amount of weight, to avoid misinterpretation of the data by the machine learning model.

A well-known data transformation technique known as the ”one-hot encoding” was used in the current study to transform the data. One-hot encoding splits a single variable with n observations and d unique values into d binary variables, each having n observations. Depending on the observation, the dichotomous binary variable is either present (1) or absent (0) (Potdar et al., 2017).

3.5 Inputs and Forecasting Based-Models

The research considered daily based evaporation, rainfall, water level, and reservoir input data, with a total of seven thousand records for each parameter. The data span a period of 20 years, beginning in the year 2000 and continuing until 2019. Because the objective of this research was to make a prediction regarding the amount of water that would flow into a reservoir, the amount of water that flowed into the reservoir was designated as the Y-variable, and the other three parameters were categorised as the X-variables.

It is the dynamical daily reservoir operating mechanism that alters the control of reservoirs proactively depending on the reservoir's water storage level, the hour of the day, and volume of water coming from the water source (river), as well as reacts to variations in the climate. The research has considered both monthly and daily events as a result. As a result, this study was able to develop a more extensive and accurate scale to evaluate the impact of dams on the regulation of downstream flow, which can be calculated using the river's level and the amount of water released.

One of the major goals of the present research is to anticipate the inflow, with rainfall serving as the primary supply to the reservoir catchment that forms the inflow, hence understanding the climate of Malaysia. It is extremely important to understand the rainfall pattern in the current research. As is the case in other parts of the world, the climate of Malaysia is altering. Numerous recent studies have pointed to global climate change as the cause of variations in the amount of rainfall and the severity of rainfall extremes in Malaysia (Pour et al., 2020; Nashwan et al., 2019; Ridwan et al., 2021; Khan et al., 2019). In the last few years, researchers have seen repeated instances of catastrophic flooding and water stress, which are clear indicators of the effects of growing rainfall extremes. The monthly observed pattern of rainfall that spans over collected data for this study is shown in Figure 3.3.

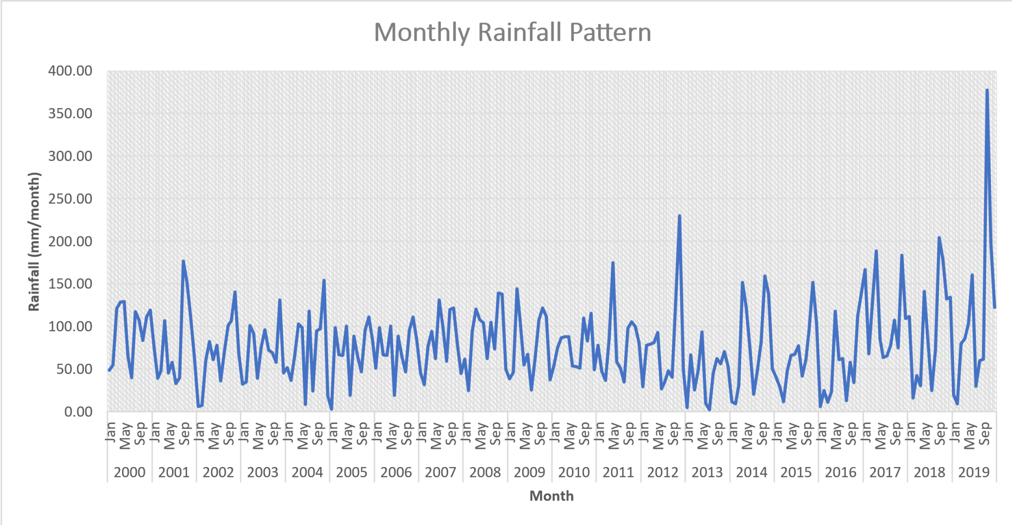


Figure 3.3: Monthly Rainfall Pattern

When the climate and hydraulic properties change due to rainfall, water bodies are affected. As the volume of water streaming into a reservoir is directly proportional to the duration and intensity of rainstorms, variations in rainfall patterns have a significant effect on reservoir parameters. Based on the findings (Li et al., 2015), it is evident that rainfall is an important factor to include when estimating daily inflow since it considerably raises forecasting accuracy.

To sum up, in the next section, the input parameters that were considered in this study and the formulation of the mathematical formula will be explained and shown in detail.

Let us assume that {} is an abbreviation that stands for precipitation time series. It is possible to recast it as a number of time vectors, such as , where (t) represents the quantity of precipitation that was measured at a particular period, which may either be monthly or daily period, or daily-time delay records provided by , by (N) denoting specific number of days, where (t) is the amount of rainfall that was recorded at a certain time.

When it comes to making an accurate prediction of inflow, one of the most important parameters is the water level, which is symbolised by the letter ‘’. The water level is measured over time and provided in the form of a time series that is written down as . Records of water levels may be reorganised to create time series as well as time-delay series, which are provided by the formula (t-n), where (t) stands for the time when each recorded date was taken.

While essential to the hydrological cycle, evaporation is complicated and hard to predict because of the wide variety of variables that affect it. For reservoir planning in dry and semiarid regions, an accurate estimate of evaporation is crucial.

As another example, researchers may refer to evaporation as a time series denoted by the {} By writing as a sequence of time vectors, we may think about it in a different way. … , where (t) represents the time in each record, expressed as a day or month.

For each model, we may draw the following mathematical relationship among the input and output vectors donated by the transfer function .

(3.1)

3.5.1 Support Vector Regression (SVR)

The main goal of SVR is to allow linear regression by non-linearly scaling the training data X into a multidimensional space (F) therefore separating the support vectors between positive and negative hyperplane. This process is denoted by Equation (3.2).

(3.2)

where:

f(x)= Expected Output

φ(x)= Input data feature

w= vector coefficient weight

b= vector offset (threshold)

The objective is to reduce the amount of variation that exists between the two groups in terms of the weight coefficient vector (w) and the offset vector threshold (b). Because of this, Equation (3.3) estimates the two coefficients by using a regularised risk function to determine w and b.

(3.3)

Equation 3.3 is made to satisfy the necessary requirements outlined in Equation 3.4.

(3.4)

In which:

C= Penalising coefficient

Size of each Tube

= Slack variables

φ(x)= Input data features

Equation 3.4 represents a bounded optimisation problem, which may be solved by utilising the primal Lagrangian form indicated in the following Equation: (3.5).

(3.5)

Equation (3.5) is subjected to Lagrangian multipliers constraints which are shown in Equation (3.6)

(3.6)

In consequence of the Equation (3.6), Equation (3.7) is a consideration of the previous multiplier.

(3.7)

Where:

K= activation functions (Kernel)

ai, ai'= multipliers of Lagrangian

b= Offset Vector (threshold)

The kernel function, K, in Equation (3.8) can enhance the mapping process. The information can be implicitly mapped into a subspace using the kernel function, which is particularly effective because it does not require complete knowledge of (). In the current study all the four kernels shown in Equation (3.8) are tested and their performance was ranked based on the statistical analysis.

(3.8)

In summary, the regression function considers the non-zero Lagrange multipliers as well as the support vectors, which are associated with the input vectors of the dataset, hence the final Equation is simplified as below:

(3.9)

3.5.2 Multilayer Perceptron Neural Network (MLPNN)

In this line of study, the studies proposed using a feed-forward neural network that consists of three hidden layers in conjunction with the back-propagation method. The MLPNN model comprises a few transfer functions, the most important of which are presented in Equations 3.10 and 3.11, respectively.

(3.10)

Where:

hj= j^{th} value Hidden Neurons

= activation Function (transfer function)

Wi= Weight Assigned for Each j^{th} Neuron

=Input Data

(3.11)

f

Where:

= k-th the Output layer value

= Transfer Function

hj= Hidden Layer

= Weight Assigned

Transfer functions from Equation (3.10) & (3.11), serve as a representation of the activation function for the output layer and the hidden layer. which are designated by the letter & , respectively. They were tested in the present research, moreover, each of the activation functions, namely the hyperbolic tan function, the rectified linear unit function, the logistic sigmoid function, and finally the identity function which are shown in the Equations below beginning with Equation (3.12) and ending with Equation (3.15), were subjected to testing.

(3.12)

(3.13)

(3.14)

(3.15)

The number of hidden layers and nodes is a crucial element to consider since the model is sensitive to the different combinations of hidden layers and nodes chosen. The optimal number of nodes cannot presently be determined using any acknowledged standard method. The optimal hidden layer (NH) node count for building a neural network is between (2nv0.5 +m) and (2n+1), where n is the total number of input nodes and m is the total number of output nodes (Fletcher and Goss, 1993). In addition, it was hypothesised by (Palani et al., 2008; Ibrahim et al., 2023; Karsoliya, 2012) that the greatest number of neurons that might exist in the hidden layer (NH) is equal to . In the current investigation, in addition to the approach of trial and error, other methods were used, which are represented in Equation (3.16).

In this Equation, represents the hidden nodes amount, and represents the number of inputs that were utilised. A basic diagram of MLPNN which illustrates the model architecture is shown in Figure 3.4.

(3.16)

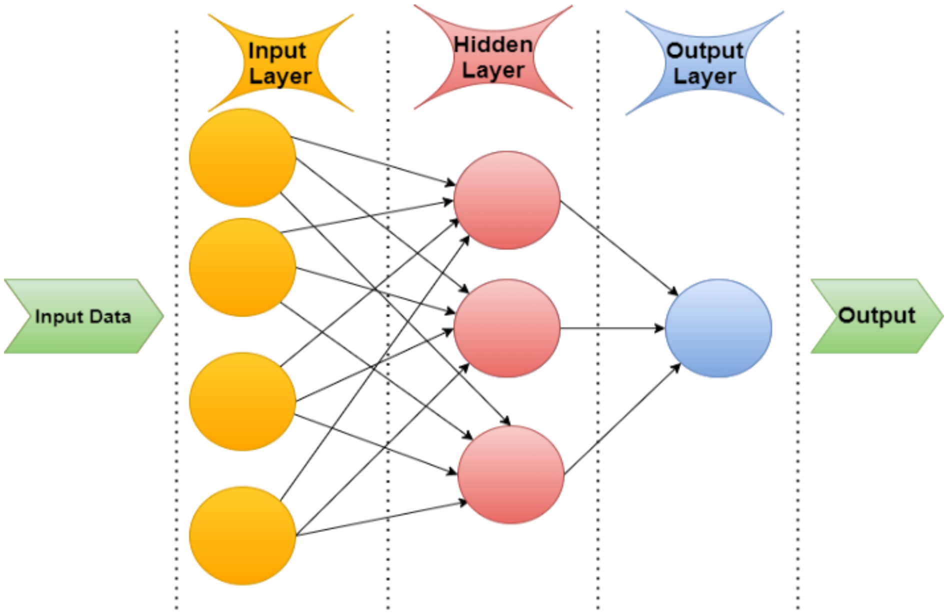


Figure 3.4: MLPNN Model Structure

3.5.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The present investigation used a version of the ANFIS model, and that version had five layers. As can be seen in Equation (3.17), in both the Takagi-Sugeno forms, there are two if-then rules that are not quite clear. Figure 3.5 presents a straightforward representation of the ANFIS, which may be used to get a deeper comprehension of the functioning of the five levels. This representation labels both the inputs and the outputs.

(3.17)

In which:

A_i and B_i = Membership Functions related to x, y

x and y = Inputs

r_i, p_i, q_i = Output parameters

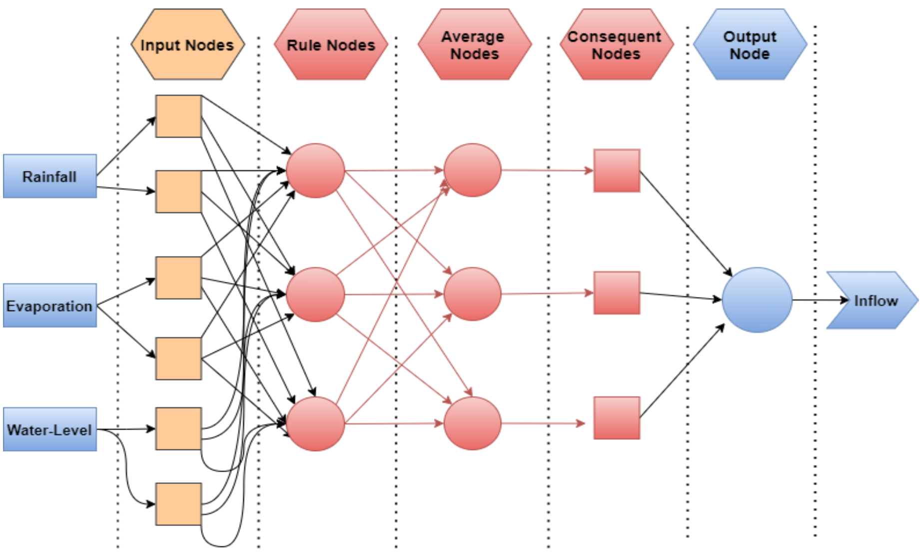


Figure 3.5: ANFIS 5-Layer Structure Model

Layer 1

The name given to this section is the "fuzzification layer." Fuzzification layer generates fuzzy clusters from input data by using membership functions. The mathematical form is shown in Equation 3.18.

(3.18)

Where:

x and y = Inputs to nodes

= linguistic fuzzy defined by the shape of membership functions

Layer 2 (Membership Rules)

A whole fuzzy set may be defined by its membership function. Triangular, Gaussian, trapezoidal, sigmoidal, and generalised bell membership computations are only a few examples. Straight-line segments are included in trapezoidal and triangular functions, although these shapes are not smooth at the corners as defined by the input parameters. Since Bell and Gaussian functions have clear notations and a smooth form, they are often used to characterise fuzzy sets.

It has an extra degree of flexibility that may be used to adjust the degree of steepness at the artificial intelligence node of crossover. Up to this date there is no optimal way to find the best membership function and it all depends on a case-to-case basis (Talpur et al., 2017; Nur Adli Zakaria et al., 2021; Sherif et al., 2021); hence, all the five membership functions were tested in this study.

(3.19)

(3.20)

(3.21)

(3.22)

(3.23)

Where:

{ai, bi, ci} = bell-shape parameters function.

Layer 3 (Normalising/Averaging Rules)

The membership strengths that were assessed in the layer before this one is normalised in this layer so that the discharging strengths of each rule may be differentiated from the overall discharging strengths of all rules.

(3.24)

Layer 4 (Consequent Layer)

In this stage, node I determines how much the ith rule contributed to the final result based on Equation 3.25.

(3.25)

Where:

= Output from Layer 3

= following variables

layer 5 (Output Layer)

At this level, a single node is responsible for calculating the overall output by adding together the portions contributed by each layer.

(3.26)

1.1.1 Extreme Gradient Boosting (XG-BOOST)

In order to find a mapping that is appropriate for the training data sets, XG-BOOST makes use of an ensemble of regression and reclassification trees, often known as CARTs. Each CART has its own unique selection rule framework in the form of a binary tree and stores an ongoing score on each leaf node. The main objective of Equation (3.27) is minimise the difference between each cart score seeking for an optimal score.

(3.27)

Whereby:

Q: differentiable loss function.

: Mapped Result.

: Predicted Inflow.

: The number of leaves in the vth cart.

: Vector that represents the scores of CART leaves.

Minimization for the objective function would not be achieved without training the CART leaves and iterating it based on Equation (3.28):

(3.28)

Where the new term in Equation (3.28) is the which stands for mapping parameter obtained from Vth cart. The minimisation objective function (3.27) is further expanded and corrected by Taylor expansion factor as shown in Equation (3.29)

(3.29)

It is possible to further reduce the complexity of Equation (3.29) by eliminating the constant components as seen in Equation (3.30).

(3.30)

In which the ohm () unit is denoted by Equation (3.31) as shown:

(3.31)

Since the ohm () parameter is directly proportional to the optimal value of the parameter () hence the optimal value is computed using the Equation (3.32):

(3.32)

Normally, it is challenging to list all possible tree architectures The greedy method, which starts with a single leaf and progressively adds branching is an alternative for the conventional method. Equation (3.33) provides the mathematical formula for the greedy algorithm.

(3.33)

A simplified illustration of XG-Boost CART leaves is shown in Figure 3.6.

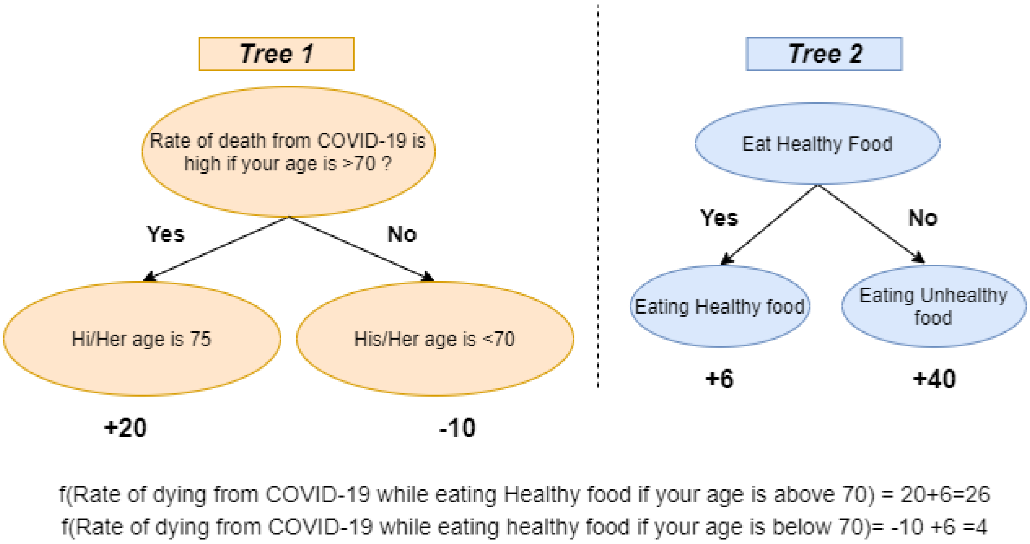


Figure 3.6: XG-BOOST Simplified Model

3.6 Optimisation and Simulation Models

During this stage of the project, there will be two phases that will address objectives two and three, respectively. Nevertheless, these two processes did not take place in the sequence that was specified there. The second goal is to create a simulation of the functioning of the reservoir, but instead of utilising historical data to determine the losses, this study will use anticipated losses instead. In contrast, objective three itself is carried out in two stages, the first of which involves optimising the reservoir operation by taking the historical losses into consideration, and the second of which involves optimising the reservoir operation by carrying out a real-time reservoir operation by integrating the forecasting models into the objective two.

Both of these stages are carried out in order to accomplish objective three. As a result, the artificial intelligence model that is based on loss forecasting operates simultaneously with the optimisation algorithm, so transforming it into a cohesive system. The goal function and the reservoir constraints that bind the reservoir operation, which is the minimum and maximum storage and release during a monthly period have been defined as shown in Figure (3.7).

These are the parameters that define the reservoir operation. In addition to that, the catchment area losses and gains have been included into the calculation used to determine the balance of the reservoir. The water deficit Equation has finally been minimised using a formulation based on a year's worth of data. In conclusion, twelve different release curves that are dependent on three different inflow situations (high, medium, and low) have been constructed as shown in Table 3.2. In the upcoming few subsections, the methodology of how each optimisation algorithm operates will be discussed.

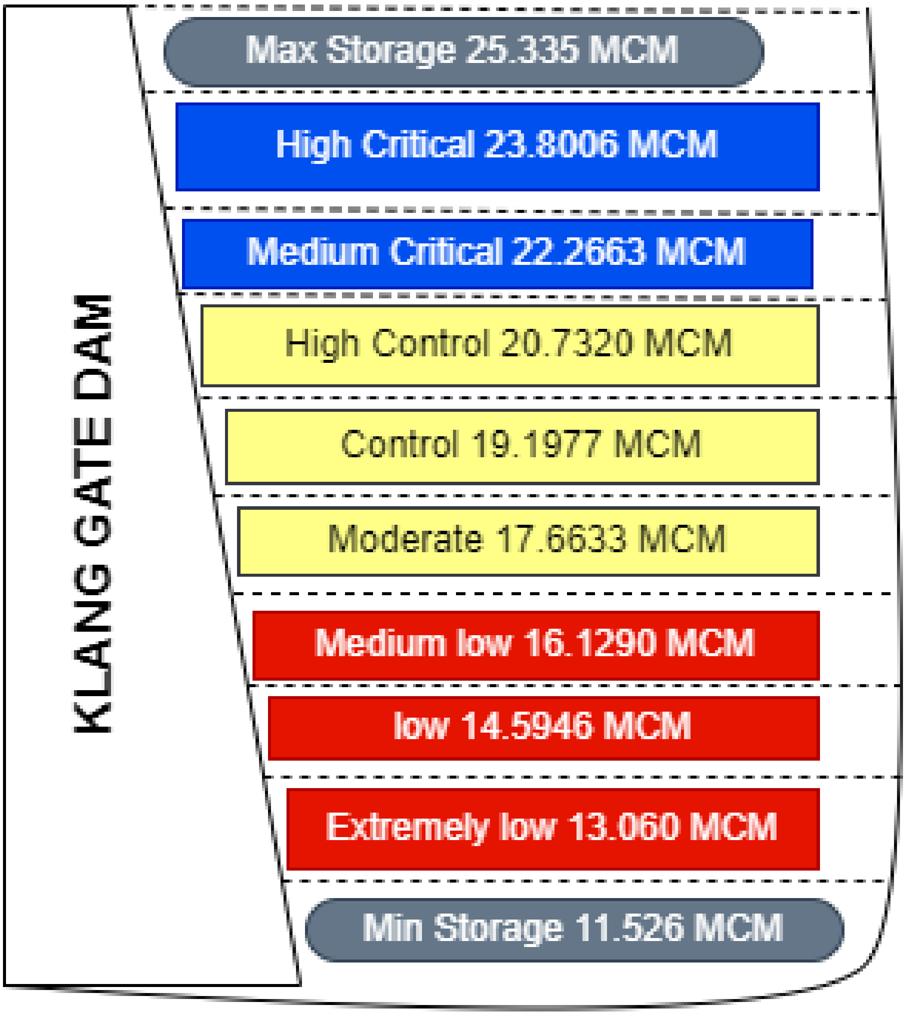


Figure 3.7: Reservoir Storage Levels

Table 3.2: Ranges of Inflow

Inflow Category	Inflow Range (MCM)
High	8.40 & above
Medium	5.21 – 8.39
Low	2.00 – 5.20

3.6.1 Real-Coded-Genetic-Algorithm (RCGA)

There are many different sorts of GA operators, which results in many different varieties of GA, such as real-coded GA (RCGA) and binary-coded GA (BCGA). When binary and real-coded GA are compared, both conceptual evidence and facts from the actual case studies show that RCGA performs better than BCGA in the majority of circumstances, especially when dealing with real-world optimisation problems (Katoch et al., 2021; Akopov et al., 2019) As a consequence, this research uses RCGA as an optimisation strategy. The RCGA runs on a few successive phases that are shown in Figure 3.8.

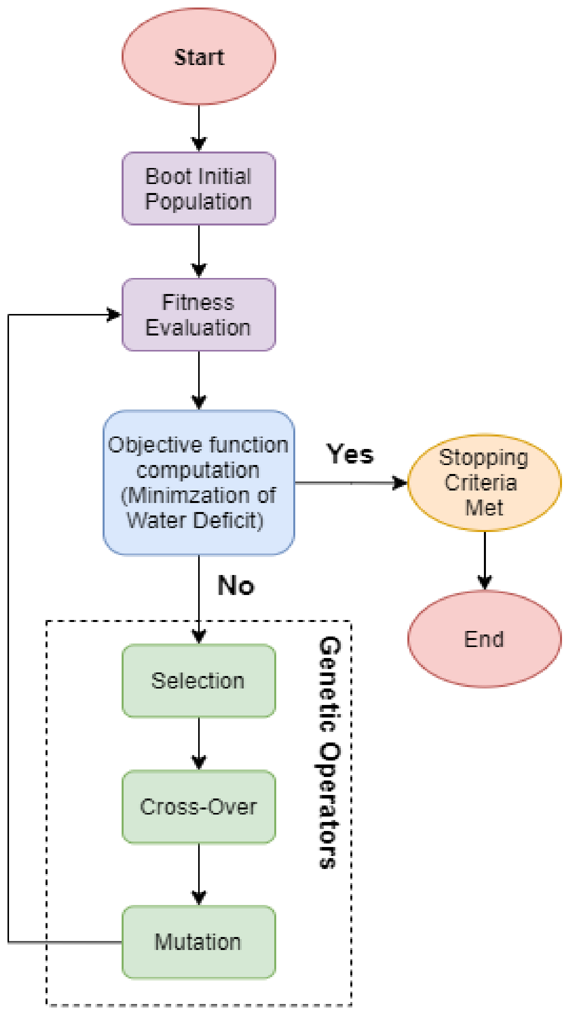


Figure 3.8: RCGA Flowchart

3.6.2 Particle Swarm Optimisation (PSO)

The PSO technique conducts its search for optimal solutions in a search space that is composed of many dimensions using a swarm of particles. Although each particle is a product of a greater system, it is nonetheless formed and influenced by both the environment in which it is located and the events that occur within its own existence. The PSO technique conducts its search for optimal solutions in a search space that is composed of many dimensions using a swarm of particles.

Although each particle is a product of a greater system, it is nonetheless formed and influenced by both the environment in which it is located and the events that occur within its own existence. Each PSO particle is given a location and velocity, and those values are updated depending on the finest particle as shown in Equation (3.34), such that the remainder of the swarm will join the best individual as shown in Equation (3.35).

(3.34)

(3.35)

Whereby = velocity; = weight of Inertia; = Constant integers; = Random variables; = particle best location achieved; = newly born member of the flock; = Existing individual member of the flock; = global best position achieved by one member of the flock. The whole process that PSO goes through in order to optimise the current problem is shown in Figure 3.9, which may be seen below.

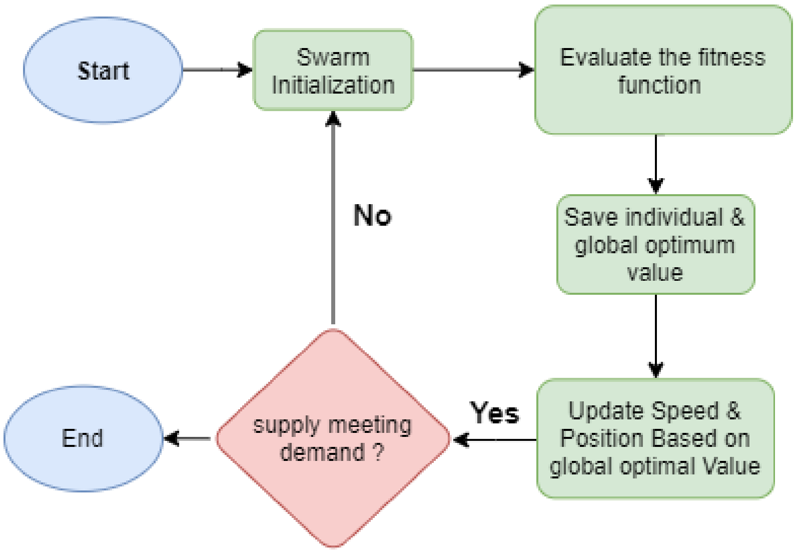


Figure 3.9: PSO Flowchart

3.6.3 Nuclear Reaction Optimiser (NRO)

As was noted earlier, the NFi and NFu phases of a physics-based algorithm that was inspired by nuclear processes are employed for utilising and analysing a search solution space. Because nuclear processes make use of several operational principles, the method in question has been given the name "nuclear reaction optimisation." The NRO approach, much like other meta-heuristic algorithms, begins by randomly constructing an evenly dispersed populations of N nuclei. This process is known as initialisation, and it is depicted by the mathematical formula in Equation (3.36)

(3.36)

Where is the initial population generated, stands for the lower and upper bond for the dth variable located in the search space, and lastly, rand stands for randomly distributed integers between 0 and 1.

The continuous production of neutrons by an earlier event is often what keeps the reaction going in a process known as chain nuclear fission. It has been suggested that the nuclear fusion of two distinct random nuclei might result in the production of hot neutrons following Equation (3.37).

(3.37)

Whereas stands for the heated neutron, 'Xi' and 'Xj' stand for the initial and final nuclei that result from the fission process, respectively. The fission process uses Gaussian distribution when locating the global best optima as shown by Equation (3.38).

(3.38)

However, there is a question which is how does the fission process computes ? The answer to this question is logged and used to compute sigma determined by calculating the difference between the value of a random nucleus formed and the best nucleus value produced. The fusion step of the algorithm comes at the very end of the process.

Nuclear fusion processes may be categorised as either hot nuclear fusion or cold nuclear fusion, depending on the conditions present throughout the ionisation event. When nuclei are heated to temperatures above that of plasma, a process known as thermal nuclear fusion may take place. This occurs when the nuclei combine to produce a heavier nucleus than the original light nuclei.

The ionisation stage is the initial step in the process of nuclear fusion. In accordance with the principles of ionisation, the NFL phase utilises this step for the purpose of investigation. The ionisation Equation is shown in the following Equation:

(3.39)

Where represents the ionized variable, whiles stands for the worst and best random variable found during the fission states. The complete process of (NRO) is shown in Figure (3.10).

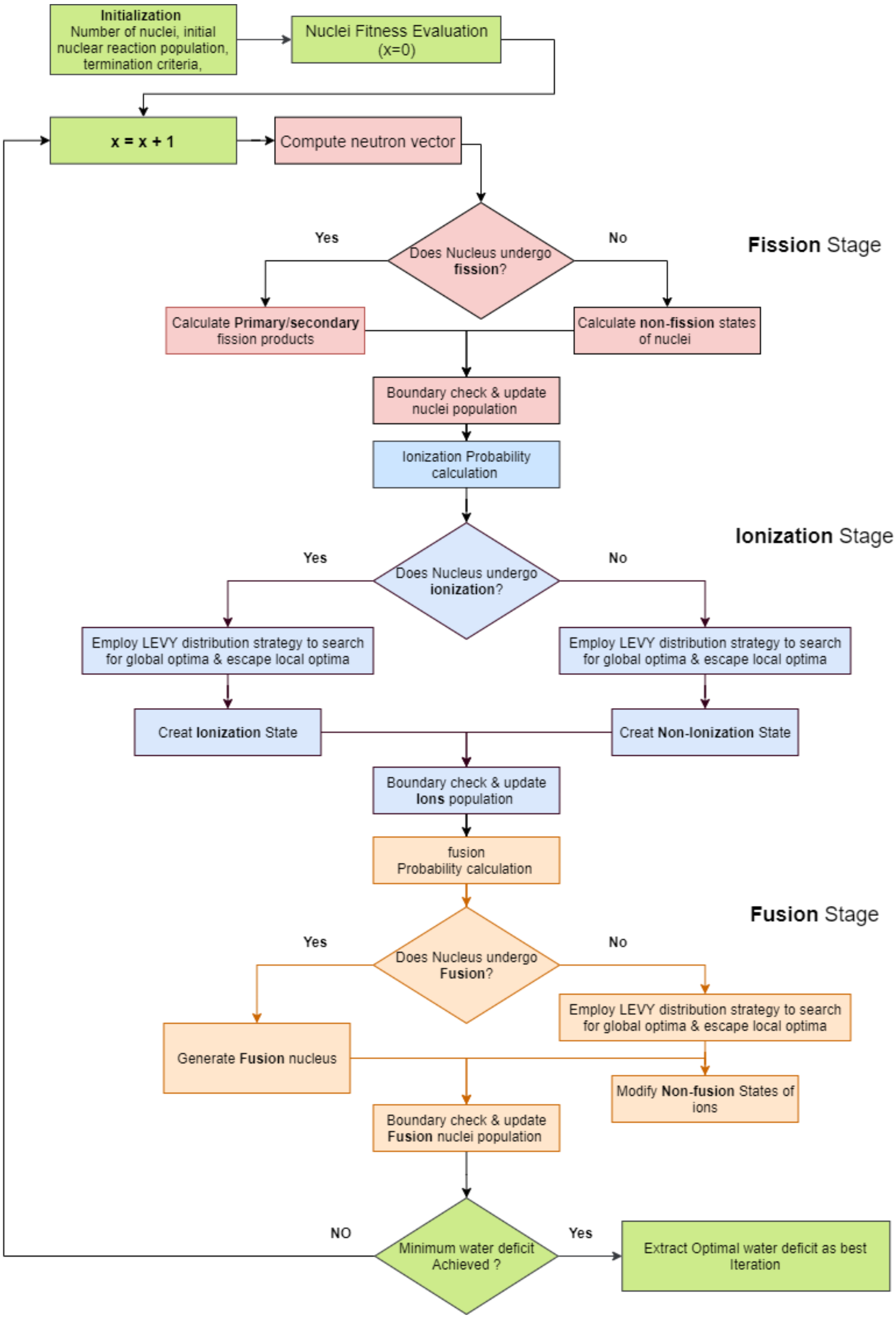


Figure 3.10: NRO Flowchart

3.6.4 Firefly Optimisation (FA)

The FA was influenced by firefly' innate behaviour. Fireflies use their accumulated energy to manifest as light in order to procreate, feed, or avoid predators. Fireflies attract attention by emitting light. The FA presupposes the next two idealised rules.

Since all fireflies are gender-neutral, their attractiveness is decided by the amount of light they flash, as quantitatively shown in Equation (3.40).

(3.40)

In which = attractiveness of fireflies; = attractiveness at the starting point when no distance moved (r = 0); = coefficient used to represent the absorption of light

The allure of fireflies is directly proportional to the intensity of their glow. Therefore, if there are two flashing fireflies in the area, the firefly that flashes less often will go toward the firefly that flashes more frequently.

The greater the gap that separates two fireflies, the less stunning and brilliant each of them will become. As a consequence of this, the migration of fireflies continues until there are no fireflies in the group that are more brightly lit. When this takes place, the fireflies will wander about in a random pattern. It is necessary to calculate Equations (3.41) and (3.42) in order to understand their motion in terms of distances and locations.

(3.41)

In which = is the cartesian distance measured among a pair of fireflies & ; & = kth dimension of the spatial coordinate of the iterated firefly for I & j respectively; d = the number of dimensions analysed to obtain the optimal solution.

(3.42)

By which and = new position of firefly I with less brightness and current position of firefly I with less brightness, respectively; = position of firefly j with more brightness; α =a randomized parameter; and stands for random variable chosen between 0 and 1. The processes that firefly goes through, from the very beginning to the very end, in order to produce an optimum solution for the optimisation issue are shown in Figure 3.11.

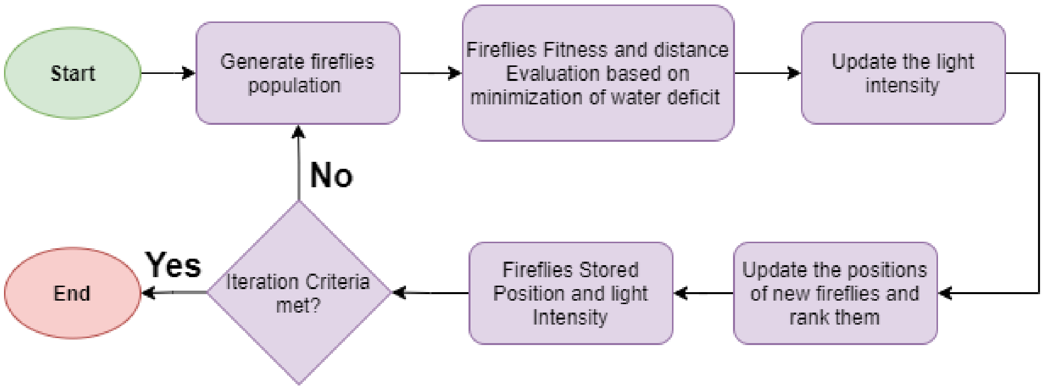


Figure 3.11: Firefly Flowchart

Determining an appropriate objective function is a crucial part of any optimisation model that aims to achieve a certain goal. The primary goal of this investigation is to reduce the possibility of water shortages in the public water supply as well as in the demand for irrigation. In order to better align with the planned operating strategy of Malaysia's water managers and planners. The formulated objective will be covered in the following section (3.7).

3.7 Objective Function and Constraints

The main objective function for this study is deemed to be achieved when the gap between the downstream water demands and water discharged is near to zero during a 12-month period. Equation (3.43) translates it into mathematical form that is understood by the optimisation models.

(3.43)

Where:

Z: Shortage of water supply

: Downstream Demand at a certain Month (t)

: Released water at a certain month (t)

The value that is calculated at any particular period for a certain amount of water that is released will have an effect on the reservoir storage value in the period that comes after it, in accordance with the continuous-discrete method. In a simpler word there must be an Equation that updates the reservoir at the beginning of new water cycle, such Equation is given by Equation (3.44).

(3.44)

The activities of the reservoir should adhere to the operational rules and follow the technical constraints. Some examples of these constraints are the reservoir water balance, the allowable boundaries of water release and storage, as well as the ultimate aim of water storage. The two constraints that were considered in the current research are the storage and release as shown in Equation (3.45) and (3.46), respectively.

(3.45)

(3.46)

Researchers have only lately created a number of limited handling strategies in order to retain the applicability of optimisation algorithms to constrained optimisation issues and avoid the algorithm search from leaving the viable zone (Zhou et al., 2019; Kulkarni et al., 2018; Chang et al., 2010). Constrained handling strategies include, but are not limited to, the use of penalty functions, repair methods, isolating the purpose from the constraints, and using particular representations and operators.

But among the EA community's most established and widely utilised strategies are penalty functions (Kumar et al., 2020). The process of penalty functions take place by converting a restricted optimisation problem into an unconstrained one by adding the constraints to the objective function in the form of a penalty term. When the solution cannot be implemented, the goal value is reduced according to a predetermined reduction factor. Two penalty functions will be considered, and the formulas are shown in Equation (3.47) and (3.48).

(3.47)

(3.48)

Where:

: Penalty Coefficients were chosen by trial & Error depending on the range.

: Minimum storage (Lower Bound)

: Maximum storage (Upper Bound)

The reservoir storage capacity limit range is associated with the penalty coefficients C1 and C2, which may be written as: (maximum and minimum storage). An iterative method is used to determine the values that should be assigned to these two variables so that they can accurately reflect the tolerance that the objective function has for the end outcome.

As a direct consequence of this, Equation (3.43) has been altered such that it now includes a correction for the objective function as shown in Equation (3.49).

(3.49)

3.8 Models Evaluation

The present study makes use of a number of statistical measures in order to assess the performance integrity of the presented technique to predict inflow and optimise the entire reservoir operation. This was done in order to accomplish the objectives of the research. These indicators are being used to analyse not just the effects of the training process but also the efficiency of the models when they are being tested

There are four primary indices that have been modified in order to evaluate the effectiveness of the optimisation methods, namely vulnerability, reliability, resiliency, and sustainability. Each and every index will be briefly explained in the next few sections

3.8.1 Root Mean Square Error (RMSE)

In hydrology, air quality, and climate research investigations, the root mean square error (RMSE) has been employed as a standard statistical tool to quantify the performance of the models (Chai and Draxler, 2014). The basic concept behind RMSE is that is computes the difference between the projected data and the historical data. The RMSE mathematical Equation is shown in Equation (3.50).

(3.50)

Where:

n = observation samples

= forecasted inflow for i-th

= observation Data

3.8.2 Median Absolute Error (MedAE)

The fact that the median absolute error is unaffected by outliers makes it a particularly intriguing statistic. The number that represents the loss is derived by finding the middle value of all the absolute deviations that exist between the goal and the forecast data. The Equation is given as follows:

(3.51)

Where:

n = observation samples

= forecasted inflow for i-th

= Observation Data

3.8.3 Mean Absolute Error (MAE)

The MAE is appropriate for use when describing mistakes that are evenly distributed because it is more common for model errors to have a uniform distribution than a non-uniform distribution. The Equation is given as below:

(3.52)

Where:

n = observation samples

= forecasted inflow for i-th

= Observation Data

3.8.4 Coefficient of determination (R)

In regression models, the coefficient of determination, which also goes by the name R^2, indicates the percentage of variance in the dependent variable that can be accounted for by the predictors that are included in the model. This coefficient is well-defined. To extend it for our forecasting models, we use the Equation below.

(3.53)

While the lower bound of the previous four statistical models is mainly limited to zero the upper bound however is not restricted to a certain value. As the sample size expands, the upper bound of RMSE (Root Mean Square Error) also expands.

This implies that even if one model is superior to the other, two models with different sample sizes might have the same RMSE. As a result, it is critical to exercise caution when comparing RMSE values from two different sample sizes. Some relations between RMSE and MAE can be presented as shown in Equation 3.54 below:

(3.54
)

3.8.5 Reliability Index

The reliability of the water demand is measured as the chance that the available water supply will be sufficient to fulfil the water demand during the duration of the simulation. In other words, Reliability is the likelihood that there will not be any failures during a certain time frame, which is often believed to be the planning period. The Equation of reliability is given by:

(3.55)

In which:

= Water Deficit

= Target demand and supplied demand

Equation (3.55) is subjected to form the main reliability Equation shown by Equation (3.56)

(3.56)

Whereby stands for the number of time intervals considered.

3.8.6 Resilience Index

The ability of a system to adjust to shifting circumstances is known as resilience. Resilience must be considered as a statistic that evaluates the ability of water management operations to respond to changing circumstances since climatic conditions are no longer constant. Resilience may also be defined as the likelihood that a system will recover after a time of failure. The resiliency formula to assess the reservoir failure probability is given by Equation (3.57).

(3.57)

3.8.7 Vulnerability

The potential impact of losses, if they materialise, is referred to as a vulnerability value. Vulnerability, in its most basic form, is an expression of the intensity of failures. One way to describe vulnerability is as (1) the typical rate of failure. (2) the average of the largest deficits during the whole of the continuous failure time, and (3) the chance of going beyond a specific deficit threshold. The vulnerability Equation adapted in the current study is based on the average failure; hence the Equation is shown as follows:

(3.58)

In which, failed models not meeting the demand, : Targeted demand, : Release generated by the metaheuristic optimisation models.

3.8.8 Sustainability Index (SI)

(Loucks and van Beek, 2017) suggested the SI as a means of quantifying the sustainability of water resources systems. Additionally, the SI was developed with the intention of making it easier to evaluate and compare various water management approaches.

The Sustainability Index (SI) is a summary index that evaluates the sustainability of water resources systems; it may be used to evaluate the sustainability for water users and to get the change in sustainability by analysing the index among many water Guidelines that have been offered. Hence, the SI could be seen as the summation of all the previous water indices as shown in Equation number (3.59).

(3.59)

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Parameter Tuning

In this section, the approach used to tune the parameters of the machine learning models and the optimisation algorithms will be discussed and reviewed. Tuning machine learning parameters is an essential step in training a machine learning model. It involves adjusting the values of the parameters that control the behaviour of the model in order to optimise its performance on a given task. There are several reasons why tuning machine learning parameters is essential, as one may frequently enhance a machine learning model 's efficiency on the problem it was developed to handle by adjusting its parameters. This may be particularly essential if the objective is crucial, like when the model is being used to make judgments that might have serious repercussions such as the forecasting task discussed in this research.

Another important reason for parameter tuning is that it helps in minimising over-fitting. Over-fitting arises when a model tries to match the training data noise instead of the underlying pattern because it is too sophisticated. This might result in inaccurate generalisation to new predicated data.

To conclude, the enhancement of parameters in a machine learning model plays a significant role in expediting the training process. This aspect holds particular significance within the context of the current case study, considering the utilization of a substantial dataset during the model's training. In the current research, two methods of tuning were considered to tune the hyper-parameters of the forecasting models which are Grid-Search Method and Cross-Validation (K-fold).

Identical to the repetitive randomised subsampling procedure, cross-validation ensures that no two test sets overlap by careful sampling. The available learning set was divided into ten k-folds distinct subgroups of roughly similar size for k-fold cross-validation. The number of resultant subsets is the "fold" in this context. The learning set's cases were randomly sampled for this division without being replaced. The training set was made up of k-1 subsets, which collectively made up the learning dataset. The performance of the model was then evaluated once it has been applied to the final subset, also known as the validation, set. Up till each of the k subsets has intervened as a validation set, this process was repeated. The results of the cross-validation sampling are shown in Table 4.1.

It is significant to highlight that there is no recognised technique for estimating the frequency of folds; rather, it is totally dependent on the kind and amount of noise in the data. The optimal range of K-folds, however, was shown to be between 4 and 19folds in all of the prior research, and it was suggested that the optimal number of K-folds is 10, especially when the computing system is confined. (Berrar, 2018; Wang et al., 2019; Arlot and Lerasle, 2012).

Table 4.1: Inflow Scenario vs Tuning Option

Inflow Scenario	Model	Tuning Option	Coefficient of Determination
Monthly	SVR	K-FOLD (10-Folds)	0.4832
		Grid-Search	0.7757
	MLP	K-FOLD (10-Folds)	0.2863
		Grid-Search	0.6501
	ANFIS	K-FOLD (10-Folds)	-0.0718
		Grid-Search	0.3894
	XG-Boost	K-FOLD (10-Folds)	0.9885
		Grid-Search	0.5885
Daily	SVR	K-FOLD (10-Folds)	0.3568
		Grid-Search	0.6492
	MLP	K-FOLD (10-Folds)	0.1598
		Grid-Search	0.5236
	ANFIS	K-FOLD (10-Folds)	-0.1983
		Grid-Search	0.2629
	XG-Boost	K-FOLD (10-Folds)	0.8620
		Grid-Search	0.4621

According to Table 4.1, each of the four model were given a trial dataset from the original database. The training amount of data was 80% out of the trial set given to each model while the testing was the remaining 20%. Table 4.1 clearly shows that the four models did not perform well when tuned with K-Fold cross-validation method. However, the suggested model's accuracy significantly improved with the addition of grid-search accompanied by a Trial-and-error approach.

The suggested models provide the greatest accuracy in forecasting the inflow by including this stage in the machine learning (ML) process. The value of the coefficient of determination in predicting monthly rainfall using K-fold varies between -0.07 to 0.9885, and for the grid-search it ranges between 0.2629 to 0.7757.

In daily inflow prediction, it ranges between -0.1983 – 0.46 for K-Fold tuning approach while for the grid search method it ranges between 0.2629 – 0.86. Based on these findings, it can be claimed that all four models are capable of forecasting inflow over a range of time horizons with a respectable degree of accuracy, and that the model's accuracy increased as additional inputs were included.

When employing a metaheuristic algorithm to address an optimisation problem, two primary stages require user input: defining the objective and assessing the model's fitness. The objective is usually well-defined in most cases and can easily be changed to suit the user demand. However, selecting the component of an evolutionary algorithm is challenging as each component contains its own variables such as mutation probability, selection size, population size, iteration rate, etc. These parameters' settings have a significant impact on the algorithm's ability to efficiently identify solutions that are close to optimal. Nevertheless, selecting the appropriate parameter values is a time-consuming process.

Fundamentally, researchers differentiate between parameter control and parameter tuning as the two primary methods of setting parameter values (Eiben et al., 2007). By parameter tuning, researchers mean the method that is frequently used, which entails determining suitable values for the parameters before the algorithm is performed and then executing the algorithm using these values, which stay constant during the iterations. Tuning these parameters has a crucial effect on the exploration and exploitation ability of the optimisation model.Exploration and exploitation are particularly significant in any optimisation method since they describe the pattern of attaining an optimum state for a system's overall iteration stages. Explore and exploit are two distinct processes in the search space. Exploration involves visiting new parts of a search space.

In particular, exploration entails searching the solution space more thoroughly, and during this stage, the algorithm makes every effort to forget its initial local point. In particular, the impact of the good local points must be lowered in the case of local optima, resulting in the algorithm emerging in a distinct direction to discover a divergent solution in the global optima.

Additionally, exploitation refers to making the most use of the advantageous population ancestors in order to develop in a certain path, thus not only causing the algorithm to converge more rapidly but also shrinking the search area. In other words, visiting earlier search spaces with the aim of pushing the search process to its limit.

A search algorithm must have a good balance of exploration and exploitation to be effective. Firstly, to address these optimisation issues, a series of exploratory runs should be carried out to determine the right values for the suggested algorithms' parameters.

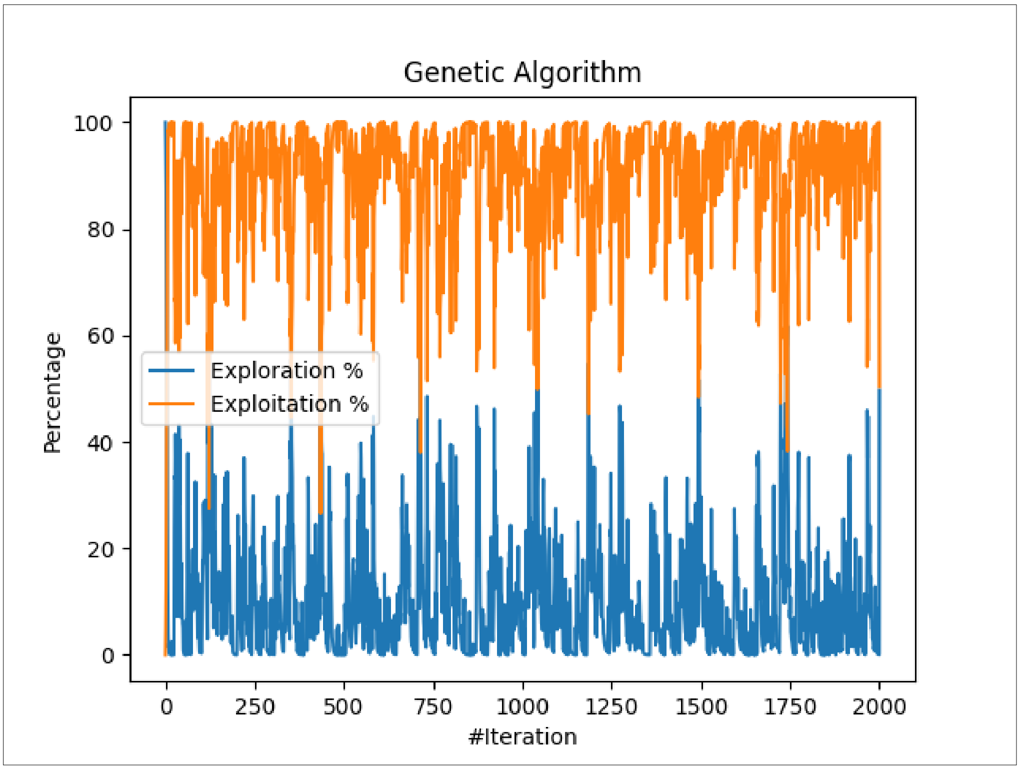
The number of iterations and population size were maintained while the rest of the parameters were adjusted until the optimum parameters have been achieved, as seen in Table 4.2. As it can be seen in Table 4.2, firefly (FA) is a parameter-rich algorithm as it contains seven major parameters that need to be tuned followed by Particle Swarm Optimisation (PSO) that has six parameters that require tuning. Genetic algorithm (GA) and nuclear reaction (NRO) have a relatively small number of tuning parameters, suggesting that they are more user-friendly and less time-consuming.

Table 4.2: Parameters Tuning

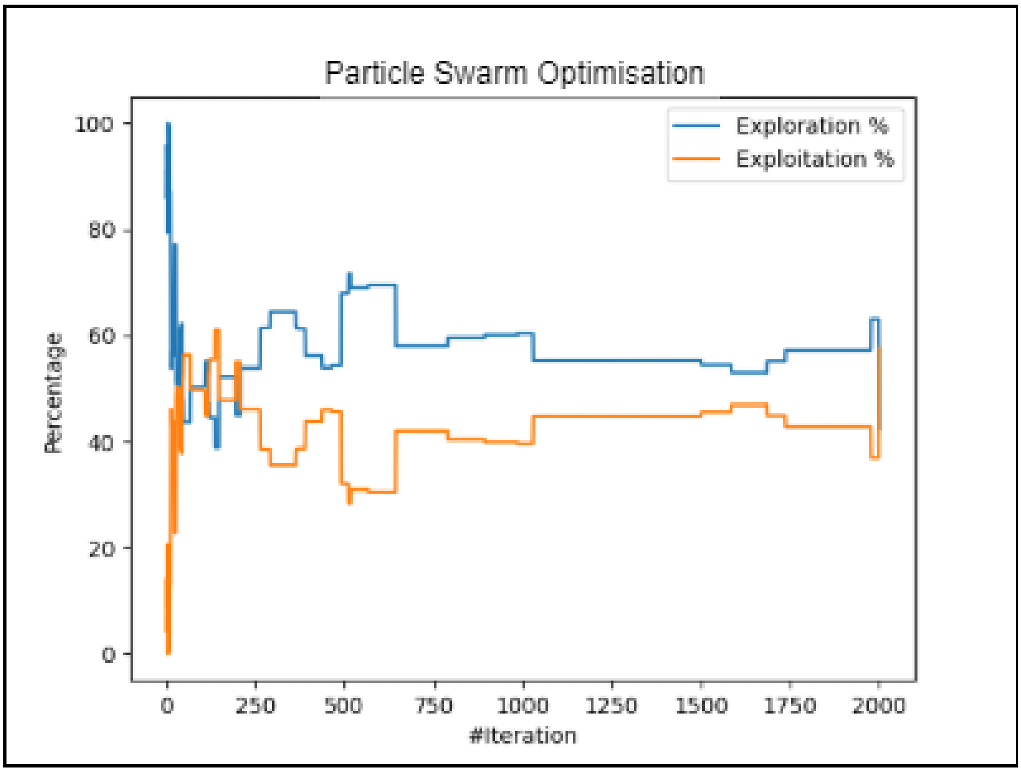
Optimisation Algorithm Model	Parameters	Values
Genetic Algorithm (GA)	Iterations	2000.00
	Population Size	10.00
	cross-over probability	1.00
	Mutation Probability	0.02
Particle Swarm Optimisation (PSO)	Iterations	2000.00
	Population Size	10.00
	C1 (Local coefficient)	2.05
	C2(Global Coefficient)	2.05
	w_min (Minimum weight of bird)	0.40
	w_max (Maximum weight of bird)	0.90
Firefly Algorithm (FA)	Iterations	2000.00
	Population Size	10.00
	γ (Gamma coefficient for light absorption)	5.50
	β (Beta coefficient for attraction)	2.00
	α (Alpha coefficient for mutation)	0.50
	α_ damp (Alpha coefficient for mutation damp rate)	0.99
	Δ (Delta coefficient for step size mutation)	0.05
Nuclear reaction optimisation (NRO)	Iterations	2000.00
	Population Size	10.00

According to the selected parameters, all four optimisation algorithms have been developed successfully and could search through the search space while maintaining a balance between the exploration and exploitation process, as seen in Figure 4.1.

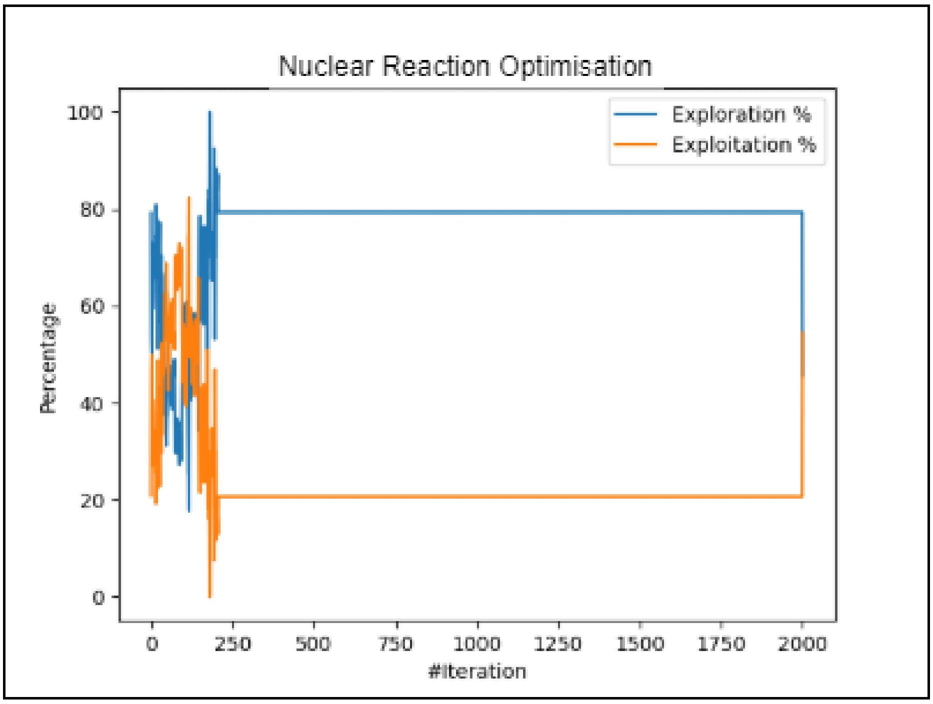
(a)


Genetic Algorithm
Percentage
#Iteration
Exploration %
Exploitation %
0
20
40
60
80
100
0
250
500
750
1000
1250
1500
1750
2000

(b)


Particle Swarm Optimisation
Percentage
#Iteration
Exploration %
Exploitation %
0
20
40
60
80
100
0
250
500
750
1000
1250
1500
1750
2000

(c)



(d)

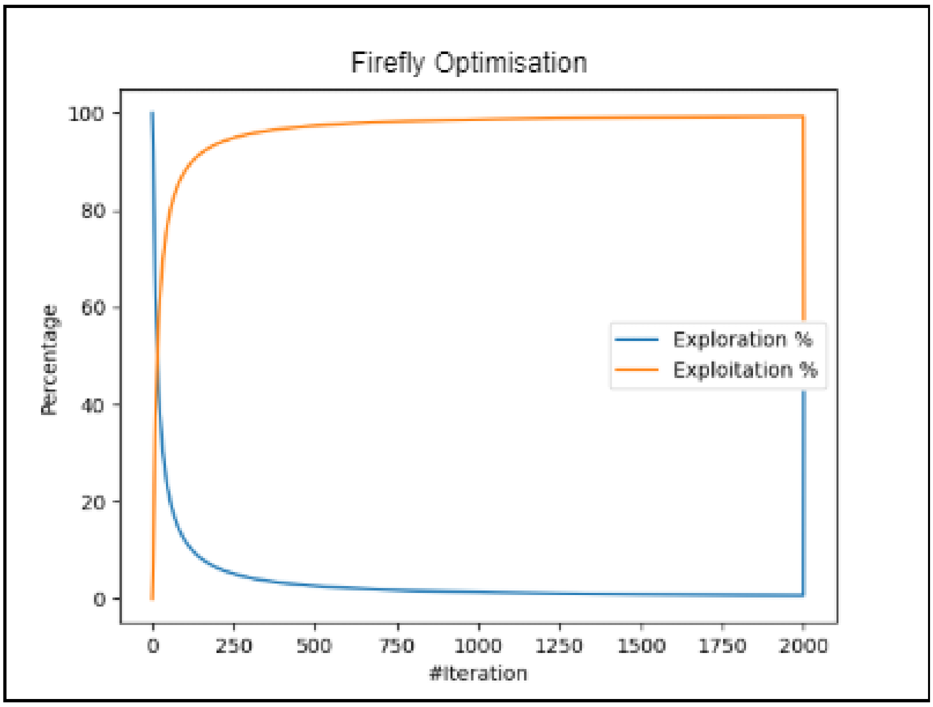


Figure 4.1: Exploration Vs Exploitation (a) Genetic algorithm, (b) Particle Swarm Optimisation, (c) Nuclear Reaction Optimisation and (d) Firefly Optimisation

As is seen in Figure 4.1, all four algorithms had a similar shape of a plot among the exploration and exploitation curves; as if they are mirroring each other, which illustrates that each algorithm was trying to balance both phenomena. An analysis of the Genetic algorithm graph shows its strong capability in searching the search space as the patterns were deep and interconnected between each other, forming a triangular ascending and descending pattern that shows that the algorithm is taking a long time in finding the local and global optima and that it tries to avoid being stuck in the local space. In a genetic algorithm, the exploration and exploitation mainly depend on the crossover operator.

To increase the likelihood of producing superior offspring, a crossover operator mixes two or more parents. The notion that knowledge sharing between good offspring individuals would result in offspring who are even better than their parents can be used to justify such a pairing. A crossover operator is viewed in this light to be more of an exploitation operator. A successful crossover operator should also produce individuals in the exploratory zone. The exploratory power of an operator is the number of possible ways it might create a new offspring individual. In many circumstances, predicting whether newly created individuals generated by crossover and/or mutation operators will fit into the search space zones, is challenging (Črepinšek et al., 2013). The Particle swarm optimisation shows a sharp pattern through the first 100 iterations. However, as it continues iterating the algorithm untangles itself, forming a clearer rectangular trend.

It is evident from the line graphs that the PSO and FA were exploitative in the majority of repetitions. An interesting observation of the NRO shows that it had a similar graphical pattern as the PSO but, with more tangles among exploration and exploitation for the first 200 iterations, and as it iterates further the pattern changes to a linear shape.

On the other hand, the Firefly graph has formed a bowl shape with an increase in the exploration and exploitation percentage until it reaches iteration number 2000, and then it drops immediately. One reason for this could be that the algorithm has not found the global optima yet and a greater number of iterations is required. FA was extremely excellent on exploration and similarly poor on exploitation, which was the two opposing qualities. This is due to Levy Flight, which makes use of the swarm's long-distance mobility to assist in escaping the local optima.

4.2 Forecasting-Based-Model (Monthly Inflow)

4.2.1 Forecasting Scenarios

Forecasts for both the long-term and short-term reservoir inflow are required for this research. As a result, several input patterns were used with the algorithms SVR, XG-BOOST, ANFIS and MLPNN, to predict the reservoir inflow. There have been seven different scenarios formulated for this study starting from scenario-1 with no time lags consideration for any input variable while for scenario-2, scenario-3 and scenario-4 time lagging was considered for water-level input variable.

Similarly, scenario-5, scenario-6 and scenario-7 had a one, three and five time-lags for historical inflow and water level as shown in Table 4.3. In general, performance metrics were used to determine which approach is better. During the testing period, all indicators were determined by comparing the output of the four models to the actual inflow records (20 years).

Table 4.3: Forecasting Scenarios

No. Of Scenarios	Inputs	Outp
1		
2		
3		
4		
5		
6		
7		

Notation:

4.2.2 Support Vector Regression (SVR)

To begin, there are many different parameters in the SVR that need to be calibrated. Nevertheless, according to the findings of our research, the kernel type is the hyper-parameter in the SVR that has the greatest impact, followed by the Gamma, coef, and epsilon. The SVR model makes use of a total of four kernels: the (RBF), the Sigmoid (SGD), linear kernel, and the polynomial kernel (POLY). Only some parameters affect each kernel; for example, the coef effect is only meaningful when the polynomial and sigmoid kernels are used together. On the other hand, the linear kernel is unaffected by either gamma or coef.

The model was initially constructed using the default (def) configuration settings for each parameter, as shown in Table 4.4. The default value for Coef and epsilon was 0.0, and 0.25 for gamma. Initially, The linear kernel has shown excellent result, while the polynomial kernel was far from accurate. However, these findings might be misleading. A value based on the (MAE), the linear method has shown (MSE), the R^2 value, and the median absolute error, which are correspondingly 42.346, 2420.5624, 0.5813, and 42.4249. During this time, the value of the polynomial has fluctuated between 60.1283, 8111.9762, -0.4032, and 41.8226, respectively. When calculated using the default settings, the model does not perform well enough to provide the desired results in their optimal form. As a result, we were able to improve the model by using a diverse set of values for each parameter. For example, the gamma value ranged from 0.00001 to 0.1, and the epsilon value ranged from 0.001 to 1.

Table 4.4: SVR Parameter Tuning

Kernel Type	Gamma	coef0	epsilon	Mean Absolute Error	Mean Square Error	R^2	Median Absolute Error
RBF	1 / n_features	0.0000	0.1000	44.5896	3163.0666	0.4529	42.0476
Sigmoid	1 / n_features	0.0000	0.1000	47.1196	3390.5640	0.4135	41.0463
linear	1 / n_features	0.0000	0.1000	42.3461	2420.5624	0.5813	42.4249
Polynomial	1 / n_features	0.0000	0.1000	60.1283	8111.9762	-0.4032	41.8226
RBF	0.0100	-	0.1095	44.1958	2564.1918	0.5565	40.8344
Sigmoid	0.0213	0.0010	0.1971	43.4654	2425.7390	0.5804	41.9831
Linear	-	-	0.3820	42.7106	2357.4203	0.5922	46.4137
Polynomial	0.0008	97.0000	0.0070	38.7946	2108.9788	0.6351	37.2362

For the RBF, the values for the gamma parameter were taken from the range of 0.001-0.1 and the values for the epsilon parameter were picked from the range of 0.001-1.0. The coef parameter was omitted since it does not have an impact on the effectiveness of the model. The model result evaluations based on the (MAE, MSE, R^2, and MDAE) displayed values that were substantially less accurate than the default values.

At first, increasing gamma and epsilon enhances the model's performance; but, after the model hits its optimal point, doing so results in a decline in performance. It was discovered that a gamma value of 0.01 and an epsilon value of 0.1095 produced the best results for the RBF kernel's set of variables. When these settings were used, the performance of the model achieved its maximum potential, displaying a value of 44.1958, 2564.1918, 0.5565, and 40.8344 for MAE, MSE, R^2, and MDAE, respectively. The performance of the model will suffer if gamma and epsilon are increased above their current values of 0.01 and 0.1095, respectively.

The sigmoid kernel's response to changes in gamma, coef, and epsilon is analogous to that of the RBF: the model's performance improves with increasing values of these parameters, but then reaches an optimal value beyond which further increases in these values actually degrades performance. For gamma, the best range of values was between 0.02-0.03, while for epsilon, it was between 0.200-0.300. Although finding the optimal value of gamma and epsilon for the model's performance might be challenging due to the limitless combinations of these two variables within a certain range, there are a number of methods that can help.

Because the epsilon is the only hyper-parameter that may affect the linear kernel, it's also known as the transparent box kernel occasionally. There are no complexities in the linear kernel, thus the name. Changing the epsilon parameter does not seem to have any effect on the model's efficiency. On average, the linear kernel's R^2 was 0.580, while the relative best performance was recorded at 42.710, 2357.420, and 0.592., and 46.413 for the MAE, MSE R^2, and MDAE, respectively.

The polynomial kernel is the most enhanced kernel since it is affected by all three hyper-parameters; nevertheless, the coef has the most significant impact on the polynomial kernel. Because of this, finding the value that is optimal for the model has been given the highest emphasis throughout the tuning process. With these parameters set to their ideal values of 0.001, 99, and 0.01, we get an R^2 of 0.6352, a (MAE) of 38.7946, a (MSE) of 2108.9788, a median absolute error of 37.2362, and an R^2 of 0.6352.

In order to acquire a full comparison of each kernel, Figure 4.2 depicts the finest coefficient of correlation achieved with each kernel vs the optimal epsilon value for each kernel. This was done so that the reader may receive this information in a clearer form.

The results of the R^2 performance metric test showed that the polynomial attained the greatest possible performance. In predicting inflow based on the monthly data series, thus, it is considered to be the most outstanding kernel, followed by linear, then sigmoid, and finally is the RBF.

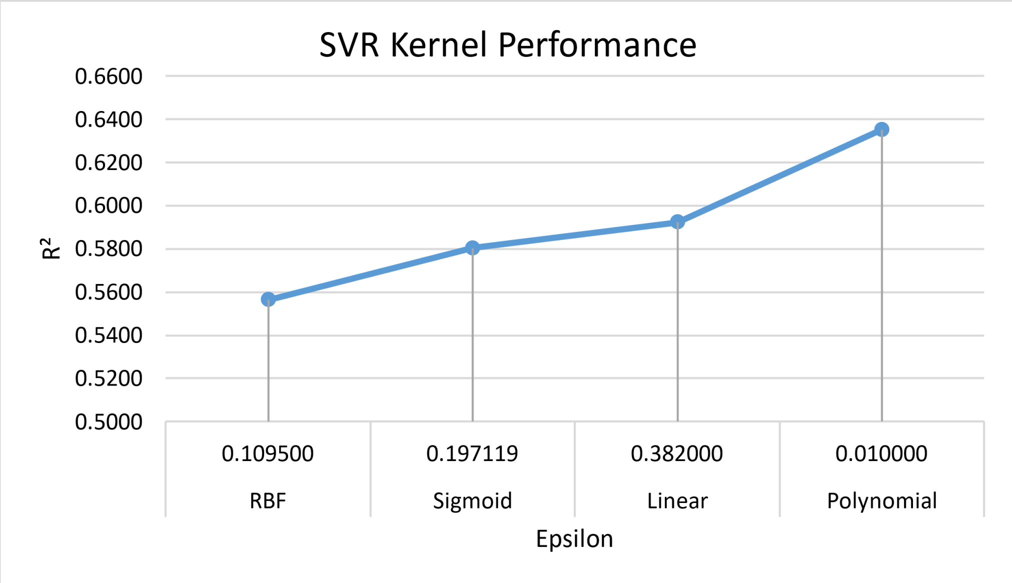


Figure 4.2: SVR Kernel Performance

4.2.3 Multilayer Perceptron Neural Network (MLPNN)

It is nearly impossible to find the ideal collection of parameters that will allow the MLPNN to operate at its highest level since it is a parametric machine learning model with several parameters that need to be adjusted. The logistic sigmoid function (log), identity function (Id), rectified linear unit activation function (relu), and hyperbolic function are the first four main nonlinear functions of the MLPNN (Tanh). Second, there are three fundamental solvers: the stochastic gradient descent (sgd), the (lbfgs) operator from the family of quasi-Newton methods, and (adam). To get the best results possible, each of these solutions is put to the test. The most important characteristics of the MLPNN are the hidden layer, the node size, the activation, and solver parts. Alpha and tol have been evaluated since, in addition to the previously listed factors, they may potentially have a little impact on the model. The optimal set of hyper-parameters is shown in Table 4.5, where tol values range from 0.0001 to 1.0000, and alpha values range from 1.0000 to 25.0000.

Table 4.5: Four Hidden Layer MLPNN Parameter Tuning

FXN	Solver	Tol	Alpha	MAE	MSE	R^2	MDAE
Id	lbfgs	0.1000	1.0000	40.5646	2236.2584	0.6132	39.2476
	sgd	0.0001	1.0000	40.0521	2242.2707	0.6121	39.7365
	adam	0.0001	1.0000	40.2090	2234.2954	0.6135	44.2946
Log	lbfgs	0.0001	2.3000	38.9555	2243.8250	0.6119	35.4898
	sgd	0.0001	1.1000	49.5325	3406.4963	0.4108	51.3071
	adam	0.0001	0.1000	45.2135	2729.8948	0.5278	42.4114
tanh	lbfgs	0.0001	3.9000	33.3901	1747.2303	0.6978	28.1564
	sgd	0.0001	5.0000	43.6977	2638.3413	0.5436	41.2286
	adam	0.0010	25.0000	43.8127	2963.4753	0.4874	40.6189
relu	lbfgs	0.0100	3.0000	37.0488	2145.2520	0.6289	27.1342
	sgd	0.0001	8.0000	39.0564	2140.1385	0.6298	41.7922
	adam	0.0001	10.0000	39.2656	2272.1355	0.6070	35.5422

With the focus to get the desired optimal outcomes, each of the four activation functions must be tuned and computed independently. Each activation function is based on a different underlying concept. In order for the solver to successfully generate the four convolution layers, it is necessary for it to also reposition the nonlinear and linear data into a feature space. Cross-validation is required for each and every one of the model's activations functions and solvers in order to get the best possible result from the model. In addition to this, the number of hidden layers and nodes has an effect on all five of the models since the three solvers and the four activation functions are all components of the same model.

The findings show inconsistent pattern across the trials and that each trial's greater performance is a result of its own unique set of factors. With a Tol of 0.10000, the "Identity" activation function with solver "lbfgs" obtained an R^2 of 0.6132; in contrast, the identical activation function with solver "adam" needed a Tol of 0.0001 to obtain an R^2 value that was close to that of "lbfgs" (0.6135).

The results demonstrate uneven patterns throughout the trials, indicating that each trial's superior performance is due to a different collection of specific circumstances. Using the identical activation function with the solver "adam" resulted in an R^2 value that was near to that of "lbfgs," but using the "Identity" activation function with the solver "lbfgs" resulted in an R^2 value of 0.6132 with a Tol of 0.1000 (0.6135).

The "logistic" activation-function has underperformed the other kernel function on average, with an R^2 of 0.5168. This is due to the fact that "sgd" and "adam" solvers both produced R^2 values lower than 0.6000. Although it was useful, the "tanh" activation function used with the "lbfgs" solver was superior. All of the "relu" activation function solvers have R^2 values over 0.6000, suggesting satisfactory performance. R^2 was 0.6978, tol was 0.0001, and alpha was very near to 4 with just 4 hidden layers and nodes. It's also the best and most precise option available. Figure 4.3 displays the visual displays to demonstrate how each activation function has affected the trend. The outcomes of each variable can be considered to be linearly related to one another.

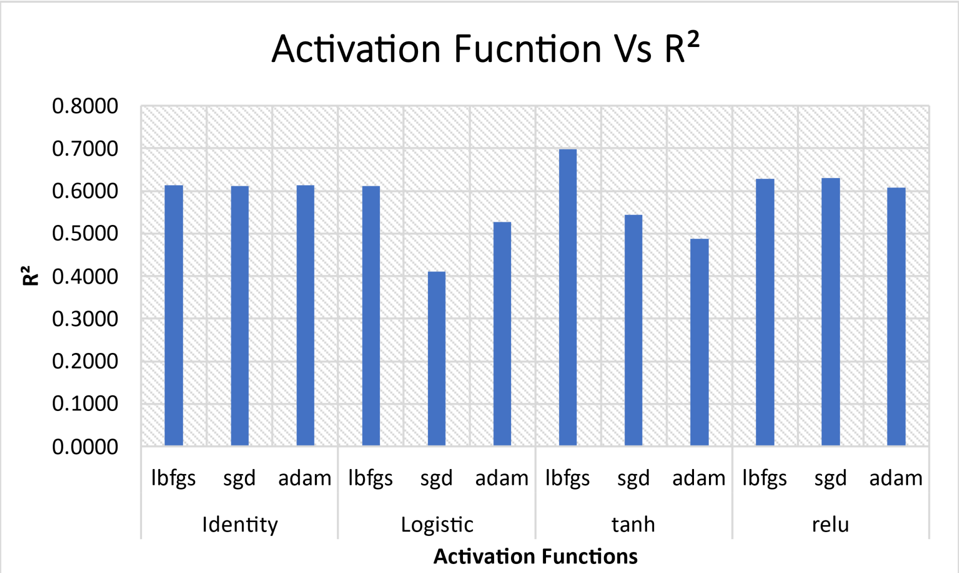


Figure 4.3: MLPNN Activation Functions vs R^2

4.2.4 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS distinguishes out from other techniques of forecasting because it is a hybrid model that incorporates aspects of fuzzy systems with neural networks. This gives it a distinct advantage over other approaches. This indicates that it has virtually precisely the same qualities as other frequently used neural network models in terms of setting the number of layers and nodes.

These characteristics include the fact that no one approach works for everyone and that the number of concealed layers is often established via a process of trial and error. This is analogous to the process of selecting the optimal neural network design before training.

There are eight distinct membership functions in ANFIS known as Gaussian Membership (GAUSSMF), Two-term Gaussian Membership (GAUSS2MF), Triangular Membership (TRIMF), Trapezoidal Membership (TRAPMF), Gaussian Bell Membership (GBELL), Sigmoidal Membership (DSIGMF), pi-shaped membership function (PIMF) and finally polynomial sigmoid membership function (PISGMF) each one was put through its paces throughout the various trials.

The outcomes of the several tests conducted with varying numbers of hidden layers produced a wide diversity of model architectures ranging from M222, M333, M444, M355, M455, M555, M655, M525, M522, M535, M545, M552, M553, M554, M556, M244, M344, M544, M666, M777, M622, M633, M644 and finally M523. Among all the 24 models M553 has shown the best results hence it has been selected for further testing as shown in Table 4.6.

The four statistical factors that were discussed before were used in making the decision to go with the improved membership function. As a result, the generalised bell function was discovered to be the most suitable membership function for the structure M553, as it had an R^2 value of 0.6244 and a mean average error (MAE) value that was quite low at 30.7858.

It was shown that as the number of hidden layers increases, the training process slows down and the results accuracy decline. This finding is in good agreement with the observation discovered by (Marjani et al., 2020; Allawi et al., 2021). Beyond M553, as the number of hidden layers increased, the overall training time increased by around two minutes for each additional hidden layer. Because of this, it is suggested that not more than 15 layers are employed to reduce the amount of time required for the calculation and get superior results.

Table 4.6: M553 Membership Tuning

Statistical Tests	Membership Functions							
	TRIMF	TRAPMF	GBELLMF	GAUSSMF	GAUSS2MF	PIMF	DSIGMF	PSIGMF
MAE	38.2310	41.4323	30.7858	42.8744	35.7744	36.9835	36.3460	36.0646
MSE	3412.1488	4364.2261	1780.3422	5494.4008	1294.6172	2056.8016	2385.1147	3552.7232
R^2	0.4177	0.3926	0.6244	0.2887	0.5110	0.4403	0.4398	0.4010
MDAE	1740.6887	1800.2192	1659.6838	1871.8942	1671.5526	1674.7755	1716.9322	1742.8262

4.2.5 Extreme Gradient Boosting (XG-BOOST)

XG-Boost is a contemporary model that requires little fine-tuning to provide significant performance gains. The XGB's versatility in hyper-parameter support makes it a powerful query language for almost any database. On the other hand, in this study, only the Gamma, learning rate, Max depth, and Reg lambda hyper-parameters had any noticeable impact on the model's performance. There is a total of 24 hyper-parameters. The default values for each parameter were used when the model was initially run. Unexpectedly, the model's efficiency under the default conditions was somewhat better than that of the other three forecasting models.

As discussed Earlier the K-Fold cross-validation technique was examined and evaluated in this research project. K-fold had a detrimental effect on the performance of the three forecasting models, rather than improving or maximising those models' capabilities. In contrast, when paired with the XG-Boost, it yielded results that were very near too perfect. Performance measures were found to have an MAE of 37.9172, MSE of 2194.5987, R^2 of 0.6204, and MDAE of 30.9001 when the model's parameters were first set to their default values and statistical analysis was performed. These values are very close to those obtained by tuning the SVR and ANFIS models, despite only using the models' default settings.

The XGB undergoes further fine-tuning via a process of trial and error. The values gamma (γ), (learning Rate), and lambda λ were all found to be within the range of 0 to 1. The XGB has one noteworthy parameter, which is the (Max depth). The (Max depth) specifies the deepest you may let your tree roots grow because XGBoost is modelled like a tree and its growth process. This is a very relevant option. Therefore, when the value of (Max depth) rises, the model will grow more complicated as a result of the addition of new layers. Because we wanted to conduct the training and testing of the model using the fewest possible values in this investigation, we specified that the maximum depth should be between 1 and 20.

The model was first updated without first transforming the input into folds. After experimenting with different values for Lambda and Max depth, it was found that 0.60 for lambda and 6.00 for depth produced the best results for the model. If these factors are manipulated more, the model's efficiency will decline. We determined that gamma= 0 is the best value, and that increasing it will decrease the performance of the model. According to the results shown in Table 4.7, the parameters gamma (γ) = 0.00, Learning rate (L_R) =0.10, Max depth =12.00, Reg lambda (λ) = 0.80, and the number of K-folds (k-25) =25.00 resulted in the best performance.

Table 4.7: XG-Boost Hyper-Parameters Tuning

γ	L_R	Max depth	λ	MAE	MSE	R^2	MDAE	Random State	K-Fold (splits)
0.25	1.00	1.00	0.0001	37.92	2194.60	0.62	30.90	1.00	-
0.00	1.00	6.00	0.60	34.82	1763.30	0.69	29.20	1.00	-
0.00	0.10	12.00	0.80	12.05	213.44	0.88	9.80	10.00	25.00

In an effort to wrap up this section, four separate reservoir inflow forecasting models were built using the monthly period scenarios. Four independent reservoir inflow forecasts models were created based on the monthly period scenarios in an effort to wrap up this section of the debate. A total of 240 samples for four different parameters—evaporation, precipitation, water level, and historical inflow—were incorporated in these models. Researchers used not one but two distinct methods, namely the Grid-Search methodology and the trial-and-error method, in order to discover the optimal parameter configurations to alter in each model in order to get more accurate predictions.

The terms "search strategy" and "search method" allude to these two different approaches. The four different statistical scores used to rate each predictive model were distinct. Last but not least, Figure 4.4 provides a fair comparison of the four models, by comparing the predicted inflows from each model to the actual predictions.

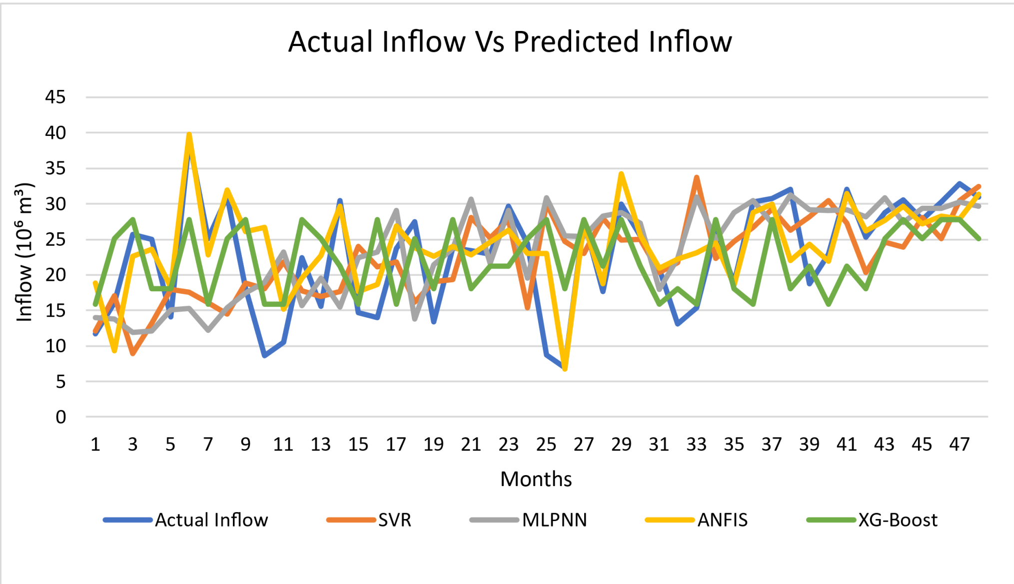


Figure 4.4: Actual Inflow vs Predicted Inflow from the Models

The maximum reservoir inflow measured in Figure 4.4 was 40xm3, whereas the lowest was 8xm3. A total of 48 months over 4 years were used to evaluate the model. Southwest of Peninsular Malaysia is where the reservoir station is located and where the data were collected. The monsoon season (SW monsoon) for this area begins in the month of MAY and lasts until around SEPTEMBER (Pour et al., 2020). In light of this, the models suggest that the maximum inflow pattern shown in Figure 4.4 will match with the pattern of rainfall, which starts to increase by the end of the fifth month and decreases by the end of the ninth month.

4.3 Forecasting-Based-Model (Daily Inflow)

In this part, the same four artificial intelligence models that were used earlier to estimate the monthly inflows are evaluated to see how well they can predict the flow into the Klang Gate reservoir based on the daily circumstances. In addition, to determine which Input variables were the most sensitive, a sensitivity analysis was run using autocorrelation. The term "autocorrelation" refers to a measurement that determines the degree of the link that exists between two successive time periods. In a time, series, there are two versions of datasets which are referred to as the time-lagged data and the original time series.

Since the gaps in time between observations are substantial and the inflow to the reservoir follows an indeterminate non-uniform distribution, it is imperative to create a time lag model. These varying time delays serve as a helpful reminder that the significance of previous observations ought should lessen with the passage of time (Luo et al., 2018), in the event that a variable is missing for a long period of time. It is possible to improve streamflow predictions by delaying data collection for a longer period of time (Yang et al., 2017).

This research agree with (Kim et al., 2019; Khazaee Poul et al., 2019) in their assertion that considerably more accurate predictions may be generated by including flow, temperature, and precipitation lag periods among the inputs. In light of this decision-making process, it is thus strongly suggested to employ lagged inflow since it is helpful in forming judgements about the functioning of the reservoir.

The correlogram that was plotted for the 60-day lag is shown here by Figure 4.5. Rainfall was chosen as the independent variable in this study since it is the mainstream of water supply to the reservoir. In addition, the elements that were being examined, however, were the water levels, evaporation, and historical inflows. In accordance with Figure 4.5, the water level was the most impactful variable, followed by the inflow, while evaporation was the least significant contributor.

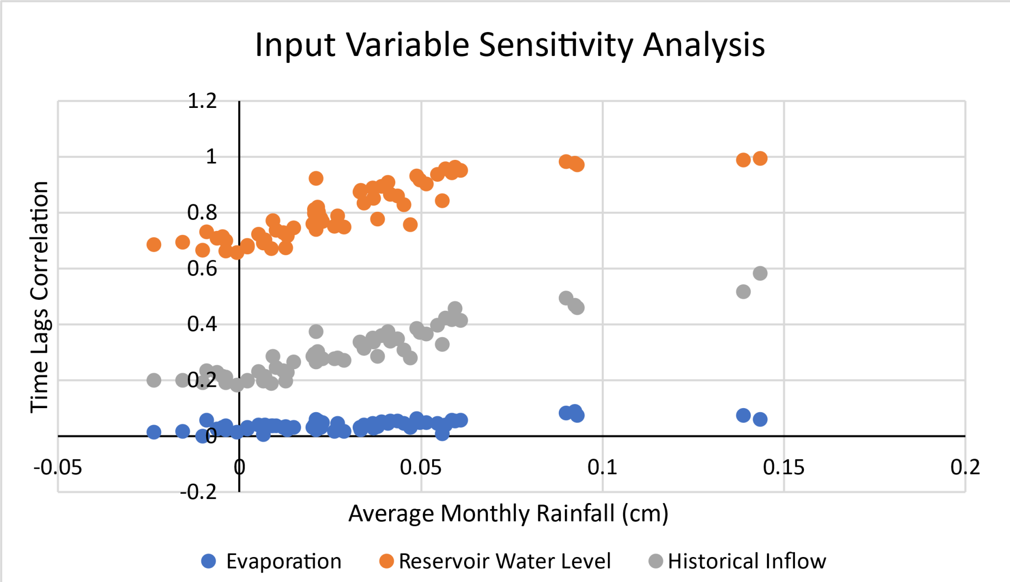


Figure 4.5: Input Variable Sensitivity Analysis

Before moving on to more complex situations, each of the four previously used models was put through its paces by being tested on the first four scenarios. The top two models were then used to estimate reservoir inflow for Scenarios 5 through 7, which used the historical inflow as an input, based on the first four Scenarios. The outcomes of each of the possible situations are shown below in Table 4.8

Table 4.8: Statistical Evaluation of All Models

Scenarios	Model	Kernel	Hidden Layers	Statistical Evaluation			
				MAE	MAE	R^2	MSE
1	SVR	Poly	-	59.9802	8919.3302	0.4240	40.1389
	MLPNN	relu (lbfgs)	(14-9)	52.5941	7130.6476	0.5395	33.9323
	ANFIS	Pi	(4-4-2)	59.6012	7985.8597	0.4247	39.2593
	XG-Boost	-	5	30.2360	5097.1204	0.7220	18.4754
2	SVR	(RBF)	-	51.1157	7241.3984	0.4814	28.3359
	MLPNN	logistic (lbfgs)	(10-5)	51.0265	6836.7206	0.5682	32.1400
	ANFIS	TRAPMF	(3-3-5)	612.1626	10125.1133	0.1681	170.943
	XG-Boost	-	6	40.6211	5486.8070	0.6457	21.1554
3	SVR	(RBF)	-	51.9611	7368.4589	0.4723	29.7794
	MLPNN	logistic (lbfgs)	(10-5)	51.0926	7586.3967	0.5208	32.0243
	ANFIS	Gaussian	(2-5-5)	530.6267	10304.8390	0.1624	172.844
	XG-Boost	-	5	41.3030	5449.3034	0.6481	23.1501

Table 4.8 (continued): Statistical Evaluation of All Models

Scenarios	Model	Kernel	Hidden Layers	Statistical Evaluation			
				MAE	MAE	R^2	MSE
4	SVR	(RBF)	-	52.7055	7424.5836	0.4682	31.0070
	MLPNN	logistic (lbfgs)	(10-5)	50.6292	7597.9790	0.5201	30.7787
	ANFIS	Gaussian	(2-2-4)	621.8024	10435.1987	0.1615	170.5209
	XG-Boost	-	5	43.1057	5892.9277	0.6194	24.0792
5	MLPNN	relu(adam)	2	3.5969	4.4547	0.9964	1.8845
	XG-Boost	-	5	0.4493	1.3707	0.9999	0.2209
6	MLPNN	relu(adam)	2	12.7200	6.1468	0.9641	9.2036
	XG-Boost	-	5	0.3863	1.7039	0.9999	0.2055
7	MLPNN	relu(adam)	2	19.8485	9.8198	0.9459	14.1924
	XG-Boost	-	5	0.3854	1.6764	0.9999	0.2020

When used on the first four scenarios, it has been demonstrated that perhaps the MLPNN and the XG-Boost are the two models that perform best in terms of the model performance indices. An R^2 of 0.53949 was obtained in scenario 1 for the MLPNN model with two hidden layers, 14 nodes in the first layer and 9 nodes in the second layer. This was a better outcome than the SVR model's R^2 of 0.42397 and the ANFIS model's R^2 of 0.42469. However, it fell short of the XG-Boost, which outperformed the other two models and achieved an R^2 score of 0.7222.

While the ANFIS performed poorly in scenarios 2 through 4, earning an R^2 of 0.16808 in scenarios 2 and 3, performance continued to deteriorate, earning an R^2 of 0.16239. The ANFIS obtained an R^2 of 0.16808 in scenario-2 for scenario-4. In scenario-4, the ANFIS got an R^2 of 0.16808 in scenario-2. In addition to this, it was still classified as the worst model for the fourth scenario since its performance is becoming worse and worse, and it had the worst R^2 score of all the different situations, which was 0.16154. Because of this, the model was given the lowest possible score.

The MLPNN and XG-Boost both performed similarly and were rated as the top two models in scenario two. However the XG-performance Boost's was slightly down from scenario one, with an R^2 of 0.6456. The MLPNN and SVR performances, on the contrary side improved by 5.3% and 13.536%, respectively.

Model performance continues to drop until scenario 4, when the MLPNN achieves its lowest score R^2 of 0.5200 and the XG-Boost reaches an R^2 of 0.6194, a drop of 14% from its best-ever R^2 value in scenario 1. SVR's performance was also going downhill, but to a lesser extent; the metric's R^2 value was 0.46. In Scenario 5, the XG-Boost achieved an R^2 of 0.9999, which was almost perfect and higher than the MLPNN's R^2 of 0.9964. For all situations 6 and 7, the XG-Boost kept up its remarkable performance, achieving an R^2 of 0.9999 while the MLPNN performance dipped to 0.9459.

After testing all the models across the seven proposed scenarios ultimately, it can be finalised that MLPNN and XG-Boost are the best two forecasting models. By using estimates for the previous day's influx, the models' predicting accuracy is greatly improved. Figure 4.6 shows a 3-D graphic depiction of each model's performances that have been described earlier under.

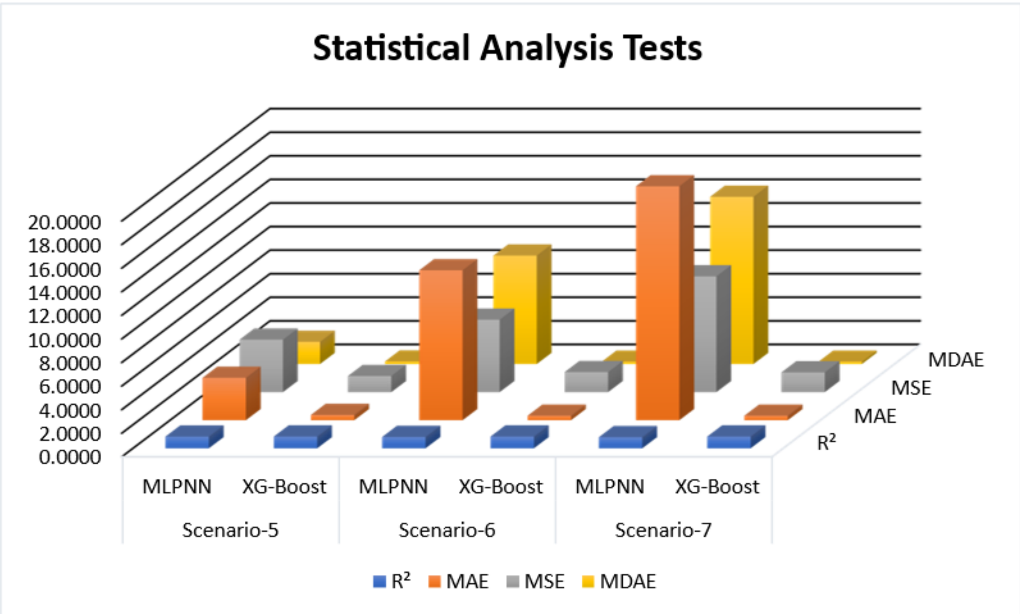


Figure 4.6: Statistical Analysis Tests for Scenarios 5, 6 and 7

To provide a brief summary of this section When dividing the data for the daily time series into their respective 7 scenarios, a separate procedure was used. Scenario 1 was the control, and there was no time lag in this scenario. Scenarios 2, 3, and 4 each had a time lag of one, three and five, respectively, where (t) denotes the total number of days in the scenario. Scenario 1 was the control, and there was no time lag in this scenario.

Following the completion of an autocorrelation function computation for each of the parameters, the water level and historical input of the time lag series were selected. Therefore, it turned into a historical inflow in situations 5 through 7, which were equivalent to scenarios 2 through 4, yet the historical input is delayed in time. In other words, the predicted influx by the model served as the basis for training the models to anticipate reservoir inflow. The XG-Boost continues to come out on top when compared to the other three models during the first four scenarios. The ANFIS received the lowest possible score, falling below both the MLPNN and the SVR, which tied for second place. In contrast, the ANFIS generated results that were somewhat better than those of the SVR in scenario 1, which placed it as the third-best model overall.

As the XG-Boost and the MLPNN were the top performers in the first four conditions, these two models were further investigated for the remaining three cases. Results were consistent with those from earlier iterations, with XG-Boost again emerging as the top contender ($R^2 = 0.9999$) and proving to be the best model overall. Based on the results of the study, XG-Boost is the best performing model, followed by MLPNN, SVR, and finally ANFIS.

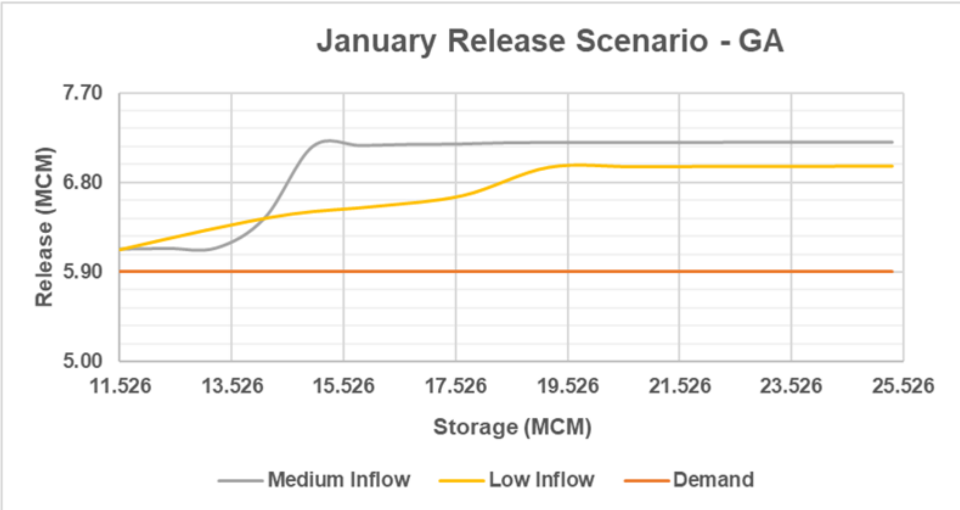
4.4 Conventional Operation System

4.4.1 Real Coded Genetic Algorithm (RCGA)

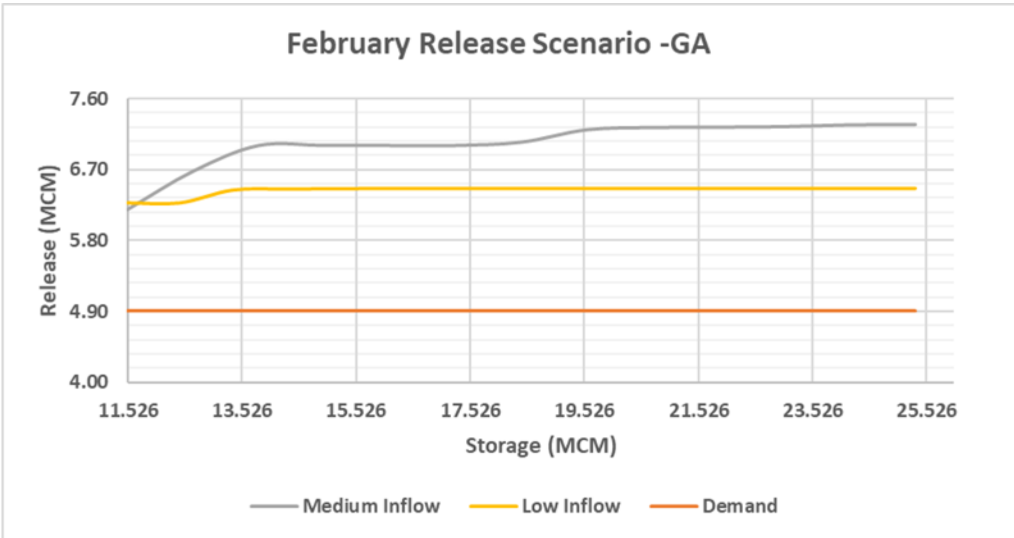
The first remark that can be derived using the 12 release curves (Figure 4.7) is that the three types of input do not always occur in the same month. This can be seen from the fact that there is a difference in the slopes of these curves. The reservoir only had flows of medium and low levels throughout the months of January, February, and July. Only high and medium levels of inflow were received by the reservoir during the months of May, October, and November.

This is consistent with the pattern of rainfall that happens in the state of Selangor, where the largest rainfall totals occur between the months of October and November, followed by the second highest rainfall totals which occur between the months of April and May (Bahar et al., 2021). During the course of the year, the genetic algorithm was successful in meeting the demand on one occasion, with the exception of the first three months (January, February, and March). For the whole month of January, the downstream demand was 5.90 MCM, whereas for both February and March it was 4.90 MCM.

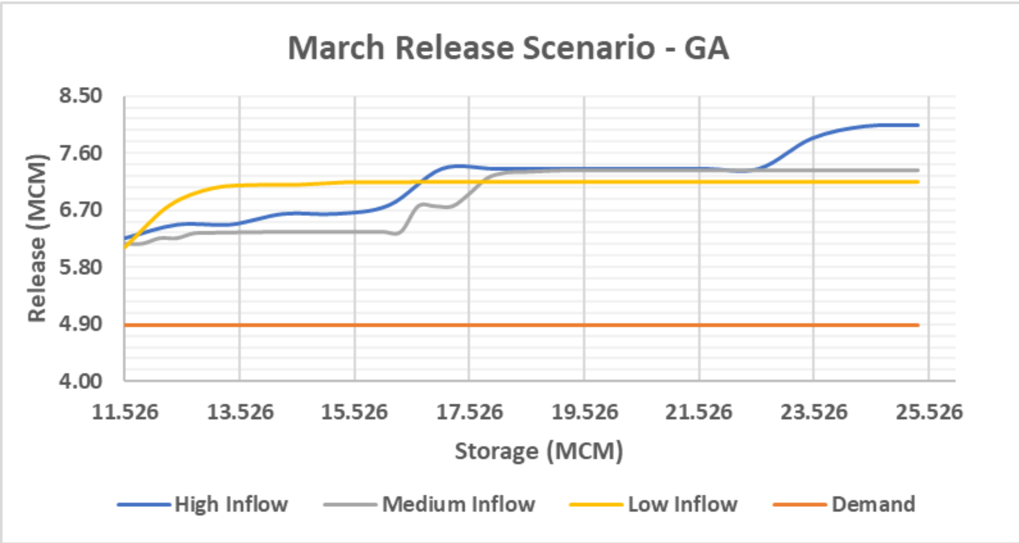
(a)


January Release Scenario - GA
Release (MCM)
7.70
6.80
5.90
5.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

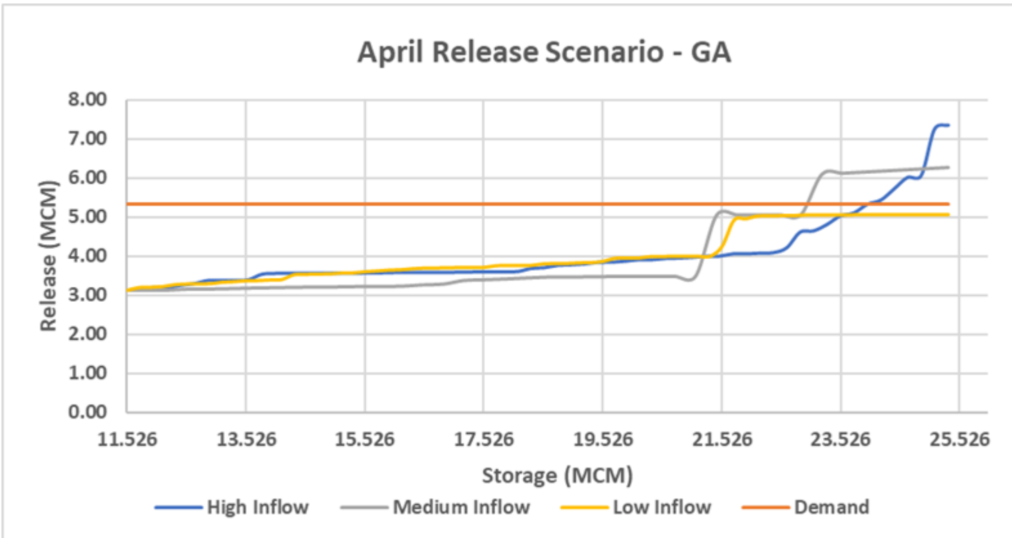
(b)


February Release Scenario -GA
Release (MCM)
7.60
6.70
5.80
4.90
4.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

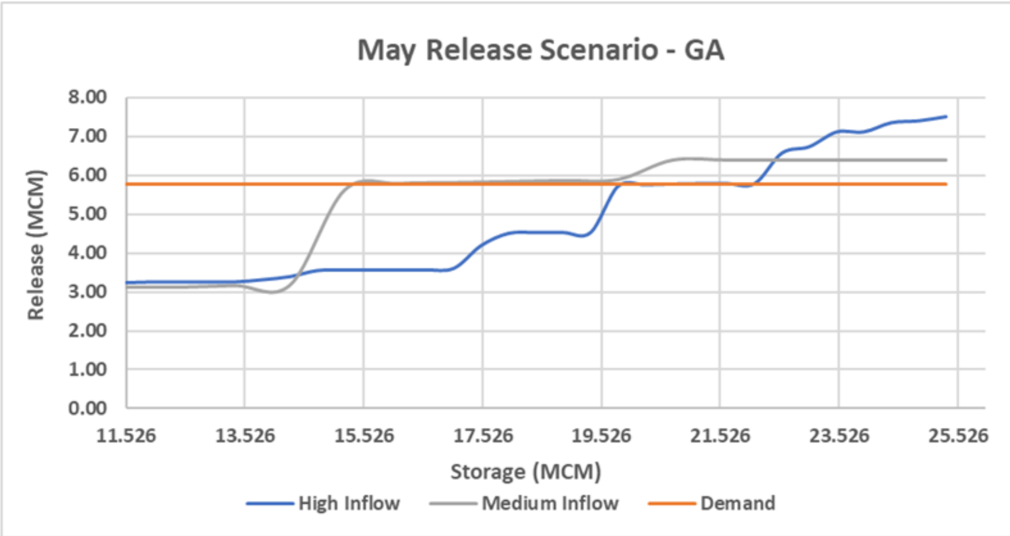
(c)


March Release Scenario - GA
8.50
7.60
6.70
5.80
4.90
4.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

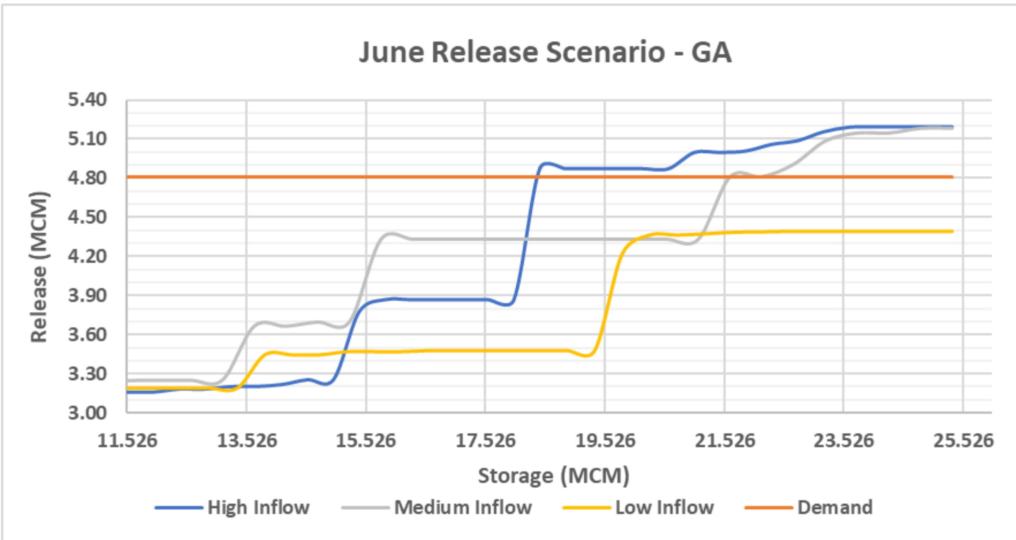
(d)


April Release Scenario - GA
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

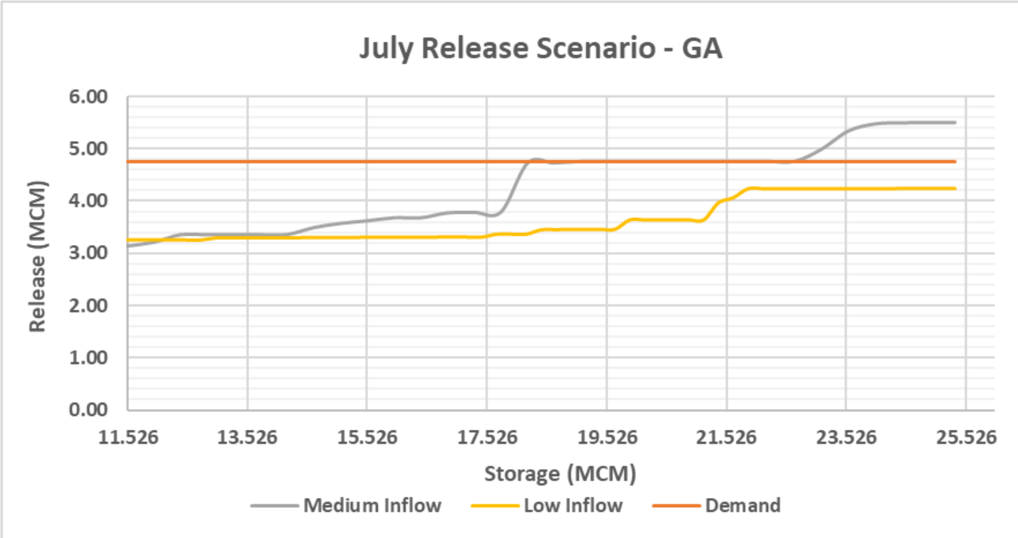
(e)


May Release Scenario - GA
Release (MCM)
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

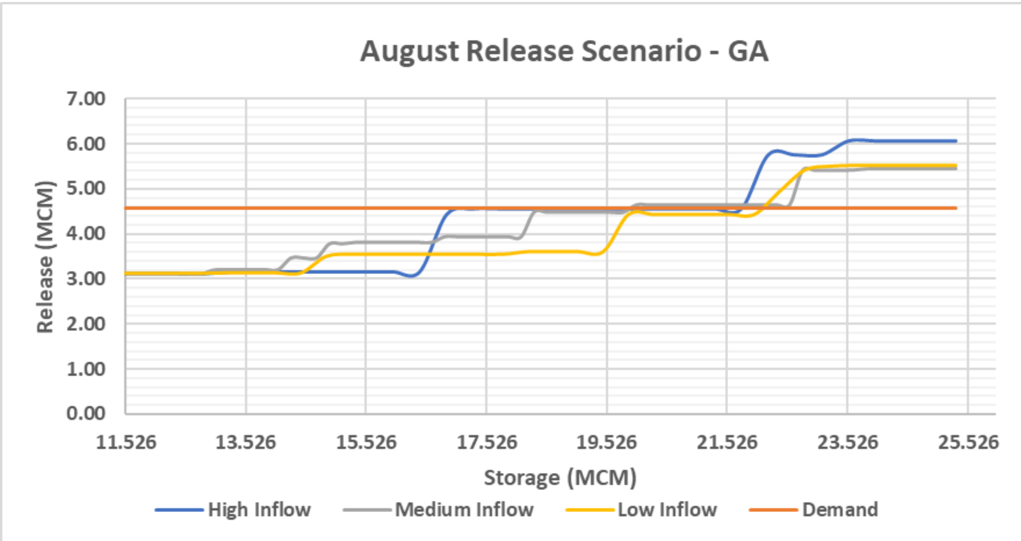
(f)


June Release Scenario - GA
Release (MCM)
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

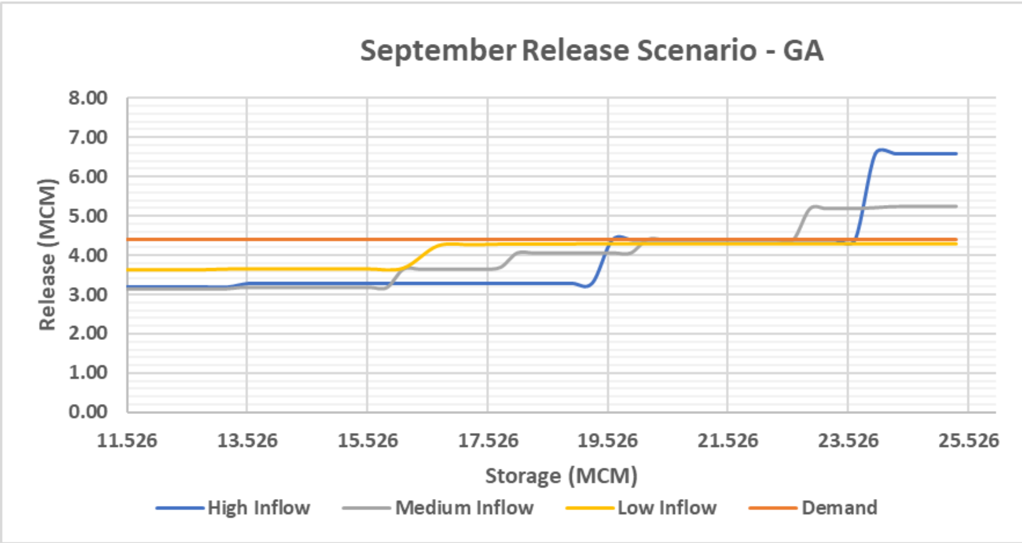
(g)


July Release Scenario - GA
Release (MCM)
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

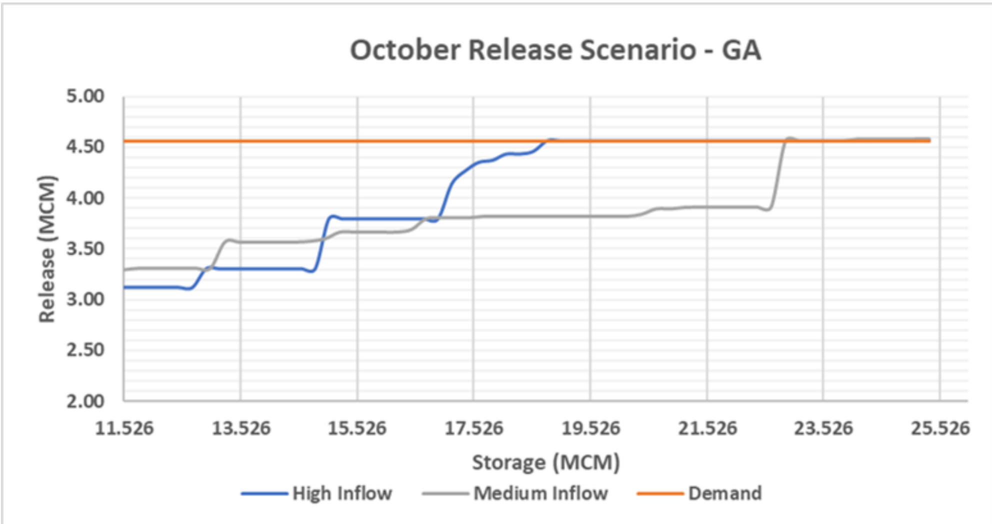
(h)


August Release Scenario - GA
Release (MCM)
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

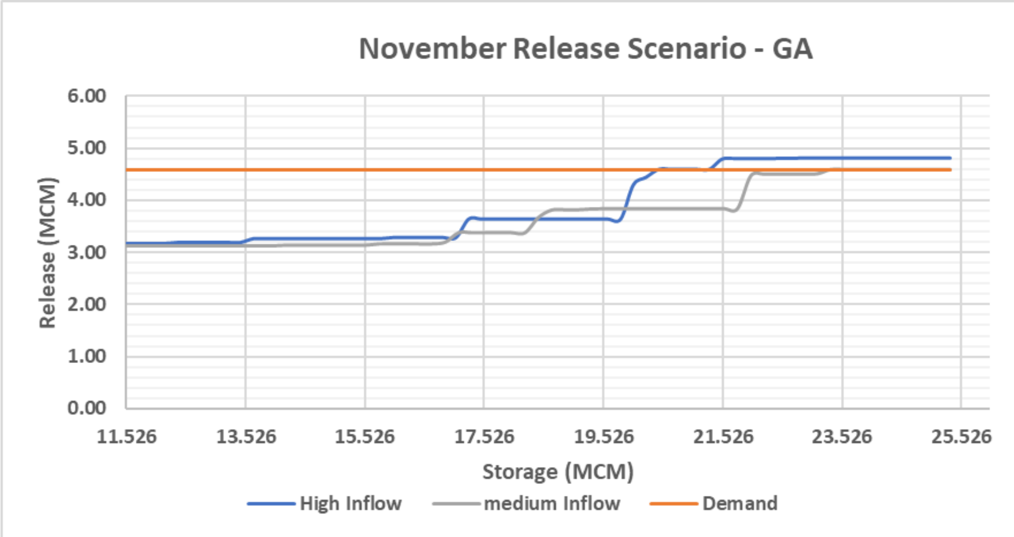
(i)


September Release Scenario - GA
Release (MCM)
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

(j)


October Release Scenario - GA
Release (MCM)
5.00
4.50
4.00
3.50
3.00
2.50
2.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

(k)



(l)

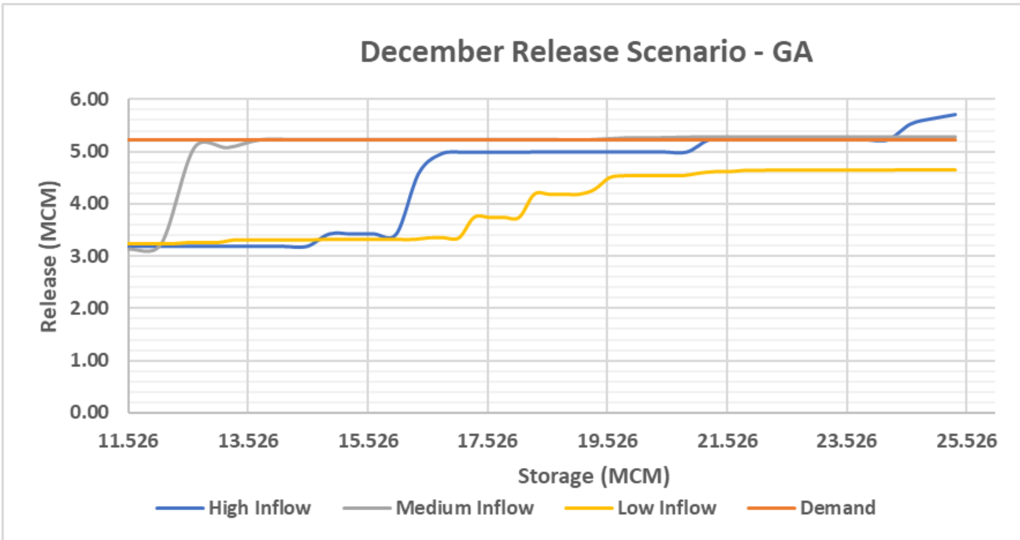


Figure 4.7: GA Release Curves for Conventional Operation System (CS) for (a) Jan - (l) Dec

As can be seen from Figure 4.7, the lowest volume of water that was released during the months of January and February was 6.13 MCM, which is about one MCM more than the demands of the downstream communities. Although the inflow was more than the required amount of water farther downstream in March, it was not enough to satisfy the demand for water. When the storage level is between 19 MCM and 23 MCM, the release will be the same regardless of whether the reservoir gets a large or medium amount of input.

This is a further point that stands out when looking at the release curve. Since the purpose of using the optimisation algorithm is to fulfil the downstream demand with no deficit, the GA has not been able to reach its optimisation goal during the first two months of its use. This is because the goal requires there to be no deficit. Moving on to the months of April and May, downstream demand was virtually exactly the same for both months; it was 5.33 MCM in April and 5.78 MCM in May, and these demands were the second and third greatest downstream requests after January.

The release curve for April demonstrates that the reservoir release was proceeding upward to satisfy the demand. When the reservoir storage level reached 24 MCM, the release that was created due to the high inflow was finally able to satisfy the demand farther downstream. They were just about able to satisfy the demand with the medium and low levels of input, but only just. However, the reservoir operation in May can be considered relatively the best operation among the twelve months because both the high and medium inflows could meet the demand consistently for a longer duration. This was the case because May had the most inflows and extended duration of any of the months. In terms of the medium inflow, it was sufficient to meet the demand when the storage level was 14.5 MCM until it reached 19.526 MCM.

On the other hand, the high inflow was sufficient to meet the demand at the storage level of 19.526 MCM and maintained its level until the storage level reached 23 MCM. After successfully completing the requirements, both inflow patterns have shown an increasing tendency.

Because it displays continuous release at certain places, generating what is known as a step function, the release curves that were created by the GA beginning in June and continuing all the way through December have a graphical shape that is comparable to the pattern of a staircase. A step function will always have the same value at the same intervals, but the value of the constant will change between intervals. The concept that the constant varies at each interval causes the jumps in the series of horizontal lines, which is formed by the constant value at each interval.

The constant value at each interval creates the sequence of horizontal line segments. This is why the graph of a step function looks like a flight of stairs. In June, the high and medium inflows could only satisfy the demand of 4.81 MCM at two or three places alone. This was the case for both categories.

Addressing the month of July, the medium inflow fulfilled the requirement when the reservoir storage was 19.526 MCM, and it maintained doing so until the storage levels were increased to 23.526 MCM. This allowed the inflow to continue serving the demands of the downstream areas (which forced the reservoir operator to release more water downstream to avoid entering the critical stage). In the month of August, the downstream demand was satisfied by all three inflows while the storage capacity of the reservoir was 15, 19, and 23 MCM, respectively, for high, medium, and low inflows.

The trend made sense when considering the fact that the high inflow was able to reduce the water deficit more quickly than both the medium and the low inflow. In September, the high and medium inflows were both able to decrease the water deficit almost exactly at the same reservoir storage level, which was 20 MCM; however, the low input could not fulfil the downstream demand.

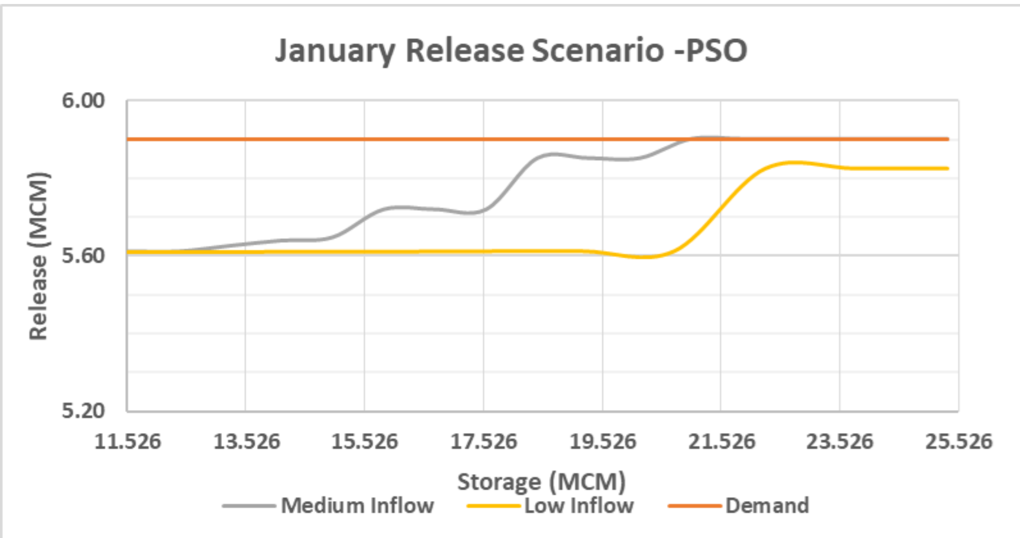
In October, the GA was able to show good performance as the high inflow reached its optimal performance by meeting the demand of releasing 4.5 MCM of water when the storage level was 18 MCM and had continued fulfilling the demand. This was made possible because the high inflow could reach its optimum performance.

When the storage level was increased to 23 MCM, the medium inflow could only match the demand by a hair's breadth. One of these constant values, more specifically the one before the last step, met the demand exactly and generated a water deficit equal to zero during the month of November when the high inflow had a perfect staircase graph with a constant release at each step size. This was the case because the high inflow occurred during November. With regard to the medium inflow, the demand was satisfied by the most recent step size, which demonstrates that the medium inflow is capable of satisfying the demand even when the reservoir levels are getting close to their maximum storage levels. In December, both the high and the medium had rather successfully satisfied the expectations.

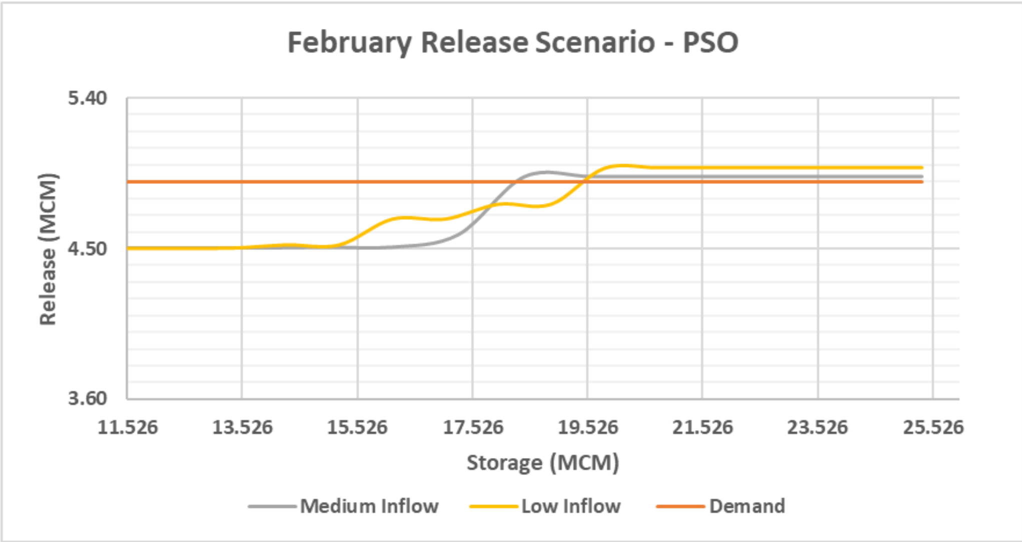
4.4.2 Particle Swarm Optimisation (PSO)

In contrast to GA, a linear pattern can be seen in the Particle swarm release curves (Figure 4.8). PSO has historically been known to demonstrate better performances when compared to GA. This is due to the fact that the computing effort required by PSO to identify high-quality solutions is less than that required by the GA (Hassan et al., 2005; Zhang et al., 2014). As a result, the curves illustrate a higher degree of linearity by showing straight lines. Particle swarm has always been known to show better performances. PSO was able to supply the demands of the downstream in approximately nine months regardless of the inflow condition; nevertheless, it was unable to satisfy the needs of the downstream in only three months (January, March, and May when the inflow was medium/low).

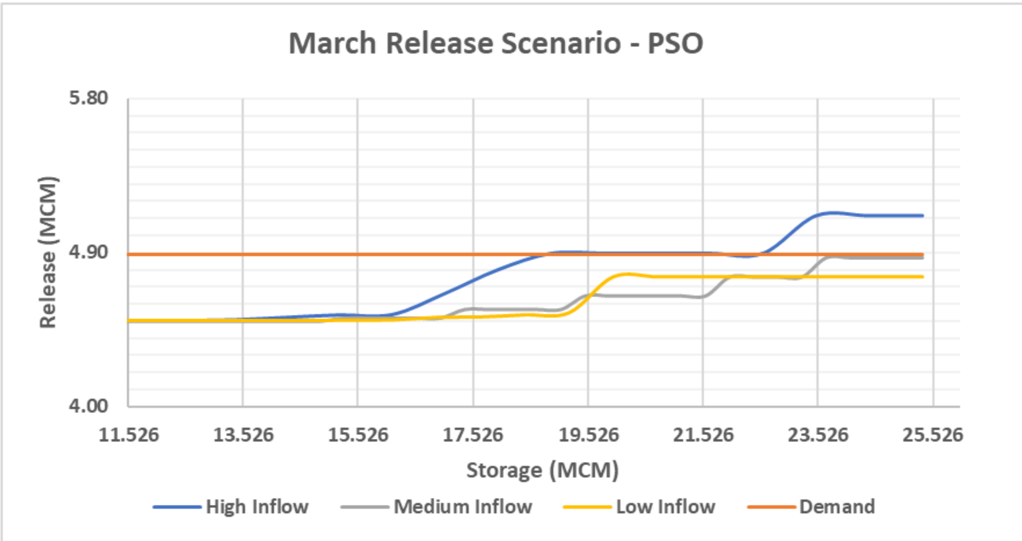
(a)



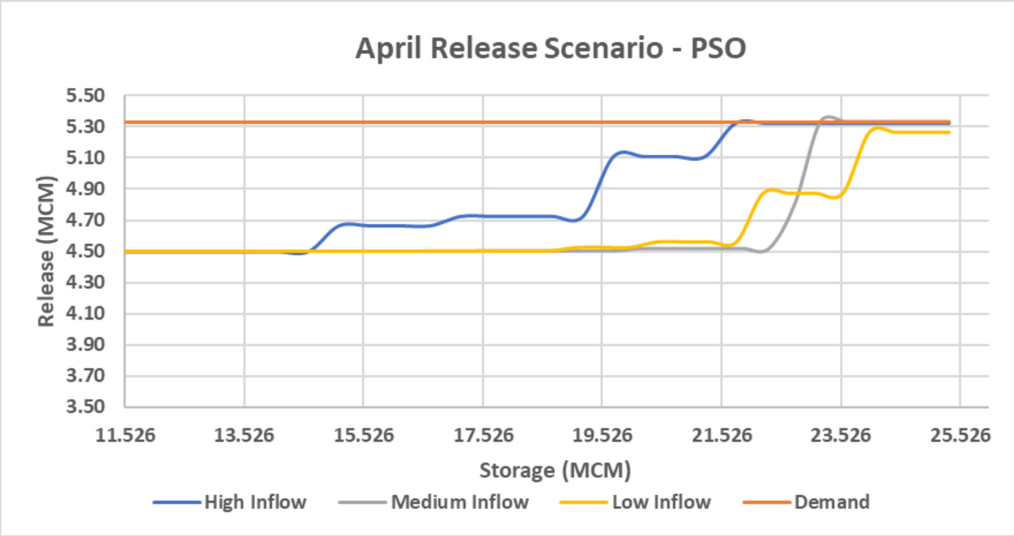
(b)


February Release Scenario - PSO
5.40
4.50
3.60
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

(c)

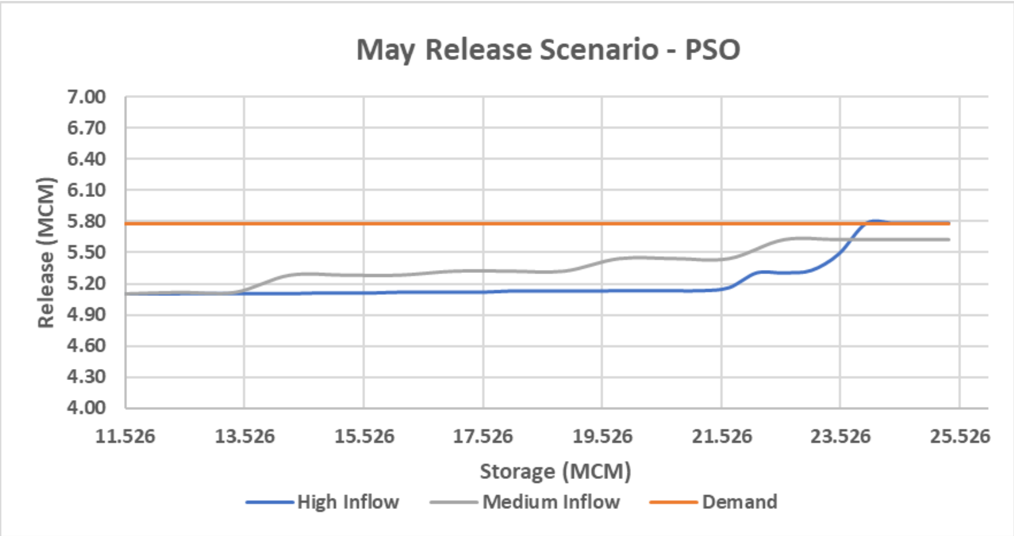

March Release Scenario - PSO
5.80
4.90
4.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

(d)



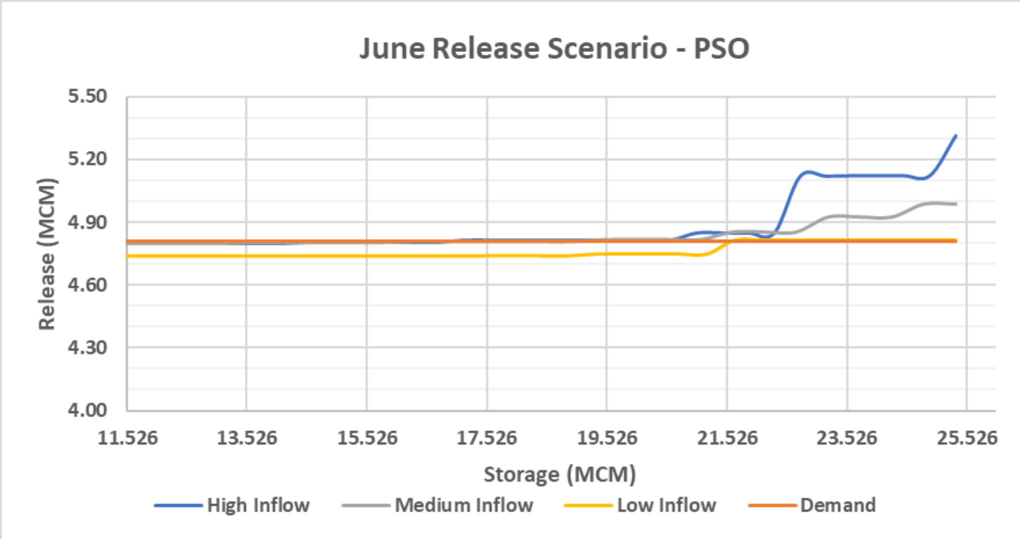
April Release Scenario - PSO
Release (MCM)
5.50
5.30
5.10
4.90
4.70
4.50
4.30
4.10
3.90
3.70
3.50
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

(e)

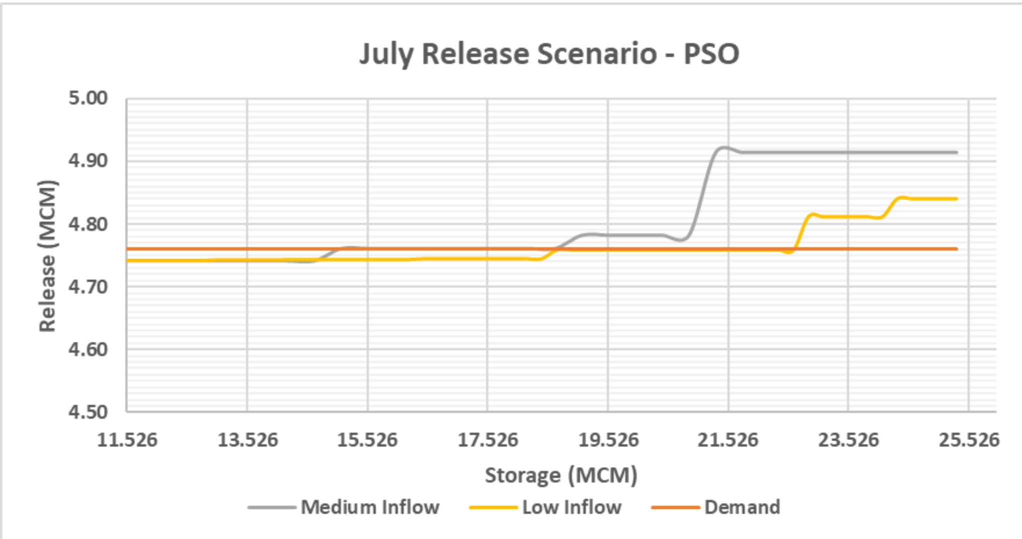


May Release Scenario - PSO
Release (MCM)
7.00
6.70
6.40
6.10
5.80
5.50
5.20
4.90
4.60
4.30
4.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

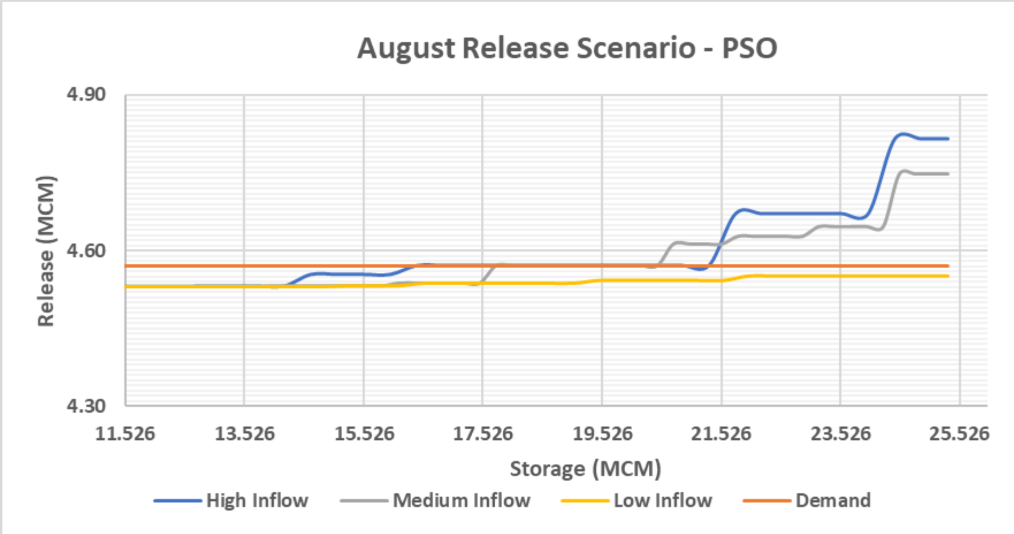
(f)


June Release Scenario - PSO
5.50
5.20
4.90
4.60
4.30
4.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

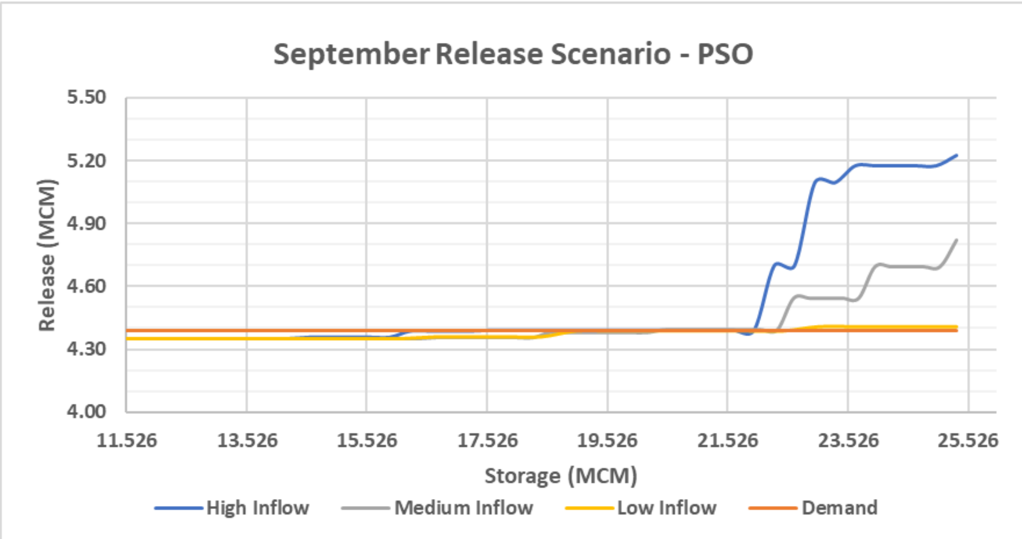
(g)


July Release Scenario - PSO
5.00
4.90
4.80
4.70
4.60
4.50
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

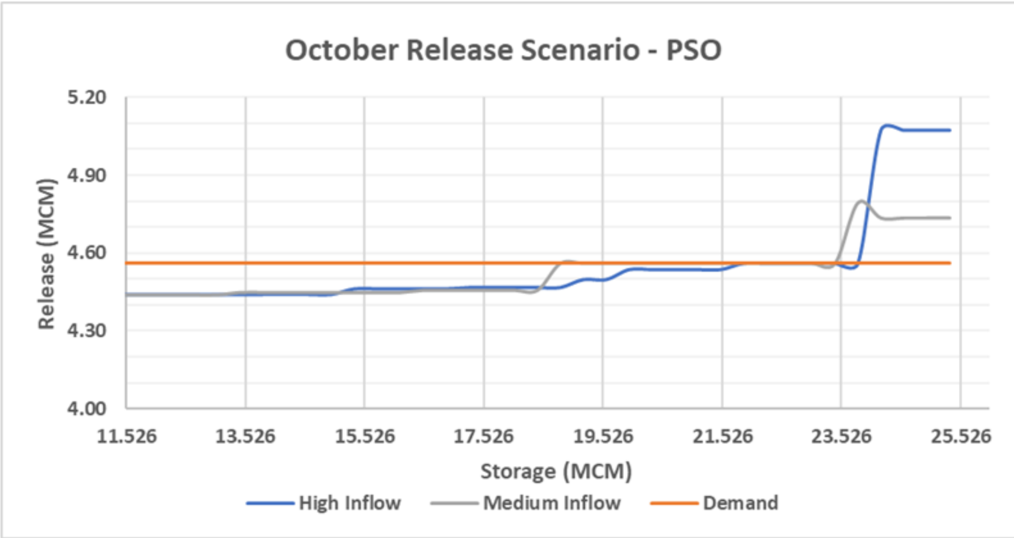
(h)


August Release Scenario - PSO
4.90
4.60
4.30
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

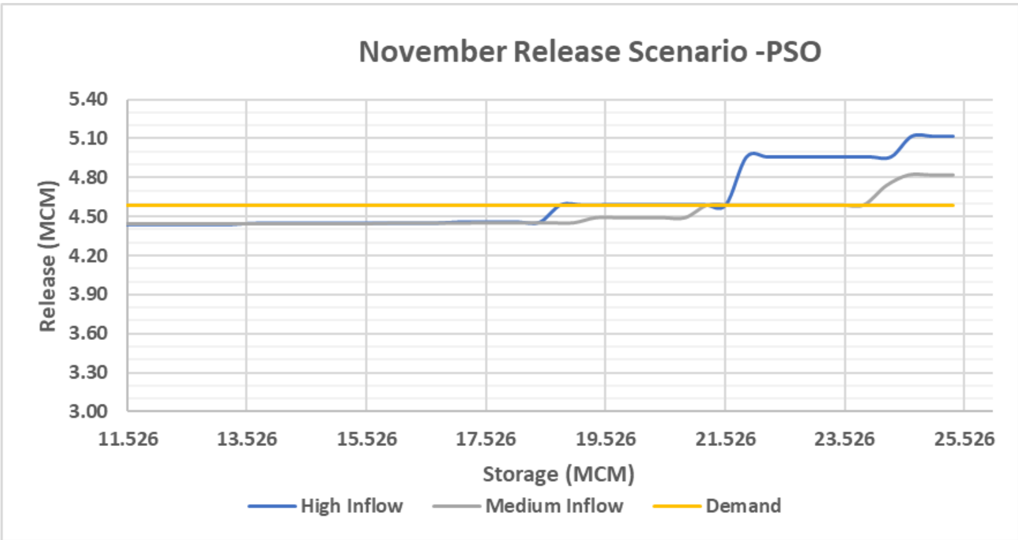
(i)


September Release Scenario - PSO
5.50
5.20
4.90
4.60
4.30
4.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

(j)


October Release Scenario - PSO
5.20
4.90
4.60
4.30
4.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

(k)


November Release Scenario -PSO
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

(l)

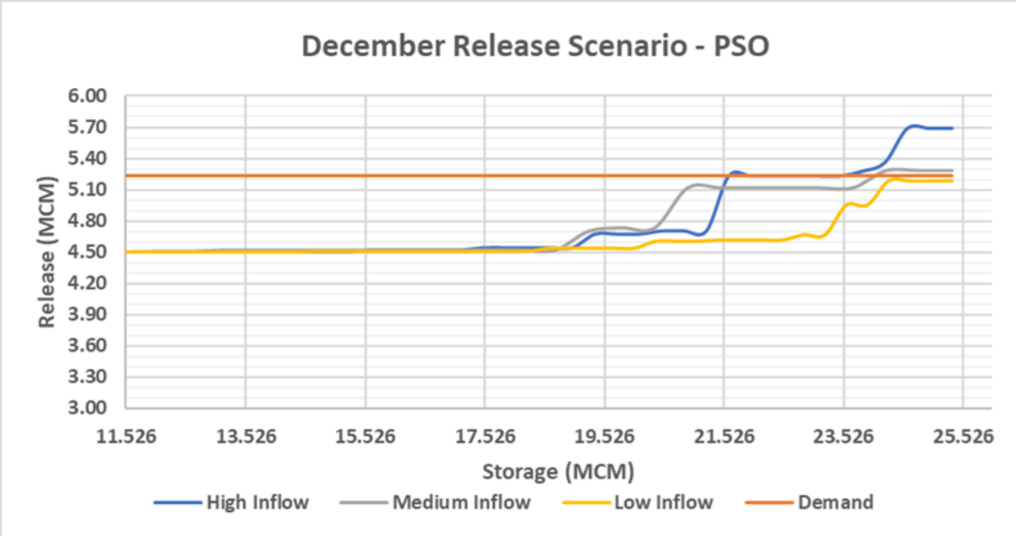


Figure 4.8: PSO Release Curves for Conventional Operation System (CS) for (a) Jan - (l) Dec

When there was a medium inflow of water into the reservoir catchment in January, PSO was able to reduce the water deficit to zero. However, when there was a low inflow of water, the reservoir was unable to satisfy demand and was only just able to increase the water level to prevent the storage level from reaching a critical level. The reservoir experienced all three types of inflow patterns throughout the month of March; thus, it was able to fulfil the requirements of the downstream community when the storage capacity of the reservoir reached 18 MCM as a result of the large amount of water that was being added to the reservoir (high inflow).

In a similar way, the PSO was able to optimise the operation of the reservoir and reduce the gap between the water that was released and the downstream from medium inflow because it released 4.90 MCM of water and continued to supply a constant release, which brought the gap down to zero. This enabled the PSO to reduce the gap between the water that was released and the downstream from medium inflow.

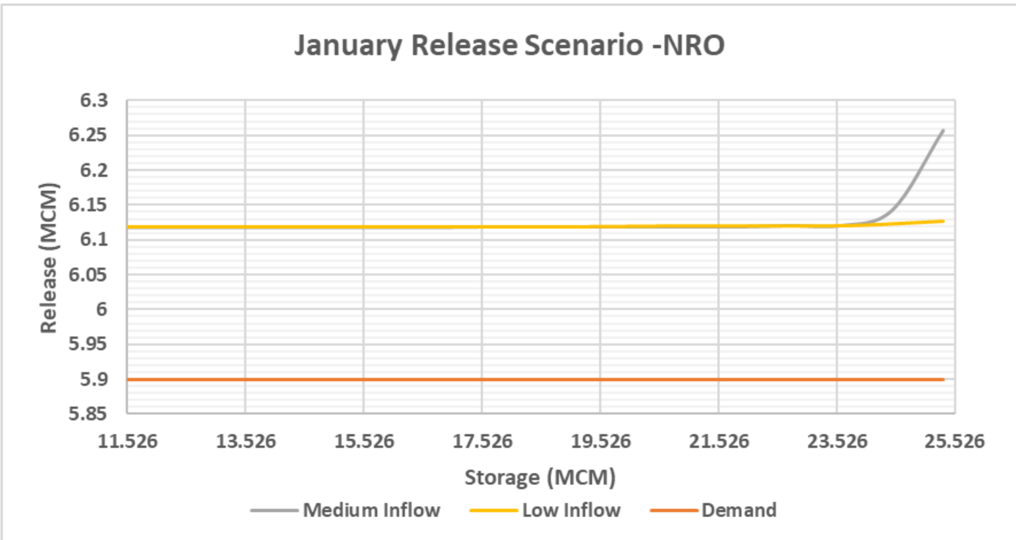
On the other hand, when there was a low level of incoming water, the reservoir let out water, but it was not enough to meet the demands of the area downstream. This was because, throughout the month, whenever there was a low level of incoming water, the reservoir was unable to release an amount of water that was sufficient to meet the downstream demand.

In May, the reservoir was able to fulfil the downstream needs of 5.80 MCM when the storage level was approximately 24 MCM, which was close to the peak storage level. On the other hand, even though the medium inflow was getting very close to achieving the optimal release by fulfilling the downstream needs, it was not able to do so. For the remainder of the year, PSO has demonstrated strong optimising ability as it has the reservoir water release that was able to meet the need from water either by intersecting the downstream line at one point and then releasing more than the downstream need from water or by remaining to supply the downstream with the volume of water that they demand precisely.

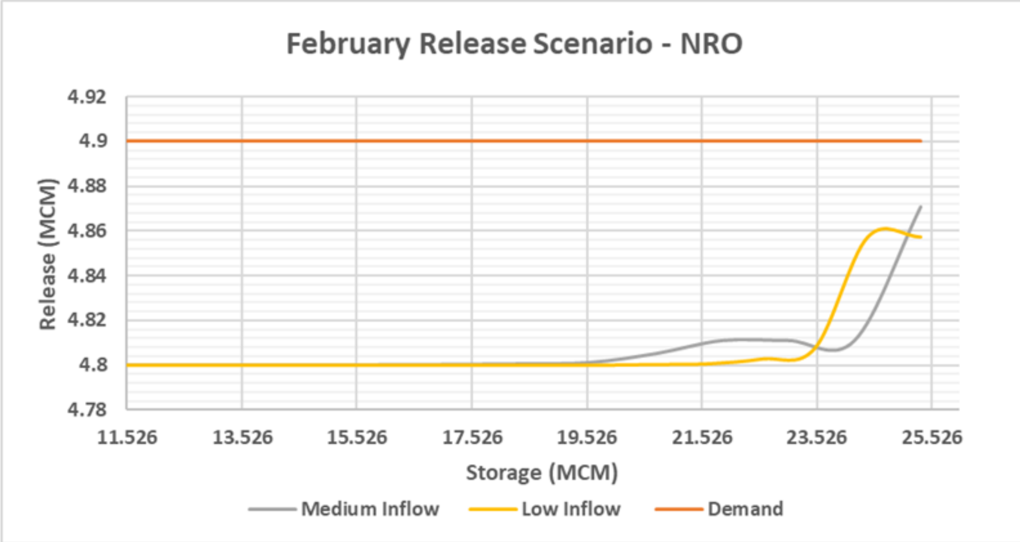
4.4.3 Nuclear Reaction Optimisation (NRO)

Given that it was just recently invented in 2019, the nuclear reaction optimisation method is a relatively new algorithm. Due to the fact that it was only evaluated on benchmark functions, it has not yet been put to the test for applications in the real world. One of the many novelties in this work was employing NRO to enhance reservoir operation; nevertheless, as shown by the release curves (Figure 4.9), which were created by NRO, its performance was substandard.

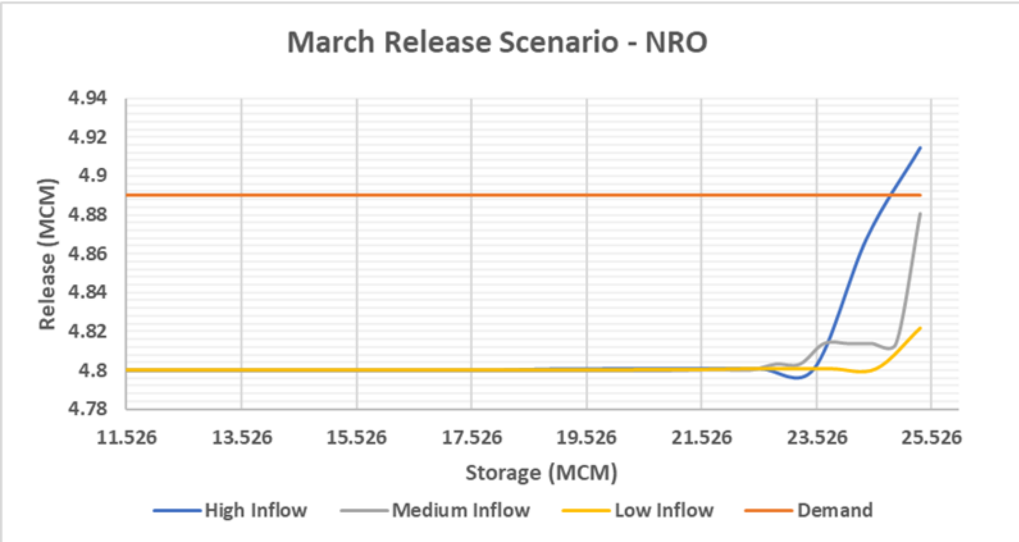
)a(



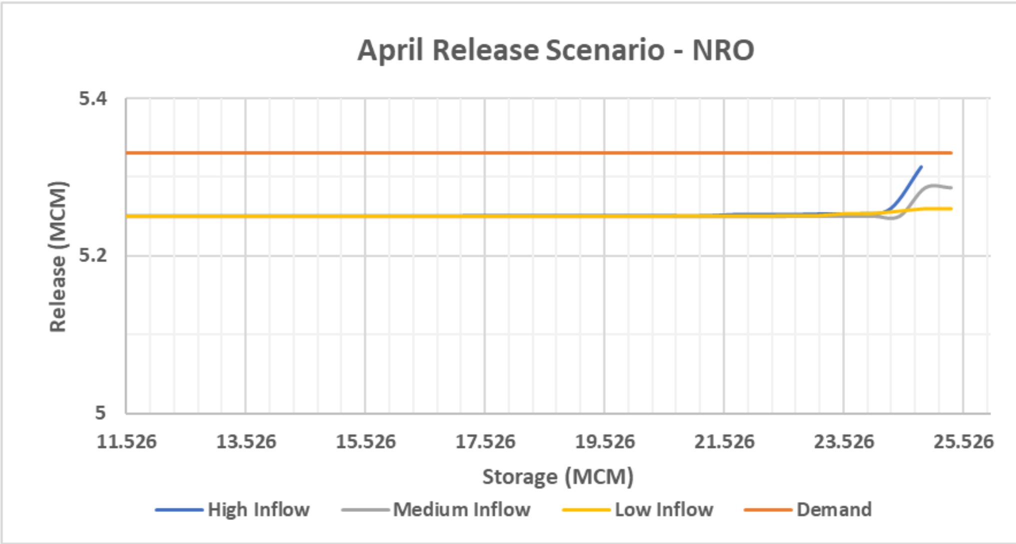
(b)


February Release Scenario - NRO
Release (MCM)
4.92
4.9
4.88
4.86
4.84
4.82
4.8
4.78
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

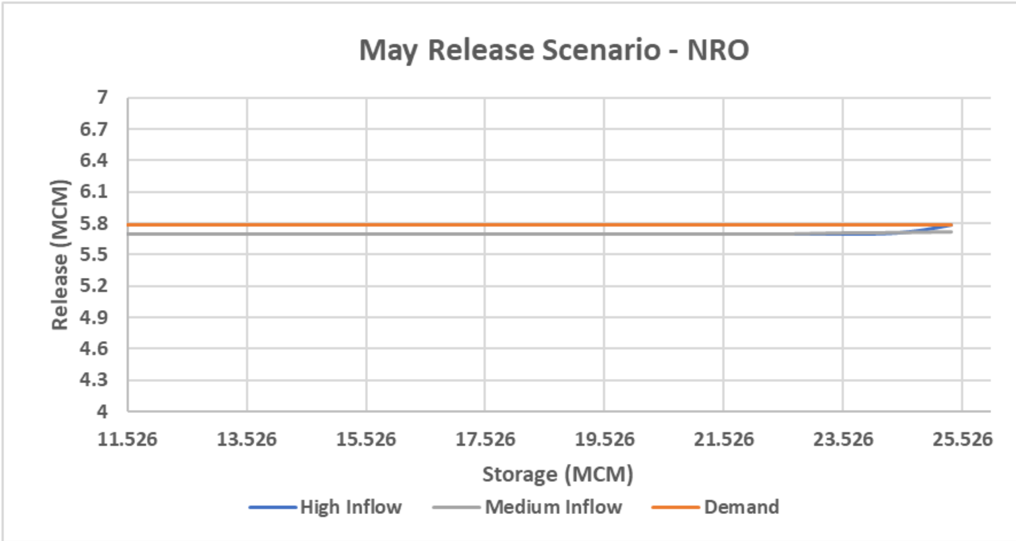
(c)


March Release Scenario - NRO
Release (MCM)
4.94
4.92
4.9
4.88
4.86
4.84
4.82
4.8
4.78
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

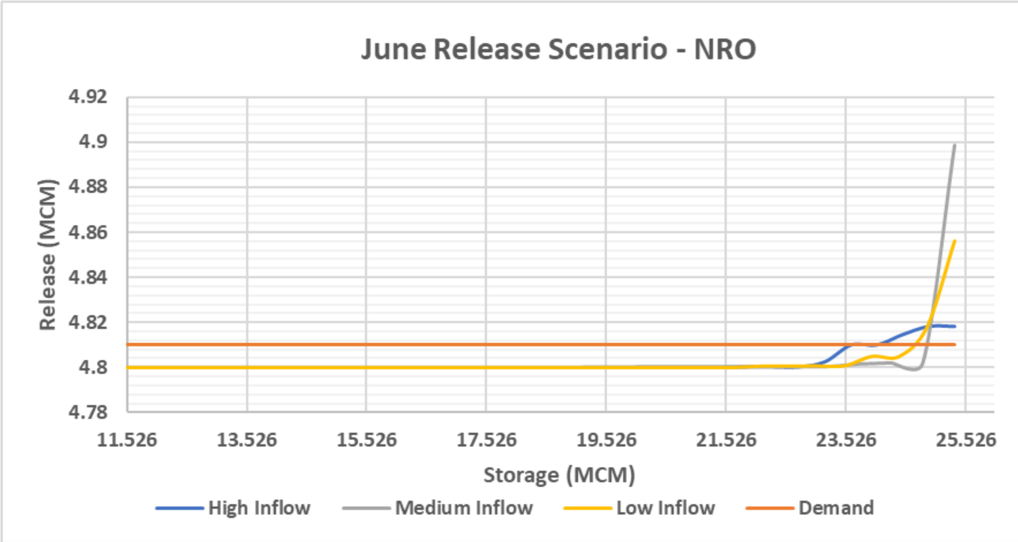
(d)


April Release Scenario - NRO
5.4
5.2
5
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

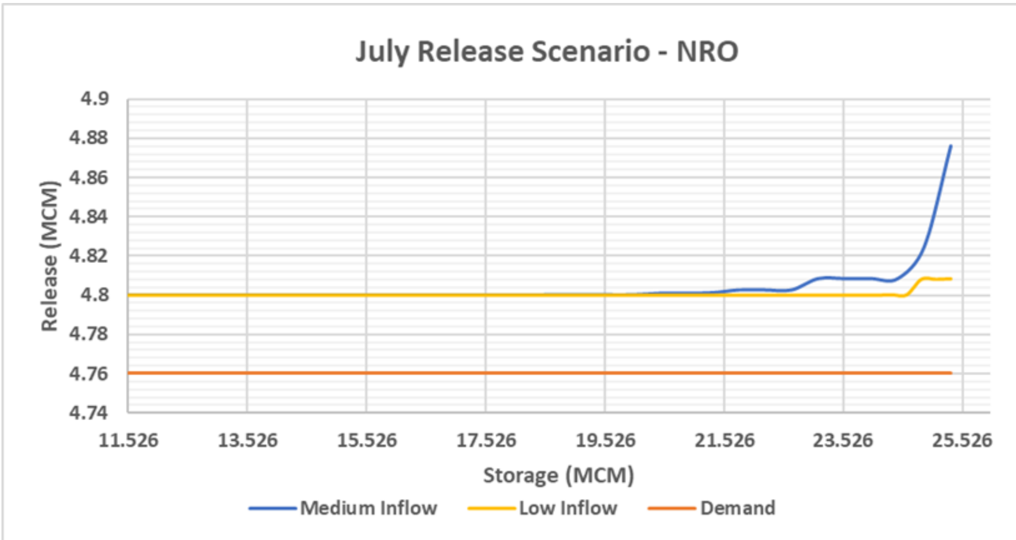
)e(


May Release Scenario - NRO
7
6.7
6.4
6.1
5.8
5.5
5.2
4.9
4.6
4.3
4
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

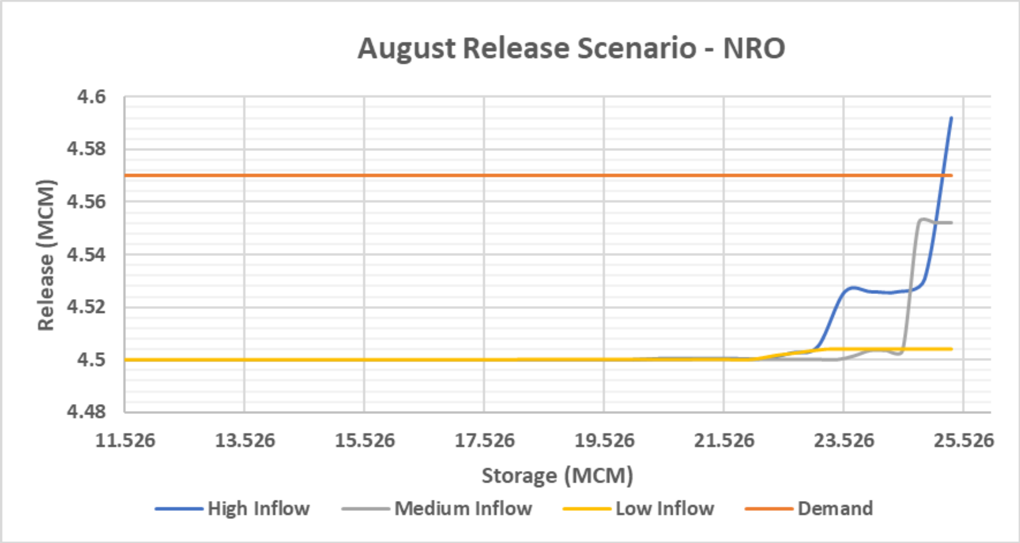
)f(


June Release Scenario - NRO
Release (MCM)
4.92
4.9
4.88
4.86
4.84
4.82
4.8
4.78
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

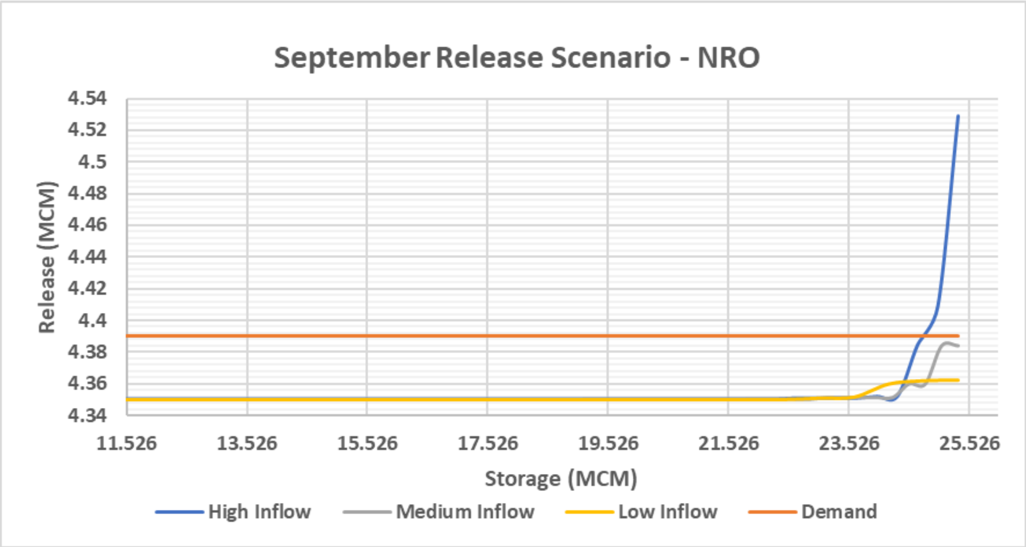
)g(


July Release Scenario - NRO
Release (MCM)
4.9
4.88
4.86
4.84
4.82
4.8
4.78
4.76
4.74
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

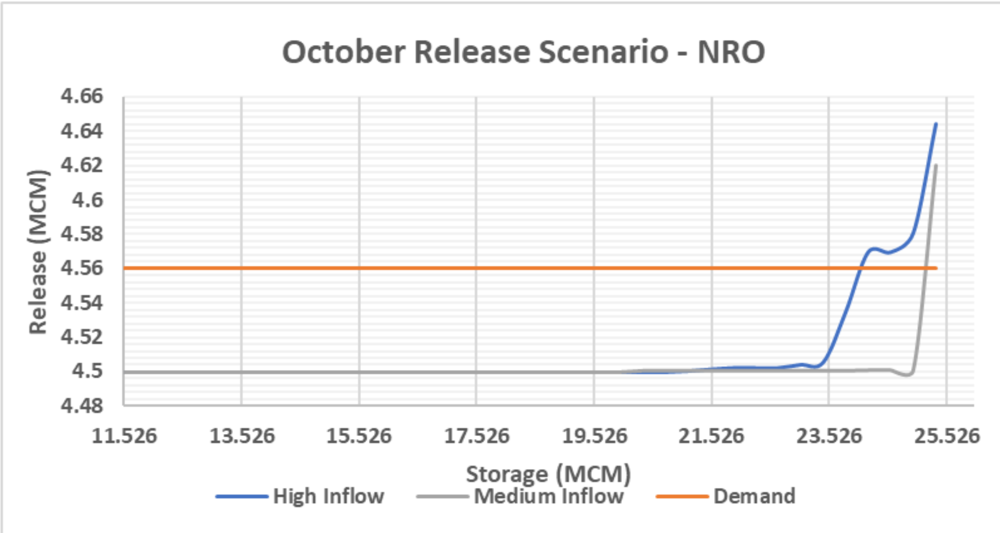
)h(


August Release Scenario - NRO
Release (MCM)
4.6
4.58
4.56
4.54
4.52
4.5
4.48
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

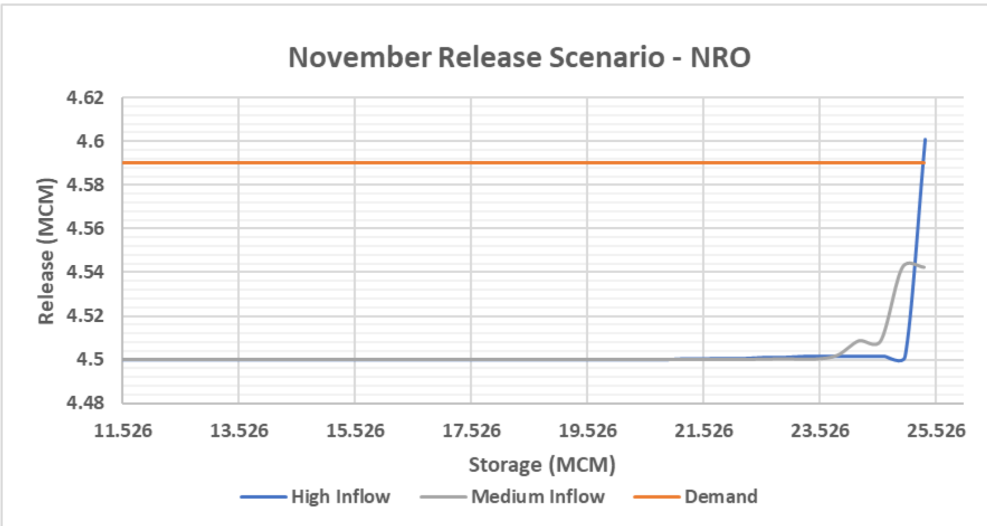
)i(


September Release Scenario - NRO
Release (MCM)
4.54
4.52
4.5
4.48
4.46
4.44
4.42
4.4
4.38
4.36
4.34
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

)j(


October Release Scenario - NRO
Release (MCM)
4.66
4.64
4.62
4.6
4.58
4.56
4.54
4.52
4.5
4.48
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

(k)


November Release Scenario - NRO
Release (MCM)
4.62
4.6
4.58
4.56
4.54
4.52
4.5
4.48
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

(l)

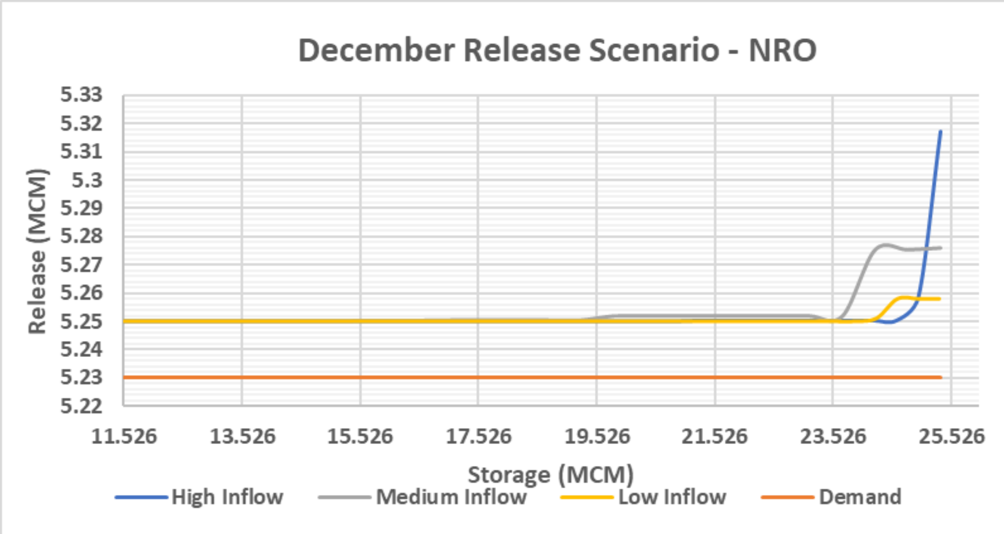


Figure 4.9: NRO Release Curves for Conventional Operation System (CS) for (a) Jan - (l) Dec

In the first two months of the year, January and February, the model did not at any time satisfy the demand that was placed on it. The NRO simulation in January which found that the reservoir produced 6.12 MCM despite the downstream demand being just 5.9 MCM. However, the NRO produced curve shows that the reservoir is lacking around 0.10 million cubic meters of water; this is despite the fact that the downstream demand in February was 4.90.

In the subsequent three months of March, April, and May, NRO simulated that the reservoir can only meet the demand when the water level is raised to fill up the reservoir or else it is not able to meet the demand.

For example, in the month of March, high and medium inflow could only satisfy the need when the water level was at 24 MCM, which demonstrates that NRO is not robust enough to meet the demand while it is being constrained and bounded by the storage and release mechanisms.

In June, it came as a surprise that the reservoir was able to discharge the same quantity of water that was needed downstream for the whole month. Additionally, it has been observed that NRO performance is inconsistent and varies. Because of this, the release pattern does not have any curvature; rather, it is a straight line, which refers to a constant amount of water that is to be released throughout the month while ignoring the demand but satisfying the constraints.

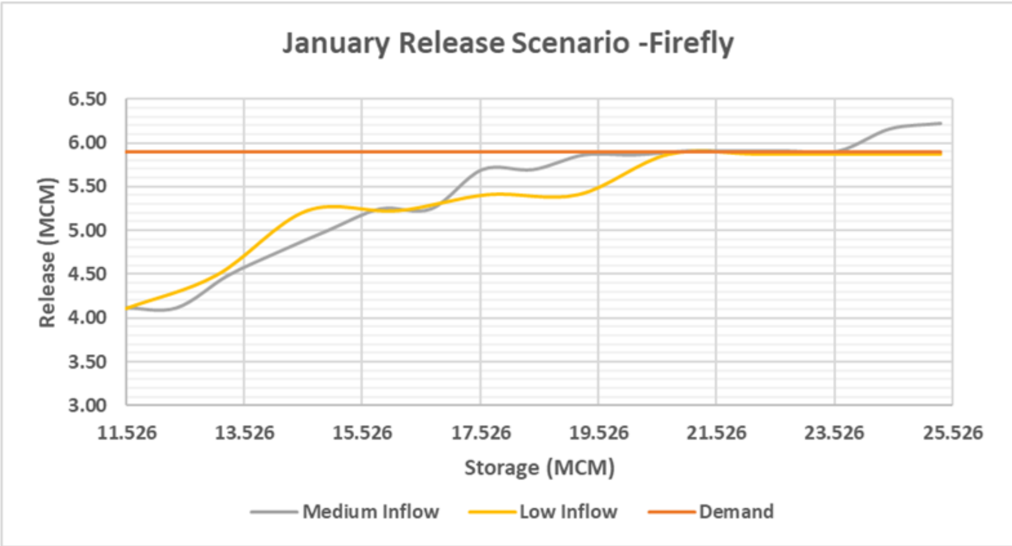
This is one more noticeable observation. NRO does not search for the global optimal solution; instead, it is stuck in the local optimal solution, which assumes that the optimal release has been found. In July and December, the model discharged water at a rate that was greater than the demand downstream, with a total of 4.8 MCM and 5.25 MCM, respectively; however, from August until November, the reservoir was unable to meet the demand in any way, except when the reservoir storage is at its maximum capacity.

4.4.4 Firefly Optimisation Algorithm (FA)

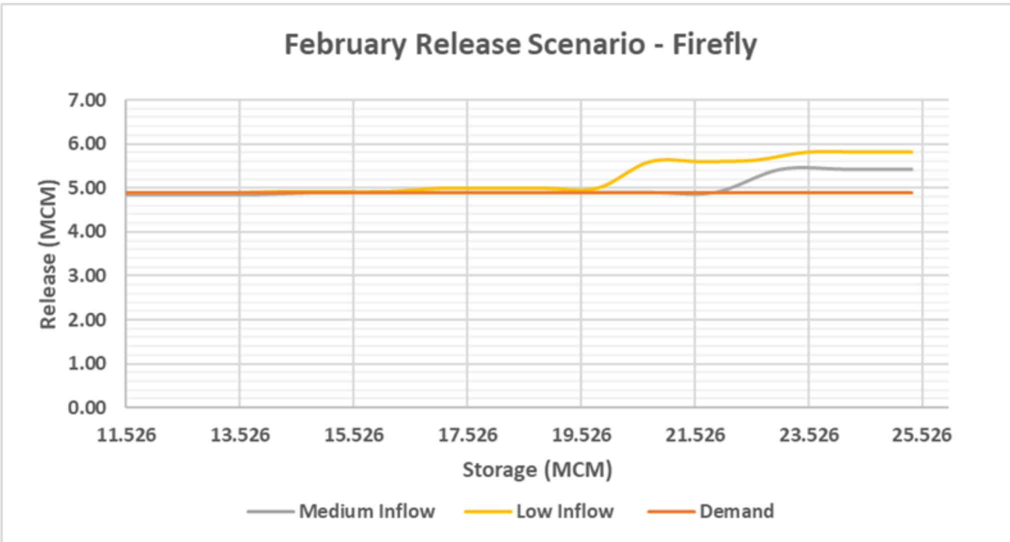
The optimisation issue for which this study seeks a solution is known as a noisy optimisation problem. Because it is based on a real-world application in which the assessment of a solution's fitness may be impacted by a variety of kinds of ambiguity, ambiguity plays a role in the evaluation. Historically, FA has been used for optimising situations that are noisy and nonlinear.

According to Figure 4.10, FA has shown outstanding performance since it has been able to solve the issue with the reservoir and keep up with the demand during the whole year. In the months of March, May, July, August, September, October, and November, and December, the model provided virtually flawless results, as it conformed exactly to the downstream demand straight line. This was a clear observation that was made from the release curves that were created by FA.

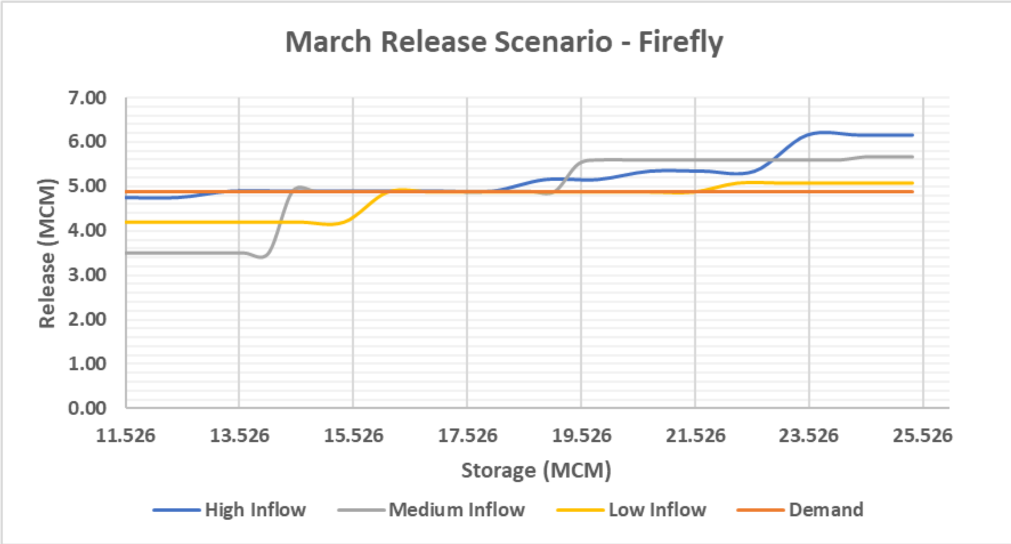
(a)


January Release Scenario -Firefly
Release (MCM)
6.50
6.00
5.50
5.00
4.50
4.00
3.50
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

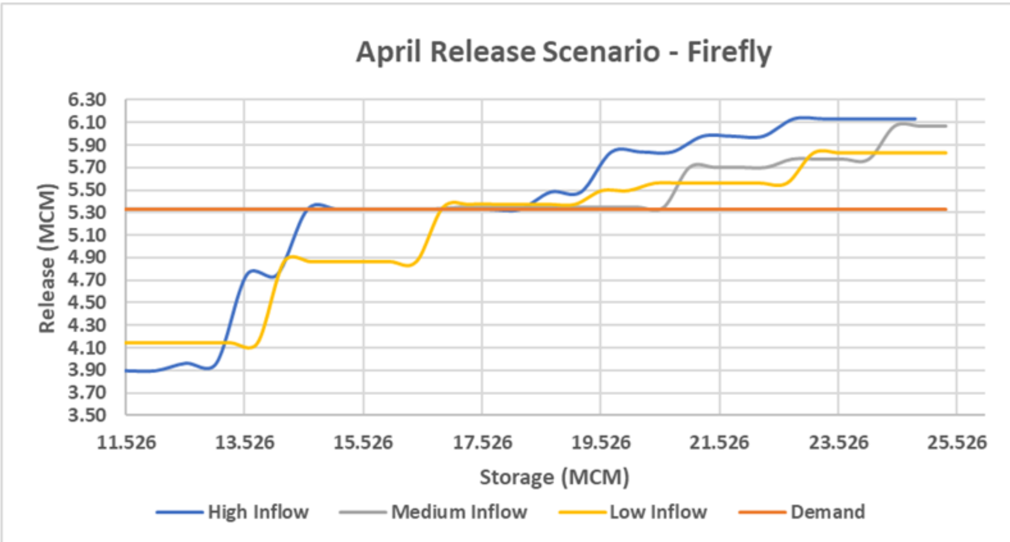
(b)


February Release Scenario - Firefly
Release (MCM)
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

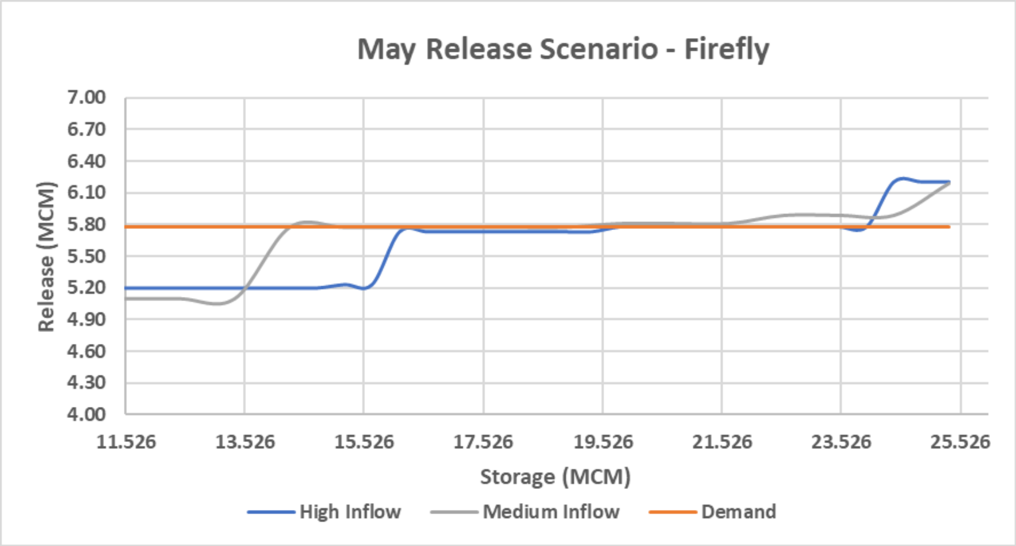
)c(


March Release Scenario - Firefly
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

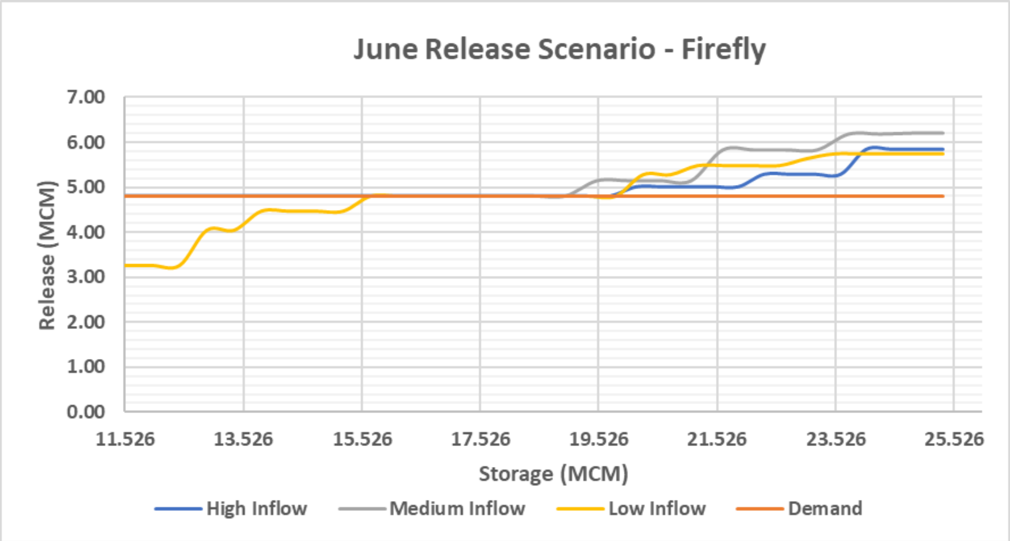
(d)


April Release Scenario - Firefly
6.30
6.10
5.90
5.70
5.50
5.30
5.10
4.90
4.70
4.50
4.30
4.10
3.90
3.70
3.50
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

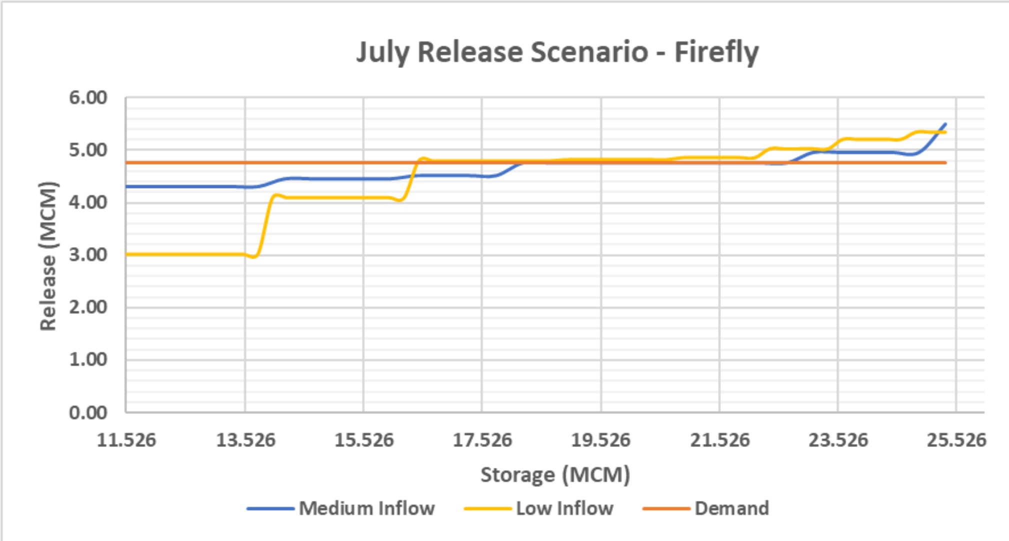
)e(


May Release Scenario - Firefly
7.00
6.70
6.40
6.10
5.80
5.50
5.20
4.90
4.60
4.30
4.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

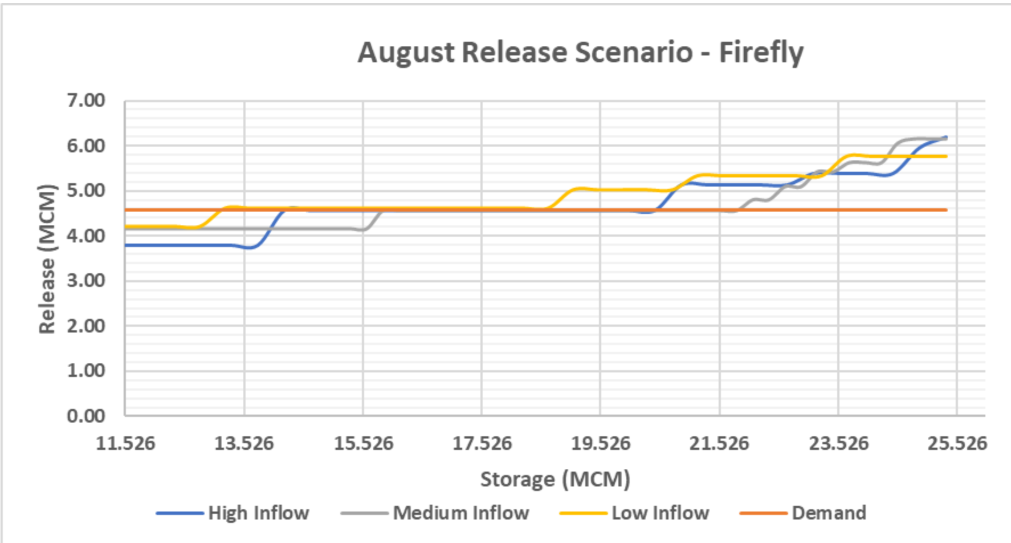
(f)


June Release Scenario - Firefly
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

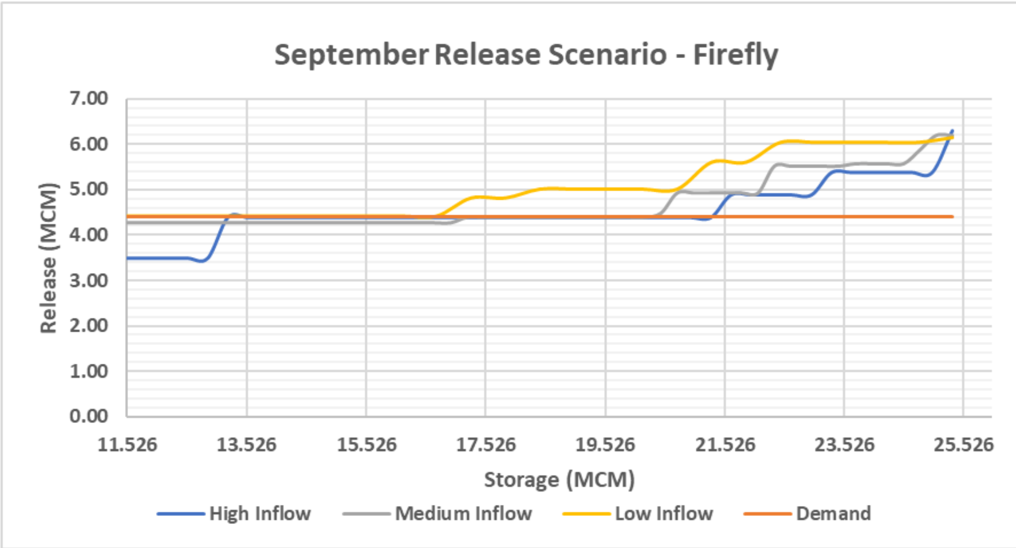
)g(


July Release Scenario - Firefly
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

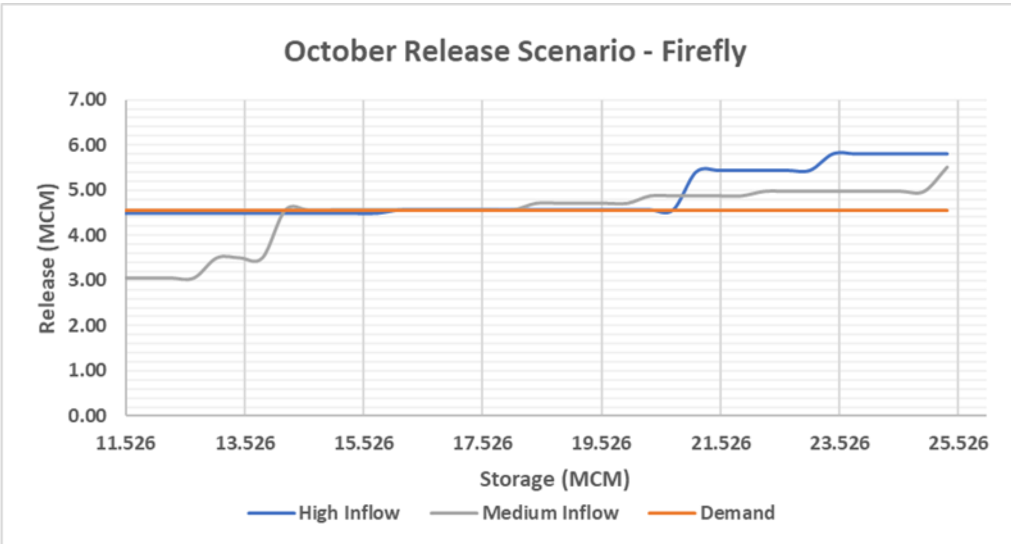
(h)


August Release Scenario - Firefly
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

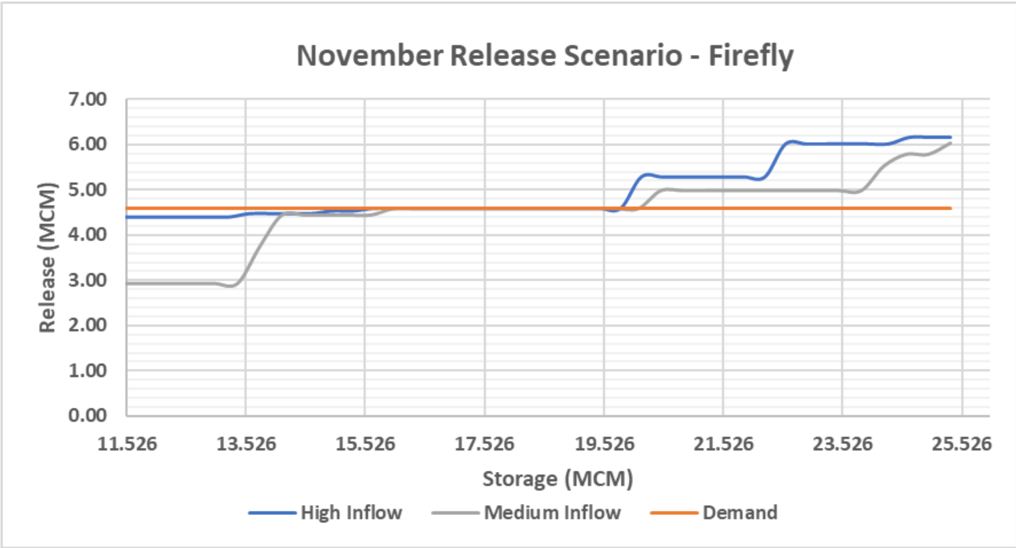
(i)


September Release Scenario - Firefly
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

(j)


October Release Scenario - Firefly
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

(k)



(l)

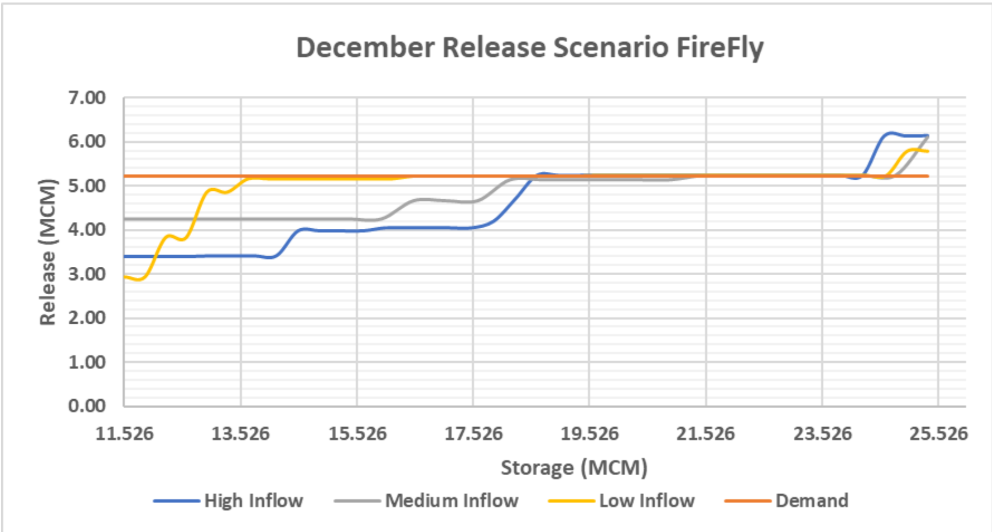


Figure 4.10: FA Release Curves for Conventional Operation System (CS) for (a) Jan - (l) Dec

The release curves that FA generated were able to meet the demand earlier than those that were generated by Genetic algorithm, Particle Swarm, and Nuclear reaction Optimisation in the majority of cases.

This was demonstrated by the fact that FA was able to meet the demand when the water level in the reservoir was raised to approximately 14 MCM. Not to mention the fact that the reservoir FA was able to meet demand in February, which was 4.9 MCM, as well as demand in May, which was higher with approximately 1 million cubic metre (5.8 MCM).

This demonstrates that the model is robust and is able to optimise the water deficit problem even when there are large variations in the needs of the downstream area. Beginning in June and continuing through December, the release curves predicted by the model were in perfect alignment with the downstream line when the level of the reservoir was roughly 15.5 MCM. Starting in June and continuing through December, the requirements for the downstream were 4.76 MCM, 4.57 MCM, 4.39 MCM, 4.56 MCM, 4.59 MCM, and 5.23 MCM, respectively.

4.5 Real-time Simulated Losses Operation System

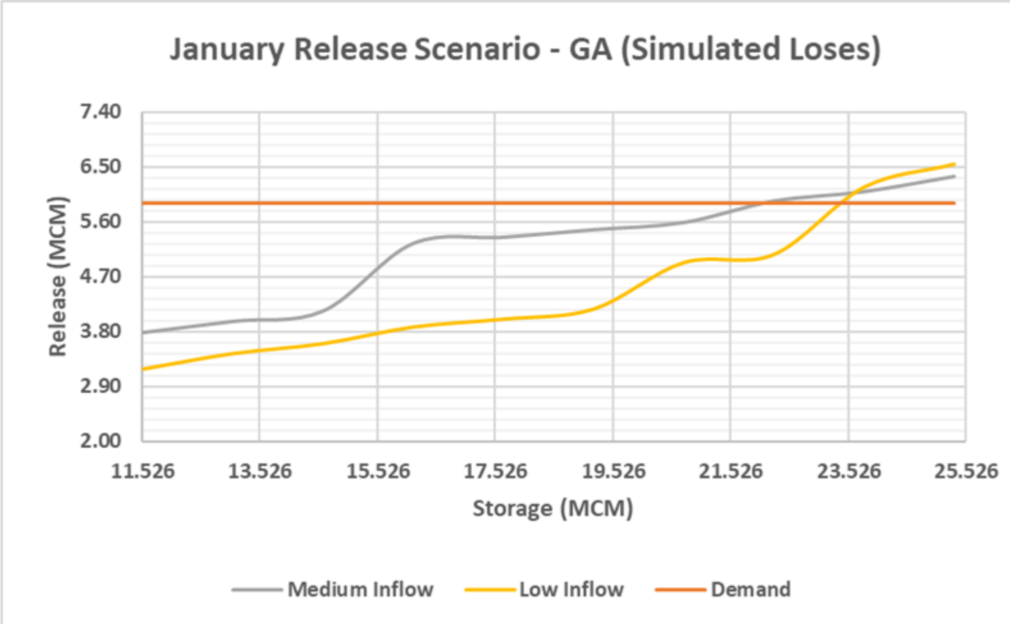
In this subsection, the existing reservoir operating system is improved by taking into account the reservoir losses in the form of a predicted output generated by XG-Boost. It is defined as a closed-loop system that will iterate and generate the release curves with minimal user interaction. In this way, it overcomes the necessity of having a reservoir operating manager available at the reservoir site to feed the model with the most recent recorded data obtained from the sensors. The optimisation systems that were developed under this section are defined as having been developed as a closed-loop system.

The first thing that can be noticed among all of the operation curves is that the shape of the line graph has become more linear in comparison to the traditional system. This indicates that the algorithms were able to reach the global solution with fewer numbers of iterations, which is the second observation that can be seen among all of the operation curves.

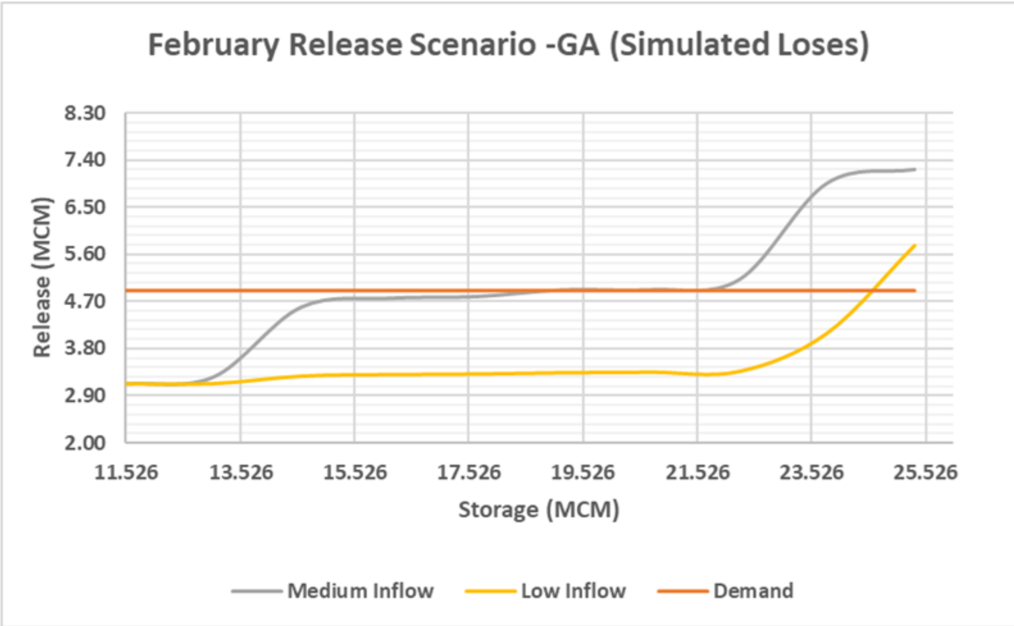
4.5.1 Genetic Algorithm (IS)

While both systems were able to fulfil demand in January when inflow was medium, the conventional system never came close to doing so when inflow was low. However, the integrated system's GA was able to meet demand at one moment in both the medium and low inflow situations when storage levels hit 22.264 MCM and 23.798 MCM, respectively, as shown in Figure 4.11. However, the storage level required to reach 19.526 MCM in medium and 24.5 MCM in low for the model to supply the downstream demand.

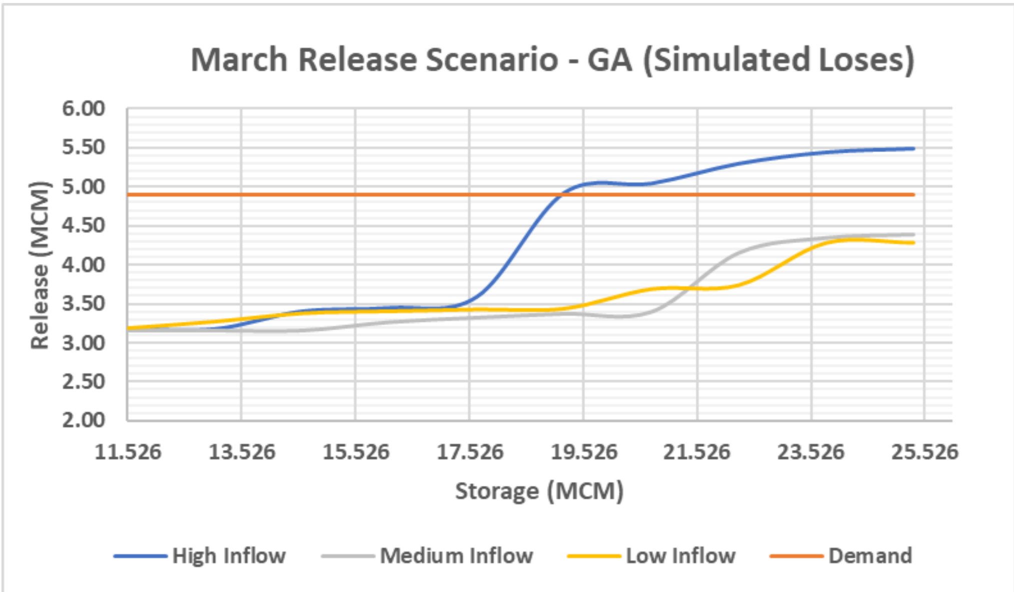
(a)


January Release Scenario - GA (Simulated Loses)
7.40
6.50
5.60
4.70
3.80
2.90
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

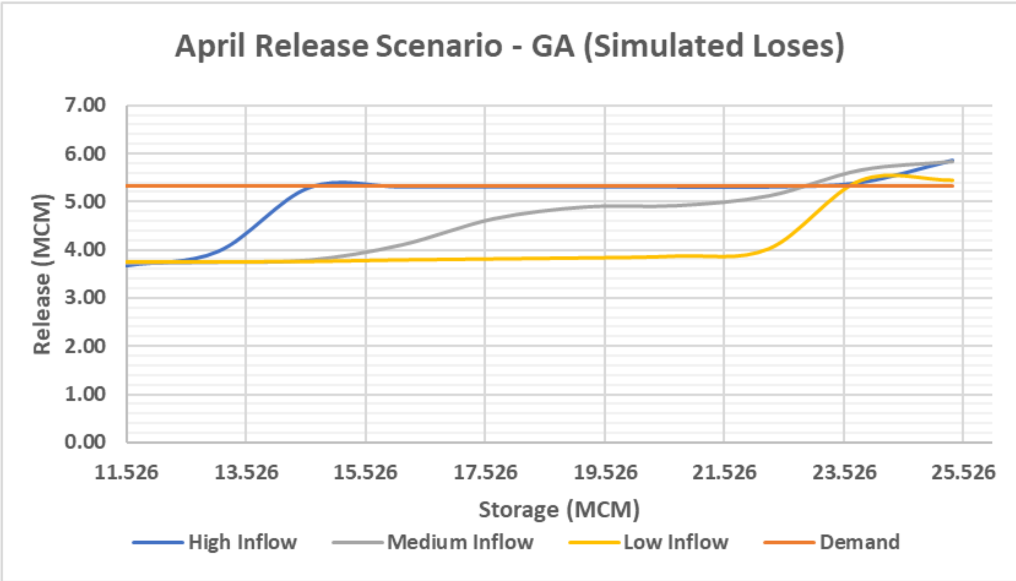
(b)


February Release Scenario -GA (Simulated Loses)
8.30
7.40
6.50
5.60
4.70
3.80
2.90
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

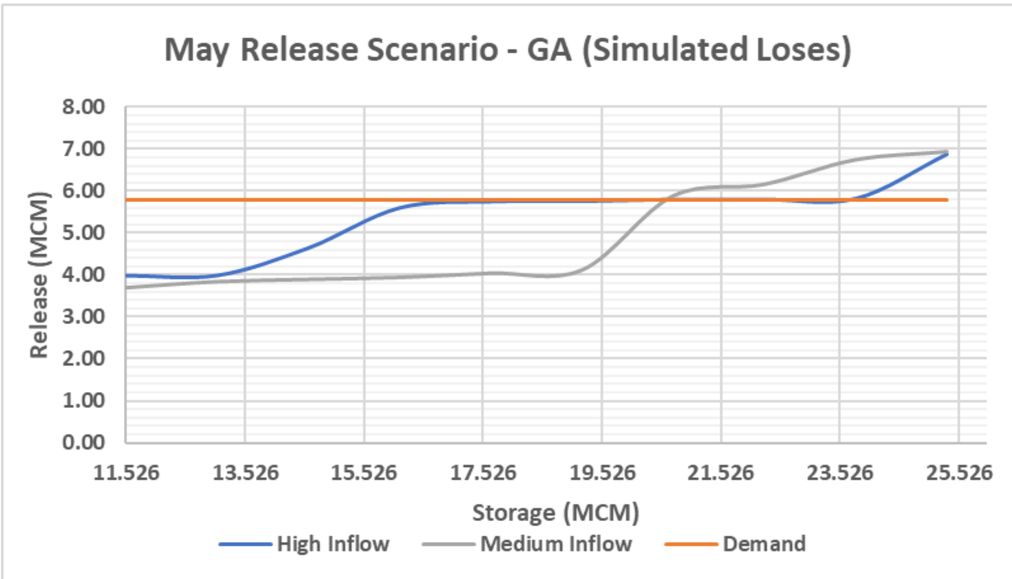
(c)


March Release Scenario - GA (Simulated Loses)
6.00
5.50
5.00
4.50
4.00
3.50
3.00
2.50
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

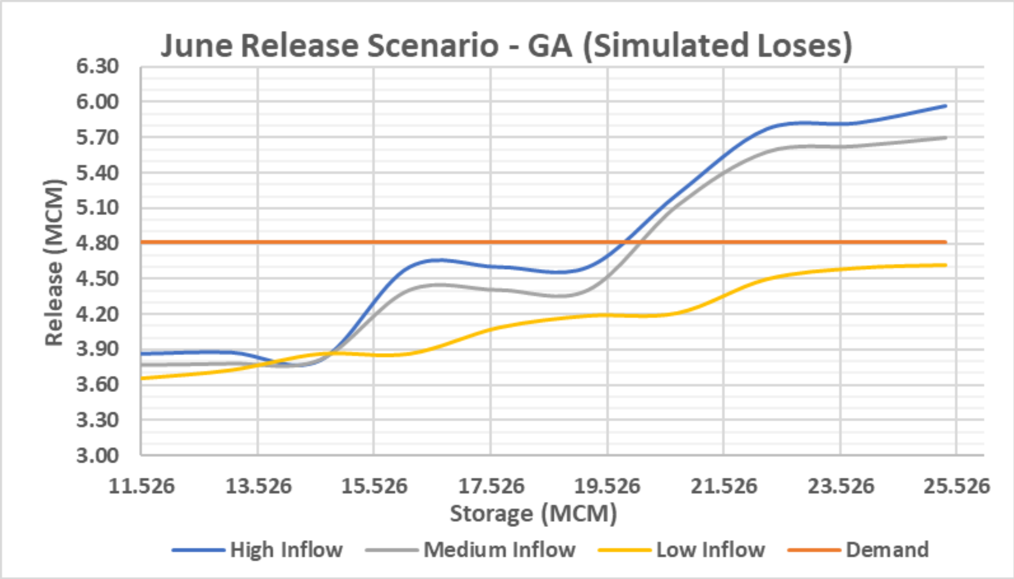
(d)


April Release Scenario - GA (Simulated Loses)
Release (MCM)
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

(e)

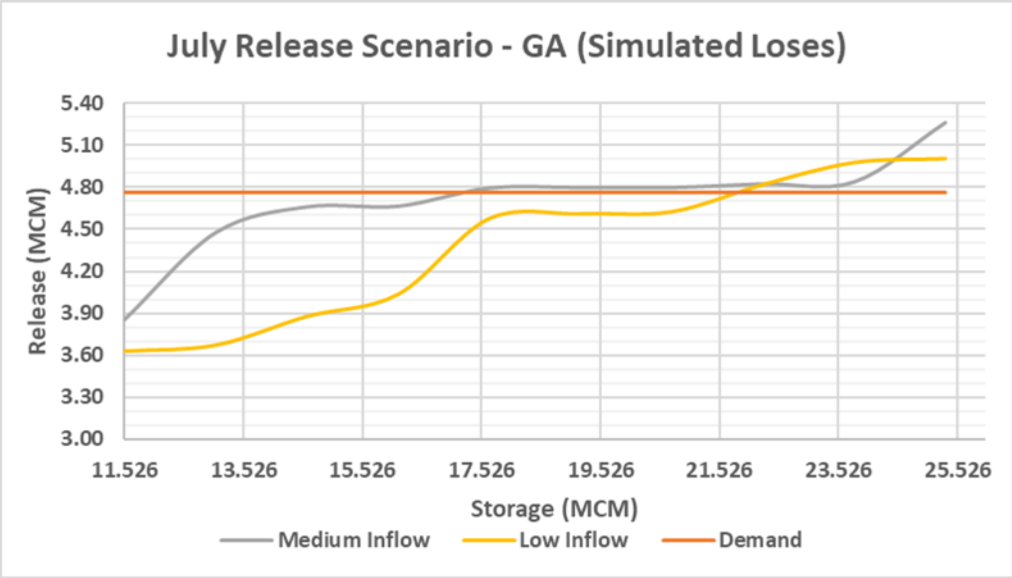

May Release Scenario - GA (Simulated Loses)
Release (MCM)
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

(f)



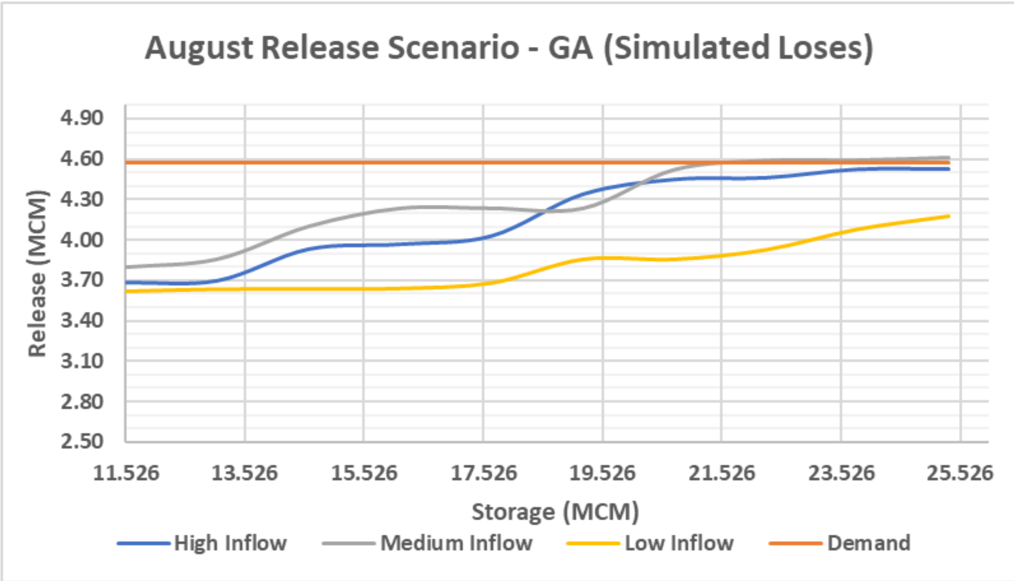
June Release Scenario - GA (Simulated Loses)
Release (MCM)
6.30
6.00
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

(g)

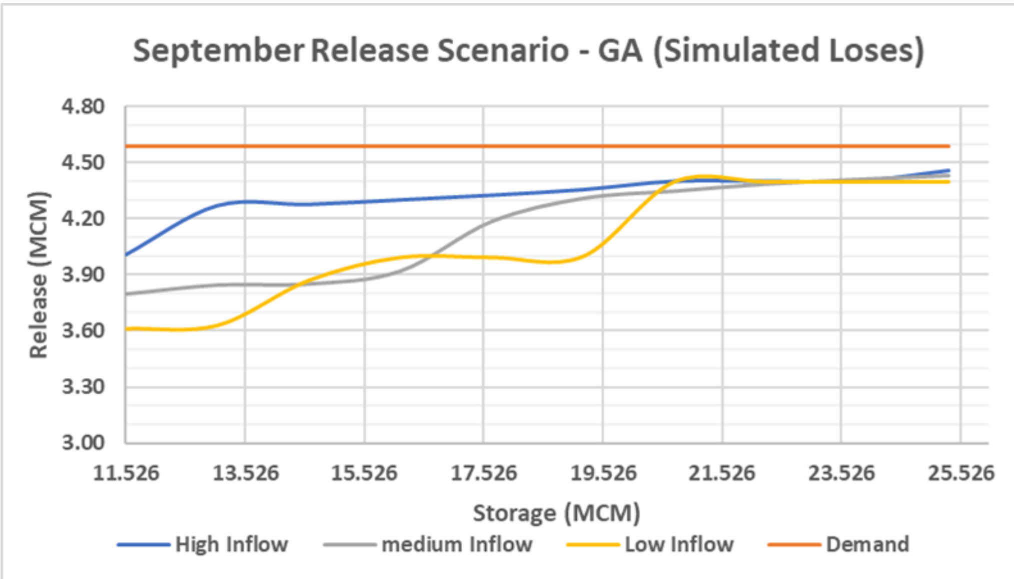


July Release Scenario - GA (Simulated Loses)
Release (MCM)
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

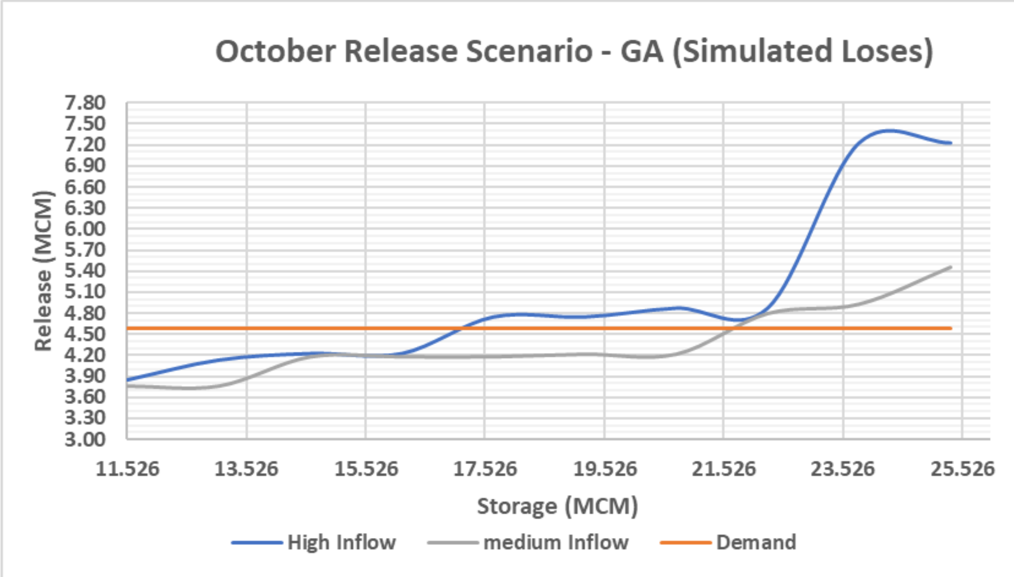
(h)


August Release Scenario - GA (Simulated Loses)
Release (MCM)
4.90
4.60
4.30
4.00
3.70
3.40
3.10
2.80
2.50
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

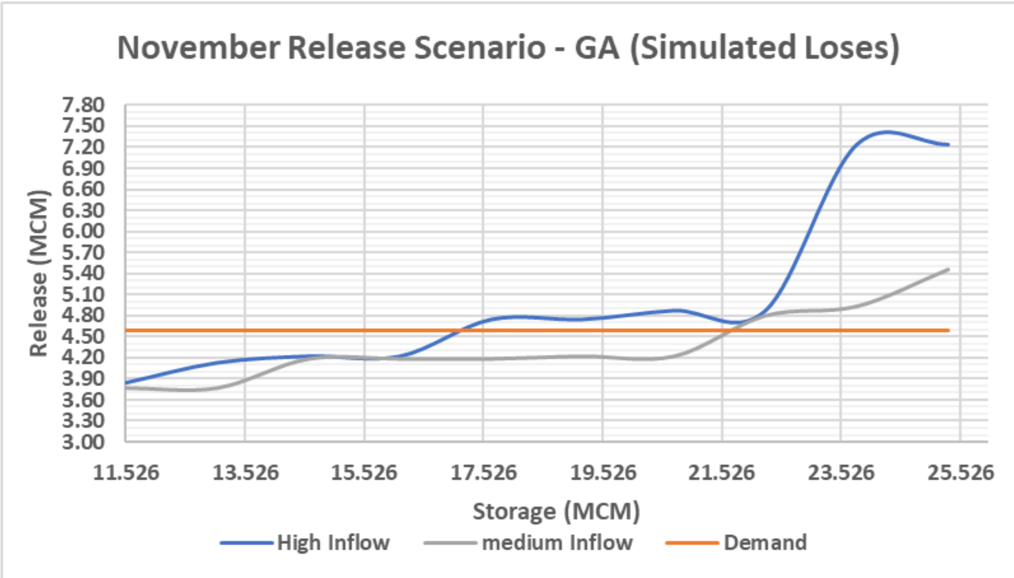
(i)


September Release Scenario - GA (Simulated Loses)
Release (MCM)
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Low Inflow
Demand

(j)


October Release Scenario - GA (Simulated Loses)
Release (MCM)
7.80
7.50
7.20
6.90
6.60
6.30
6.00
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Demand

(k)


November Release Scenario - GA (Simulated Loses)
Release (MCM)
7.80
7.50
7.20
6.90
6.60
6.30
6.00
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Demand

(l)

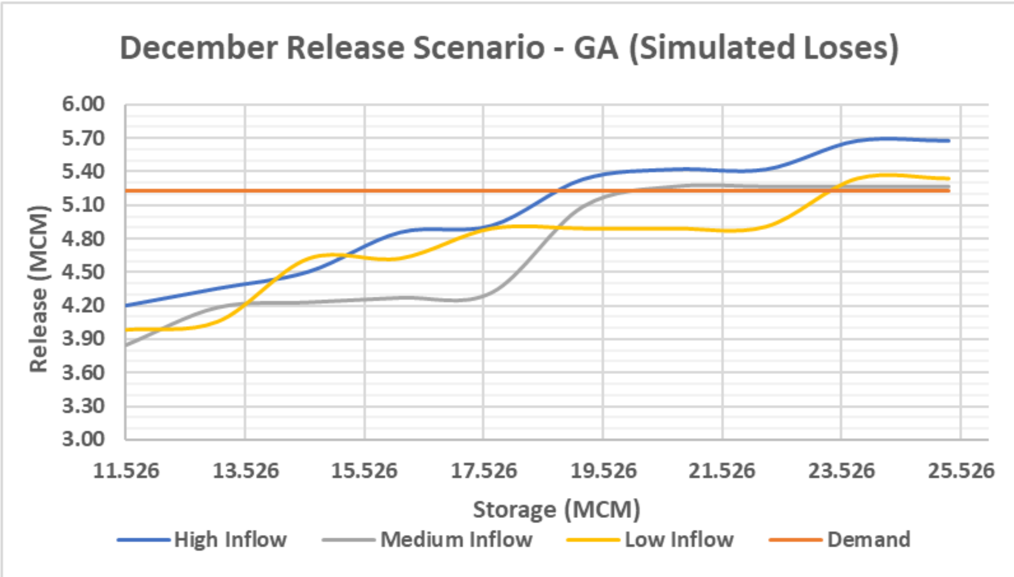


Figure 4.11: GA Release Curves for Integrated System (IS) for (a) Jan - (l) Dec

In February, GA fared somewhat better in the integrated system as it was able to meet the demand in both circumstances. The model did not function well in March because it was unable to meet the requirements of the downstream processes, with the exception of situations in which the reservoir received strong inflow and the reservoir storage was increased to over 18 MCM.

In the months of April and May, the model was able to suit the requirements of the community since it was able to provide the demand regardless of the circumstances. When the reservoir's storage reached the level of 14 MCM in April, when there was a large inflow, the reservoir could supply the needs when it kept its release at the same level of 5.3 MCM.

In addition, the model successfully met demand even during periods of medium and low input when the water level in the reservoir was down to around 22 MCM. In spite of the difference in storage level, the pattern of the release curve under high inflow circumstances in May was quite similar to the one attained in April. This was the case despite the fact that the model satisfied the demand in May when the storage level was increased to around 16.5 MCM.

Next in line, the model was not sensitive enough or accurate enough during the month of June. As a result, the reservoir was compelled to release more water than was necessary since the model could not fulfil the demand under any circumstances.

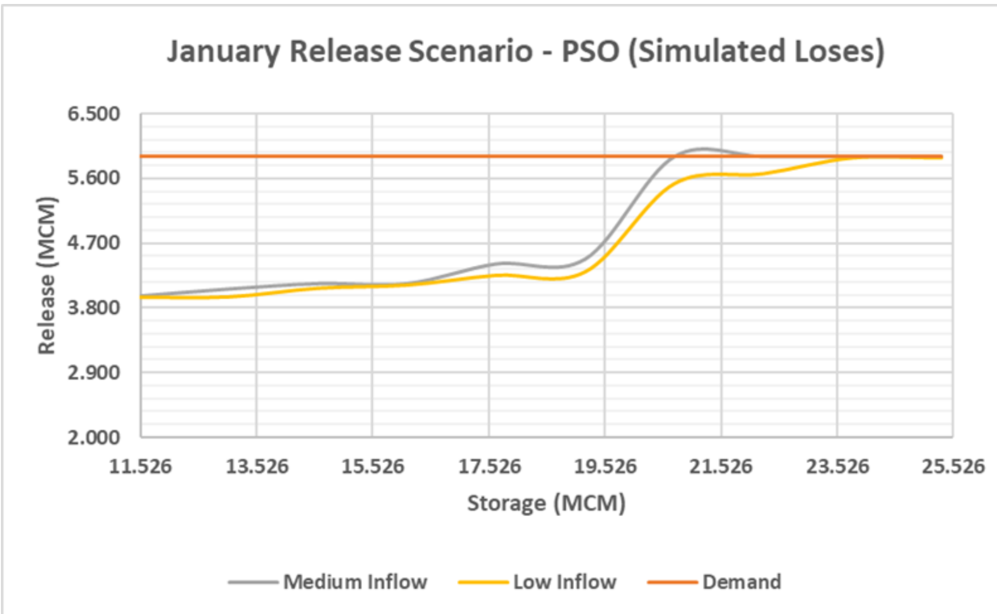
The amount of water that is discharged from the reservoir during times of intense inflow is always greater than the demand. This holds true from July until December (with the exception of September, when the model was unable to satisfy any downstream needs under any inflow conditions), so the overall conclusion is as follows: when there was a medium amount of incoming water, practically all of the discharges were able to meet the demand even if there was limited storage.

Nevertheless, when there was a low inflow, the volume of water that was released by the reservoir was only just enough to meet the exact demand. This was due to the fact that the variable changes (low inflow) immediately impacted both the capacity of the reservoir and the amount of water that was released.

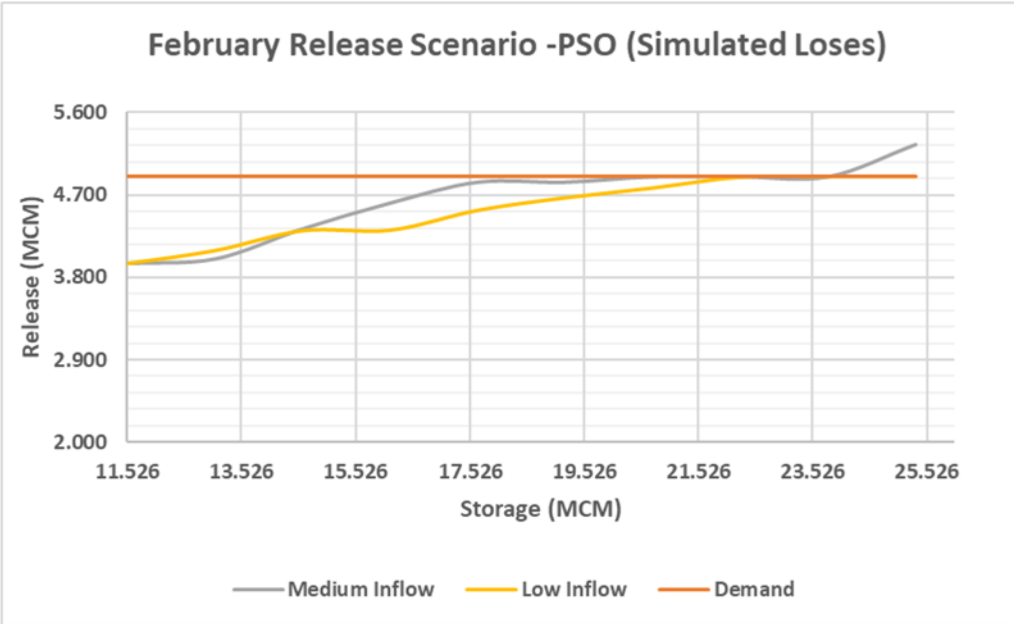
4.5.2 Particle Swarm (IS)

A cursory review of the rule curves shown in Figure 4.12 indicated that the low storage capacity phase of the monthly release curves had a serious lack of available water. This was discovered when the curves were examined. The integrated PSO model suggested emptying the reservoir with a high storage capacity of a significant amount of water (an oversupply) during times of high inflow in order to keep the reservoir storage within a safe range and doing the opposite during times of low inflow in order to satisfy the demand for the water.

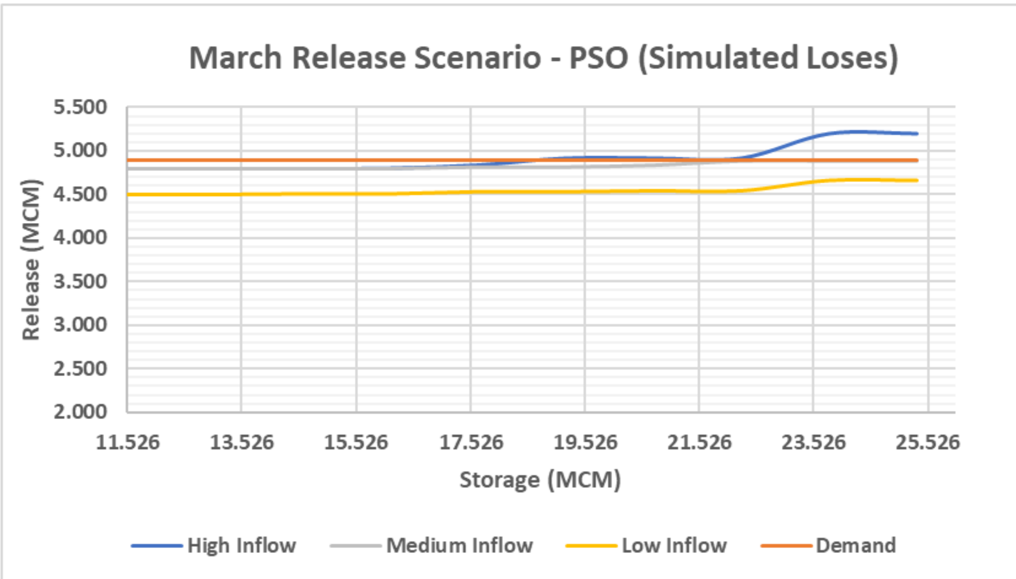
(a)



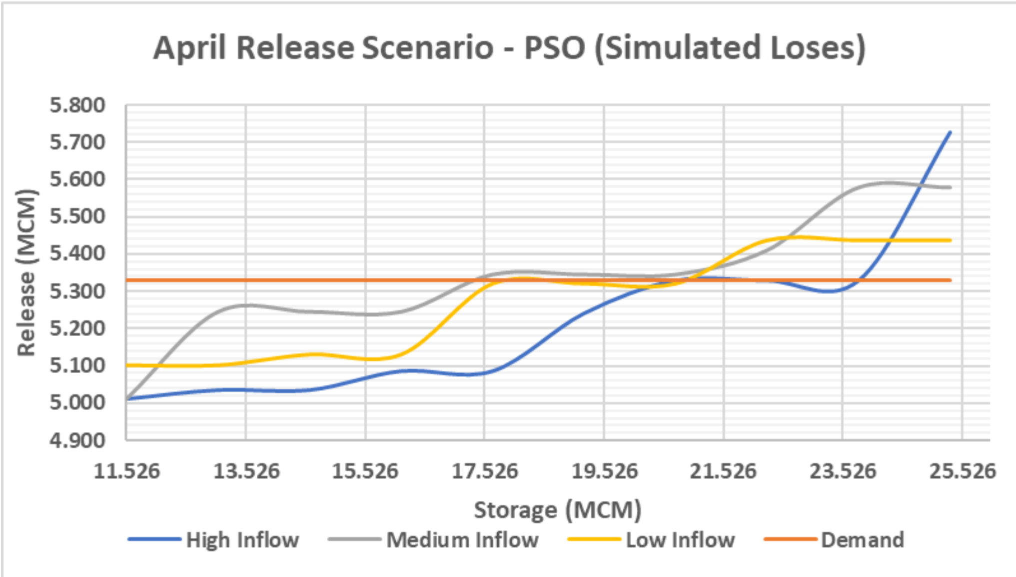
(b)


February Release Scenario -PSO (Simulated Loses)
5.600
4.700
3.800
2.900
2.000
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

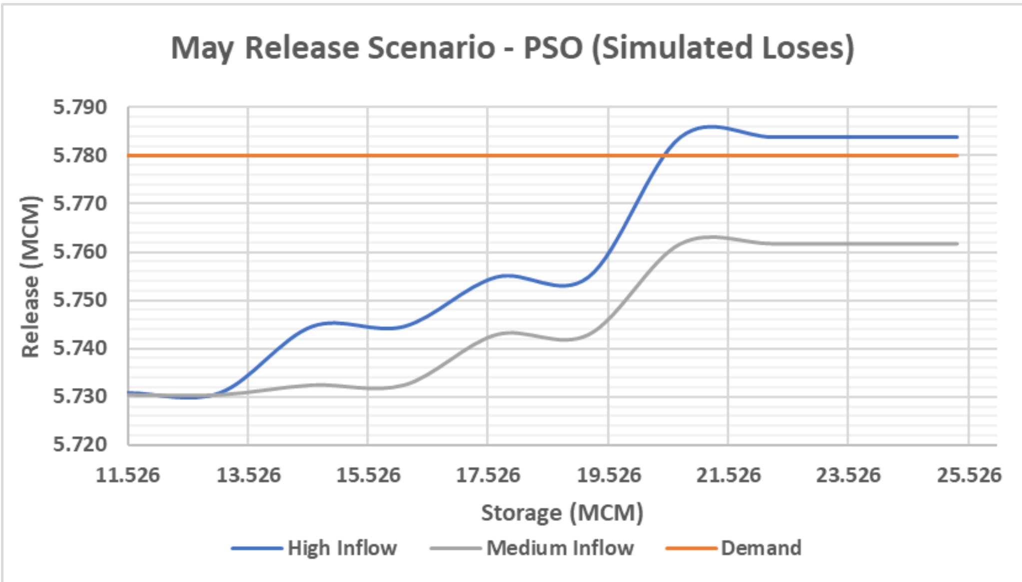
(c)


March Release Scenario - PSO (Simulated Loses)
5.500
5.000
4.500
4.000
3.500
3.000
2.500
2.000
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

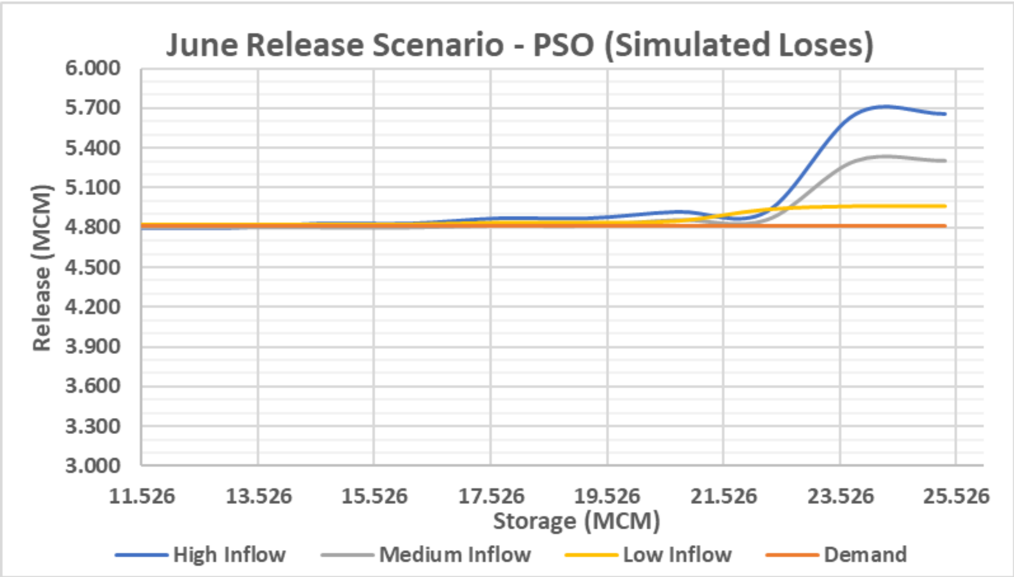
(d)


April Release Scenario - PSO (Simulated Loses)
Release (MCM)
5.800
5.700
5.600
5.500
5.400
5.300
5.200
5.100
5.000
4.900
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

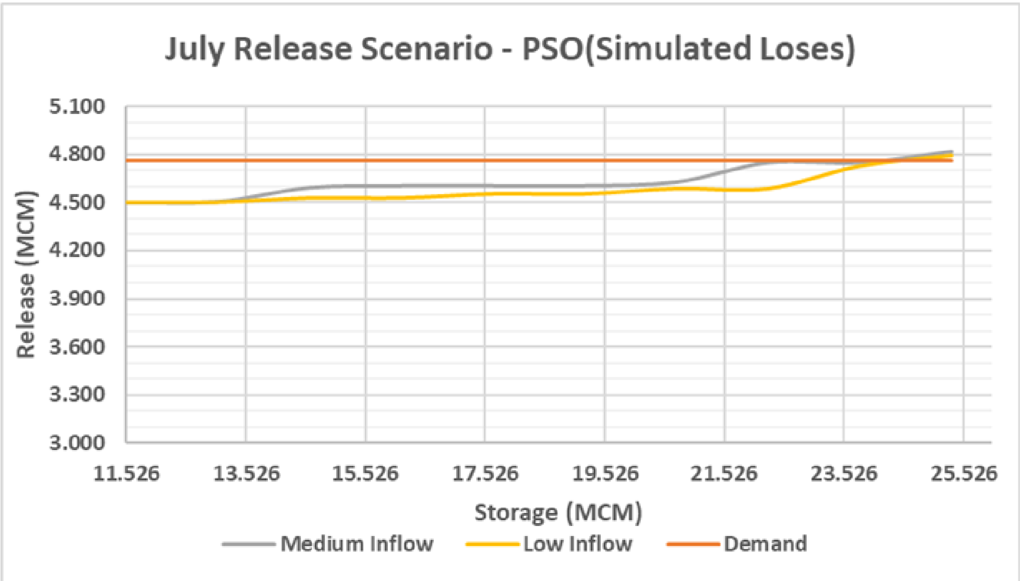
(e)


May Release Scenario - PSO (Simulated Loses)
Release (MCM)
5.790
5.780
5.770
5.760
5.750
5.740
5.730
5.720
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

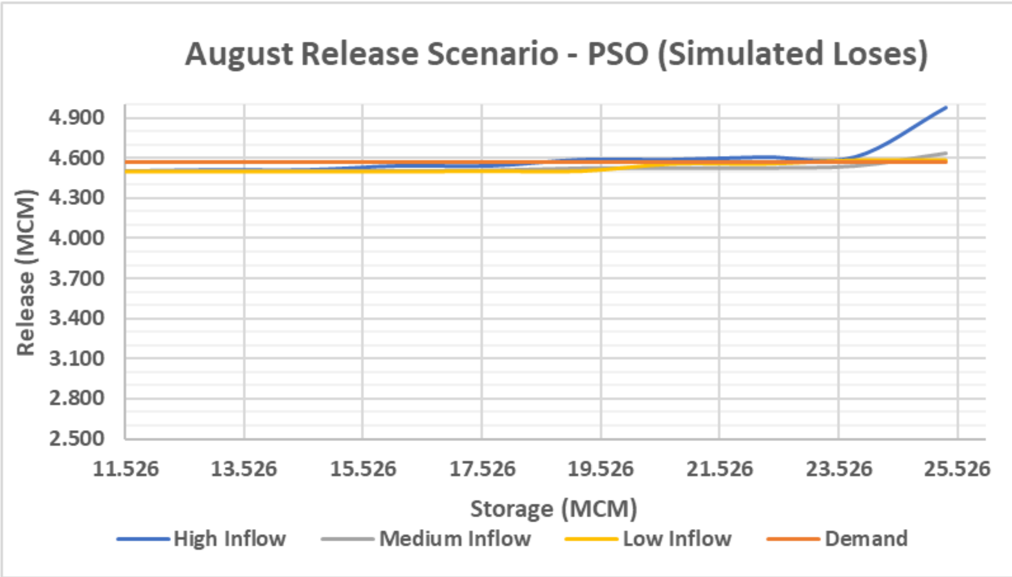
(f)


June Release Scenario - PSO (Simulated Loses)
Release (MCM)
6.000
5.700
5.400
5.100
4.800
4.500
4.200
3.900
3.600
3.300
3.000
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

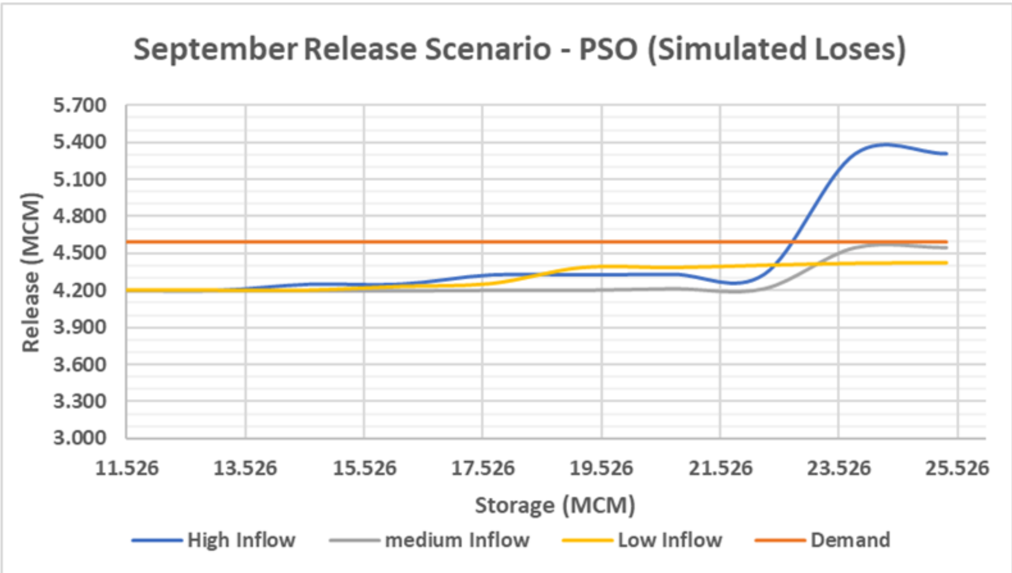
(g)


July Release Scenario - PSO(Simulated Loses)
Release (MCM)
5.100
4.800
4.500
4.200
3.900
3.600
3.300
3.000
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

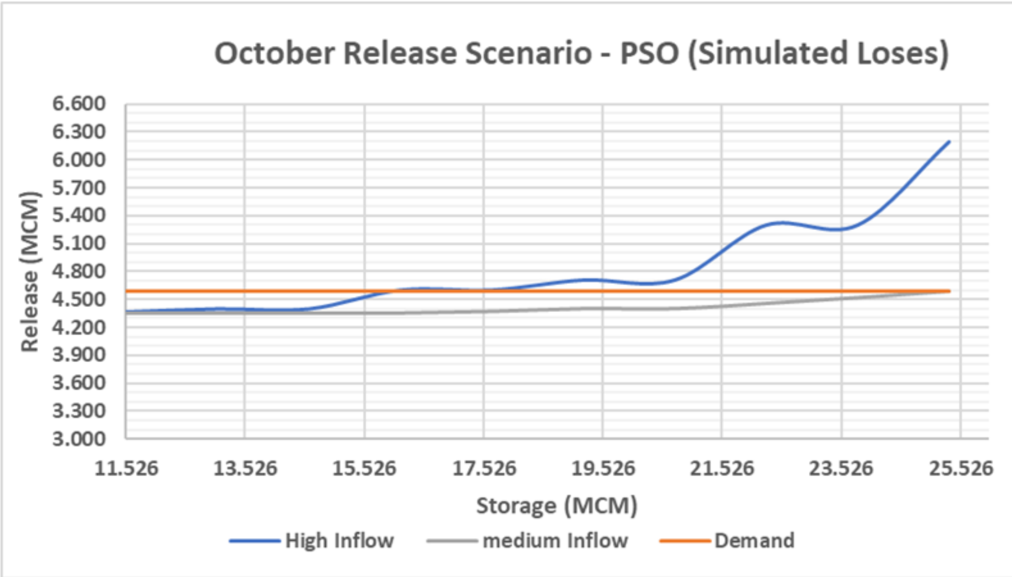
(h)


August Release Scenario - PSO (Simulated Loses)
Release (MCM)
4.900
4.600
4.300
4.000
3.700
3.400
3.100
2.800
2.500
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

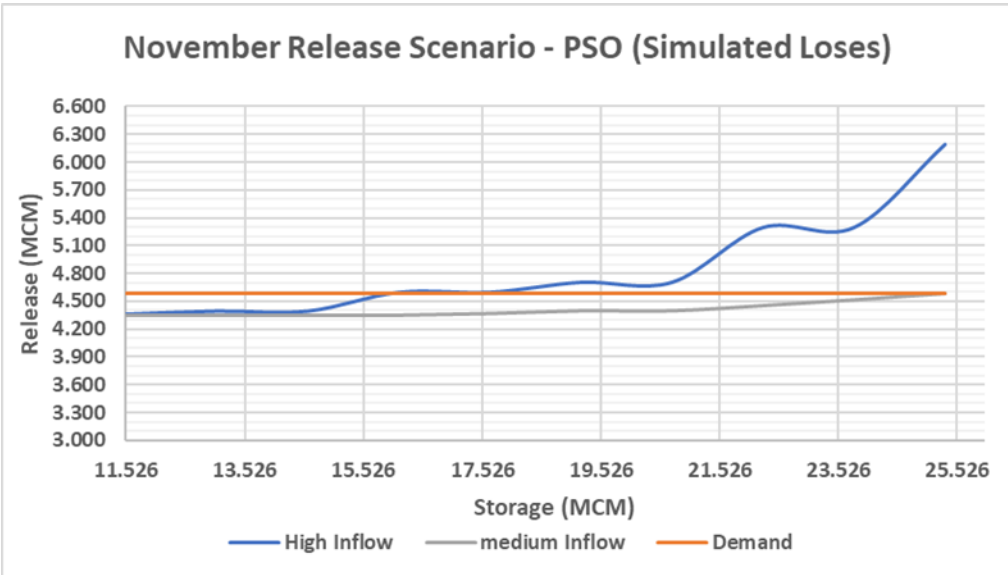
(i)


September Release Scenario - PSO (Simulated Loses)
Release (MCM)
5.700
5.400
5.100
4.800
4.500
4.200
3.900
3.600
3.300
3.000
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Low Inflow
Demand

(j)


October Release Scenario - PSO (Simulated Loses)
Release (MCM)
6.600
6.300
6.000
5.700
5.400
5.100
4.800
4.500
4.200
3.900
3.600
3.300
3.000
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Demand

(k)


November Release Scenario - PSO (Simulated Loses)
Release (MCM)
6.600
6.300
6.000
5.700
5.400
5.100
4.800
4.500
4.200
3.900
3.600
3.300
3.000
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Demand

(l)

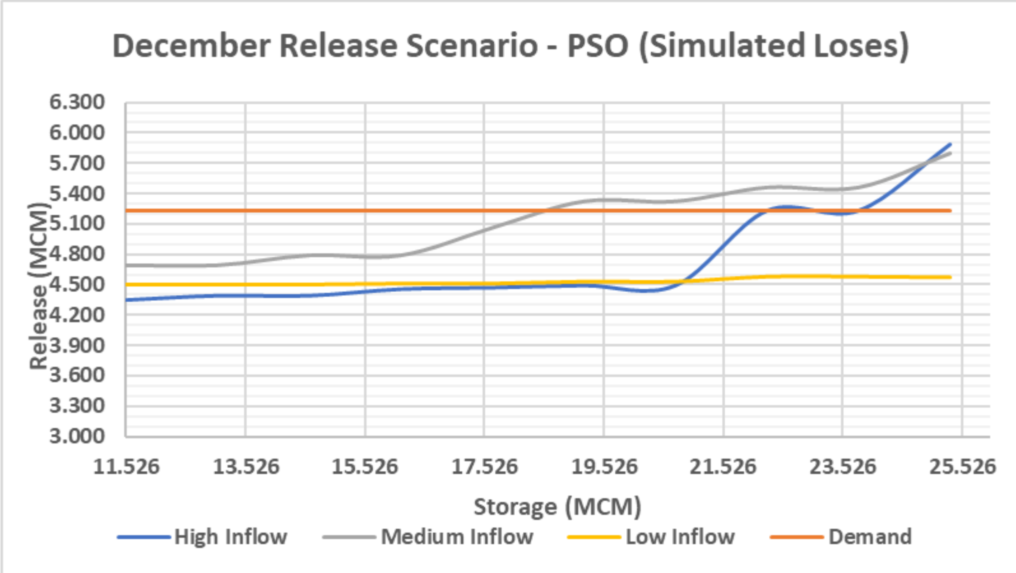


Figure 4.12: PSO Release Curves for Integrated System (IS) for (a) Jan - (l) Dec

The results shown in Figure 4.12 suggested that the Loss integrated PSO operation rule that was proposed might perhaps reduce the water deficit in a more efficient manner. The results also showed that the planned PSO performed substantially better than GA in which it was able to satisfy the demands of the downstream area in 9 out of 12 months regardless of the inflow conditions. March, May, and December were the three months in which PSO was unable to supply the demand downstream. The demand downstream was 4.50 MCM, 5.78 MCM, and 5.3 MCM correspondingly during those three months.

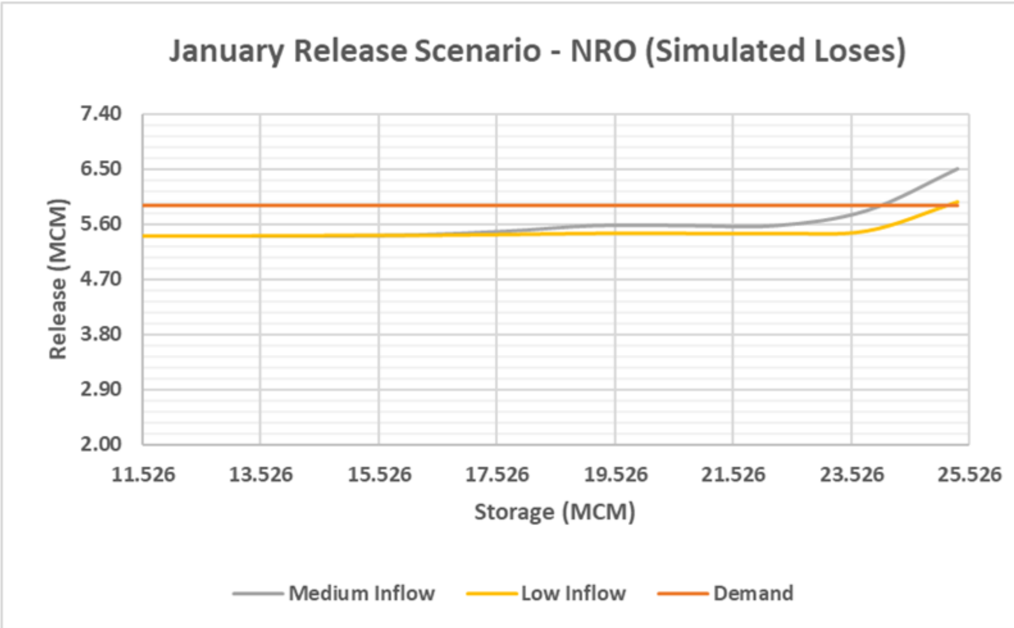
In addition to its strong performance, the PSO algorithm is a viable choice for optimal water distribution in times of scarcity and drought due to its low sensitivity to the starting population and rapid reaction time in comparison to other heuristic algorithms. This makes the PSO algorithm a good option for optimal water distribution in times of scarcity and drought (such as genetic algorithms). As a result of the results acquired by PSO, it is possible to deduce that the PSO model has a high capacity to produce satisfying outcomes for optimising reservoir operations in a short period of time. This can be derived from the fact that the PSO model was successful in obtaining the findings.

4.5.3 Nuclear Reaction Optimisation (IS)

In comparison to the traditional NRO, the performance of the NRO that has been integrated with the losses modelling system has been marginally improved, as shown in Figure 4.13. Utilizing the differential operator of each parameter, the NRO develops the ion that efficiently balances the exploitation and exploration capacities in proportion to the risk of ionisation at this stage.

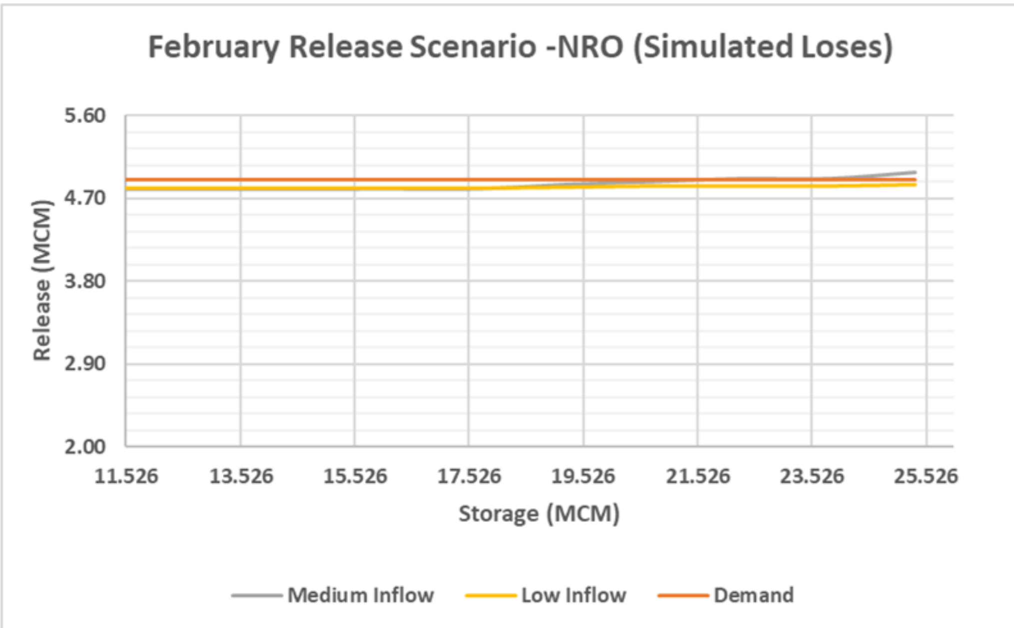
In order for the NRO to manufacture fusion products during the fusion phase, it employs differential operators of whole ions in addition to modelling exploitation and exploration while taking into consideration the likelihood of fusion. The integration of losses has prompted NRO to employ the Levy flight variants to seek at random and be able to escape the local minimum at each step. This was done as a result of the motivation provided by the integration of losses.

(a)



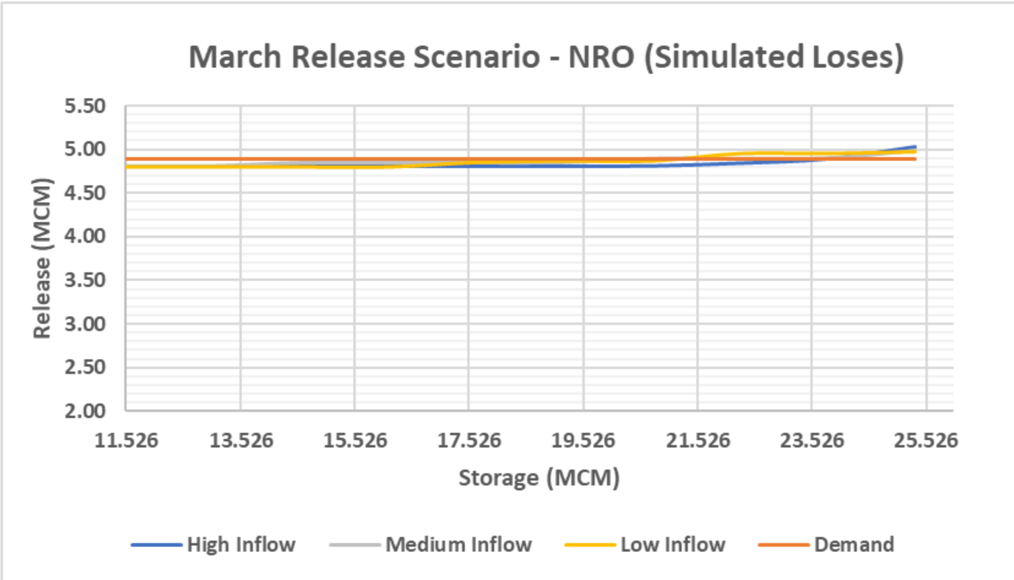
January Release Scenario - NRO (Simulated Loses)
7.40
6.50
5.60
4.70
3.80
2.90
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

(b)

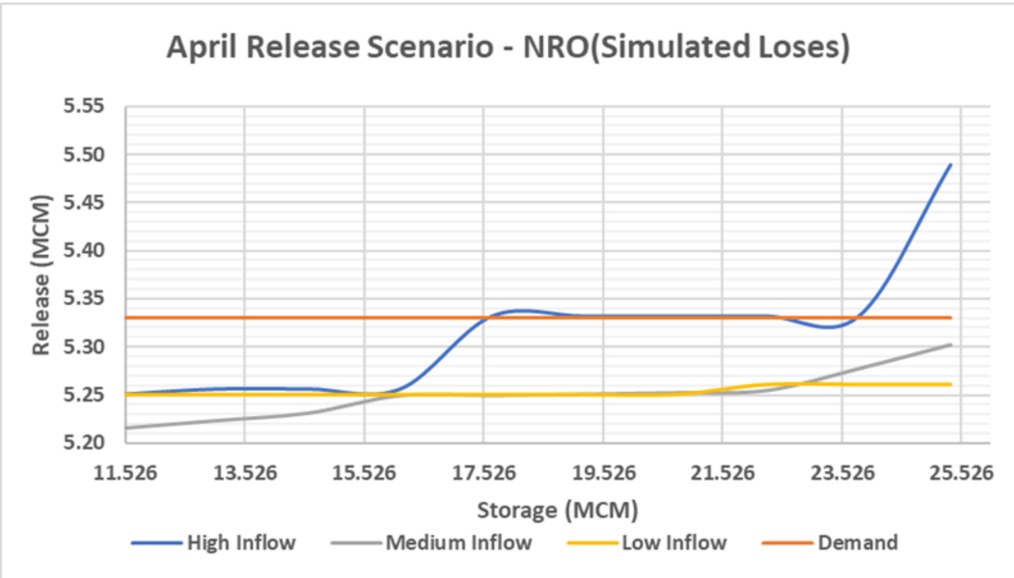


February Release Scenario -NRO (Simulated Loses)
5.60
4.70
3.80
2.90
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

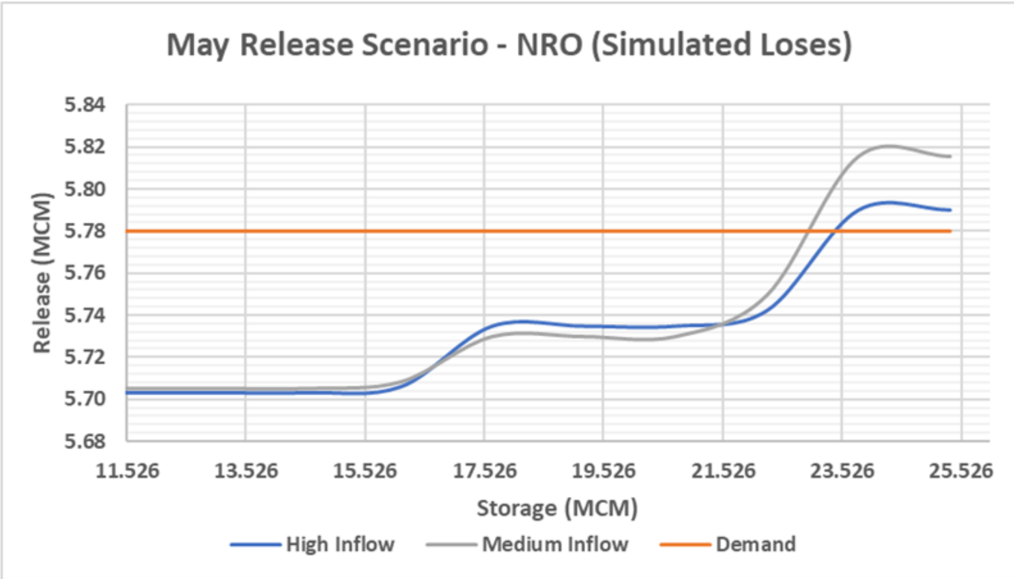
(c)


March Release Scenario - NRO (Simulated Loses)
5.50
5.00
4.50
4.00
3.50
3.00
2.50
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

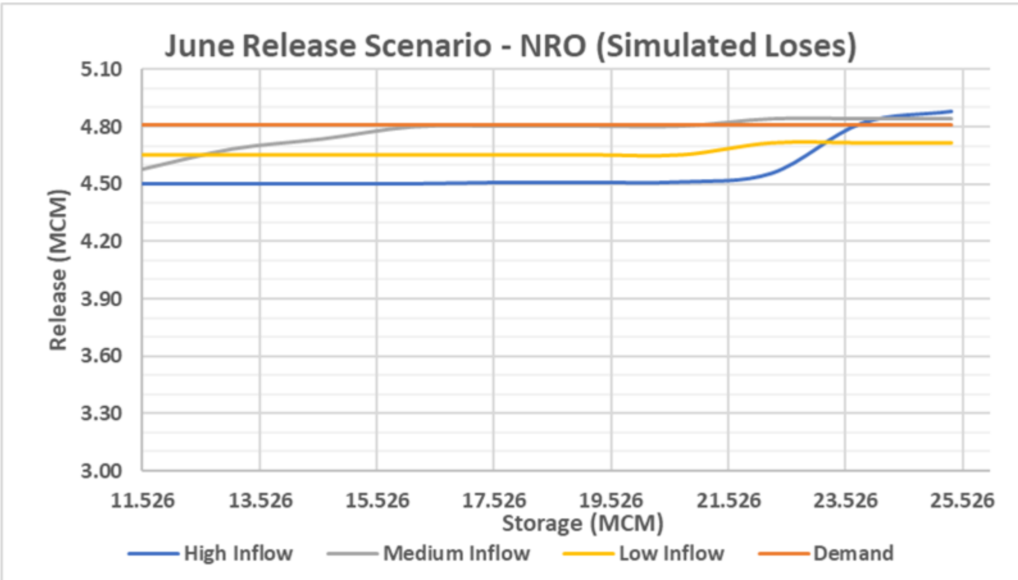
(d)


April Release Scenario - NRO(Simulated Loses)
5.55
5.50
5.45
5.40
5.35
5.30
5.25
5.20
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

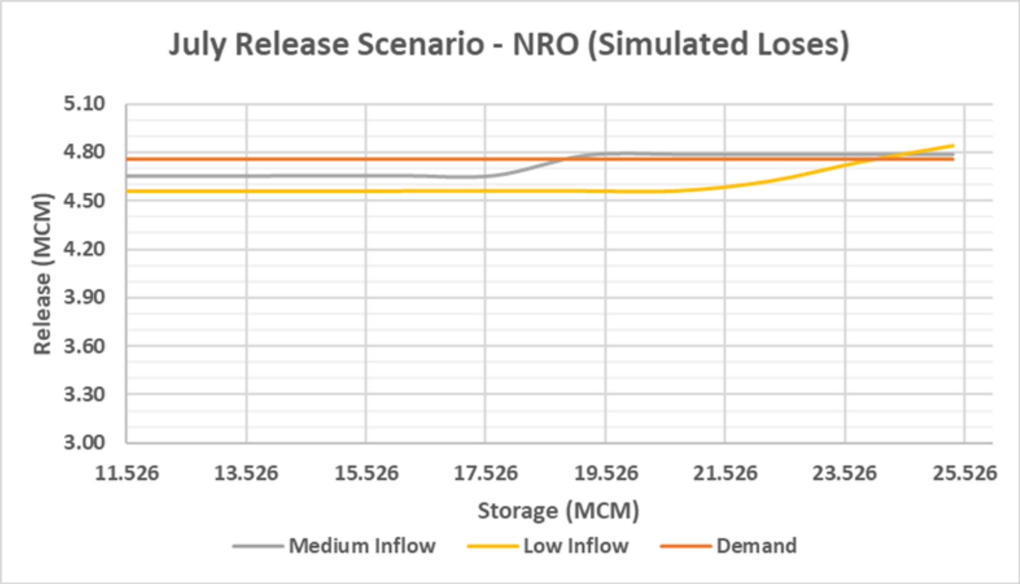
(e)


May Release Scenario - NRO (Simulated Loses)
5.84
5.82
5.80
5.78
5.76
5.74
5.72
5.70
5.68
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

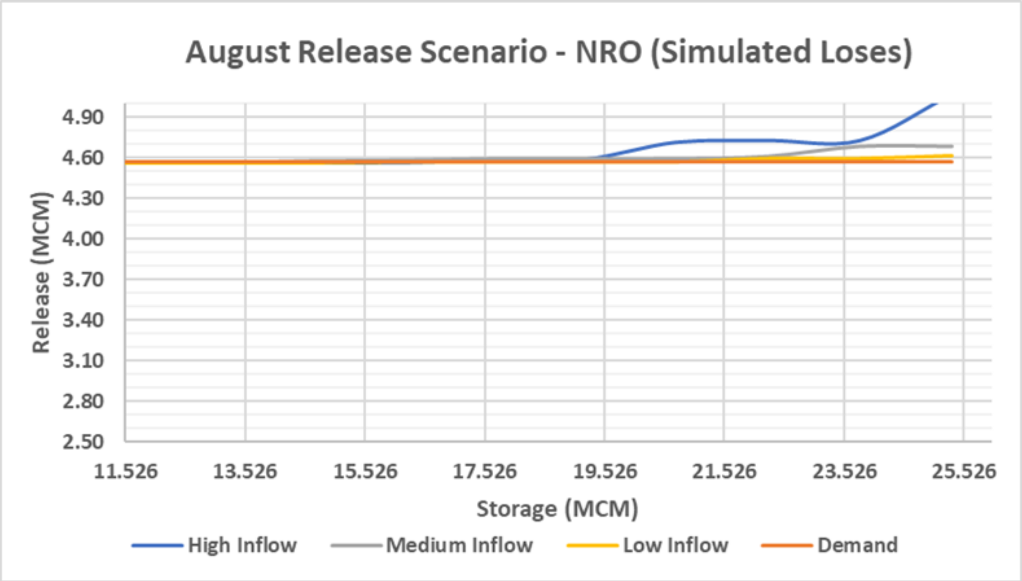
(f)


June Release Scenario - NRO (Simulated Loses)
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

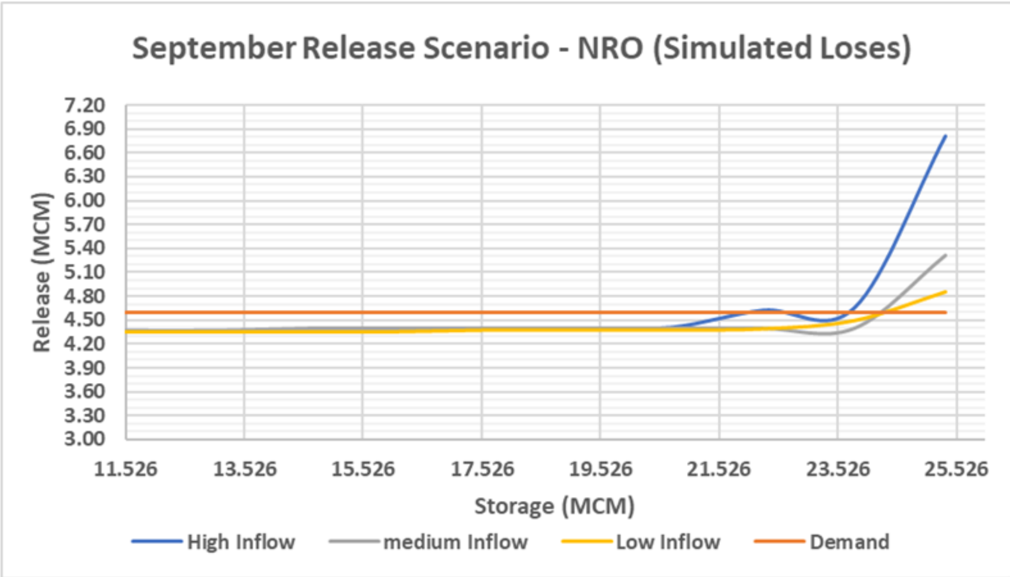
(g)


July Release Scenario - NRO (Simulated Loses)
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

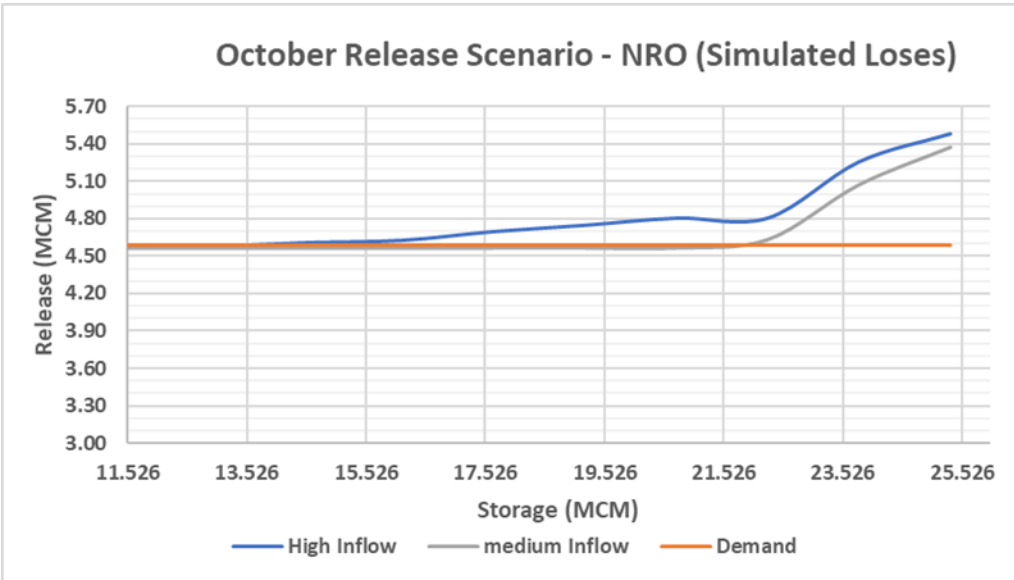
(h)


August Release Scenario - NRO (Simulated Loses)
4.90
4.60
4.30
4.00
3.70
3.40
3.10
2.80
2.50
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

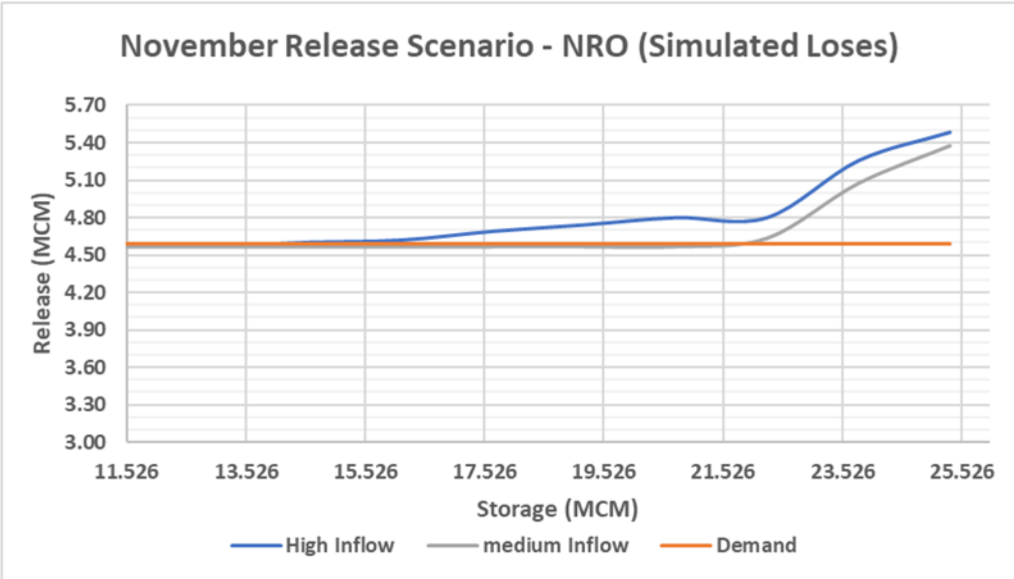
(i)


September Release Scenario - NRO (Simulated Loses)
Release (MCM)
7.20
6.90
6.60
6.30
6.00
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Low Inflow
Demand

(j)


October Release Scenario - NRO (Simulated Loses)
Release (MCM)
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Demand

(k)



(l)

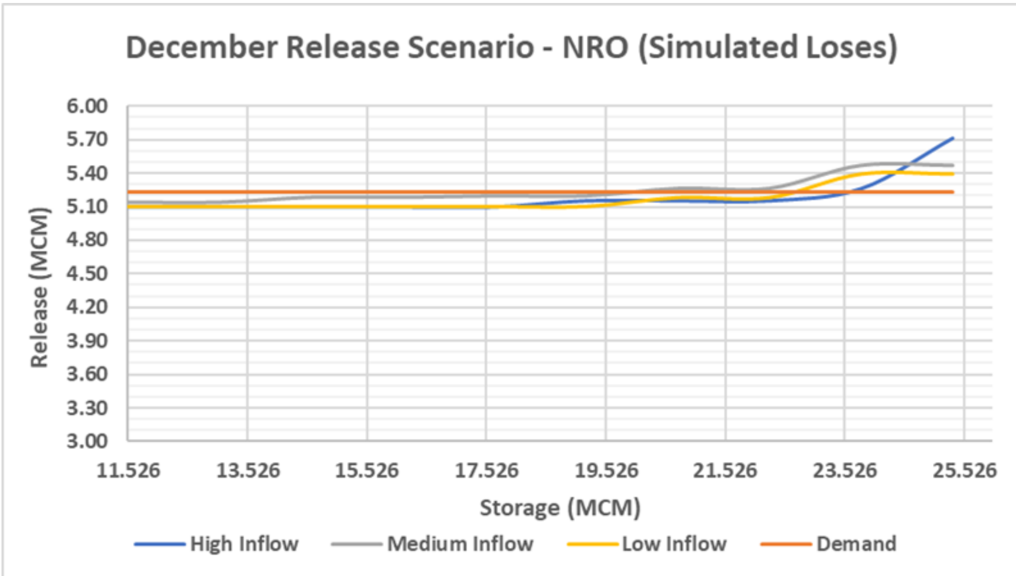


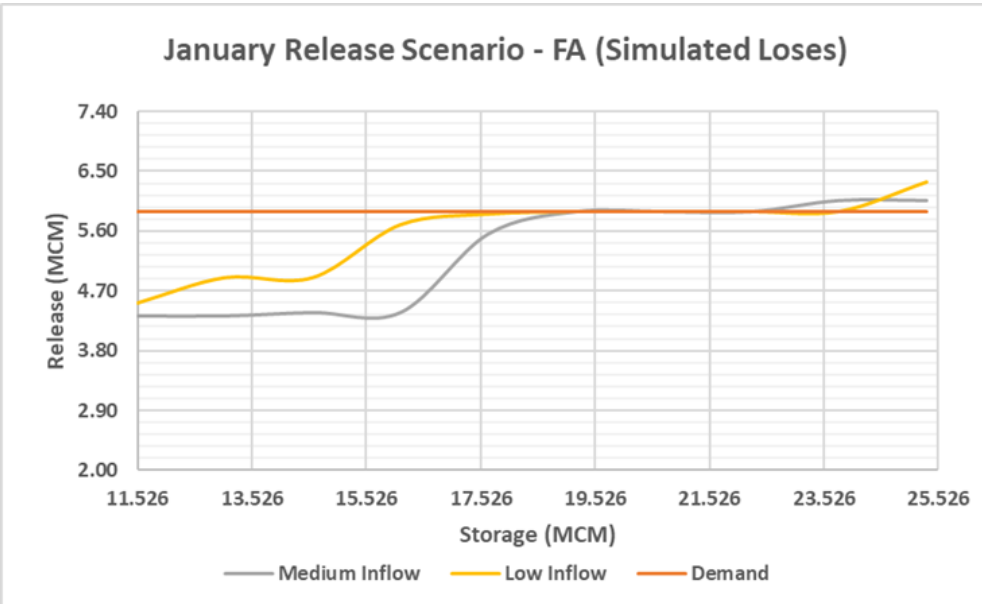
Figure 4.13: NRO Release Curves for Integrated System (IS) for (a) Jan - (l) Dec

4.5.4

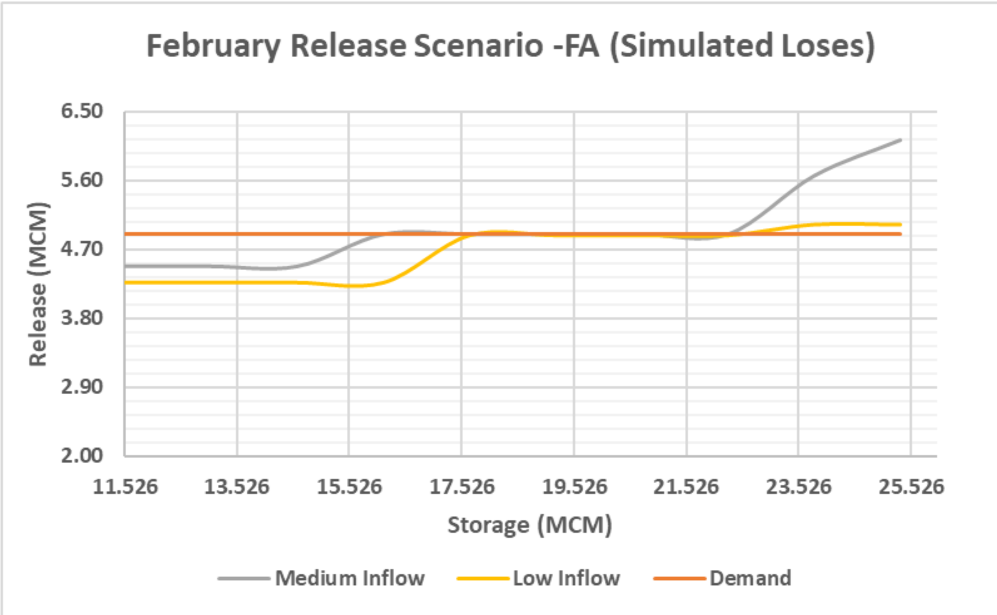
4.5.5 Firefly Optimisation (FA)

It is clear from looking at all of the graphs that were produced by FA in Figure 4.14 that the model once again proves its powerful optimisation skill since it was able to satisfy the demand every month despite the fact that the inflow was relatively modest in both the Integrated and Conventional System. Since firefly lacks velocity in addition to various other variables, it may be able to avoid reaching local maxima. At the conclusion of each cycle, the output will consist of the solution that is deemed to be the most appropriate. After making use of these outputs as its initial population input, the acquired population is then enhanced and promoted by FA throughout each iteration of the process.

(a)

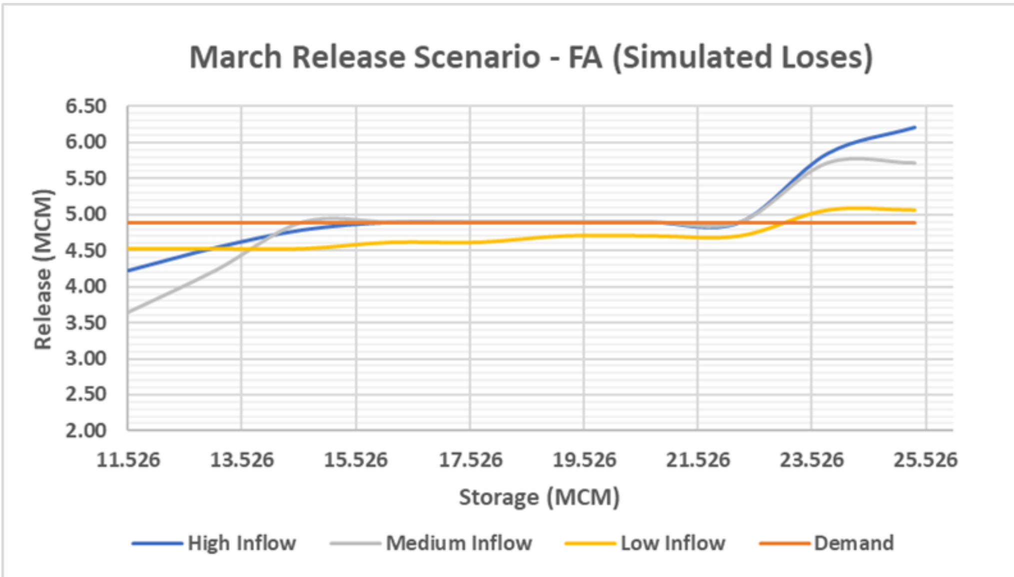


(b)



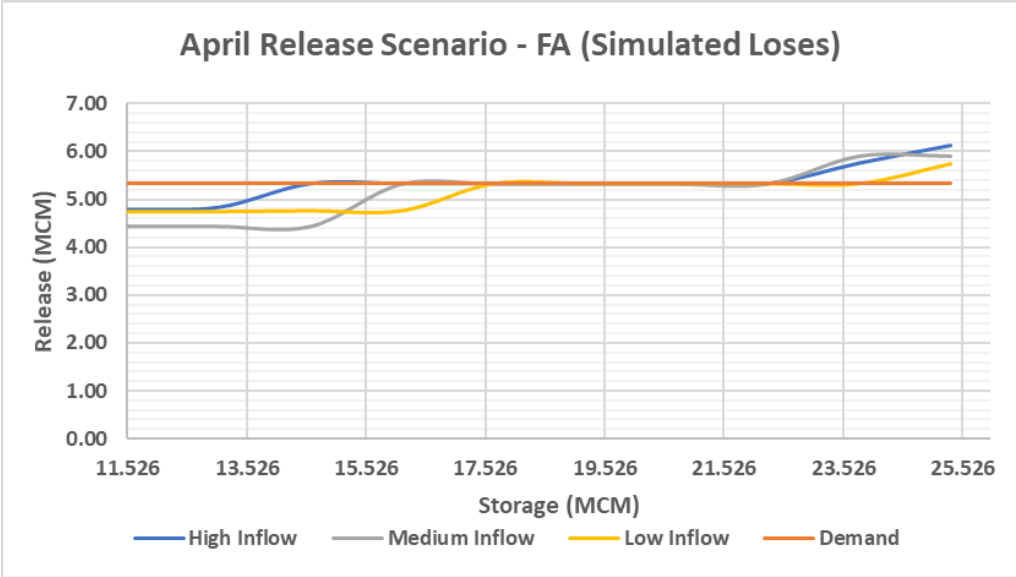
February Release Scenario -FA (Simulated Loses)
6.50
5.60
4.70
3.80
2.90
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

(c)

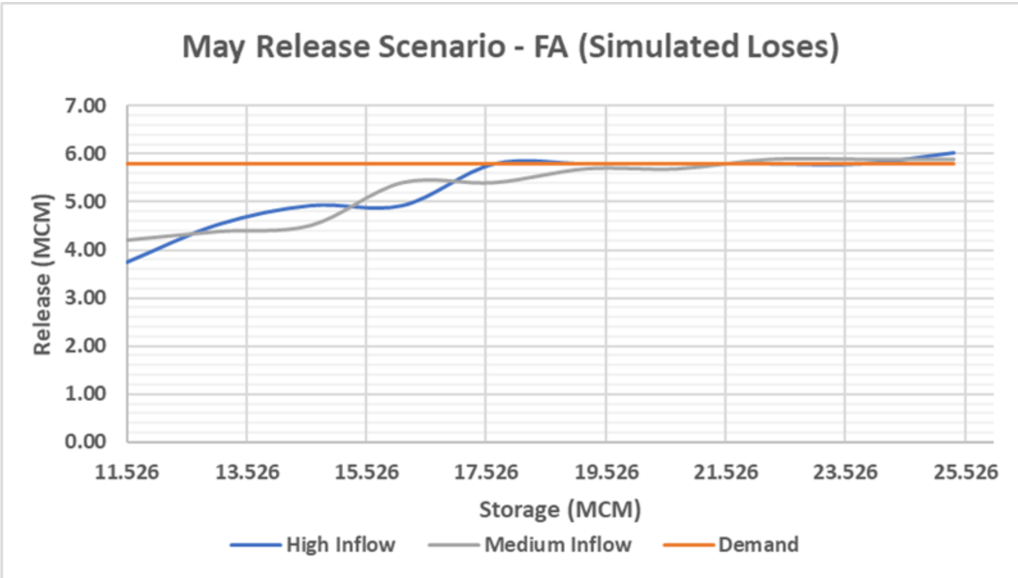


March Release Scenario - FA (Simulated Loses)
6.50
6.00
5.50
5.00
4.50
4.00
3.50
3.00
2.50
2.00
Release (MCM)
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

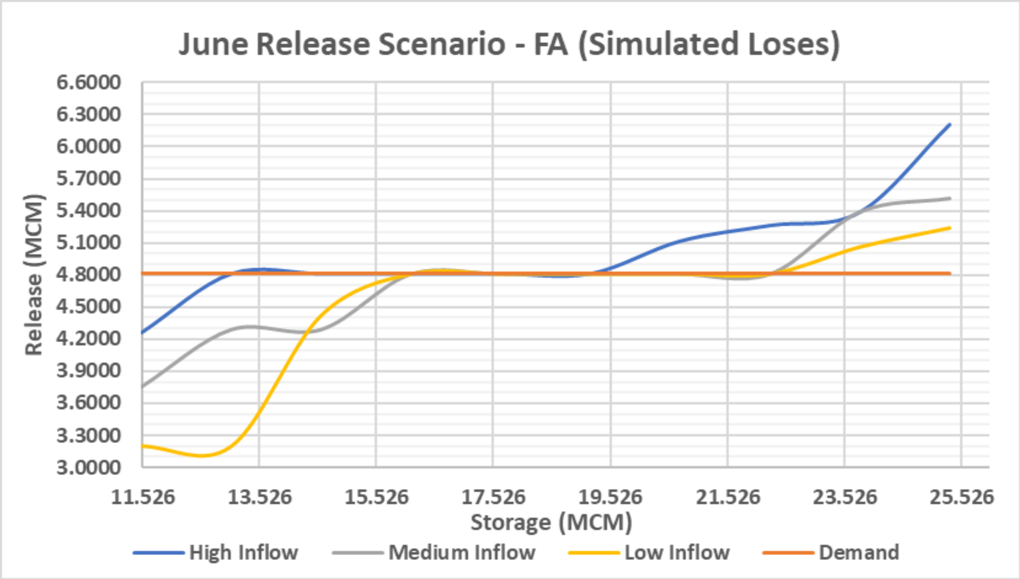
(d)


April Release Scenario - FA (Simulated Loses)
Release (MCM)
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

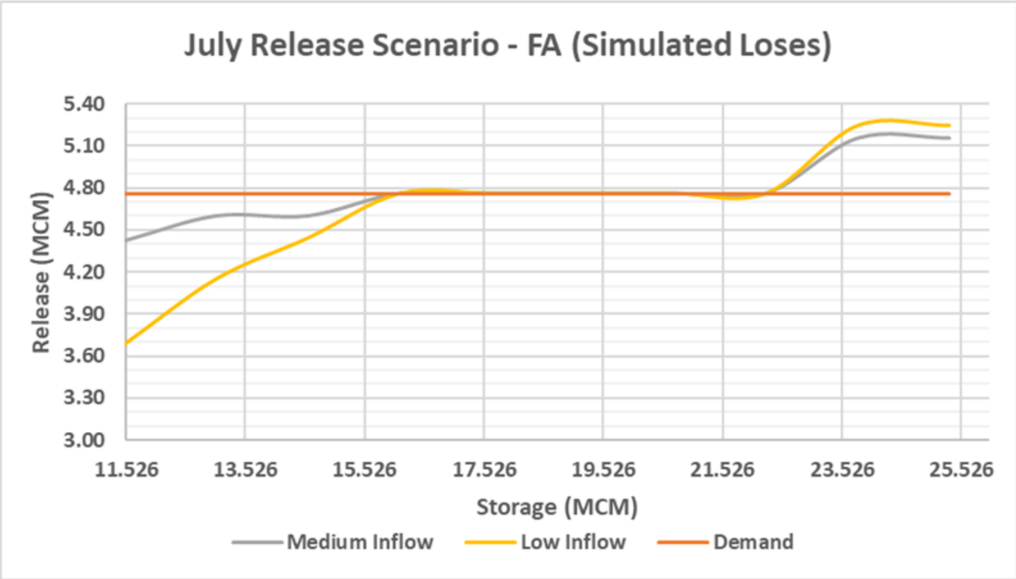
(e)


May Release Scenario - FA (Simulated Loses)
Release (MCM)
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Demand

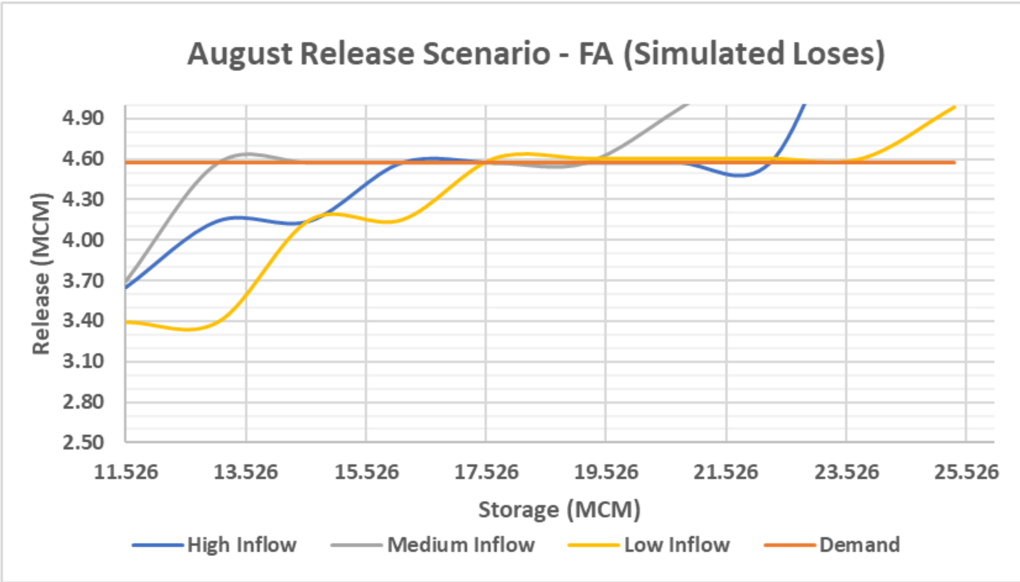
(f)


June Release Scenario - FA (Simulated Loses)
Release (MCM)
6.6000
6.3000
6.0000
5.7000
5.4000
5.1000
4.8000
4.5000
4.2000
3.9000
3.6000
3.3000
3.0000
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

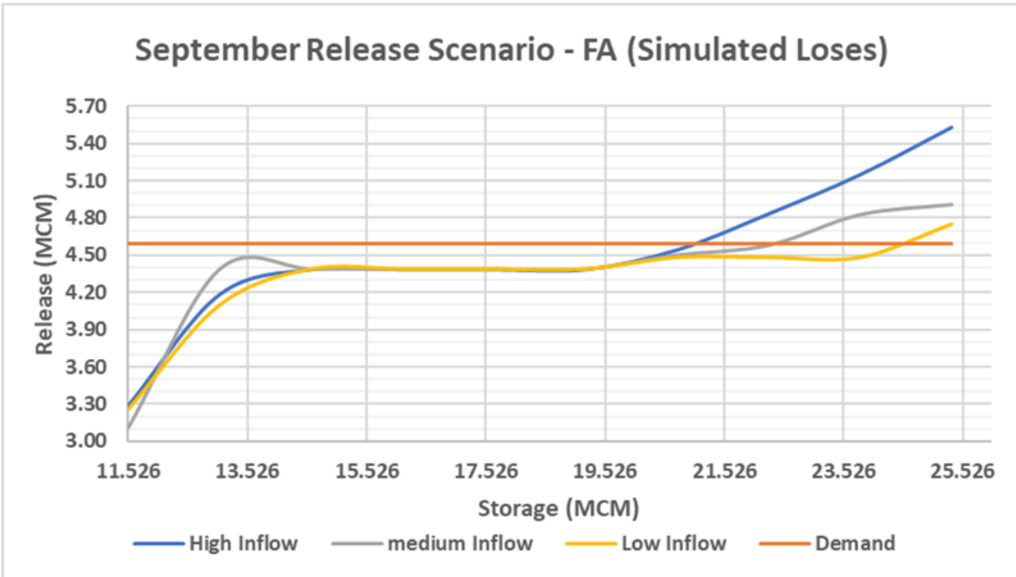
(g)


July Release Scenario - FA (Simulated Loses)
Release (MCM)
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
Medium Inflow
Low Inflow
Demand

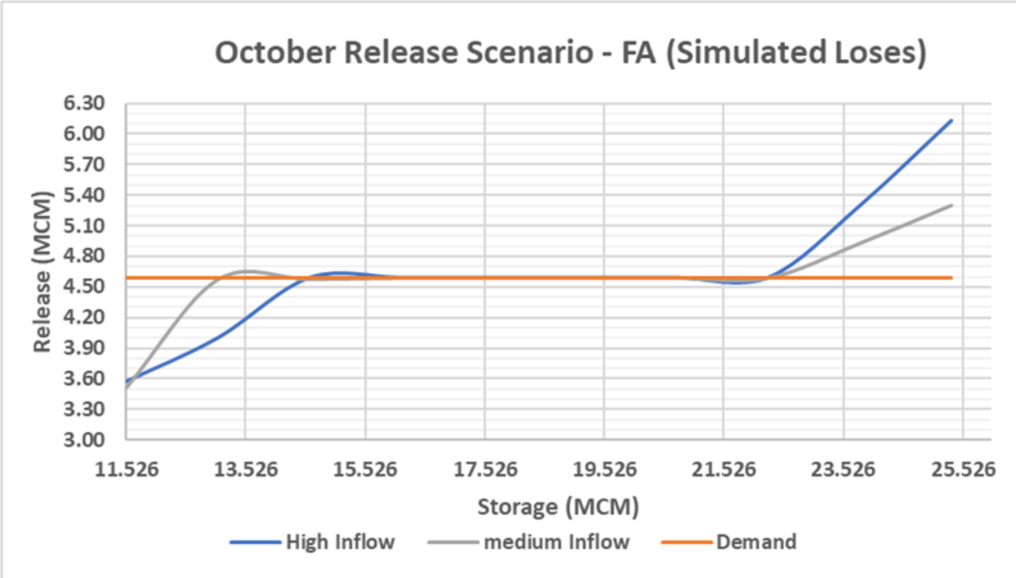
(h)


August Release Scenario - FA (Simulated Loses)
Release (MCM)
4.90
4.60
4.30
4.00
3.70
3.40
3.10
2.80
2.50
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
Medium Inflow
Low Inflow
Demand

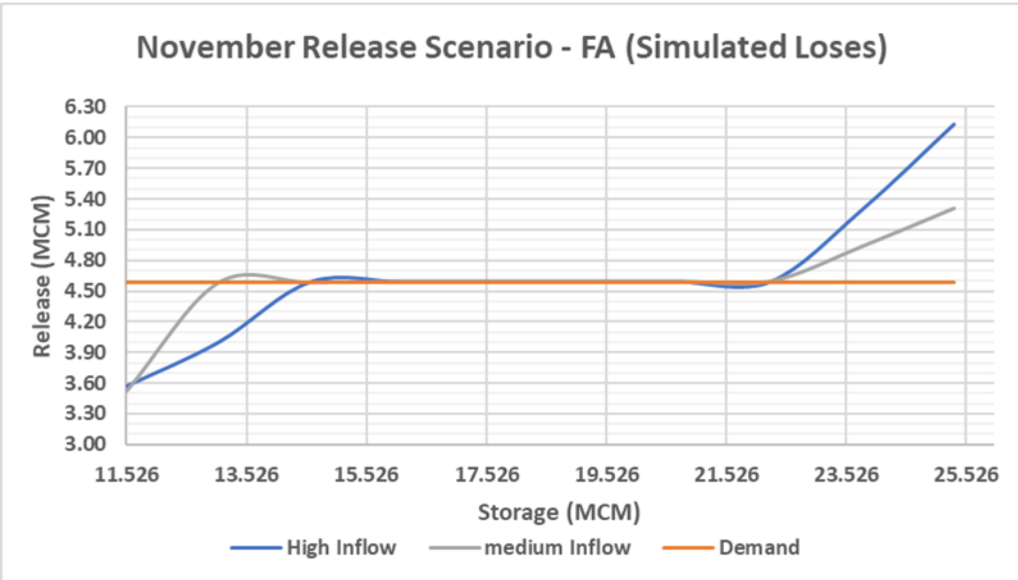
(i)


September Release Scenario - FA (Simulated Loses)
Release (MCM)
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Low Inflow
Demand

(j)


October Release Scenario - FA (Simulated Loses)
Release (MCM)
6.30
6.00
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Demand

(k)


November Release Scenario - FA (Simulated Loses)
Release (MCM)
6.30
6.00
5.70
5.40
5.10
4.80
4.50
4.20
3.90
3.60
3.30
3.00
11.526
13.526
15.526
17.526
19.526
21.526
23.526
25.526
Storage (MCM)
High Inflow
medium Inflow
Demand

(l)

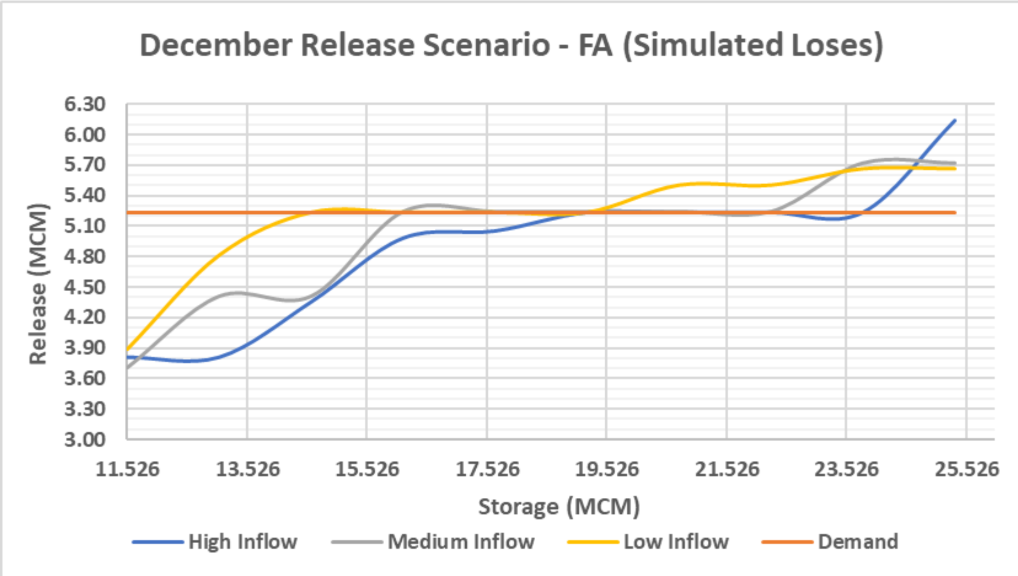


Figure 4.14: FA Release Curves for Integrated System (IS) for (a) Jan – (l) Dec

4.6 Reservoir Performance Assessment

During the process of analysing the proposed optimisation algorithms, several distinct performance statistical metrics were tested and evaluated. In addition, during this investigation, two distinct methods have been taken into consideration.

The first approach, known as the standard approach, considers the availability of precise forecasting for the reservoir's intake. The second method incorporates the anticipated model accuracy for the reservoir's intake throughout the simulation period. This method is based on the second technique. In order to evaluate how well the operational rules perform when the reservoir system is being operated, the following novel technique, which involves running the reservoir system under situations as close to real life as possible, is advised. In this Section, four indices were used to assist the metaheuristic algorithm models: Reliability, Resilience, Vulnerability and finally sustainability index which indicates how sustainable the produced release curves are.

4.6.1 Reliability Index (Rel)

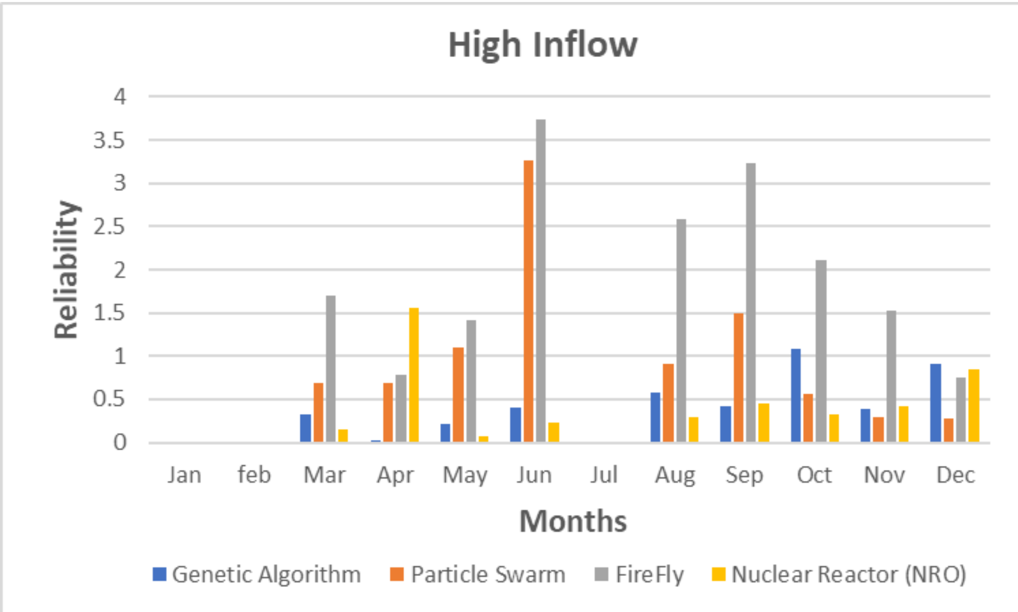
In the reservoir management approach, many widely used statistical performance measures are utilised in conjunction with a reliability index to assess the efficiency of the release operating strategy. This is done so in order to determine whether or not the plan is successful. These criteria are derived from the extensive range of potential scenarios that may arise throughout the course of its operational lifetime. The term "reliability" in reference to a reservoir relates to the measuring of an output from an activity.

According to Figure 4.15, all four models' reservoir releases have been evaluated to see how accurate the output of each model is in each of the three inflow conditions. The results of these tests are shown in the given Figure. Each model needed to operate based on two distinct systems in order to optimise the reservoir's functioning.

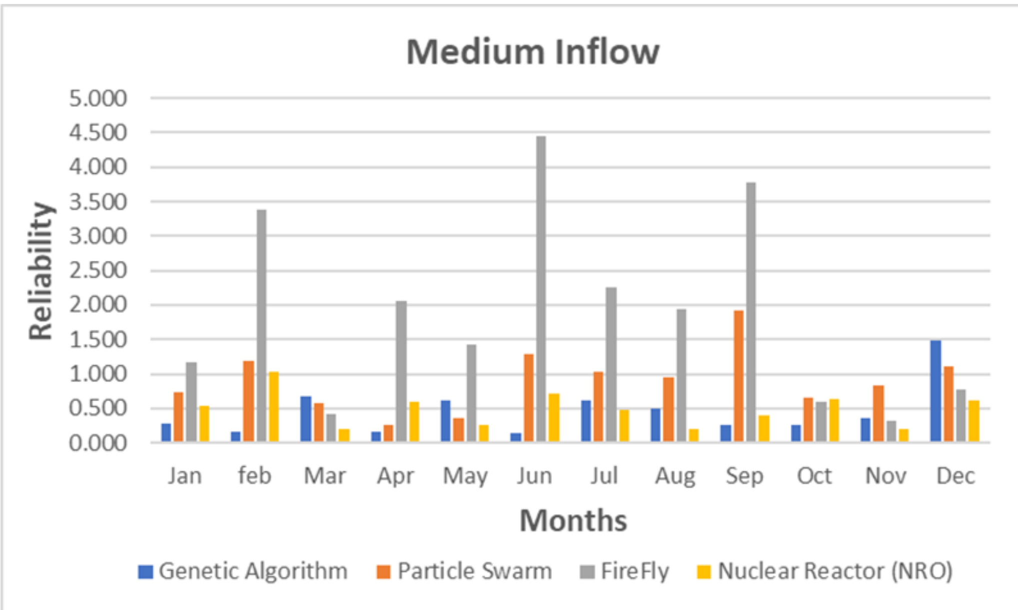
The first one was called the conventional system (cs), and in it, the reservoir losses were not recreated by any machine learning model; instead, they were left as historical data.

This system was used. The second system is known as an integrated system (IS), and it is the one in which the optimisation model combines the losses that were simulated by XG-Boost while simultaneously optimising the reservoir operation.

(a)



(b)



(c)

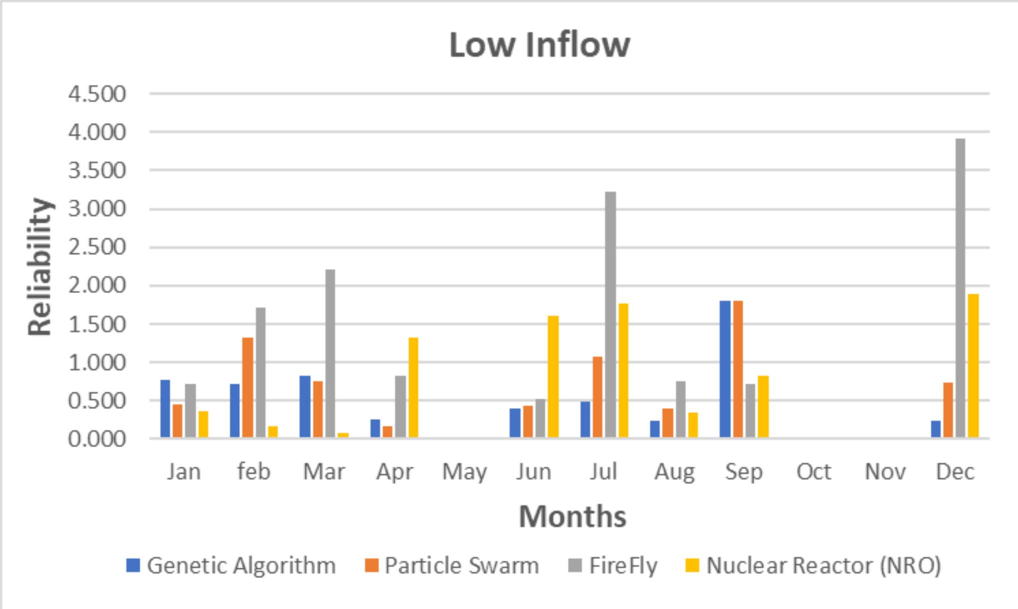


Figure 4.15: Reliability Indices for (a) High, (b) Medium and (c) Low

Figure 4.15 demonstrates that the firefly algorithm is a superior model because the release values that it generates are more reliable than the release values produced by any of the other models in the conventional and the integrated loss system. The firefly algorithm achieved its highest reliable score of 3.6 and 4.4 in June for high and medium inflows, respectively, with a slight disparity between the scores achieved by the integrated system and the conventional system. Except for the months of April and December, Firefly was able to get points for good dependability in every month when there was substantial inflow.

However, when the reservoir had a medium inflow, fireflies achieved the most ideal outcomes virtually every month, with the exception of March and December, when their scores were below 1 in both systems. When the water volume received in the reservoir catchment was low, the firefly model was still able to produce excellent results, which allowed it to outperform all of the other models in nearly every month, with the exception of April and September, when it scored a reliability point that was less than 1.0.

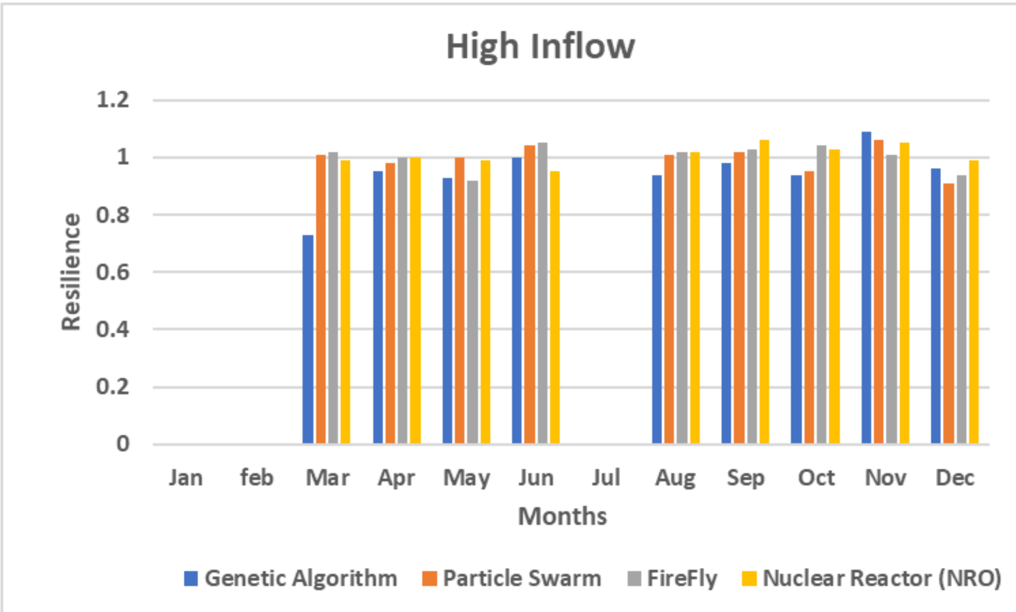
Out of the four algorithms, it can be said that the firefly algorithm is the one that is the most trustworthy. This may be stated in a nutshell. In addition to this, the integrated losses that are simulated by XG-Boost considerably increase the model's dependability.

4.6.2 Resiliency Index (Res)

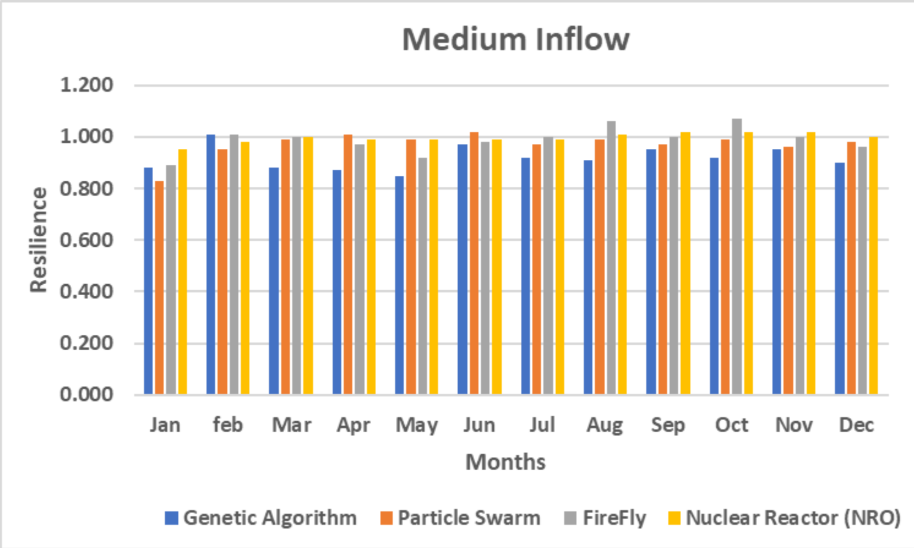
The resilience index is a measurement that determines how fast a system can recover from a failure or a deficit period when it is faced with either situation. The failure sequence will have an effect on the performance of the system; for instance, if a 12-month operation period fails over four periods, the performance will be reduced. To put it another way, a failure in one period is always followed by a period in which there is no failure: but a failure in four times in a row results in a different conclusion.

Figure 4.16 demonstrates that the release curve that was generated by Firefly performed better than the other order rule curves, achieving a low resilience index overall. This was the case despite the fact that the Nuclear Reaction algorithm achieved a low resilient score in the month of March, April, May, June, August, Sept, Oct and November during periods of high inflow only. However, firefly still scores low resilient score in medium and low inflow. In fact, the resilience index was not a clear performance as it did not provide us with a straightforward ranking; hence vulnerability index is needed in this case.

(a)



(b)



Medium Inflow
1.200
1.000
0.800
0.600
0.400
0.200
0.000
Resilience
Jan
feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months
Genetic Algorithm
Particle Swarm
FireFly
Nuclear Reactor (NRO)

(c)

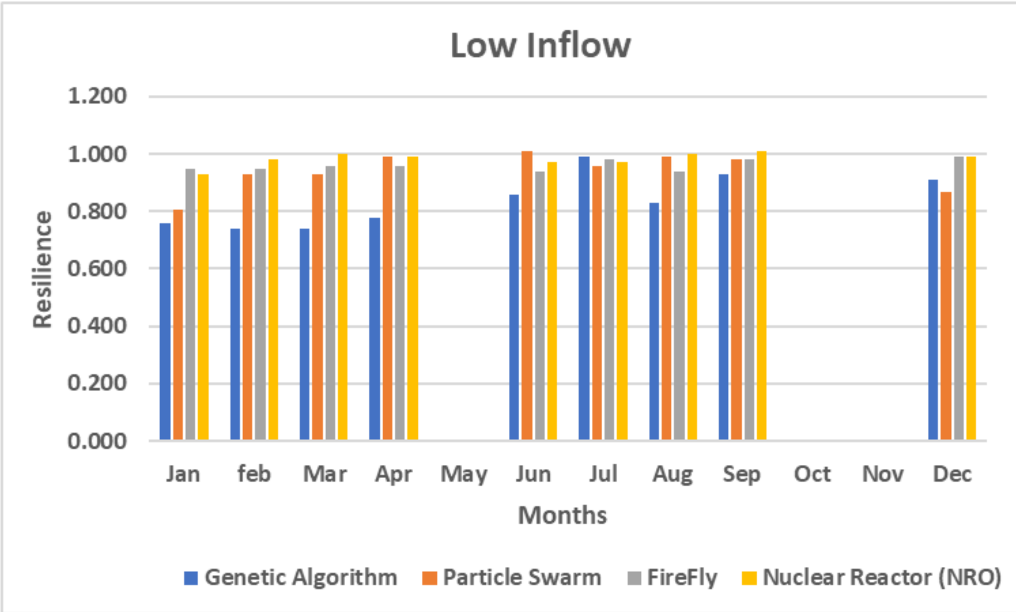


Figure 4.16: Resilience Index for (a) High, (b) Medium and (c) Low

4.6.3 Vulnerability Index

In this context, vulnerability refers to the potential severity of a failure if it does take place. Even when there is a low likelihood of anything going wrong, it is still essential to consider the worst-case scenario and plan accordingly. Efforts to improve a system's dependability include attempting to break it in some way so that it cannot function properly.

Free yet, there are not many systems that can be constructed so big or have so many backups that failures are never an option. After a certain point, it is better to expend effort making the implications of failure less severe and more acceptable than it is to attempt to eradicate the chance of failure. When floods do occur, it is possible to lessen the financial toll they take by taking precautions such as flood-proofing buildings, investing in early warning systems, and obtaining flood insurance. Other measures that may be taken to reduce the financial impact of flooding include keeping structures away from floodplains and repurposing areas at risk of flooding as agricultural land, natural preserves, and parks.

Using a method that was susceptible to error, the performance of the models was analysed, and the results exposed the severe shortage scenario. According to the data, firefly optimisation gave much better results than monthly optimisation did in terms of vulnerability index performance (lower value). To put it another way, the firefly optimisation was effective in determining the worst failure's worst magnitude. This unequivocally illustrates that the Firefly optimisation outperformed in terms of vulnerability, which ultimately led to developing a superior release plan for efficient operation, as seen in Figure 4.17.

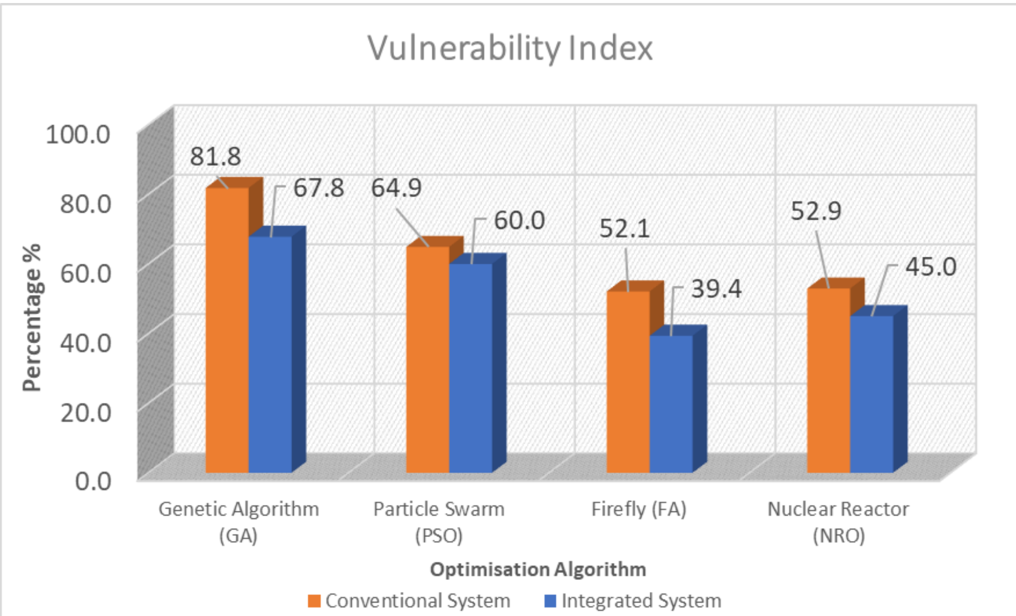


Figure 4.17: Vulnerability Index

Comparing the results of the traditional system, which got 50%, with those of the firefly model, which got 39% overall, it can be observed that the firefly model integrated with reservoir losses was the model with the least amount of vulnerability overall. In contrast, the Genetic Algorithm was the model with the highest level of vulnerability among the integrated and conventional systems.

This was due to the fact that the vulnerability percentage obtained for the integrated system was approximately 65%, whereas the vulnerability percentage obtained for the conventional system was approximately 80%. Overall, it is possible to draw the conclusion that adding reservoir losses into optimisation models has enhanced them, making them less fragile and leading to the acquisition of more accurate release curves.

4.6.4 Sustainability Index (SI)

Using the SI outlined in this work, one is able to evaluate the degree to which different water management practices are environmentally friendly. The SI defines strategies that, in the future, will either maintain or enhance the basin's intended features for its water management. When there are trade-offs between several performance criteria such as the one seen in this study as the performance of each model varies in terms of resiliency and reliability, the SI makes it much simpler to analyse, compare, and create adaptive solutions that will enhance water management.

The scientific community has relied on the SI because it concisely outlines key performance characteristics of water management while excluding irrelevant or irrelevantly complex components (Ye et al., 2018; de O. Vieira and Sandoval-Solis, 2018; Sandoval-Solis et al., 2011). In order to give a comparison between the four models, the average performance that was achieved from each category of inflow for each model was computed, and the results are displayed in Table 4.9.

Table 4.9: Sustainability Index for Optimisation Models

Model	Conventional System (CS)			Integrated System (IS)		
	Inflow					
	High	Medium	Low	High	Medium	Low
GA	0.057	0.071	0.111	0.139	0.153	0.193
PSO	0.123	0.113	0.072	0.205	0.195	0.154
FA	0.241	0.331	0.207	0.323	0.413	0.289
NRO	0.045	0.06	0.128	0.127	0.142	0.210

When the reservoir losses were incorporated into the system, the capabilities of the model were enhanced overall. Additionally, this gave the system additional credibility when optimising the operation at the end of each month and syncing it with the beginning of the following month.

The Firefly model, which was developed by integrating the reservoir losses, demonstrates how it is sustainable in handling any inflow changes that may occur as it obtained a scoring value of 0.413, 0.323, and 0.289 for medium, high and low inflow respectively. Furthermore, it demonstrates that it is capable of being used as sustainable tool for effectively optimising reservoir release, as it has obtained the highest average points. The PSO achieves a score of 0.205, 0.195, and 0.154, respectively, for high, medium, and low inflows, which places it in the second-place position in the ranking. The Genetic Algorithm (GA) has once again shown superior performance in contrast to the Nuclear Reaction (NRO), which has fallen to the very bottom of the rating.

In a nutshell, the results presented in this research, which sought to determine the optimal optimisation algorithm integrated with machine learning model, show that the particle swarm algorithm frequently exhibits better performance than the genetic algorithms and nuclear reaction optimisation. The innovative firefly algorithm, on the other hand, outperforms the PSO, GA and NRO in terms of efficacy and the speed at which the desired results are obtained. This shows that the firefly method has the potential to be excellent at solving difficult complex problems.

CHAPTER 5

CONCLUSIONS AND FUTURE RECOMMENDATIONS

5.1 Conclusions

The study presented in this thesis focused on developing a credible solution to the issue of predicting reservoir inflow at a dam in Peninsular Malaysia while getting a precise and trustworthy reservoir operating system. Empirical measurements and actual inflow computations are now faced with significant obstacles due to several factors, such as cost, technological viability, and the required prerequisite trained labour. Positive results were obtained when the linear water-balance equation and black-box machine learning models were intercepted, but they were both qualitatively and quantitatively data hungry.

The research described in this thesis intended to construct a reliable machine learning model for reservoir inflow forecasting in Peninsular Malaysia with minimal data needs to optimise, simulate, and was assessed at the Klang gate Dam, the oldest water dam in Malaysia. The strategy that was used to accomplish this employed the Multi-layer perceptron neural network (MLPNN), Support vector regression (SVR), Adaptive neuro fuzzy inference system (ANFIS) and the extreme gradient boosting (XG-BOOST) as the four machine learning models to forecast and simulate the reservoir inflow and operation. Additionally, there are four metaheuristic algorithms known as the genetic algorithm (GA), particle swarm optimisation algorithm (PSO), firefly optimisation algorithm (FA) and the nuclear reaction optimisation algorithm (NRO) for the purpose of optimising the reservoir operations to generate an optimised release curve.

Four objectives drove the undertaking of this extensive research, the findings from each objective are given as follows. The most important conclusions drawn from Objective-1 are as follows:

- As much as the hyper-parameters of each machine learning model plays a critical role, there is no standard method to obtain an optimal combination among the model parameters and the trial-and-error along with the grid-search and cross-validation are for now, the best available strategies to search for the optimum parameters.
- MLPNN is a super-rich parametric model and obtaining the best combination of parameters is a challenging process. SVR is sensitive to the number of folds in the cross-validation stage and XG-Boost is the easiest model to be tuned, while ANFIS requires a huge amount of training data to obtain positive outcomes.
- Rainfall is the most important input variable when forecasting reservoir inflow as the highest volume of water that flows into the reservoir is contributed by rainfall.
- Time-lags are necessary when forecasting inflow as it expands the forecasting search space, therefore providing a wider range of forecasted inflow.

- XG-BOOST forecasting abilities are superior to the other three models as it has obtained an R^2 value of 0.7220, while MLPNN is a good competitor to XG-BOOST scoring an R^2 value of 0.5392 which was superior to SVR which had an R^2 value of 0.4247 and ANFIS with an R^2 of 0.4240.

Optimising reservoir operation was never an easy task due to the involvement of many unknown factors and the thousands, if not millions, of expected release scenarios. Nevertheless, researchers were constantly seeking ways to solve such tasks. The second objective of this research was to optimise reservoir operation and obtain a satisfactory resolution hence the main conclusion points that may be drawn from objective-2 are as below:

- Real coded Genetic Algorithm is a super-simple and user-friendly algorithm as it does not require intensive parameter tuning. PSO and FA are totally opposite to GA as they have various hyper-parameters.
- NRO is a newly developed novel algorithm, and it is yet to be tested in real case studies, especially in environmental and hydrology fields.
- FA was able to optimise the reservoir performance nearly obtaining a perfect operation with minimal or zero water deficit across all twelve months.
- PSO avoids falling in local optima and consistently searching for the global optima; hence the release output differs in each iteration. While NRO gets trapped in the local minima assuming the best release scenario has been achieved therefore the release curves produced by NRO are mainly linear with minimal increments.

In objective-3, it is noteworthy that the overall and total deficit capacities for each optimisation model changed when the reservoir was managed using the predicted inflow data or under more actual conditions. It is helpful to obtain the real behaviour of the optimisation in managing a reservoir system if the reservoir inflow forecasting is integrated alongside the simulation of reservoir losses throughout the duration of operating a reservoir system. The third objective achieved by this study was to simulate the reservoir operation by integrating the losses in the form of evaporation and adding the forecasted inflow as an input hence developing an autonomic/semi-real-time reservoir operating system.

Finally, even though each objective was presented separately;' all the objectives are however, interrelated. Hence, the performance of the reservoir that was optimised in objective-3 requires an assessment to assess the performance and thus therefore objective-4 is necessary. In objective-4, the risk analysis assessment has been conducted on the reservoir by testing each model output which reflects the reservoir condition based on the reliability, resiliency, vulnerability, and sustainability index. It is important to note that there was some variation among the results. Hence, it was challenging to identify the best model but the sustainability index which considers all the other three analytical tests could wrap up the results by ranking firefly as the most sustainable algorithm followed by particle swarm, Genetic algorithm and finally the nuclear reaction optimisation.

In a nutshell, the strategies that were provided in our research and the optimisation system are generally applicable to different case studies, and they are flexible enough to allow for a specialized design to address any specific water resources planning issue. All the forecasting-based model and the objective function are robust and flexible for any alterations. Furthermore, this research study has proffered a solution to the dilemma of forecasting an accurate reservoir inflow and optimising the reservoir operation, by developing an autonomic computing program that runs on integrating the forecasted inflow from the machine learning model along with the optimisation algorithm to generate an optimal release scenario.

5.2 Future Recommendations

Despite the fact that the forecasting models used in this study have a high level of accuracy, the work that is being presented here does have certain shortcomings that might be remedied by a series of follow-up actions. Due to the reservoir data's profoundly stochastic nature, certain processes necessitate supplementary improvements through pre-processing methods such as the Wavelet Algorithm (Rahman et al., 2020). The wavelet function decompresses time series data into sub-components which provide more information at different resolution levels, that results in improving the accuracy of forecasting models. Although the suggested Algorithms performed admirably in determining the ideal operation strategy for the hydropower production, more research into alternative optimisation algorithms is necessary since they may provide an even more effective operation Guideline.

As a future recommendation, it is suggested that downstream demand shall be treated as an unpredictable variable rather than a fixed variable throughout the month to further provide robustness to the research field by updating the objective function hence resulting in a broader sets of output which does not just provide a complete picture on the reservoir operation but a detailed information on how each variable contributes to reservoir operation.

In addition, reservoir operations could be tested in real-time with optimisation of operations reported at the time rather than relying on historical data. Other variations of each algorithm, such as the Self-organizing Hierarchical PSO with Time-Varying Acceleration Coefficients (HPSO TVAC) and the Gaussian Distribution Firefly and other hybrid versions of the GA, should be tested for comparative purposes.

REFERENCES

Ababaei, B., Mirzaei, F., Sohrabi, T. and Araghinejad, S., 2013. Reservoir Daily Inflow Simulation Using Data Fusion Method. *Irrigation and Drainage*, 62(4), pp.468–476.

Abdul-Rahman, O.A., Munetomo, M. and Akama, K., 2013. An adaptive parameter binary-real coded genetic algorithm for constraint optimisation problems: Performance analysis and estimation of optimal control parameters. *Information Sciences*, 233, pp.54–86. Available at: http://dx.doi.org/10.1016/j.ins.2013.01.005.

Adil Hashmi, A.H., 2013. Firefly Algorithm for Unconstrained Optimisation. *IOSR Journal of Computer Engineering*, 11(1), pp.75–78. Available at: http://www.iosrjournals.org/iosr-jce/papers/Vol11-issue1/M01117578.pdf.

Ahmad, S.K. and Hossain, F., 2019. A generic data-driven technique for forecasting of reservoir inflow: Application for hydropower maximization. *Environmental Modelling and Software*, 119(June), pp.147–165. Available at: https://doi.org/10.1016/j.envsoft.2019.06.008.

Ahmed, E.-S.N. and Amr, E.-S., 2019. Daily Forecasting of Dam Water Levels Using Machine Learning. *International Journal of Civil Engineering and Technology*, 10(6), pp.314–323. Available at: http://www.iaeme.com/IJCIET/index.asp314http://www.iaeme.com/ijciet/issues.asp?JType=IJCIET&VType=10&IType=6http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=10&IType=6.

Ahsan Md, Mahmud, M., Saha, Pritom, Gupta, Kishor, Siddique, Zahed 2021. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. Technologies, 9(3), p.52. Available at: https://www.mdpi.com/2227-7080/9/3/52.

Akbari, M. and Afshar, A., 2014. Similarity-based error prediction approach for real-time inflow forecasting. *Hydrology Research*, 45(4–5), pp.589–602. Available at: https://iwaponline.com/hr/article/45/4-5/589/1258/Similaritybased-error-prediction-approach-for.

Akopov, A.S., Beklaryan, L.A., Thakur, M. and Verma, B.D., 2019. Parallel multi-agent real-coded genetic algorithm for large-scale black-box single-objective optimisation. *Knowledge-Based Systems*, 174, pp.103–122. Available at: https://doi.org/10.1016/j.knosys.2019.03.003.

Al-Aqeeli, Y.H. and Mahmood Agha, O.M.A., 2020. Optimal Operation of Multi-reservoir System for Hydropower Production Using Particle Swarm Optimisation Algorithm. *Water Resources Management*, 34(10), pp.3099–3112. Available at: https://link.springer.com/10.1007/s11269-020-02583-8.

Allawi, M.F., Jaafar, O., Mohamad Hamzah, F., Mohd, N.S. Nuruol Syuhadaa Mohd & Ravinesh C. Deo& Ahmed El-Shafie, 2018. Reservoir inflow forecasting with a modified coactive neuro-fuzzy inference system: a case study for a semi-arid region. *Theoretical and Applied Climatology*, 134(1–2), pp.545–563.

Allawi, M.F., Jaafar, O., Mohamad Hamzah, F., Abdullah, S.M.S, 2018. Review on applications of artificial intelligence methods for dam and reservoir-hydro-environment models. *Environmental Science and Pollution Research*, 25(14), pp.13446–13469.

Allawi, M.F., Hussain, I.R., Salman, M.I. and El-Shafie, A., 2021. Monthly inflow forecasting utilizing advanced artificial intelligence methods: a case study of Haditha Dam in Iraq. *Stochastic Environmental Research and Risk Assessment*, 35(11), pp.2391–2410. Available at: https://doi.org/10.1007/s00477-021-02052-7.

Allawi, M.F., Jaafar, O., Mohamad Hamzah, F. and El-Shafie, A., 2019. Novel reservoir system simulation procedure for gap minimization between water supply and demand. *Journal of Cleaner Production*, 206, pp.928–943.

Arlot, S. and Lerasle, M., 2012. Choice of V for V-Fold Cross-Validation in Least-Squares Density Estimation. *Journal of Machine Learning Research*, 17(1), pp.1–50. Available at: http://arxiv.org/abs/1210.5830.

Arora, S. and Singh, S., 2019. Butterfly optimisation algorithm: a novel approach for global optimisation. *Soft Computing*, 23(3), pp.715–734. Available at: https://doi.org/10.1007/s00500-018-3102-4.

Babel, Mukand Shrestha, Sangam, Kawasaki, Akiyuki, Tripathi, Nitin, 2016. Assessment of Climate Change Impact on Reservoir Inflows Using Multi Climate-Models under RCPs—The Case of Mangla Dam in Pakistan. Water, 8(9), p.389. Available at: http://www.mdpi.com/2073-4441/8/9/389.

Bahar, A.F., Yusop, Z., Alias, N.E. and Ramli, M.W.A., 2021. Influence of Dam to Rainfall-Run-off Response in a Tropical Climate – A Case Study of Selangor River Basin, Malaysia. *IOP Conference Series: Materials Science and Engineering*, 1153(1), p.012004.

Bao, Y., Xiong, T. and Hu, Z., 2014. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing*, 129, pp.482–493. Available at: http://dx.doi.org/10.1016/j.neucom.2013.09.010.

Benmouiza, K. and Cheknane, A., 2019. Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. *Theoretical and Applied Climatology*, 137(1–2), pp.31–43.

Berrar, D., 2018. Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(January 2018), pp.542–545.

Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24(2), pp.123–140. Available at: http://link.springer.com/10.1007/BF00058655.

Chai, T. and Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), pp.1247–1250. Available at: https://gmd.copernicus.org/articles/7/1247/2014/.

Chang, F.-J., Wang, Y.-C. and Tsai, W.-P., 2016. Modelling Intelligent Water Resources Allocation for Multi-users. *Water Resources Management*, 30(4), pp.1395–1413. Available at: http://link.springer.com/10.1007/s11269-016-1229-6.

Chang, L.C., Chang, F.J., Wang, K.W. and Dai, S.Y., 2010. Constrained genetic algorithms for optimizing multi-use reservoir operation. *Journal of Hydrology*, 390(1–2), pp.66–74. Available at: http://dx.doi.org/10.1016/j.jhydrol.2010.06.031.

Chen, D., Chen, Q., Leon, A.S. and Li, R., 2016. A Genetic Algorithm Parallel Strategy for Optimizing the Operation of Reservoir with Multiple Eco-environmental Objectives. *Water Resources Management*, 30(7), pp.2127–2142. Available at: http://link.springer.com/10.1007/s11269-016-1274-1.

Chen, H. tao, Wang, W. chuan, Chen, X. nan and Qiu, L., 2020. Multi-objective reservoir operation using particle swarm optimisation with adaptive random inertia weights. *Water Science and Engineering*, 13(2), pp.136–144. Available at: https://doi.org/10.1016/j.wse.2020.06.005.

Chen, T. and Guestrin, C., 2016. XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 13 August 2016 ACM, New York, NY, USA, pp. 785–794.

Cheng, Chun-tian, Niu, Wen-jing, Feng, Zhong-kai, Shen, Jian-jian, Chau, Kwok-wing, 2015. Daily Reservoir Run-off Forecasting Method Using Artificial Neural Network Based on Quantum-behaved Particle Swarm Optimisation. Water, 7(12), pp.4232–4246. Available at: http://www.mdpi.com/2073-4441/7/8/4232.

Chitsaz, N. and Banihabib, M.E., 2015. Comparison of Different Multi Criteria Decision-Making Models in Prioritizing Flood Management Alternatives. *Water Resources Management*, 29(8), pp.2503–2525. Available at: http://link.springer.com/10.1007/s11269-015-0954-6.

Cinar, U.K., 2019. Combining Domain Knowledge & Machine Learning: Making Predictions using Boosting Techniques. *Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence*. 26 October 2019 ACM, New York, NY, USA, pp. 9–13.

Çolak, A.B., 2021. An experimental study on the comparative analysis of the effect of the number of data on the error rates of artificial neural networks. *International Journal of Energy Research*, 45(1), pp.478–500. Available at: https://onlinelibrary.wiley.com/doi/10.1002/er.5680.

Črepinšek, M., Liu, S.-H. and Mernik, M., 2013. Exploration and exploitation in evolutionary algorithms: A Survey. *ACM Computing Surveys*, 45(3), pp.1–33. Available at: https://dl.acm.org/doi/10.1145/2480741.2480752.

Dashti Latif, Sarmad, Najah Ahmed, Ali, Sherif, Mohsen, Sefelnasr, AhmedEl-Shafie,, 2021. Reservoir water balance simulation model utilizing machine learning algorithm. Alexandria Engineering Journal, 60(1), pp.1365–1378. Available at: https://doi.org/10.1016/j.aej.2020.10.057.

Dettinger, M., 2011. Climate Change, Atmospheric Rivers, and Floods in California - A Multimodel Analysis of Storm Frequency and Magnitude Changes1. *JAWRA Journal of the American Water Resources Association*, 47(3), pp.514–523. Available at: https://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2011.00546.x.

Damian, Dibie Chidubem, 2019. A Critical Review on Artificial Intelligence Models in Hydrological Forecasting How Reliable are Artificial Intelligence Models. *International Journal of Engineering Research and*, V8(07). Available at: https://www.ijert.org/a-critical-review-on-artificial-intelligence-models-in-hydrological-forecasting-how-reliable-are-artificial-intelligence-models.

de O. Vieira, E. and Sandoval-Solis, S., 2018. Water resources sustainability index for a water-stressed basin in Brazil. *Journal of Hydrology: Regional Studies*, 19(January), pp.97–109. Available at: https://doi.org/10.1016/j.ejrh.2018.08.003.

Di Zhanga, Junqiang Lina,Qidong Penga, Dongsheng Wangb, Tiantian Yang, 2018. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. Journal of Hydrology, 565(May), pp.720–736. Available at: https://doi.org/10.1016/j.jhydrol.2018.08.050.

Dornpunya, Pheeranat; Rittima Areeya, Phankamolsil Yutthana, Tabucanon Allan Sriratana, Talaluxmana Yutthana, 2021. The Predictability Of Reservoir Inflow Prediction Model For Sirikit Dam Using XGBoost Machine Learning Algorithm. The 5th International Conference on Water Resources Engineering. 2021

Drucker, Harris; Surges Chris J.C, Kaufman Linda, Smola Alex, Vapnik Vladimir, 1997. Support vector regression machines. Advances in Neural Information Processing Systems, 1, pp.155–161.

Eiben, A.E., Michalewicz, Z., Schoenauer, M. and Smith, J.E., 2007. Parameter Control in Evolutionary Algorithms. In: *Parameter Setting in Evolutionary Algorithms*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 19–46.

El Harraki, W., Ouazar, D., Bouziane, A. and Hasnaoui, D., 2021. Optimisation of reservoir operating curves and hedging rules using genetic algorithm with a new objective function and smoothing constraint: application to a multipurpose dam in Morocco. *Environmental Monitoring and Assessment*, 193(4), pp.1–17. Available at: https://doi.org/10.1007/s10661-021-08972-9.

Fallah-Mehdipour, E., Bozorg Haddad, O. and Mariño, M.A., 2012. Real-Time Operation of Reservoir System by Genetic Programming. *Water Resources Management*, 26(14), pp.4091–4103. Available at: http://link.springer.com/10.1007/s11269-012-0132-z.

Fletcher, D. and Goss, E., 1993. Forecasting with neural networks. *Information & Management*, 24(3), pp.159–167. Available at: https://linkinghub.elsevier.com/retrieve/pii/037872069390064Z.

García, S., Luengo, J. and Herrera, F., 2015. *Data Preprocessing in Data Mining*, Springer International Publishing, Cham.

Gardner, M.. and Dorling, S.., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14–15), pp.2627–2636. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1352231097004470.

Garousi-Nejad, I., Bozorg-Haddad, O. and Loáiciga, H.A., 2016a. Modified Firefly Algorithm for Solving Multireservoir Operation in Continuous and Discrete Domains. *Journal of Water Resources Planning and Management*, 142(9), pp.1–15. Available at: https://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000644.

Garousi-Nejad, I., Bozorg-Haddad, O., Loáiciga, H.A. and Mariño, M.A., 2016b. Application of the Firefly Algorithm to Optimal Operation of Reservoirs with the Purpose of Irrigation Supply and Hydropower Production. *Journal of Irrigation and Drainage Engineering*, 142(10), p.04016041. Available at: https://ascelibrary.org/doi/10.1061/%28ASCE%29IR.1943-4774.0001064.

Ghose, D.K., 2018. Measuring Discharge Using Back-Propagation Neural Network: A Case Study on Brahmani River Basin. *Intelligent Engineering Informatics, Advances in Intelligent Systems and Computing*, 695, pp.591–598. Available at: http://link.springer.com/10.1007/978-981-10-7566-7.

Goldberg, D.E. and Deb, K., 1991. *A Comparative Analysis of Selection Schemes Used in Genetic Algorithms*, Morgan Kaufmann Publishers, Inc.

Guo, Jun; Jianzhong Zhou, Hui Qin, Qiang Zou, Qingqing Li College, 2011. Monthly streamflow forecasting based on improved support vector machine model. Expert Systems with Applications, 38(10), pp.13073–13081. Available at: http://dx.doi.org/10.1016/j.eswa.2011.04.114.

Gutiérrez, F. and Dracup, J.A., 2001. An analysis of the feasibility of long-range streamflow forecasting for Colombia using El Niño–Southern Oscillation indicators. *Journal of Hydrology*, 246(1–4), pp.181–196.

Available at: https://linkinghub.elsevier.com/retrieve/pii/S0022169401003730.

Hadiyan, P.P., Moeini, R. and Ehsanzadeh, E., 2020. Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir. *Sustainable Computing: Informatics and Systems*, 27, p.100401. Available at: https://doi.org/10.1016/j.suscom.2020.100401.

Hamdan, A.M. and Zaki, M., 2016. Long-Term Estimation of Water Losses Through Evaporation from Water Surfaces of Nasser Lake Reservoir, Egypt. *International Journal of Civil & Environmental Engineering IJCEE-IJENS*, 16(05). Available at: http://ijens.org/Vol_16_I_05/160905-7878-IJCEE-IJENS.pdf.

Hansun, S., 2013. A new approach of moving average method in time series analysis. *2013 Conference on New Media Studies (CoNMedia)*. November 2013 IEEE, pp. 1–4.

Hassan Ramchoun, Mohammed Amine Janati Idrissi, Youssef Ghanou, Mohamed Ettaouil, 2016. Multilayer Perceptron: Architecture Optimisation and Training. International Journal of Interactive Multimedia and Artificial Intelligence, 4(1), p.26. Available at: http://www.ijimai.org/journal/node/907.

Hassan, R., Cohanim, B., de Weck, O. and Venter, G., 2005. A Comparison of Particle Swarm Optimisation and the Genetic Algorithm. *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. 18 April 2005 American Institute of Aeronautics and Astronautics, Reston, Virigina, pp. 1138–1150.

He, Jin-ping; Jiang, Zhen-xiang; Zhao Cheng; Peng, Zheng-quan; Shi, Yu-qun; 2018. Cloud-Verhulst hybrid prediction model for dam deformation under uncertain conditions. *Water Science and Engineering*, 11(1), pp.61–67. Available at: https://doi.org/10.1016/j.wse.2018.03.002.

He, Z., Wen, X., Liu, H. and Du, J., 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, 509, pp.379–386. Available at: http://dx.doi.org/10.1016/j.jhydrol.2013.11.054.

Hill, T., Marquez, L., O'Connor, M. and Remus, W., 1994. Artificial neural network models for forecasting and decision making. *International Journal of Forecasting*, 10(1), pp.5–15. Available at: https://linkinghub.elsevier.com/retrieve/pii/0169207094900450.

Hipni, Afiq , El-shafie Ahmed, Najah Ali, Karim, Othman Abdul, Hussain Aini, Mukhlisin Muhammad, 2013. Daily Forecasting of Dam Water Levels: Comparing a Support Vector Machine (SVM) Model With Adaptive Neuro Fuzzy Inference System (ANFIS). Water Resources Management, 27(10), pp.3803–3823. Available at: http://link.springer.com/10.1007/s11269-013-0382-4.

Holland, J.H., 1992. Genetic Algorithms. *Scientific American*, 267(1), pp.66–73. Available at: http://www.jstor.org/sTable/24939139.

Hossain, M.S. and El-shafie, A., 2014. Performance analysis of artificial bee colony (ABC) algorithm in optimizing release policy of Aswan High Dam. *Neural Computing and Applications*, 24(5), pp.1199–1206. Available at: http://link.springer.com/10.1007/s00521-012-1309-3.

Hounmenou, C.G., 2021. A Formalism of the General Mathematical Expression of Multilayer Perceptron Neural Networks. , (May), pp.1–12.

Hussain, K., Salleh, M.N.M., Cheng, S. and Shi, Y., 2019. On the exploration and exploitation in popular swarm-based metaheuristic algorithms. *Neural Computing and Applications*, 31(11), pp.7665–7683. Available at: https://doi.org/10.1007/s00521-018-3592-0.

Ibrahim Karim; Sherif Mostafa Hassan, Huang Yuk Feng, Ahmed Ali Najah, Koo Chai Hoon, El-Shafien Ahmed, 2023. Forecasting multi-step-ahead reservoir monthly and daily inflow using machine learning models based on different scenarios. Applied Intelligence, 53(9), pp.10893–10916. Available at: https://link.springer.com/10.1007/s10489-022-04029-7.

Igel, C., Hansen, N. and Roth, S., 2007. Covariance Matrix Adaptation for Multi-objective Optimisation. *Evolutionary Computation*, 15(1), pp.1–28. Available at: https://direct.mit.edu/evco/article/15/1/1-28/1257.

Jang, J.-S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), pp.665–685. Available at: http://ieeexplore.ieee.org/document/256541/.

Jang, J., 1991. Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm. *Proceedings of the 9th National Conference on Artificial Intelligence*, 91, pp.762–767. Available at: http://www.aaai.org/Library/AAAI/1991/aaai91-119.php.

Jothiprakash, V. and Magar, R.B., 2012. Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. *Journal of Hydrology*, 450–451, pp.293–307. Available at: http://dx.doi.org/10.1016/j.jhydrol.2012.04.045.

Kalteh, A.M., 2013. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers and Geosciences*, 54, pp.1–8. Available at: http://dx.doi.org/10.1016/j.cageo.2012.11.015.

Karaboga, D. and Kaya, E., 2019. Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. *Artificial Intelligence Review*, 52(4), pp.2263–2293. Available at: https://doi.org/10.1007/s10462-017-9610-2.

Karami Hojat, Ehteram Mohammad, Mousavi Sayed-Farhad, Farzin Saeed, Kisi Ozgur, El-Shafie Ahmed, 2019. Optimisation of energy management and conversion in the water systems based on evolutionary algorithms. Neural Computing and Applications, 31(10), pp.5951–5964. Available at: https://doi.org/10.1007/s00521-018-3412-

6.

Karim Sherif Mostafa Hassan Ibrahim, Yuk Feng Huang, Ali Najah Ahmed, Chai Hoon Koo, Ahmed El-Shafie, 2021. A review of the hybrid artificial intelligence and optimisation modelling of hydrological streamflow forecasting. Alexandria Engineering Journal. Available at: https://doi.org/10.1016/j.aej.2021.04.100.

Karsoliya, S., 2012. Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. *International Journal of Engineering Trends and Technology*, 3(6), pp.714–717.

Katoch, S., Chauhan, S.S. and Kumar, V., 2021. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80(5), pp.8091–8126. Available at: http://link.springer.com/10.1007/s11042-020-10139-6.

Kennedy, J. and Eberhart, R., 1995. Particle swarm optimisation. *Proceedings of ICNN'95 - International Conference on Neural Networks*. April 1995 IEEE, pp. 1942–1948.

Khan Najeebullah, Pour Sahar Hadi, Shahid Shamsuddin, Ismail Tarmizi, Ahmed Kamal, Chung Eun Sung, Nawaz Nadeem, Wang Xiaojun, 2019. Spatial distribution of secular trends in rainfall indices of Peninsular Malaysia in the presence of long term persistence. *Meteorological Applications*, 26(4), pp.655–670. Available at: https://onlinelibrary.wiley.com/doi/10.1002/met.1792.

Khan, Y. and Chai, S.S., 2017. Ensemble of ANN and ANFIS for water quality prediction and analysis - a data driven approach. *Journal of Telecommunication, Electronic and Computer Engineering*, 9(2–9), pp.117–122.

Khazaee Poul, A., Shourian, M. and Ebrahimi, H., 2019. A Comparative Study of MLR, KNN, ANN and ANFIS Models with Wavelet Transform in Monthly Stream Flow Prediction. *Water Resources Management*, 33(8), pp.2907–2923. Available at: http://link.springer.com/10.1007/s11269-019-02273-0.

Kim, Chong; 2019. The use of large-scale climate indices in monthly reservoir inflow forecasting and its application on time series and artificial intelligence models. *Water (Switzerland)*, 11(2).

Kisi, O., Shiri, J. and Nikoofar, B., 2012. Forecasting daily lake levels using artificial intelligence approaches. *Computers & Geosciences*, 41, pp.169–180. Available at: http://dx.doi.org/10.1016/j.cageo.2011.08.027.

Kulkarni, O., Kulkarni, N., Kulkarni, A.J. and Kakandikar, G., 2018. Constrained cohort intelligence using static and dynamic penalty function approach for mechanical components design. *International Journal of Parallel, Emergent and Distributed Systems*, 33(6), pp.570–588. Available at: http://dx.doi.org/10.1080/17445760.2016.1242728.

Kumar, Abhishek; Wu Guohua, Ali Mostafa Z, Mallipeddi Rammohan, Suganthan Ponnuthurai Nagaratnam, 2020. A test-suite of non-convex constrained optimisation problems from the real-world and some baseline results. Swarm and Evolutionary Computation, 56(April), p.100693. Available at: https://linkinghub.elsevier.com/retrieve/pii/S2210650219308946.

Kumar, V. and Kumar, D., 2021. A Systematic Review on Firefly Algorithm: Past, Present, and Future. *Archives of Computational Methods in Engineering*, 28(4), pp.3269–3291. Available at: https://doi.org/10.1007/s11831-020-09498-y.

Krishna, Leela; K., UmaMahesh, N. V. and Srinivasa Prasad, A., 2018. Optimal multipurpose reservoir operation planning using Genetic Algorithm and Non Linear Programming (GA-NLP) hybrid approach. *ISH Journal of Hydraulic Engineering*, 24(2), pp.258–265. Available at: https://doi.org/10.1080/09715010.2017.1422191.

Leijnen, S. and Veen, F. van, 2020. The Neural Network Zoo. *IS4SI 2019 Summit*. 12 May 2020 MDPI, Basel Switzerland, p. 9.

Lewis, Sara M. Wong, Choong Hay Owens, Avalon C S Fallon, Candace Jepsen, Sarina Thancharoen, Anchana Wu Chiahsiung, De Cock Raphael, 2020. A Global Perspective on Firefly Extinction Threats. BioScience, 70(2), pp.157–167.Availableat:https://academic.oup.com/bioscience/article/70/2/157/571507.

Li, Pei-Hao; Kwon Hyun-Han, Sun Liqiang, Lall Upmanu, Kao Jehng-Jung, 2010. A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan. International Journal of Climatology, 30(8), pp.1256–1268. Available at: http://doi.wiley.com/10.1002/joc.1954.

Li, Xuan; Huang Tinglin, Ma Weixing, Sun Xin, Zhang Haihan, 2015. Effects of rainfall patterns on water quality in a stratified reservoir subject to eutrophication: Implications for management. Science of the Total Environment, 521–522, pp.27–36. Available at: http://dx.doi.org/10.1016/j.scitotenv.2015.03.062.

Lin, G.F., Chen, G.R., Huang, P.Y. and Chou, Y.C., 2009. Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *Journal of Hydrology*, 372(1–4), pp.17–29. Available at: http://dx.doi.org/10.1016/j.jhydrol.2009.03.032.

Lin, J.-Y., CHENG, C.-T. and CHAU, K.-W., 2006. Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4), pp.599–612. Available at: https://www.tandfonline.com/doi/full/10.1623/hysj.51.4.599.

Liong, S.-Y. and Sivapragasam, C., 2002. FLOOD STAGE FORECASTING WITH SUPPORT VECTOR MACHINES. *Journal of the American Water Resources Association*, 38(1), pp.173–186. Available at: https://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2002.tb01544.x.

Liu, Hongrui; Sun Yuanyuan, Yin Xinan, Zhao Yanwei, Cai Yanpeng,Yang Wei, 2020. A reservoir operation method that accounts for different inflow forecast uncertainties in different hydrological periods. Journal of Cleaner Production, 256, p.120471. Available at: https://doi.org/10.1016/j.jclepro.2020.120471.

Liu, Hao, Zhang, X.W. and Tu, L.P., 2020. A modified particle swarm optimisation using adaptive strategy. *Expert Systems with Applications*, 152, p.113353. Available at: https://doi.org/10.1016/j.eswa.2020.113353.

Liu, Yongqi; Qin Hui, Zhang Zhendong, Yao Liqiang, Wang Yongqiang, Li Jie,Liu Guanjun, Zhou Jianzhong, 2019. Deriving reservoir operation rule based on Bayesian deep learning method considering multiple uncertainties. Journal of Hydrology, 579(October), p.124207. Available at: https://doi.org/10.1016/j.jhydrol.2019.124207.

Loucks, D.P. and van Beek, E., 2017. *Water Resource Systems Planning and Management*, Springer International Publishing, Cham.

Luo, Yonghong; Cai Xiangrui, Zhang Ying, Xu Jun, Yuan Xiaojie,, 2018. Multivariate time series imputation with generative adversarial networks. Advances in Neural Information Processing Systems, 2018-Decem(NeurIPS), pp.1596–1607.

Łukasik, S. and Żak, S., 2009. Firefly Algorithm for Continuous Constrained Optimisation Tasks. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 97–106.

Maity, R. and Nagesh Kumar, D., 2009. Hydroclimatic influence of large-scale circulation on the variability of reservoir inflow. *Hydrological Processes*, 23(6), pp.934–942. Available at: http://jamsb.austms.org.au/courses/CSC2408/semester3/resources/ldp/abs-guide.pdf.

Marjani, A., Babanezhad, M. and Shirazian, S., 2020. Application of adaptive network-based fuzzy inference system (ANFIS) in the numerical investigation of Cu/water nanofluid convective flow. *Case Studies in Thermal Engineering*, 22(September), p.100793. Available at: https://doi.org/10.1016/j.csite.2020.100793.

Minu, K.K., Lineesh, M.C. and Jessy John, C., 2010. Wavelet neural networks for nonlinear time series analysis. *Applied Mathematical Sciences*, 4(49–52), pp.2485–2495.

Mirjalili Seyedali, Gandomi Amir H, Mirjalili Seyedeh Zahra, Saremi Shahrzad, Faris, Hossam, Mirjalili Seyed Mohammad, 2017. Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. Advances in Engineering Software, 114, pp.163–191. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0965997816307736.

Mirjalili, S., Mirjalili, S.M. and Hatamlou, A., 2016. Multi-Verse Optimizer: a nature-inspired algorithm for global optimisation. *Neural Computing and Applications*, 27(2), pp.495–513. Available at: http://link.springer.com/10.1007/s00521-015-1870-7.

Moeeni, H., Bonakdari, H., Fatemi, S.E. and Zaji, A.H., 2017. Assessment of

Stochastic Models and a Hybrid Artificial Neural Network-Genetic Algorithm Method in Forecasting Monthly Reservoir Inflow. *INAE Letters*, 2(1), pp.13–23. Available at: http://link.springer.com/10.1007/s41403-017-0017-9.

Müller K. R., Smola A. J., Rätsch G., Schölkopf B., Kohlmorgen J, Vapnik V, 1997. Predicting time series with support vector machines. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 999–1004.

Muluye, G.Y. and Coulibaly, P., 2007. Seasonal reservoir inflow forecasting with low-frequency climatic indices: a comparison of data-driven methods. *Hydrological Sciences Journal*, 52(3), pp.508–522. Available at: http://w

Nashwan, M.S., Ismail, T. and Ahmed, K., 2019. Non-stationary analysis of extreme rainfall in Peninsular Malaysia. *Journal of Sustainability Science and Management*, 14(3), pp.17–34.

Nguyen, P.K.-T., Chua, L.H.-C., Talei, A. and Chai, Q.H., 2018. Water level forecasting using neuro-fuzzy models with local learning. *Neural Computing and Applications*, 30(6), pp.1877–1887. Available at: http://link.springer.com/10.1007/s00521-016-2803-9.

Niu, Wen-jing; Feng, Zhong-kai; Zeng, Ming; Feng, Bao-fei;, Min, Yao-wu;, Cheng, Chun-tian; and Zhou, Jian-zhong, 2019. Forecasting reservoir monthly run-off via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. Applied Soft Computing Journal, 82, p.105589. Available at: https://doi.org/10.1016/j.asoc.2019.105589.

Noori, R.; A.R. Karbassi, A. Moghaddamnia, D. Han, M.H. Zokaei-Ashtiani, A. Farokhnia, M. Ghafari Gousheh, 2011. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. Journal of Hydrology, 401(3–4), pp.177–189. Available at: http://dx.doi.org/10.1016/j.jhydrol.2011.02.021.

Nourani, V., Hosseini Baghanam, A., Adamowski, J. and Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. *Journal of Hydrology*, 514, pp.358–377. Available at: http://dx.doi.org/10.1016/j.jhydrol.2014.03.057.

Nur Adli Zakaria, M., Abdul Malek, M., Zolkepli, M. and Najah Ahmed, A., 2021. Application of artificial intelligence algorithms for hourly river level forecast: A case study of Muda River, Malaysia. *Alexandria Engineering Journal*, 60(4), pp.4015–4028. Available at: https://doi.org/10.1016/j.aej.2021.02.046.

Obienu, A.C., 2018. Autonomic Computing : An Evidence of Better Management Technology. , (February), pp.941–954.

Oyebode, O. and Stretch, D., 2019. Neural network modeling of hydrological systems: A review of implementation techniques. *Natural Resource Modeling*, 32(1).

Palani, S., Liong, S.Y. and Tkalich, P., 2008. An ANN application for water quality forecasting. *Marine Pollution Bulletin*, 56(9), pp.1586–1597.

Panagoulia, D., Tsekouras, G.J. and Kousiouris, G., 2017. A multi-stage methodology for selecting input variables in ANN forecasting of river flows. *Global Nest Journal*, 19(1), pp.49–57.

Panchal, G., Ganatra, A., Kosta, Y.P. and Panchal, D., 2011. Behaviour Analysis of Multilayer Perceptronswith Multiple Hidden Neurons and Hidden Layers. *International Journal of Computer Theory and Engineering*, (January), pp.332–337.

Potdar, K., S., T. and D., C., 2017. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. *International Journal of Computer Applications*, 175(4), pp.7–9.

Pour, S.H., Wahab, A.K.A. and Shahid, S., 2020. Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. *Atmospheric Research*, 233, p.104720. Available at: https://doi.org/10.1016/j.atmosres.2019.104720.

Pratap, A Deb, K., Agarwal, S. and Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), pp.182–197. Available at: http://ieeexplore.ieee.org/document/996017/.

Priyon, Gus; Muhammad Ridwan, Ahmad Jais Alias1, Riza Atiq O. K. Rahmat, Azmi Hassan & Mohd. Alauddin Mohd. Ali1, 2005. Generation of Fuzzy Rules with Subtractive Clustering. Jurnal Teknologi, 43(1).

Reddy, M.J. and Kumar, D.N., 2006. Optimal reservoir operation using multi-objective evolutionary algorithm. *Water Resources Management*, 20(6), pp.861–878.

Revelle, C., Joeres, E. and Kirby, W., 1969. The Linear Decision Rule in Reservoir Management and Design: 1, Development of the Stochastic Model. *Water Resources Research*, 5(4), pp.767–777.

SaberChenari, K., Abghari, H. and Tabari, H., 2016. Application of PSO algorithm in short-term optimisation of reservoir operation. *Environmental Monitoring and Assessment*, 188(12). Available at: http://dx.doi.org/10.1007/s10661-016-5689-1.

Sakiur Rahman, A.T.M.; Takahiro Hosono, John M. Quilty, Jayanta Das, Amiya Basak, 2020. Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms. Advances in Water Resources, 141(May), p.103595. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0309170819312618.

Sandoval-Solis, S., McKinney, D.C. and Loucks, D.P., 2011. Sustainability Index for Water Resources Planning and Management. *Journal of Water Resources Planning and Management*, 137(5), pp.381–390.

Santhanam, R., Uzir, N., Raman, S. and Banerjee, S., 2017. Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets. *International Journal of Control Theory and Applications*, 9(March), pp.651–662.

Sauhats, Antans; Roman Petrichenko, Zane Broka, Karlis Baltputnis, Dmitrijs Sobolevskis Institute, 2016. ANN-based forecasting of hydropower reservoir inflow. *2016 57th International Scientific Conference on Power and Electrical Engineering of Riga Technical University, RTUCON 2016*, pp.2–7.

Seo, Y., Kim, S., Kisi, O. and Singh, V.P., 2015. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520, pp.224–243. Available at: http://dx.doi.org/10.1016/j.jhydrol.2014.11.050.

Shareef, H., Ibrahim, A.A. and Mutlag, A.H., 2015. Lightning search algorithm. *Applied Soft Computing Journal*, 36, pp.315–333. Available at: http://dx.doi.org/10.1016/j.asoc.2015.07.028.

Sharma, M. and Kaur, P., 2021. A Comprehensive Analysis of Nature-Inspired Meta-Heuristic Techniques for Feature Selection Problem. *Archives of Computational Methods in Engineering*, 28(3), pp.1103–1127. Available at: https://doi.org/10.1007/s11831-020-09412-6.

She, N. and Basketfield, D., 2005. Long Range Forecast of Streamflow Using Support Vector Machine. *Impacts of Global Climate Change*. July 2005 American Society of Civil Engineers, Reston, VA, pp. 1–9.

Sheela, K.G. and Deepa, S.N., 2013. Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Mathematical Problems in Engineering*, 2013(10), pp.1–11. Available at: http://www.hindawi.com/journals/mpe/2013/425740/.

Shensa, M.J., 1992. The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Transactions on Signal Processing*, 40(10), pp.2464–2482. Available at: http://ieeexplore.ieee.org/document/157290/.

Sifuzzaman, M., Islam, M.R. and Ali, M.Z., 2009. Application of Wavelet Transform and its Advantages Compared to Fourier Transform. *Journal of Physical Sciences*, 13, pp.121–134.

Simon, D., 2008. Biogeography-Based Optimisation. *IEEE Transactions on Evolutionary Computation*, 12(6), pp.702–713. Available at: https://link.springer.com/10.1007/978-981-16-6624-7_47.

Igor Škrjanc, Jose Antonio Iglesias, Araceli Sanchis, Daniel Leite, Edwin Lughofer, Fernando Gomide, 2019. Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A Survey. Information Sciences, 490, pp.344–368. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0020025519302713.

Sörensen, K., Sevaux, M. and Glover, F., 2018. A History of Metaheuristics. In: *Handbook of Heuristics*. Springer International Publishing, Cham, pp. 791–808.

Specht, 1988. Probabilistic neural networks for classification, mapping, or associative memory. *IEEE International Conference on Neural*

Networks. 1988 IEEE, pp. 525–532 vol.1.

Storn, N, R. and Price, K., 1997. Differential Evolution – A Simple and Efficient Heuristic for Global Optimisation over Continuous Spaces RAINER. *Global Optimisation*, 11(3), pp.341–359. Available at: http://link.springer.com/10.1071/AP09004.

Sun, Jun; Bin Feng and Wenbo Xu, 2004. Particle swarm optimisation with particles having quantum behaviour. *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*. 2004 IEEE, pp. 325–331.

Svozil, D., Kvasnicka, V. and Pospichal, J., 1997. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39(1), pp.43–62. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0169743997000610.

Talpur, N., Salleh, M.N.M. and Hussain, K., 2017. An investigation of membership functions on performance of ANFIS for solving classification problems. *IOP Conference Series: Materials Science and Engineering*, 226(1).

Taud, H. and Mas, J.F., 2018. Multilayer Perceptron (MLP). , pp.451–455.

Tayebiyan, A., Mohammed Ali, T.A., Ghazali, A.H. and Malek, M.A., 2016. Optimisation of Exclusive Release Policies for Hydropower Reservoir Operation by Using Genetic Algorithm. *Water Resources Management*, 30(3), pp.1203–1216. Available at: http://link.springer.com/10.1007/s11269-015-1221-6.

Terzi, Ö. and Ergin, G., 2014. Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. *Neural Computing and Applications*, 25(1), pp.179–188. Available at: http://link.springer.com/10.1007/s00521-013-1469-9.

Tripathi, S., Srinivas, V.V. and Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology*, 330(3–4), pp.621–640. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0022169406002368.

Tsai, W.P., Chang, F.J., Chang, L.C. and Herricks, E.E., 2015. AI techniques for optimizing multi-objective reservoir operation upon human and riverine ecosystem demands. *Journal of Hydrology*, 530, pp.634–644. Available at: http://dx.doi.org/10.1016/j.jhydrol.2015.10.024.

Valizadeh, N., 2011. Daily water level forecasting using adaptive neuro-fuzzy interface system with different scenarios: Klang Gate, Malaysia. *International Journal of the Physical Sciences*, 6(32), pp.7379–7389. Available at:http://www.academicjournals.org/IJPS/abstracts/abstracts/abstract2011/2Dec/Valizadeh et al.htm.

Wang, S., Chen, Z. and Chen, S., 2019. Applicability of deep neural networks on production forecasting in Bakken shale reservoirs. *Journal of Petroleum Science and Engineering*, 179(September 2018), pp.112–125. Available at: https://doi.org/10.1016/j.petrol.2019.04.016.

Wang, W., Nie, X. and Qiu, L., 2010. Support Vector Machine with Particle Swarm Optimisation for Reservoir Annual Inflow Forecasting. *2010 International Conference on Artificial Intelligence and Computational Intelligence*. October 2010 IEEE, pp. 184–188.

Wang, W.C., Chau, K.W., Cheng, C.T. and Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374(3–4), pp.294–306. Available at: http://dx.doi.org/10.1016/j.jhydrol.2009.06.019.

Wanie M. Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali Najah Ahmed, Ahmed El-Shafi. Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Engineering Journal, 12(2), pp.1651–1663. Available at: https://doi.org/10.1016/j.asej.2020.09.011.

Wei, Zhenglei; Huang, Changqiang; Wang, Xiaofei; Han, Tong; Li, Yintong, 2019. Nuclear Reaction Optimisation: A Novel and Powerful Physics-Based Algorithm for Global Optimisation. IEEE Access, 7, pp.66084–66109. Available at: https://ieeexplore.ieee.org/document/8720256/.

Wu, Y. chen and Feng, J. wen, 2018. Development and Application of Artificial Neural Network. *Wireless Personal Communications*, 102(2), pp.1645–1656. Available at: https://doi.org/10.1007/s11277-017-5224-x.

Yang, Tiantian; Ata Akbari Asanjan, Edwin Welles, Gao, Xiaogang; Soroosh Sorooshian, and Liu Xiaomang, 2017. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resources Research, 53(4), pp.2786–2812. Available at: http://doi.wiley.com/10.1002/2017WR020482.

Yang, X.-S., 2009a. Firefly Algorithms for Multimodal Optimisation. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 169–178.

Yang, X.-S., 2009b. Harmony Search as a Metaheuristic Algorithm. In: *Music-Inspired Harmony Search Algorithm*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–14.

Yang, X.-S., 2013. Multiobjective firefly algorithm for continuous optimisation. *Engineering with Computers*, 29(2), pp.175–184. Available at: http://link.springer.com/10.1007/s00366-012-0254-1.

Yang, X.S. and Deb, S., 2010. Engineering optimisation by cuckoo search. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(4), pp.330–343.

Ye, Quanliang; :Li Yi; Zhuo, La; Zhang, Wenlong; Xiong, Wei; Wang, Chao; Yang, Peifang; 2018. Optimal allocation of physical water resources integrated with virtual water trade in water scarce regions: A case study for Beijing, China. Water Research, 129, pp.264–276. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0043135417309521.

Yu, Xiang; Wang Yuhao, Wu Lifeng, Chen Genhua, Wang Lei,Qin, Hui, 2018. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energy Conversion and Management, 164(February), pp.102–111. Available at: https://doi.org/10.1016/j.enconman.2018.02.087

Yu, X., Wang, Y., Wu, L., Chen, G., Wang, L., Qin, H., 2020. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. Journal of Hydrology, 582, p.124293. Available at: https://doi.org/10.1016/j.jhydrol.2019.124293.

Zaher Mundher Yaseen, Ahmed El-shafie, Othman Jaafar, Haitham Abdulmohsin Afan, Khamis Naba Sayl, 2015. Artificial intelligence based models for stream-flow forecasting: 2000-2015. Journal of Hydrology, 530, pp.829–844. Available at: http://dx.doi.org/10.1016/j.jhydrol.2015.10.038.

Zhang, G., Eddy Patuwo, B. and Y. Hu, M., 1998. Forecasting with artificial neural networks: *International Journal of Forecasting*, 14(1), pp.35–62. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0169207097000447.

Zhang, X., Peng, Y., Xu, W. and Wang, B., 2019. An Optimal Operation Model for Hydropower Stations Considering Inflow Forecasts with Different Lead-Times. *Water Resources Management*, 33(1), pp.173–188. Available at: http://link.springer.com/10.1007/s11269-018-2095-1.

Zhao, W., Wang, L. and Zhang, Z., 2019. Atom search optimisation and its application to solve a hydrogeologic parameter estimation problem. *Knowledge-Based Systems*, 163, pp.283–304. Available at: https://doi.org/10.1016/j.knosys.2018.08.030.

Zhongbo Jiang, Yunzhong Zhang, Shuanghu Geng, Simin Wang, Hao Sang, Guoqing, 2014. An adaptive particle swarm optimisation algorithm for reservoir operation optimisation. Applied Soft Computing Journal, 18, pp.167–177. Available at: http://dx.doi.org/10.1016/j.asoc.2014.01.034.

Zhou, H., Zeng, X. and Hong, Y., 2019. Adaptive Exact Penalty Design for Constrained Distributed Optimisation. *IEEE Transactions on Automatic Control*, 64(11), pp.4661–4667. Available at: https://ieeexplore.ieee.org/document/8657710/.

Zor, K., Timur, O. and Teke, A., 2017. A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting. *2017 6th International Youth Conference on Energy (IYCE)*. June 2017 IEEE, pp. 1–7.

Zounemat-Kermani, M. and Teshnehlab, M., 2008. Using adaptive neuro-fuzzy inference system for hydrological time series prediction. *Applied Soft Computing*, 8(2), pp.928–936. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1568494607000828.

LIST OF PUBLICATIONS

(a) KSMH Ibrahim, YF Huang, AN Ahmed, CH Koo, A El-Shafie (2022) Forecasting multi-step-ahead reservoir monthly and daily inflow using machine learning models based on different scenarios. *Appl. Intell.*, 53 (9), pp. 10893-10916.

(b) KSMH Ibrahim, YF Huang, AN Ahmed, CH Koo, A El-Shafie (2021) A review of the hybrid artificial intelligence and optimisation modelling of hydrological streamflow forecasting. *Alex. Eng. J.*, 61 (1), pp. 279-303

(c) Hassan, K.S.M., Huang, Y.F., Koo, C.H., Weng, T.K., Ahmed, A.N., Elshafie, A.H.K.A. (2022). Forecasting of Reservoir Inflow Using Machine Learning—Case Study: Klang Gate Dam Reservoir. In: Al-Emran, M., Al-Sharafi, M.A., Al-Kabi, M.N., Shaalan, K. (eds) *Proceedings of International Conference on Emerging Technologies and Intelligent Systems. ICETIS 2021.* Lecture Notes in Networks and Systems, vol 322. Springer, Cham.