A Framework for Optimal Reservoir Operation to Improve Downstream Aquatic Environment

Application to Nakdong River Basin in South Korea

Kim, J.

DOI
10.4233/uuid:be007bb6-fcee-4814-87b4-d7a14b691547

Publication date
2023

Document Version
Final published version

Citation (APA)
Kim, J. (2023). *A Framework for Optimal Reservoir Operation to Improve Downstream Aquatic Environment: Application to Nakdong River Basin in South Korea*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:be007bb6-fcee-4814-87b4-d7a14b691547

A Framework for Optimal Reservoir Operation to Improve Downstream Aquatic Environment:

Application to Nakdong River Basin in South Korea

Jongchan Kim

A Framework for Optimal Reservoir Operation to Improve Downstream Aquatic Environment: Application to Nakdong River Basin in South Korea

Jongchan Kim

Cover design
DALL·E (https://labs.openai.com)

A Framework for Optimal Reservoir Operation to Improve Downstream Aquatic Environment: Application to Nakdong River Basin in South Korea

DISSERTATION

Submitted in fulfillment of the requirements of
the Board for Doctorates of Delft University of Technology
and
of the Academic Board of the IHE Delft
Institute for Water Education
for
the Degree of DOCTOR

to be defended in public on
Thursday, 14 September 2023 at 12:30 hours
in Delft, the Netherlands

by

Jongchan KIM
Master of Science in Civil Engineering,
Kyungpook National University, South Korea
born in Ulsan, South Korea

This dissertation has been approved by the promotor and copromotor.

Composition of the doctoral committee:

Rector Magnificus TU Delft	chairperson
Rector IHE Delft	vice-chairperson
Prof.dr. D.P. Solomatine	IHE Delft / TU Delft, promotor
Dr. A. Jonoski	IHE Delft, copromotor

Independent members:

Dr.ir. M.M. Rutten	TU Delft
Prof.dr.ir. P. Willems	KU Leuven, Belgium
Prof.dr. M.E. McClain	IHE Delft / TU Delft
Prof.dr. Z. Kapelan	TU Delft
Prof.dr.ir R. Uijlenhoet	TU Delft, reserve member

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

Published by IHE Delft Institute for Water Education
www.un-ihe.org
ISBN 978-90-73445-52-9

To my dearly beloved family and K-water

This dissertation is based on the Ph.D. Research Proposal entitled “Modeling for Improving Water Environment Coupled with Optimal Reservoir Operations of Upper Nakdong River Basin, South Korea”.

ACKNOWLEDGEMENTS

Studying for a PhD degree abroad in my forties was a big challenge for me. In particular, the COVID-19 pandemic came as a greater difficulty for me and my family to adapt to our lives in a foreign country. Nevertheless, the invaluable help from many individuals has made me address difficult situations and complete this dissertation. I would like to express my sincere gratitude and appreciation to the many individuals who have helped me along the way in completing this dissertation.

First and foremost, I am deeply grateful to my supervisors, Professor Dimitri Solomatine and Dr. Andreja Jonoski, whose unwavering support, guidance, and mentorship throughout my PhD studies have been instrumental in shaping my research trajectory. Their insightful feedback has been invaluable in effectively responding to the challenges that I have faced. I would also like to say my thank to Professor Peter Goethals at Ghent University for participating in my supervisory team and providing me with valuable feedback on my research. In addition, I fully appreciate the kindness of the staff and colleagues at IHE Delft. Furthermore, I would like to thank Professor Kun-Yeun Han who laid the foundation for my academic knowledge in the water sector, and Dr. Byunghyun Kim of Kyungpook National University in South Korea for allowing me to supervise an MSc student online.

I want to say that I love my wife Miran and my two sons, Museong and Muchan. I am thankful that they have given me their love and support although they sometimes had a difficult time here in the Netherlands. Their considerable support and great patience have been the cornerstone of my success in finishing this dissertation. Additionally, I would like to show my love and thanks to my parents, whose love and encouragement have been the driving force behind the pursuit of my goal.

I would like to express my heartfelt appreciation to Ja-ho Koo and his family. Ja-ho Koo played a pivotal role in helping me overcome the problems with which I was confronted during my study. His willingness to listen to my concerns, share his expertise, and offer practical solutions has been highly valuable for publishing my research papers.

Last but not least, I owe a debt of heartful gratitude to K-water and all my colleagues for allowing me to study in the Netherlands including financial support. The unstinting support from K-water for my academic and professional growth has been the major force for completing my PhD course. I hope that this dissertation will contribute to K-water's work in practice.

Once again, my sincere thanks to everyone who has contributed to my academic journey. Your unwavering support, encouragement, guidance, and love have made me complete this dissertation.

SUMMARY

In the water sector, issues concerning the aquatic environment have been extensively discussed due to climate change. In particular, water quality problems such as harmful cyanobacterial blooms (CyanoHABs) in rivers have arisen in South Korea since 2012. The Korean government constructed 16 weirs in the rivers during the Four Major Rivers Restoration Project. These weirs were built to more effectively use water resources in the rivers. Many environmental activists, however, have claimed that the weirs have caused water quality problems of CyanoHABs in the rivers. These CyanoHABs can be threats to the water environment while harming human health and aquatic ecosystems since CyanoHABs produce toxic substances such as microcystins.

To address the problems of these CyanoHABs, many researchers have conducted studies on predictive models for CyanoHABs. A predictive model using a data-driven approach can be useful in exploring the main factors affecting CyanoHABs at a specific location. However, these studies have not focused on preventing the occurrence of CyanoHABs but only on predicting their occurrence. If these studies are designed to link with a practical method for reducing the frequency of CyanoHABs, viable strategies can be proposed to effectively control CyanoHABs. Therefore, detailed considerations are required concerning the prevention or mitigation of CyanoHABs.

Reservoir operation can be a solution for reducing the problem of CyanoHABs in a downstream river. For example, discharging more water from upstream reservoirs can flush CyanoHABs downstream. However, the risk of water shortage can be increased in a reservoir if it is operated for improving water quality downstream. This is because reservoirs were typically designed for management of water quantity such as water supply. To use limited water resources in a reservoir to reduce the frequency of CyanoHABs downstream, optimal reservoir operations are necessary that simultaneously consider both the quantity and the quality of water.

This study focused on establishing a practical framework for the optimal operation of upstream reservoirs to address the problem of CyanoHABs in a downstream river. Furthermore, the applicability of this framework was demonstrated using observational data related to the quantity and quality of the upstream reservoirs in the study area, the Nakdong River basin of South Korea. The framework was established by incorporating three models: a machine learning model, a river water quality model, and an optimization model for reservoir operation.

The first step of the framework applies the machine learning model to predict the occurrence or the nonoccurrence of CyanoHABs at the location of the Chilgok Weir, using input data from that same location. Chilgok Weir is a target downstream location

on the Nakdong River, about 135 km downstream of an upstream boundary where conditions are controlled by the operation of two upstream reservoirs (Andong and Imha). In the event of a prediction of the occurrence of CyanoHABs, the optimization model simulates the decision variables regarding the quantity and quality of water released from the upstream reservoirs. This optimization process is aimed at reducing the incidence of CyanoHABs at Chilgok Weir. The next step employs the river water quality model to simulate the dynamics of a water quality parameter which is a main factor of CyanoHABs by using the optimization results as upstream boundary conditions. The final step is a process for confirming whether CyanoHABs would not occur at Chilgok Weir based on the simulation results of the river water quality model using the machine learning model.

The machine learning model for the first step of this research was developed by using four classification algorithms: k-Nearest Neighbor (k-NN), Decision Tree (DT), Logistic Regression (LR), and Support Vector Machine (SVM). To build the predictive model for CyanoHABs with high accuracy, input features were first selected by applying ANOVA (Analysis of Variance) and solving a multi-collinearity problem. Next, an oversampling method was adopted to overcome the problem of having an imbalanced dataset on CyanoHABs. Consequently, a model applying the k-NN algorithm ensured high accuracy of more than 80% in predicting the occurrence or nonoccurrence of CyanoHABs at Chilgok Weir. This model was developed by using average air temperature and nitrate nitrogen (NO_3-N) as input features.

The river water quality model using HEC-RAS was built to simulate the dynamics of NO_3-N, a parameter that emerged from the first step as one of the main factors for the occurrence of CyanoHABs at Chilgok Weir. By applying this river water quality model, the fate and transport of NO_3-N were analyzed under different scenarios based on variations in the quantity and quality of water at the upstream boundary. The simulation results showed how different aspects of the NO_3-N dynamics downstream can be influenced and controlled, depending on flow rate and NO_3-N concentration upstream. Thus, to formulate a strategy for reducing the incidence of CyanoHABs based on the control of NO_3-N, a quantitative analysis for the NO_3-N dynamics downstream should be performed in advance by using the river water quality model.

The optimization model for reservoir operation produced the simulation results used as the upstream boundary conditions of the river water quality model. The objective functions for the optimization process were formulated in terms of both the quantity and quality of water released from the upstream reservoirs (Andong and Imha). The decision variables for water quantity were constrained based on the optimal joint operation of the two reservoirs. The decision variable for water quality was the NO_3-N concentration which is the main factor for the occurrence of CyanoHABs at Chilgok Weir by considering the use of a selective withdrawal facility of the Imha Reservoir.

The applicability of the framework was demonstrated by the simulation results using observational data for the study area. The simulation results based on the framework confirmed that the frequency of CyanoHABs would be decreased compared to the number of days when CyanoHABs were actually observed at Chilgok Weir. Hence, this framework can support the decision-making of reservoir operation in practice to create a favorable aquatic environment in a downstream river by reducing the frequency of CyanoHABs downstream. In particular, the framework is a novelty in terms of efficiency since it can be a part of a solution to the problem of CyanoHABs without using an additional amount of water from an upstream reservoir.

SAMENVATTING

In de watersector zijn kwesties met betrekking tot het aquatisch milieu als gevolg van klimaatverandering uitgebreid besproken. Met name problemen met de waterkwaliteit, zoals schadelijke cyanobacteriële bloei in rivieren, doen zich sinds 2012 voor in Zuid-Korea. De Koreaanse regering heeft 16 stuwen in de rivieren aangelegd tijdens het Four Major Rivers Restoration Project. Deze stuwen zijn gebouwd om de watervoorraden in de rivieren effectiever te gebruiken. Veel milieuactivisten hebben echter beweerd dat de stuwen waterkwaliteitsproblemen hebben veroorzaakt door schadelijke cyanobacteriële bloei in de rivieren. Deze schadelijke cyanobacteriële bloei kan een bedreiging vormen voor het watermilieu en tegelijkertijd schadelijk zijn voor de menselijke gezondheid en aquatische ecosystemen, aangezien schadelijke cyanobacteriële bloei giftige stoffen zoals microcystines produceert.

Om de problemen van deze schadelijke cyanobacteriële bloei aan te pakken, hebben veel onderzoekers studies uitgevoerd naar voorspellende modellen voor schadelijke cyanobacteriële bloei. Een voorspellend model met een datagestuurde benadering kan nuttig zijn bij het onderzoeken van de belangrijkste factoren die schadelijke cyanobacteriële bloei op een specifieke locatie beïnvloeden. Deze onderzoeken waren echter niet gericht op het voorkomen van schadelijke cyanobacteriële bloei, maar alleen op het voorspellen van het optreden ervan. Als deze onderzoeken zijn ontworpen om te koppelen aan een praktische methode om de frequentie van schadelijke cyanobacteriële bloei te verminderen, kunnen levensvatbare strategieën worden voorgesteld om schadelijke cyanobacteriële bloei effectief te bestrijden. Daarom zijn gedetailleerde overwegingen vereist met betrekking tot het voorkomen of verminderen van schadelijke cyanobacteriële bloei.

Reservoirbeheer kan een oplossing zijn om het probleem van schadelijke cyanobacteriële bloei in een stroomafwaarts gelegen rivier te verminderen. Door bijvoorbeeld meer water uit stroomopwaartse reservoirs te lozen, kunnen schadelijke cyanobacteriële bloei stroomafwaarts worden weggespoeld. Het risico op watertekort kan echter worden vergroot in een reservoir als het wordt gebruikt om de waterkwaliteit stroomafwaarts te verbeteren. Dit komt omdat reservoirs doorgaans zijn ontworpen voor het beheer van de waterkwantiteit, zoals de watervoorziening. Om beperkte waterbronnen in een reservoir te gebruiken om de frequentie van schadelijke cyanobacteriële bloei stroomafwaarts te verminderen, zijn optimale reservoiroperaties nodig die tegelijkertijd rekening houden met zowel de kwantiteit als de kwaliteit van het water.

Deze studie was gericht op het opzetten van een praktisch raamwerk voor de optimale werking van stroomopwaartse reservoirs om het probleem van schadelijke cyanobacteriële bloei in een stroomafwaartse rivier aan te pakken. Bovendien werd de toepasbaarheid van dit raamwerk aangetoond met behulp van observatiegegevens met

betrekking tot de kwantiteit en kwaliteit van de stroomopwaartse reservoirs in het studiegebied, het stroomgebied van de Nakdong-rivier in Zuid-Korea. Het raamwerk is tot stand gekomen door drie modellen op te nemen: een machine learning-model, een rivierwaterkwaliteitsmodel en een optimalisatiemodel voor de werking van reservoirs.

De eerste stap van het raamwerk past het machine learning-model toe om het al dan niet optreden van schadelijke cyanobacteriële bloei op de locatie van de Chilgok-stuw te voorspellen, met behulp van invoergegevens van diezelfde locatie. Chilgok Weir is een beoogde stroomafwaartse locatie aan de Nakdong-rivier, ongeveer 135 km stroomafwaarts van een stroomopwaartse grens waar de omstandigheden worden beheerst door de werking van twee stroomopwaartse reservoirs (Andong en Imha). Bij een voorspelling van het optreden van schadelijke cyanobacteriële bloei simuleert het optimalisatiemodel de beslissingsvariabelen met betrekking tot de kwantiteit en kwaliteit van vrijkomend water uit de bovenstroomse reservoirs. Dit optimalisatieproces is gericht op het verminderen van de incidentie van schadelijke cyanobacteriële bloei bij Chilgok Weir. De volgende stap maakt gebruik van het rivierwaterkwaliteitsmodel om de dynamiek van een waterkwaliteitsparameter te simuleren die een hoofdfactor is van schadelijke cyanobacteriële bloei door de optimalisatieresultaten te gebruiken als stroomopwaartse randvoorwaarden. De laatste stap is een proces om te bevestigen of er geen schadelijke cyanobacteriële bloei zou optreden bij Chilgok Weir op basis van de simulatieresultaten van het rivierwaterkwaliteitsmodel met behulp van het machine learning-model.

Het machine learning-model voor de eerste stap van dit onderzoek is ontwikkeld met behulp van vier classificatie-algoritmen: k-Nearest Neighbor (k-NN), Decision Tree (DT), Logistic Regression (LR) en Support Vector Machine (SVM). Om het voorspellende model voor schadelijke cyanobacteriële bloei met hoge nauwkeurigheid te bouwen, werden eerst invoerkenmerken geselecteerd door ANOVA (Analysis of Variance) toe te passen en een multi-collineariteitsprobleem op te lossen. Vervolgens werd een oversampling-methode toegepast om het probleem van een onevenwichtige dataset over schadelijke cyanobacteriële bloei op te lossen. Bijgevolg zorgde een model dat het k-NN-algoritme toepast voor een hoge nauwkeurigheid van meer dan 80% bij het voorspellen van het al dan niet optreden van schadelijke cyanobacteriële bloei bij Chilgok Weir. Dit model is ontwikkeld door de gemiddelde luchttemperatuur en nitraatstikstof (NO_3-N) als invoerkenmerken te gebruiken.

Het rivierwaterkwaliteitsmodel met behulp van HEC-RAS is gebouwd om de dynamiek van NO_3-N te simuleren, een parameter die uit de eerste stap naar voren kwam als een van de belangrijkste factoren voor het optreden van schadelijke cyanobacteriële bloei bij Chilgok Weir. Door dit rivierwaterkwaliteitsmodel toe te passen, werden het lot en het transport van NO_3-N geanalyseerd onder verschillende scenario's op basis van variaties in de kwantiteit en kwaliteit van het water aan de stroomopwaartse grens. De simulatieresultaten lieten zien hoe verschillende aspecten van de NO_3-N-dynamiek

stroomafwaarts kunnen worden beïnvloed en geregeld, afhankelijk van het debiet en de NO_3-N-concentratie stroomopwaarts. Om een strategie te formuleren voor het verminderen van de incidentie van schadelijke cyanobacteriële bloei op basis van de beheersing van NO_3-N, zou dus vooraf een kwantitatieve analyse van de NO_3-N-dynamiek stroomafwaarts moeten worden uitgevoerd met behulp van het rivierwaterkwaliteitsmodel.

Het optimalisatiemodel voor de werking van het reservoir leverde de simulatieresultaten op die werden gebruikt als de stroomopwaartse randvoorwaarden van het rivierwaterkwaliteitsmodel. De doelfuncties voor het optimalisatieproces zijn geformuleerd in termen van zowel de kwantiteit als de kwaliteit van het vrijkomende water uit de bovenstroomse reservoirs (Andong en Imha). De beslissingsvariabelen voor waterkwantiteit waren beperkt op basis van de optimale gezamenlijke werking van de twee reservoirs. De beslissingsvariabele voor de waterkwaliteit was de NO_3-N-concentratie, die de belangrijkste factor is voor het optreden van schadelijke cyanobacteriële bloei bij Chilgok Weir door het gebruik van een selectieve onttrekkingsfaciliteit van het Imha-reservoir te overwegen.

De toepasbaarheid van het raamwerk werd aangetoond door de simulatieresultaten met behulp van observatiegegevens voor het studiegebied. De simulatieresultaten op basis van het raamwerk bevestigden dat de frequentie van schadelijke cyanobacteriële bloei zou afnemen in vergelijking met het aantal dagen dat schadelijke cyanobacteriële bloei daadwerkelijk werd waargenomen bij Chilgok Weir. Daarom kan dit raamwerk de besluitvorming over reservoirbeheer in de praktijk ondersteunen om een gunstig aquatisch milieu in een stroomafwaartse rivier te creëren door de frequentie van schadelijke cyanobacteriële bloei stroomafwaarts te verminderen. Het raamwerk is met name nieuw in termen van efficiëntie, omdat het een deel van een oplossing kan zijn voor het probleem van schadelijke cyanobacteriële bloei zonder een extra hoeveelheid water uit een stroomopwaarts reservoir te gebruiken.

CONTENTS

Acknowledgements ..... **vii**

Summary ..... **ix**

Samenvatting ..... **xiii**

Contents ..... **xvii**

1 Introduction **1**

1.1 Motivation 2

1.2 Research Gaps 3

1.2.1 Prediction of occurrence of harmful cyanobacterial blooms 3

1.2.2 Optimal reservoir operation considering the quantity and quality of water 4

1.3 Research Objective and Questions 4

1.4 Research Methodology 5

1.5 Outline 6

2 Description of the Study Area **9**

2.1 Introduction 10

2.2 Study Area 10

2.3 Data Availability for the Study Area 13

3 A Machine Learning Approach to the Prediction of Cyanobacterial Blooms **15**

3.1 Introduction 16

3.2 Materials 19

3.2.1 Data collection 19

3.2.2 Data preprocessing 21

3.3 Methods 25

3.3.1 Analysis of variance (ANOVA) for feature selection 25

3.3.2 Multi-collinearity 25

3.3.3 Classification algorithms of machine learning 25

3.3.4 Oversampling using SMOTE (Synthetic Minority Oversampling Technique) 26

3.3.5 Training, cross-validation, and test for the dataset 26

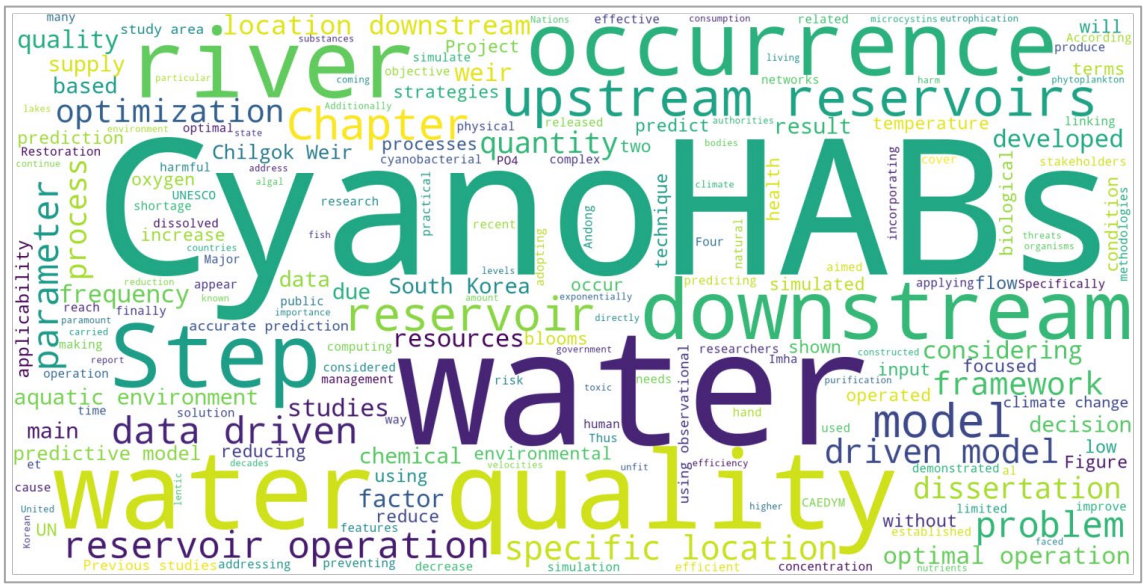
3.3.6 Model evaluation 27
3.3.7 Summary of the modelling procedure 28
3.4 Results 29
3.4.1 Determination of the modelling cases 29
3.4.2 Accuracy of the models 34
3.4.3 Summary of the modelling results 37
3.5 Discussion and Conclusions 38
4 River Water Quality Modelling for Nitrate Nitrogen Control Using HEC-RAS ... 41
4.1 Introduction 42
4.2 Materials and Methods 45
4.2.1 Model description 45
4.2.2 Data for HEC-RAS model 46
4.2.3 Data preparation 49
4.2.4 Experimental setup 51
4.3 Results 54
4.3.1 Calibration and validation for unsteady flow 54
4.3.2 Calibration and validation for NO_3-N dynamics 57
4.3.3 NO_3-N dynamics according to variation in water quantity 63
4.3.4 NO_3-N dynamics according to variation in water quality 68
4.3.5 Guidelines for design of strategies to control NO_3-N downstream 70
4.4 Discussion 72
4.5 Conclusions 74
5 Optimal Reservoir Operation to Mitigate Cyanobacterial Blooms Downstream .. 75
5.1 Introduction 76
5.2 Modelling Methods 78
5.2.1 Data preparation 78
5.2.2 Machine learning model 79
5.2.3 Optimization model 79
5.2.4 River water quality model 81
5.3 Experimental Setup 82

5.3.1 Procedure 82
5.3.2 Experimental cases for optimization 84
5.4 Results and Discussion 87
5.4.1 Simulation test 87
5.4.2 Optimization results 88
5.4.3 NO_3-N dynamics at Chilgok Weir 97
5.4.4 Prediction of the occurrence of CyanoHABs at Chilgok Weir 100
5.5 Conclusions 102
6 Conclusions 105
6.1 Reflections on Research Questions 106
6.2 Research Outcomes 107
6.2.1 Scientific perspective 108
6.2.2 Environmental and social impact 108
6.3 Limitations and Recommendations 109
6.3.1 Research Limitations 109
6.3.2 Recommendations for further studies 110
References 113
List of Acronyms 129
List of Tables 135
List of Figures 137
About the author 141

1

INTRODUCTION

This chapter introduces the motivation for this dissertation, the gaps shown in previous studies, the objective of this research considering the research gaps, the research methodology, and the outline of this dissertation.



Parts of this chapter are based on Kim, J., Jonoski, A., Solomatine, D. P., and Goethals, P. L. M.: Decision support framework for optimal reservoir operation to mitigate cyanobacterial blooms in rivers, submitted to *Sustainability* in 2023 (under review).

1.1 MOTIVATION

According to a recent report by United Nations Water (UN-Water), problems of water quality will continue to occur due to climate change in the coming decades. (UNESCO and UN-Water, 2020). This is because climate change can cause an increase in water temperatures, a decrease in dissolved oxygen levels, and a reduction in the efficiency of natural purification processes (UNESCO and UN-Water, 2020). Water quality management is of paramount importance since water quality is directly related to the aquatic environment and public health. Thus, environmental authorities in many countries have focused on how to address problems with water quality such as harmful algal blooms (or cyanobacterial blooms).

Harmful cyanobacterial blooms (CyanoHABs) are known to be threats to the aquatic environment (Paerl and Otten, 2013) and human health (Carmichael and Boyer, 2016; Falconer and Humpage, 2005; Falconer, 2005). CyanoHABs can decrease the amount of oxygen in water, which can harm fish and other organisms living in an aquatic environment (Gobler, 2020). Additionally, these CyanoHABs produce toxic substances such as microcystins (Carmichael and Boyer, 2016; Falconer and Humpage, 2005; Falconer, 2005) making the water unfit for human consumption. CyanoHABs appear when phytoplankton exponentially increase in lentic water bodies with low flow velocities such as lakes and reservoirs due to eutrophication (Jankowiak et al., 2019; Park et al., 2021a; Xu et al., 2015; Zhao et al., 2019), a condition in which the concentration of nutrients in the water is higher than that in the natural state.

In particular, South Korea has been faced with the problem of CyanoHABs in rivers since 2012. In 2012, the Korean government carried out the Four Major Rivers Restoration Project during which 16 weirs were constructed in the middle of rivers (Song and Lynch, 2018). Environmental activists have claimed that the weirs have resulted in low water quality and the frequent occurrence of CyanoHABs in the rivers because the weirs have caused low flow velocity. On the other hand, those who have supported the project have argued that these weirs have not been the cause of CyanoHABs and the water quality of the river has rather improved after the construction of the weirs. These arguments among stakeholders have led to some social conflicts in South Korea.

Reservoir operation can be a solution for alleviating the problem of CyanoHABs in a downstream river. For instance, discharging more water from upstream reservoirs can flush CyanoHABs downstream (Kim et al., 2022c; Lee and Baek, 2022). However, the reservoir operation considering the improvement of water quality downstream may increase the risk of water shortage since reservoirs were generally designed for the management of water quantity such as water supply. To reduce the problem of CyanoHABs downstream without the risk of water shortage, reservoirs should be optimally operated considering both the quantity and quality of water at the same time.

As water resources in reservoirs are limited, available water resources must be efficiently used from the different perspectives of stakeholders. Agencies for reservoir operation usually exploit water resources in reservoirs considering the factor of water quantity. On the contrary, environmental organizations can insist that reservoirs should be operated toward preventing the occurrence of CyanoHABs to support sustainable development in terms of the aquatic environment and public health. To cope with these conflicting interests, a framework for optimal reservoir operation should be established and applied to improve the water environment by reducing the frequency of occurrence of CyanoHABs downstream without disrupting the water supply.

1.2 Research Gaps

1.2.1 Prediction of occurrence of harmful cyanobacterial blooms

As part of the solution to problems of water quality, many researchers in the water sector have been conducting studies on CyanoHABs (Rousso et al., 2020). Particularly, research studies in South Korea have attempted to accurately predict the occurrence of CyanoHABs in the rivers since the Four Major Rivers Restoration Project (Yi et al., 2018; Pyo et al., 2021; Park et al., 2021b; Ahn et al., 2021; Kim et al., 2020). However, the processes of occurrence of CyanoHABs are complex because the occurrence of CyanoHABs cannot be explained only with mathematical equations or chemical reactions. Specifically, since CyanoHABs appear due to not only external factors such as flow characteristics, climate conditions, and water quality parameters but also internal factors such as chemical and biological processes, accurate prediction of their occurrence is a challenging task (Kim et al., 2017; Rousso et al., 2020).

According to Rousso et al. (2020), forecasting and predictive models for CyanoHABs were developed in various forms depending on modelling techniques. Previous studies focused on not only process-based models such as DYRESM-CAEDYM, ELCOM-CAEDYM, and WASP but also data-driven models based on techniques such as artificial neural networks, decision trees, and Bayesian networks (Rousso et al., 2020). While process-based models provide mechanisms of CyanoHABs for accurate predictions, these models require inputs for multiple parameters (Rousso et al., 2020) and spend considerable computing time (Yang et al., 2021). On the other hand, data-driven models produce output with some main predictors (Rousso et al., 2020) and less computing power (Yang et al., 2021). In this respect, the data-driven models can be more effective and efficient for decision-makers who pursue clear strategies for addressing the problems confronted.

A predictive model adopting a data-driven approach can be used to understand the main factors for the occurrence of CyanoHABs at a specific location. Thus, accurate prediction

of CyanoHABs is an essential precondition for devising viable strategies to effectively control CyanoHABs. Nonetheless, the previous studies were mostly aimed at not preventing the occurrence of CyanoHABs but predicting their occurrence. Ultimately, the predictive model for CyanoHABs needs to be linked with a practical method for reducing the frequency of CyanoHABs.

1.2.2 Optimal reservoir operation considering the quantity and quality of water

Reservoir operations have been conventionally performed to use the water resources in a reservoir considering quantitative needs, such as water supply. As problems with water quality have become increasingly serious due to climate change, researchers have been exploring ways to improve the water quality downstream using limited water resources in reservoirs away from conventional approaches to reservoir operation. For example, two recent studies (Saadatpour et al., 2021; Saadatpour et al., 2020) suggested optimal strategies for reservoir operation in consideration of temperature or dissolved oxygen (DO) downstream as well as water quantities. In addition, Yosefipoor et al. (2022) proposed the optimal operation of a reservoir without a failure of the water supply to minimize the violations of phosphate (PO_4) and Iron (Fe) concentrations in a downstream river.

However, there have been few studies on reservoir operation that has considered biological parameters of water quality (Omer, 2020) such as CyanoHABs in a downstream river. Previous studies have not focused on CyanoHABs but on physical or chemical parameters (Omer, 2020) such as temperature, DO, and PO_4. This can be because simulating biological parameters of water quality is more complex compared to physical or chemical parameters.

1.3 Research Objective and Questions

To link reservoir operation with CyanoHABs downstream, the complicated process of the occurrence of CyanoHABs should be considered in advance when making a decision on the reservoir operation. Hence, a systematic framework will offer an intimate connection between the optimal operation of upstream reservoirs to reduce the frequency of CyanoHABs and the prediction of the occurrence of CyanoHABs in a downstream river.

The main objective of this dissertation is to establish a practical framework for the optimal operation of upstream reservoirs for addressing the problem of CyanoHABs in a downstream river and to demonstrate the applicability of the framework. To achieve this aim, the following questions are addressed:

i. What is an effective and efficient way to predict the occurrence of CyanoHABs at a specific location downstream in terms of linking with the operation of upstream reservoirs?

ii. How can a river water quality model be developed to simulate the fate and transport of water quality parameters involved in CyanoHABs to cover a river reach between upstream reservoirs and a specific location downstream?

iii. What optimization process for the operation of upstream reservoirs should be set up, for simultaneously considering both the quantity and quality of water downstream?

iv. How can the optimal operation of upstream reservoirs be coupled to a predictive model for CyanoHABs and a river water quality model?

1.4 RESEARCH METHODOLOGY

A general framework should be first established for an optimal operation of upstream reservoirs to reduce the incidence of CyanoHABs at a specific location downstream, as shown in Figure 1.1. The framework is comprised of six steps as follows:

i. Step 1 applies a data-driven model to predict the occurrence or nonoccurrence of CyanoHABs at a specific location using observational data, which are input features associated with CyanoHABs.

ii. In the event of a prediction of the occurrence of CyanoHABs (Step 1), the depth distribution of water quality in the upstream reservoirs is simulated in Step 2. The water quality parameters of this simulation are consistent with the input features of the data-driven model of Step 1.

iii. Step 3 involves an optimization process in which the decision variables include the quantity and quality of water released from the upstream reservoirs. The objective functions and constraints of this optimization are aimed at decreasing the frequency of CyanoHABs at the specific location and satisfying the water demand downstream.

iv. In Step 4, the water quality at the specific location downstream is simulated using a river water quality model by incorporating the optimization results from Step 3 as upstream boundary conditions.

v. Step 5 is a process for confirming whether CyanoHABs would not occur at the specific location downstream based on the water quality simulated in Step 4 by applying the data-driven model of Step 1.

vi. If the prediction result from Step 5 indicates that CyanoHABs would not occur at the specific location downstream, the upstream reservoirs will be finally operated using the optimization results from Step 3.

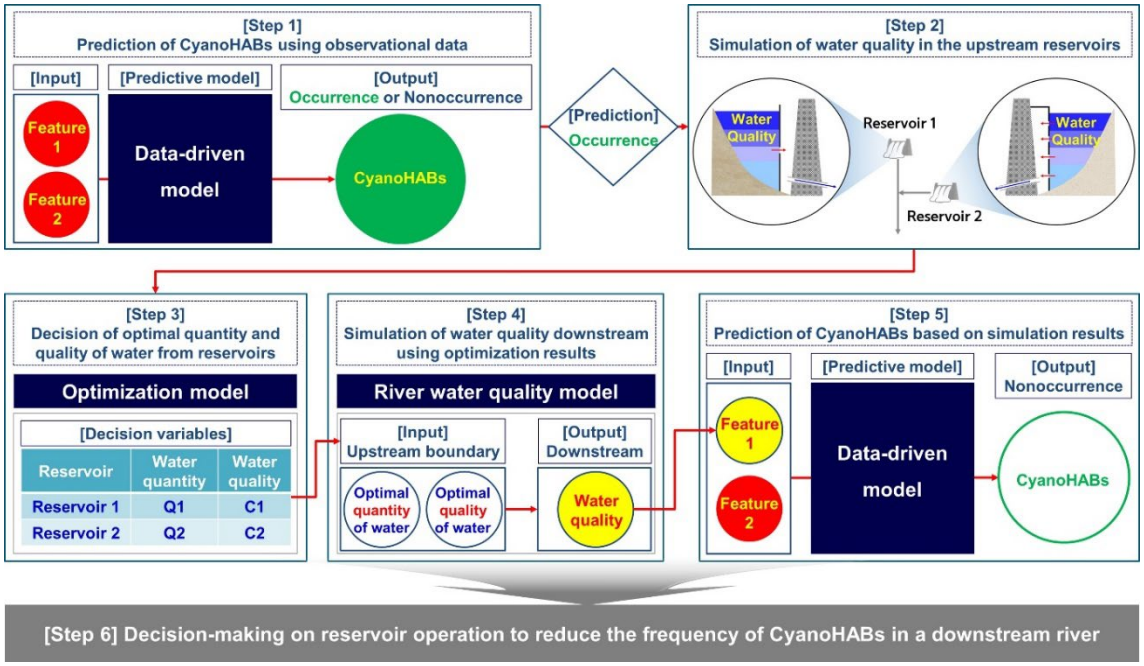


Figure 1.1. Framework for optimal reservoir operation to reduce the frequency of CyanoHABs at a specific location downstream

This dissertation focuses on developing and applying a data-driven model in Steps 1 and 5, an optimization model in Step 3, and a river water quality model in Step 4. Although the simulation of the water quality of reservoirs in Step 2 is not included in this dissertation, the applicability of this framework is demonstrated using observational data related to the quantity and quality of the upstream reservoirs in the study area. This dissertation covers the study area of the upper reach of the Nakdong River in South Korea, including two upstream reservoirs (Andong and Imha reservoirs) and the Chilgok Weir downstream.

1.5 OUTLINE

This dissertation consists of seven chapters. **Chapter 2** explains the study area and data availability for this dissertation. Chapters 3 to 6 present the research methodologies developed and the results employing the methodologies, as shown in Figure 1.2. Specifically, **Chapter 3** introduces a data-driven model developed by adopting a machine learning technique for predicting the occurrence of CyanoHABs at Chilgok Weir. In

Chapter 4, a river water quality model is developed to simulate the dynamics of a water quality parameter which is a main factor significantly influencing the occurrence of CyanoHABs at Chilgok Weir. **Chapter 5** specifies an optimization process to decide the quantity and quality of water released from two upstream reservoirs (Andong and Imha) in terms of reducing the frequency of CyanoHABs at Chilgok Weir. The applicability of the framework shown in Figure 1.1 is also demonstrated in Chapter 5 by using observational data and linking the data-driven model in Chapter 3, the river water quality model in Chapter 4, and the optimization model. Finally, **Chapter 6** draws conclusions incorporating the limitations of this dissertation and the recommendations for further studies.

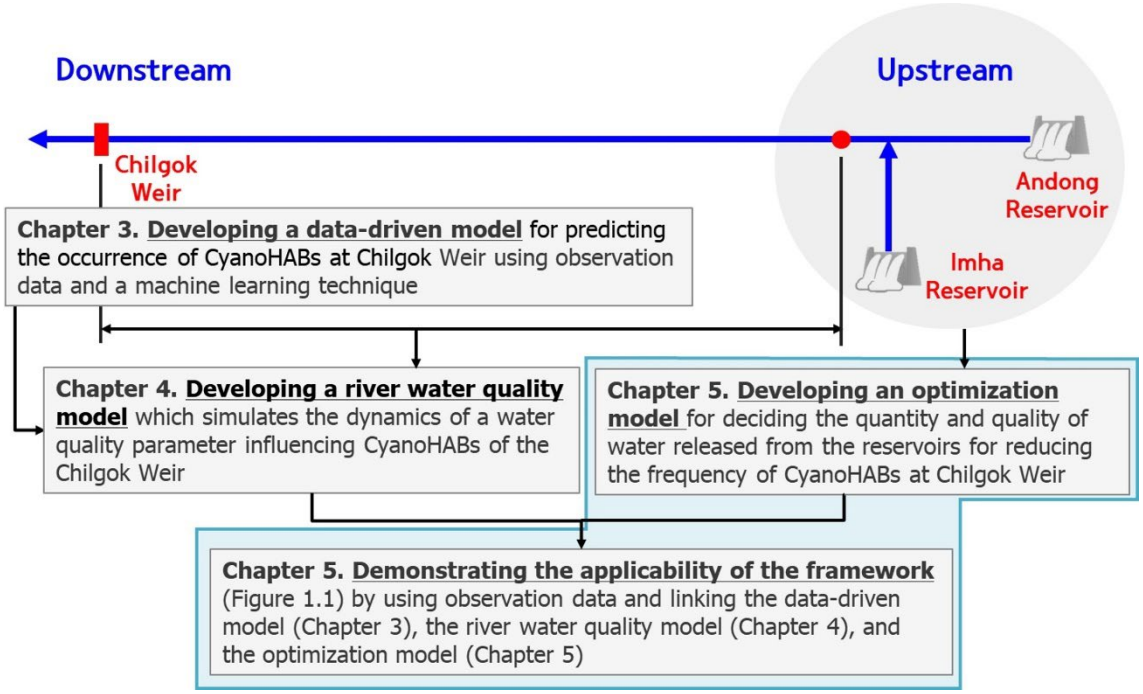
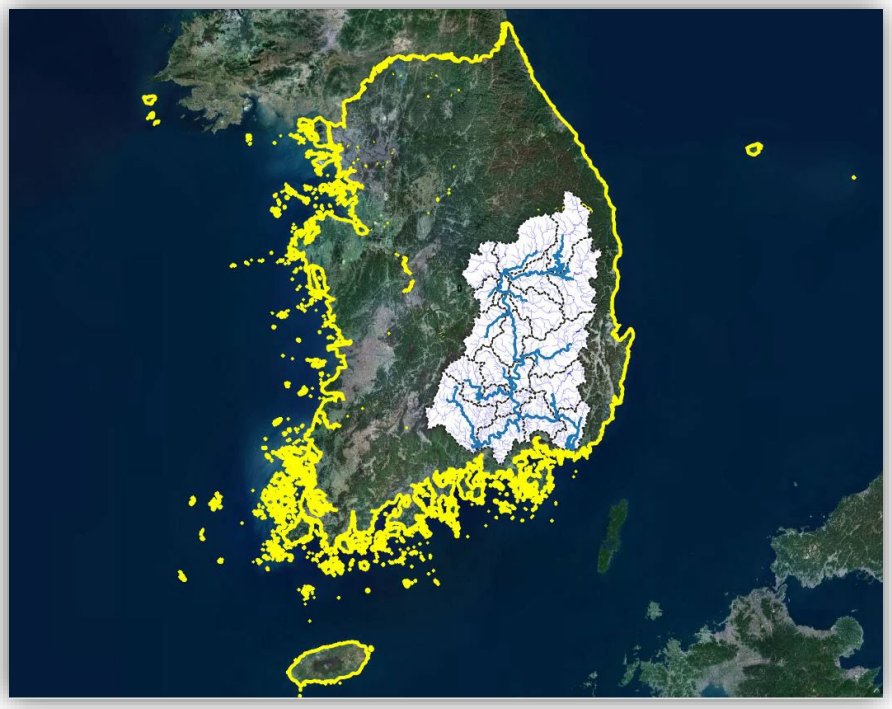


Figure 1.2. Outline of Chapters 3 to 5

2

DESCRIPTION OF THE STUDY AREA

This chapter describes the study area (the upper reach of the Nakdong River) including the principal reasons for the selection of the study area and the data availability for the study area.



Parts of this chapter are based on the following research papers:

Kim, J., Jonoski, A., Solomatine, D. P., and Goethals, P. L. M.: Water quality modelling for nitrate nitrogen control using HEC-RAS: Case study of Nakdong River in South Korea, *Water*, 15, doi:10.3390/w15020247, 2023.

Kim, J., Jonoski, A., Solomatine, D. P., and Goethals, P. L. M.: Decision support framework for optimal reservoir operation to mitigate cyanobacterial blooms in rivers, submitted to *Sustainability* in 2023 (under review).

2.1 INTRODUCTION

In South Korea, the average annual precipitation over the past 52 years (1967–2018) is 1252mm, of which 55.4% is concentrated in summer. In addition, since 63% of the land area is mountainous, rainwater fast flows into the sea due to the steep slopes (Korean Government, 2021). In these unfavorable conditions of water resources management, South Korea has been faced with problems with water quality such as CyanoHABs in rivers, in particular since 2012 after the Four Major Rivers Restoration Project (Song and Lynch, 2018). Thus, the Korean government has been trying to implement policies on water resources management, in particular for improving water quality.

By introducing the Framework Act on Water Management in 2019, the importance of water quality management in South Korea has been asserted more than before. This Act enabled the implementation of a policy on integrated water resources management concerning both the stability of water supply and the improvement of water quality. Before the Act was brought in, the Ministry of Land, Infrastructure, and Transport and the Ministry of Environment were responsible for the quantity and quality of water, respectively. Under the Act, the Ministry of Environment became a government agency to manage water resources in terms of both the quantity and quality of water in South Korea (Lee, 2019).

In consideration of the importance of water quality management in South Korea, this dissertation can propose a practical measure for improving the water environment in a river. By connecting modelling and optimization tools, this measure is aimed at efficiently using limited water resources in reservoirs and effectively reducing the incidence of CyanoHABs. The applicability of this measure can be demonstrated using observation data for a study area. In South Korea, the Nakdong River is applicable as the study area because it has the largest number of large reservoirs and weirs compared with other major rivers. Furthermore, the Nakdong River has confronted the problem of CyanoHABs.

2.2 STUDY AREA

The Nakdong River is the longest in South Korea, with a length of 510 km. The water quality of the Nakdong River has been a matter of concern to environmental authorities since the Nakdong River has been used as a major source of drinking water in adjacent cities (Lee et al., 2018). The special importance of water quality management in the Nakdong River has arisen from the phenol spill accident that happened in 1991 (Kim et al., 1994). Moreover, research studies have investigated quantitative changes in the water quality of the Nakdong River since 2012 after the Four Major Rivers Restoration Project (Jo et al., 2022; Lee et al., 2018; Park et al., 2018b).

In the Nakdong River basin, there are ten multipurpose dams and eight weirs as shown in Figure 2.1. The upper reach of the Nakdong River was selected as a study area, covering the Andong Reservoir, the Imha Reservoir, and the Chilgok Weir. The Andong Reservoir is situated furthest upstream, while the Imha Reservoir is located in the Banbyeoncheon River, a tributary of the Nakdong River. The Chilgok Weir is located approximately 140 km downstream from the Andong Dam.

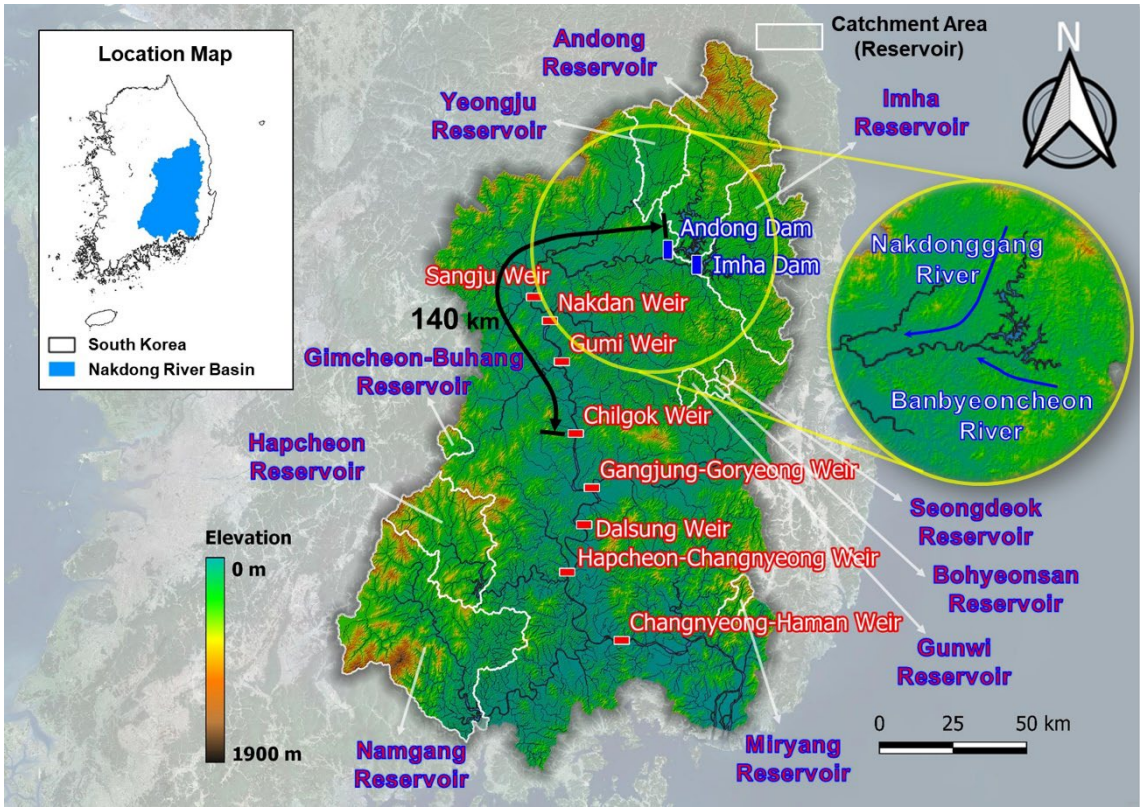


Figure 2.1. Location and schematization of the study area

The Andong Reservoir and the Imha Reservoir are connected by a water transfer tunnel for joint operation in terms of water supply, flood control, and water quality management (Jeong et al., 2020; Park et al., 2017a). The Imha Reservoir in particular has a Selective Withdrawal Facility (SWF), so the water quality can be controlled when the water in the reservoir is released downstream (Lee et al., 2007). Figure 2.2 shows the water transfer tunnel and the SWF. Table 2.1 shows the details of the Andong and Imha reservoirs (Park and Chung, 2014).

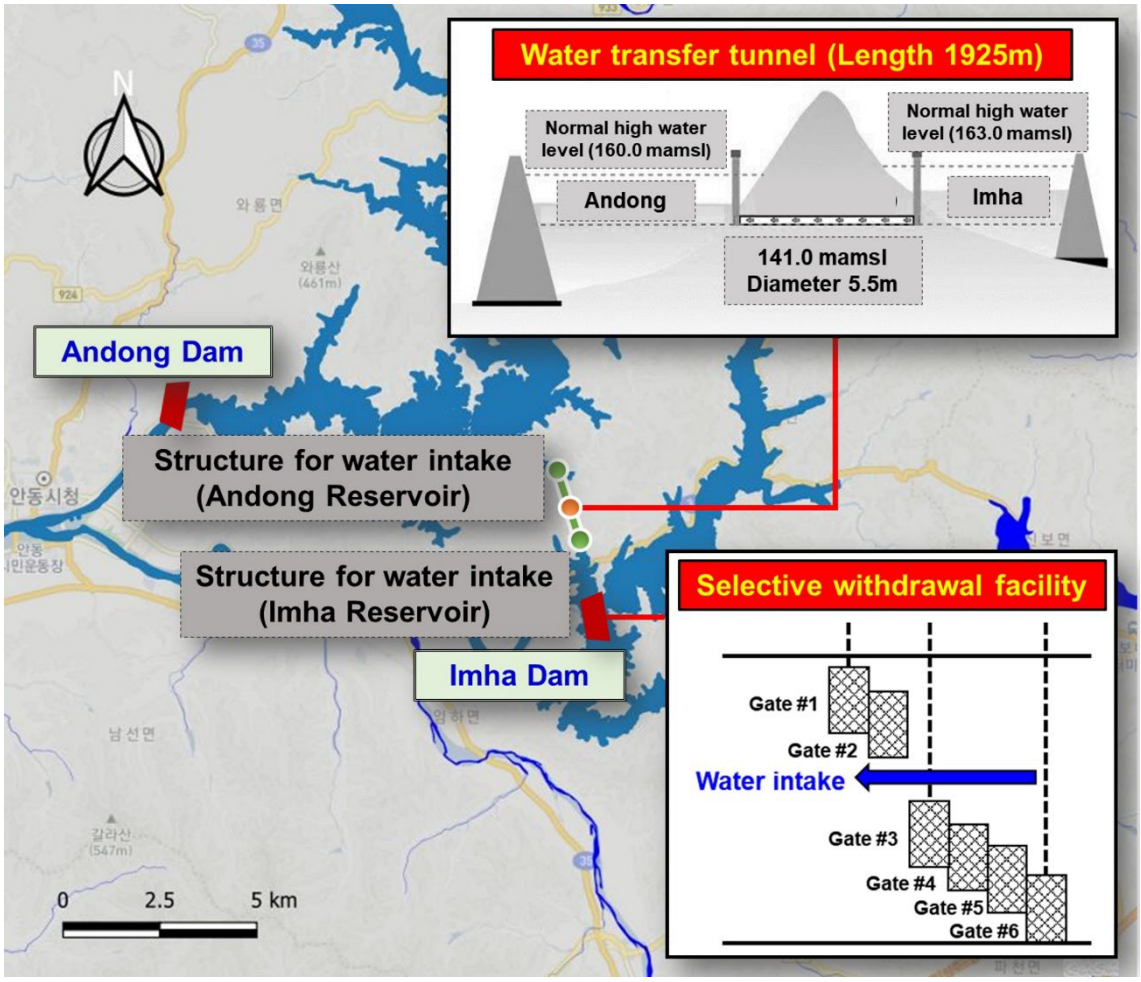


Figure 2.2. Water transfer tunnel and Selective Withdrawal Facility (SWF)

Table 2.1. Details about Andong and Imha reservoirs

Reservoir	Andong	Imha
Area of catchment (km^2)	1584	1361
Height of dam (m)	83.0	73.0
Length of dam (m)	612.0	515.0
Normal high water level (mamsl)	160.0	163.0
Effective storage volume (10^6 m^3)	1000.0	424.0

There are four weirs in the study area, including the Sangju Weir, the Nakdan Weir, the Gumi Weir, and the Chilgok Weir. Given that the intake facilities for drinking water are located between two cross sections of the Gumi Weir and the Chilgok Weir (Lee et al., 2014), the water quality for this district should be managed properly. Table 2.2 shows the details of the four weirs (Bae and Seo, 2021; Jo et al., 2022). The water level of each weir is usually maintained at each water level specified for management (Kim and Shin, 2021) through the operation of the gates.

Table 2.2. Details about Sangju, Nakdan, Gumi, and Chilgok weirs

Weir	Sangju	Nakdan	Gumi	Chilgok
Area of catchment (km^2)	7407	9221	9557	11040
Height (m)	11.0	11.5	11.0	11.8
Length (m)	335.0	286.0	374.3	400.0
Water level for management (mamsl)	47.0	40.0	32.5	25.5
Storage volume (10^6 m^3)	27.4	34.7	52.7	75.3

There are three principal reasons for the selection of this study area. First, a joint operation can be conducted for the two reservoirs. This joint operation makes the amount of the water supply from each reservoir flexible in conditions of meeting the sum of water demand of both reservoirs. Secondly, the Imha Reservoir is equipped with an SWF (Lee et al., 2007; Kim et al., 2022b), which enables the control of the quality of water released to the downstream river. This SWF is one of the important factors in this study in terms of improving the aquatic environment downstream. Finally, the Chilgok Weir is close to the intake facilities for drinking water (Lee et al., 2014), making the management of water quality at Chilgok Weir a critical issue. A monitoring station for water quality data, including cyanobacterial cell density, is located 500 m upstream of the Chilgok Weir (Park et al., 2021a).

2.3 Data Availability for the Study Area

For the studies of this dissertation, data related to water quantity, water quality, and climate of the study area are necessary. These data can be collected from the Water Resources Management Information System, the Water Environment Information System, and the Open MET Data Portal of South Korea (Kim et al., 2021). The Act on the Investigation, Planning, and Management of Water Resources states that the institutions

dedicated to hydrological investigations have to operate information systems to efficiently manage data for water resources. The Ministry of Environment forms a national network to periodically monitor water quality and manages water quality data through an information system under the Water Environment Conservation Act. The Korea Meteorological Administration runs an information system for meteorological data and provides the data to citizens under the Weather Act. All data are publicly available from the information systems operated under these Acts.

Korea Water Resources Corporation (K-water) has operated multi-purpose reservoirs and large weirs. To effectively operate the reservoirs and the weirs, K-water acquires and manages the data such as water level, inflow, outflow, and rainfall at 10-minute, hourly, and daily intervals for the reservoirs and the weirs. All observational data are open to the public on the website of K-water and shared with the Water Resources Management Information System.

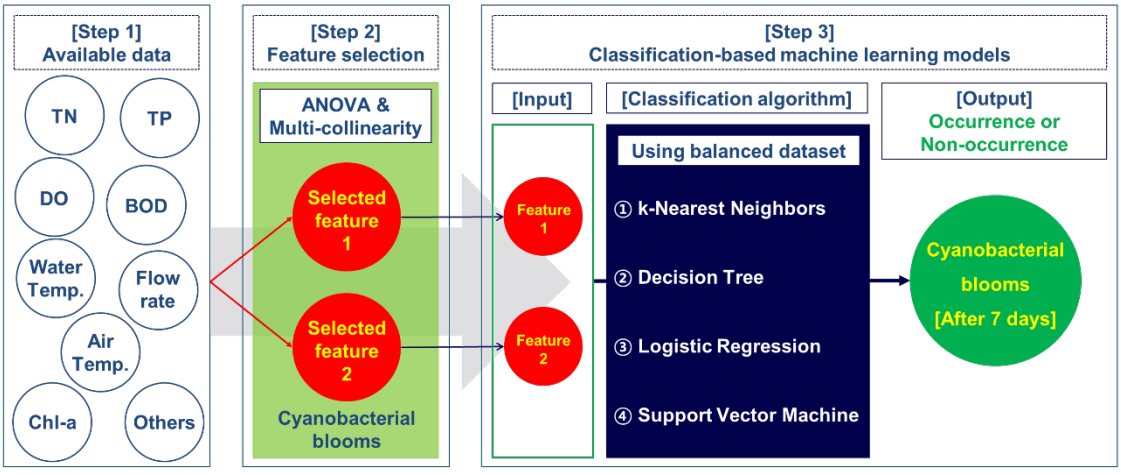
Obtaining water quality data is more difficult than collecting data on water quantity because most water quality data are generally collected in situ and obtained in laboratory experiments. In South Korea, the National Institute of Environmental Research (NIER) monitors the conditions of water quality and aquatic ecosystems in public water bodies. In addition, these data are available from the Water Environment Information System.

While hydrological or hydraulic data and meteorological data can be collected on a daily basis, water quality data are available on a weekly (48 or 36 times a year) or monthly basis (Kim et al., 2021). To resolve the problem of difference in the time interval between data, weekly data can be interpolated to convert them into daily data by using an interpolation method such as a step function (James, 2016; Mcintyre and Wheater, 2004).

3

A MACHINE LEARNING APPROACH TO THE PREDICTION OF CYANOBACTERIAL BLOOMS

This chapter presents how to develop machine learning models with high performance using two input features for predicting the occurrence of cyanobacterial blooms at Chilgok Weir of the Nakdong River, South Korea.



This chapter is extracted from Kim, J., Jonoski, A., and Solomatine, D. P.: A classification-based machine learning approach to the prediction of cyanobacterial blooms in Chilgok Weir, South Korea, *Water*, 14, doi:10.3390/w14040542, 2022.

ABSTRACT

Cyanobacterial blooms appear by complex causes such as water quality, climate, and hydrological factors. This study aims to present machine learning models to effectively and efficiently predict occurrences of these complicated cyanobacterial blooms. The dataset was classified into two, three, or four classes based on cyanobacterial cell density after a week, which was used as the target variable. We developed 96 machine learning models for the Chilgok Weir using four classification algorithms: k-Nearest Neighbor, Decision Tree, Logistic Regression, and Support Vector Machine. In the modelling methodology, we first selected input features by applying ANOVA (Analysis of Variance) and solving a multi-collinearity problem as a process of feature selection, which is a method for removing features irrelevant to a target variable. Next, we adopted an oversampling method to resolve the problem of having an imbalanced dataset. Consequently, the best performance was achieved for models using datasets divided into two classes, with an accuracy of 80% or more. Comparatively, we confirmed low accuracy of approximately 60% for models using datasets divided into three classes. Moreover, while we produced models with overall high accuracy when using logCyano (logarithm of cyanobacterial cell density) as a feature, several models in combination with air temperature and NO_3-N (nitrate nitrogen) using two classes also demonstrated more than 80% accuracy. We concluded that accurate classification-based machine learning models could be developed with two input features related to cyanobacterial blooms. This proved that we could make effective and efficient models with a low number of inputs.

3.1 INTRODUCTION

Harmful Algal Blooms (HABs) have appeared due to pollution of aquatic environments, and increasingly due to climate change, which has been a cause for the increase in water temperature (Tong et al., 2021; Park et al., 2021a). There are increasing concerns that the combined environmental factors of uncontrolled pollution and climate change (particularly higher temperatures) may lead to more frequent and more severe HABs (Kosten et al., 2012; Paerl and Huisman, 2009; Paerl and Scott, 2010). HABs have been negatively affecting not only the aquatic environment but also human health because they produce toxic substances (Smith and Daniels, 2018) such as microcystin (Plaas and Paerl, 2021; Ho and Goethals, 2020). The serious problems of HABs can be recognized through studies that showed that algal blooms (or cyanobacterial blooms) caused fish death (Kim et al., 2020; Song and Lynch, 2018) and human liver disease (Lee et al., 2019). The challenges for water management in preventing or minimizing HABs are linked to the complexity of the HAB processes (including identification of main conditioning factors), their site-specificity, and associated difficulties in their prediction (Paerl, 2017; Wurtsbaugh et al., 2019).

Cyanobacterial blooms appear when phytoplankton proliferates massively in lentic water such as lakes, reservoirs, or ponds, owing to eutrophication and stratification (Park et al., 2021a; Ahn et al., 2021). In other words, an increase in the nutrients and a rise in the water temperature of stagnant water bodies can bring about cyanobacterial blooms (Park et al., 2017b). In this regard, many people in South Korea have argued that the weirs, which were built during the Four Major Rivers Restoration Project from 2009 to 2012, have decreased the flow velocity in the rivers (Park et al., 2021a), causing the appearance of cyanobacterial blooms (Park et al., 2017b) and the environmental problems such as water pollution and ecological disturbance (Song and Lynch, 2018).

As part of the solution to the HABs problem, various studies were carried out on identifying the cause of cyanobacterial blooms and predicting their occurrence (Rousso et al., 2020). Many previous studies revealed that water quality factors (e.g., water temperature, suspended solids, dissolved oxygen, nutrients such as nitrogen compounds, phosphorus, etc.), climatic conditions (e.g., air temperature, rainfall, etc.), and hydrological factors (e.g., discharge, water level, etc.) were significant causes of the cyanobacterial blooms (Park et al., 2017b; Rousso et al., 2020). However, finding only one or two specific causes of cyanobacterial blooms is not easy because they appear by complicated biological processes in addition to external factors (Kim et al., 2017; Park et al., 2017b). In order to predict the occurrence of these cyanobacterial blooms, many previous studies have applied data-driven models, using Artificial Neural Networks (ANN), Decision Tree (DT), etc., as well as process-based models such as DYRESM/ELCOM-CAEDYM (Rousso et al., 2020).

Numerous earlier studies attempted to predict cyanobacterial blooms accurately by developing process-based models that mathematically provide the mechanism of the blooms (Rousso et al., 2020). Nevertheless, the process-based models require considerable input (Rousso et al., 2020) and computing time (Yang et al., 2021) as they all involve related factors such as water quality, climate, and flow rate. On the other hand, data-driven models using machine learning or deep learning produce output by taking less running time (Yang et al., 2021) and only some main factors (Rousso et al., 2020). Some research proved that the data-driven models employing techniques such as Random Forest (RF) (Zeng et al., 2017; Yajima and Derot, 2018), Support Vector Machine (SVM) (Zeng et al., 2017), ANN (Zeng et al., 2017), and Extreme Learning Machine (ELM) (Yi et al., 2018) ensured high accuracy in predicting the real-valued output such as cyanobacterial cell density (Yajima and Derot, 2018) or Chlorophyll-a concentration (Chl-a), which is a proxy index for the cyanobacterial blooms (Kim et al., 2017; Yajima and Derot, 2018; Yi et al., 2018). Additionally, recent studies were conducted on the machine learning models that forecast the cyanobacterial blooms in the type of the binary (Kim et al., 2020) (e.g., occurrence/non-occurrence) or the ordinal data (Mellios et al., 2020; Park et al., 2021b) (e.g., low/medium/high) using classification methods.

However, for those classification-based machine learning models, we need to consider at least two prerequisites, which were often overlooked in earlier studies. The first one is to pre-select input features of the model based on the theoretical knowledge regarding cyanobacterial blooms. By going through the process of this feature selection, we can improve the efficiency and accuracy of the model (Yajima and Derot, 2018; Gnana et al., 2016). Nonetheless, feature selection that is not derived from the physical or biological processes related to target variables may give poor performances to the models (Jiang et al., 2020; Moreido et al., 2021; Rousso et al., 2020). Moreover, the pre-selection of features without considering the statistical characteristics such as multi-collinearity can be an obstacle to developing a robust model (Al-Abadi et al., 2020; Yoo and Cho, 2019).

The other prerequisite is a balanced dataset, which is essential for high performances of the classification models (Shin et al., 2017) using nominal or ordinal data (Raschka and Mirjalili, 2017). The balanced dataset ensures the even distribution of two or more classification data without being biased toward one classification. In the raw dataset of cyanobacterial blooms, non-occurrence data generally outweigh the occurrence data (Kim et al., 2020; Shin et al., 2017; Choi et al., 2019). Therefore, the performance of the models tends to become low if the imbalanced dataset of the cyanobacterial blooms is used as it is (Shin et al., 2017). We need to correct the imbalance of the dataset with an oversampling technique (Choi et al., 2019).

The main objective of this study is to develop optimal classification-based machine learning models for effectively and efficiently predicting occurrences of cyanobacterial blooms through the process of feature selection and the oversampling of datasets. Specifically, we (i) derive significant input features using the datasets of a specific point called Chilgok Weir in South Korea, (ii) present which processes need to take place to reduce the number of input features as much as possible, which is required to develop machine learning models efficiently, (iii) identify how the target variables should be classified to improve model performance, and (iv) find out the optimal combinations of input features and four classification algorithms such as k-Nearest Neighbor (k-NN), DT, Logistic Regression (LR), and SVM. The concrete results from this research will introduce some novelty for decision-makers, who need intuitive and effective strategies for dealing with this problem, using models that can predict cyanobacterial blooms as accurately as possible with a few features. For example, decision-makers in charge of reservoir operations will be able to release more and cleaner water from an upstream reservoir if our model suggests that cyanobacterial blooms will appear in a downstream river a week later.

3.2 MATERIALS

3.2.1 Data collection

We selected the Chilgok Weir, completed in June 2012, as the target location for this study (see Figure 2.1). There are two reasons for the selection of this Chilgok Weir. The first reason is the availability of the datasets related to the cyanobacterial blooms of the Chilgok Weir. The Algae Alert System in South Korea (Park et al., 2021a), based on cyanobacterial cell density as shown in Table 3.1 (Park et al., 2021b; Park et al., 2021a; Kim et al., 2020), has been operated by the National Institute of Environmental Research (NIER) to ensure the safety of drinking water (Park et al., 2021b). One of the observation stations is at the upstream point about 500 m away from the Chilgok Weir (Park et al., 2021a). The second is the location of the observation station. The station of the Chilgok Weir is located furthest upstream among three stations involved in the Algae Alert System on the mainstream of the Nakdong River. The location of the station enables us to consider as few factors as possible influencing the cyanobacterial blooms. In general, there are more factors in the downstream points affecting the occurrence of cyanobacterial blooms, such as the inflow of pollutants from tributaries or sewage treatment plants (Yi et al., 2018; Park et al., 2021a).

Table 3.1. Criteria for algae alert in South Korea

Stage	Cyanobacterial Cell Density (cells mL^{-1})
Caution	≥1000
Warning	≥10,000
Outbreak	≥1,000,000

For the Chilgok Weir point, we acquired datasets including water quality factors, climatic conditions, and hydrological factors known as the causes (or influencing factors) of the cyanobacterial blooms. Cyanobacteria-related water quality data, meteorological data, and hydrological data are open to the public by NIER, Korea Meteorological Administration (KMA), and Korea Water Resources Corporation (K-water), respectively (Kim et al., 2020; Park et al., 2021a; Park et al., 2017b; Yi et al., 2018; Park et al., 2021b; Shin et al., 2017; Kim et al., 2019a; Ahn et al., 2021). Table 3.2 shows the feature, the frequency, and the source of each dataset. Regarding the cyanobacteria-related water quality, the harmful cyanobacteria include four genera: *Microcystis*, *Aphanizomenon*, *Anabaena*, and *Oscillatoria* spp. (Park et al., 2021a; Kim et al., 2020; Park et al., 2021b).

Table 3.2. List of features

Data/Frequency /Source	Feature	Description	Unit
Water quality data/Weekly /NIER	Cyano	Cyanobacterial cell density	cells mL^{-1}
	WT	Water temperature	℃
	pH	Hydrogen ion concentration	-
	DO	Dissolved oxygen	mg L^{-1}
	Chl-a	Chlorophyll a	mg m^{-3}
	BOD	Biochemical oxygen demand	mg L^{-1}
	COD	Chemical oxygen demand	mg L^{-1}
	SS	Suspended solids	mg L^{-1}
	TN	Total nitrogen	mg L^{-1}
	TP	Total phosphorus	mg L^{-1}
	N/P	TN/TP ratio	-
	TOC	Total organic carbon	mg L^{-1}
	EC	Electrical conductivity	μS cm^{-1}
	TotalColiform	Total coliforms	100 mL^{-1}
	TDN	Total dissolved nitrogen	mg L^{-1}
	NH_3-N	Ammonia nitrogen	mg L^{-1}
	NO_3-N	Nitrate nitrogen	mg L^{-1}
	TDP	Total dissolved phosphorus	mg L^{-1}
	PO_4-P	Phosphate phosphorus	mg L^{-1}
	FecalColiform	Fecal coliforms	-
Meteorological data/Daily /KMA	AT	Average air temperature	℃
	LT	Lowest air temperature	℃
	HT	Highest air temperature	℃
	MaxSolarRad	Maximum amount of solar radiation for one hour	MJ m^{-2}
	DaySolarRad	Total amount of solar radiation	MJ m^{-2}
Hydrological data/Daily /K-water	WeirLevel	Water level of weir	mamsl
	StorageVolume	Storage volume of weir	10^6 m^3
	Rainfall	Rainfall in weir catchment area	mm
	Inflow	Weir inflow	m^3 s^{-1}
	Outflow	Weir outflow	m^3 s^{-1}

3.2.2 Data preprocessing

Across a range of regions, there were previous studies in which the forecast horizons were set from real-time to as long as one month or more depending on the objective of developing the models and the frequency of the used datasets (Rousso et al., 2020). In this study, we used a week as the forecast horizon because the frequency of the cyanobacteria-related water quality dataset is on a weekly basis as shown in Table 3.2. The raw dataset consisted of 378 instances with the cyanobacterial cell density after a week (Cyano(t+1)) as a target variable and 30 input features including the current cell density (Cyano(t)) as shown in Table 3.3. The period for the used data was from August 2012 to December 2020.

Table 3.3. Nine-year mean, minimum, median, and maximum values for each feature in the raw dataset (378 instances)

Category	Feature	Mean	Minimum	Median	Maximum
Input features	Cyano(t)	2976	0	165	112,735
	WT	16.8	0.7	17.5	33.6
	pH	8.1	6.5	8.1	9.6
	DO	10.4	1.6	10.1	16.4
	Chl-a	20.1	2.3	15.45	87.2
	BOD	1.9	0.4	1.8	5.0
	COD	5.9	3.5	5.8	10.5
	SS	7.6	1.5	6.3	44.9
	TN	2.674	1.089	2.686	4.396
	TP	0.043	0.011	0.034	0.198
	N/P	81.5	12.7	76.6	255.5
	TOC	4.1	2.6	4.0	7.9
	EC	288	124	286	596
	TotalColiform	8219	2	264	340,000
	TDN	2.513	1.078	2.532	4.125
	NH_3-N	0.113	0.003	0.091	0.790
	NO_3-N	1.971	0.530	1.996	3.330
	TDP	0.024	0.003	0.018	0.125
	PO_4-P	0.011	0.000	0.004	0.105

Table 3.3. Cont.

Category	Feature	Mean	Minimum	Median	Maximum
Input features	FecalColiform	428	0	12	21,750
	AT	15.5	–4.8	15.8	32.5
	LT	10.8	–8.9	10.8	27.6
	HT	21.0	–1.0	21.9	38.0
	MaxSolarRad	2.39	0.19	2.54	3.74
	DaySolarRad	15.56	0.69	15.465	31.02
	WeirLevel	25.52	25.02	25.56	25.86
	StorageVolume	75.321	68.181	75.930	79.005
	Rainfall	2.290	0.000	0.023	57.263
	Inflow	112.867	3.604	67.733	1147.669
	Outflow	113.379	8.098	69.004	1140.136
Target variable (Output feature)	Cyano(t+1)	2903	0	165	112,735

The machine learning models applying classification algorithms require a nominal or an ordinal data type for target variables (Raschka and Mirjalili, 2017). In order to compare the performance of each model depending on the number of classes, we made three groups by classifying the target variable (Cyano(t+1)) based on the Algae Alert System, as presented in Table 3.1. For the first group, the dataset was classified into four classes (Normal, Caution, Warning, and Outbreak), which was the same as the Algae Alert System, and it was named Group1. It had an imbalanced dataset as it comprised 269 Normals, 83 Cautions, 26 Warnings, and zero Outbreaks. We made the other two groups (Group2 and Group3) by dividing the dataset into two classes (e.g., (Kim et al., 2020)) (Normal/Occurrence) for Group2 and three classes (e.g., (Mellios et al., 2020; Park et al., 2021b)) (None/Normal/Occurrence) for Group3. As a result, Group1, Group2, and Group3 consisted of four, two, and three classes, respectively. We used these three groups to ensure which classification of the cyanobacterial cell density provided us with a better model with reference to performance. Table 3.4 shows how each group was specified in terms of cyanobacterial cell density.

Table 3.4. Classification framework for each group

Group	Class	Cyano(t+1)	Number
Group1	Normal	<1000	269
	Caution	≥1000	83
	Warning	≥10,000	26
	Outbreak	≥1,000,000	0
Group2	Normal	<1000	269
	Occurrence	≥1000	109
Group3	None	0	136
	Normal	<1000	133
	Occurrence	≥1000	109

In addition, the logarithmic transformation of base 10 was applied to Cyano(t) to convert the skewed distribution of the raw dataset into normal distribution as much as possible (Choi et al., 2019). Consequently, it was named as logCyano. Here, when Cyano(t) was zero, logCyano was also given zero. Furthermore, we applied standardization to 30 input features according to Equation (3.1). It is one of the feature scaling methods, which is an important preprocessing task in machine learning modelling (Vien et al., 2021; Raschka and Mirjalili, 2017).

$$x' = \frac{x - \mu}{\sigma} \qquad (3.1)$$

where x' is a standardized value, x is an observation data for a specific feature from the raw dataset, μ is the mean of the whole observation data for the feature, σ is its standard deviation.

Table 3.5 shows the values of the input features after preprocessing using the logarithmic transformation and the standardization of input features, together with the classified target variable.

Table 3.5. Input features (including the mean, minimum, median, and maximum values for each feature) after preprocessing the dataset using logarithmic transformation and standardization, together with the classified target variable

Category	Feature	Mean	Minimum	Median	Maximum
Input features	logCyano	0.000	–1.200	0.224	2.046
	WT		–1.974	0.085	2.071
	pH		–3.340	0.032	3.193
	DO		–3.115	–0.077	2.144
	Chl-a		–1.194	–0.310	4.514
	BOD		–2.066	–0.148	4.237
	COD		–2.316	–0.100	4.430
	SS		–1.240	–0.280	7.533
	TN		–2.451	0.018	2.661
	TP		–1.219	–0.329	6.019
	N/P		–1.602	–0.114	4.048
	TOC		–1.784	–0.133	4.468
	EC		–2.312	–0.041	4.324
	TotalColiform		–0.308	–0.299	12.455
	TDN		–2.357	0.031	2.646
	NH_3-N		–1.213	–0.243	7.458
	NO_3-N		–2.425	0.042	2.288
	TDP		–1.045	–0.307	4.957
	PO_4-P		–0.570	–0.355	5.068
	FecalColiform		–0.252	–0.245	12.520
	AT		–2.286	0.024	1.907
	LT		–2.182	0.000	1.861
	HT		–2.372	0.088	1.826
	MaxSolarRad		–2.567	0.171	1.569
	DaySolarRad		–2.103	–0.014	2.186
	WeirLevel		–3.042	0.219	2.030
	StorageVolume		–3.250	0.277	1.677
	Rainfall		–0.349	–0.346	8.381
	Inflow		–0.751	–0.310	7.111
	Outflow		–0.723	–0.305	7.054
Target variable	Each class of three groups (Group1, Group2, and Group3) based on Cyano(t+1)				

3.3 METHODS

3.3.1 Analysis of variance (ANOVA) for feature selection

To build robust learning models, we need feature selection, leading to the elimination of redundant and irrelevant features (Gnana et al., 2016). This helps prevent overfitting, enhance model performance, and increase the running speed of a model (Gnana et al., 2016). The filter approach, one of the feature selection methods (Gnana et al., 2016), uses techniques such as ANOVA, which is widely used in statistical studies. One-way ANOVA determines whether differences between two or more classes are statistically significant through a comparison of variances between classes (Gradilla-Hernandez et al., 2020). Through one-way ANOVA, we can judge that the difference is significant when the F value is large enough and the p-value is less than 0.05 (Gradilla-Hernandez et al., 2020; Peng et al., 2020). In this study, features with valid F and p values (F value > 50, p-value < 0.05) by one-way ANOVA were selected to develop the machine learning models for three groups formed by using the categorical variables (Wu et al., 2021). This is because significant F and p values mean the features have a high correlation with the categorical target variable.

3.3.2 Multi-collinearity

Multi-collinearity arises when the inter-correlation between input features is strong (Yoo and Cho, 2019; Al-Abadi et al., 2020). It can be a problem in statistical analysis such as regression as it distorts the prediction results of the model (Yoo and Cho, 2019; Al-Abadi et al., 2020). For classification-based machine learning, the multi-collinearity problem can be addressed as part of feature selection (e.g., (Al-Abadi et al., 2020; Xu et al., 2021; Zhou et al., 2021; Nagawa et al., 2021)). In this study, features with weak inter-correlation are candidates to be selected. To be specific, the features with high inter-correlation are removed after correlation analysis using all the features selected through one-way ANOVA. As a result, we could achieve the purpose of feature selection, such as warding off the overfitting of the model, by having only the minimum number of features (Tousi et al., 2021).

3.3.3 Classification algorithms of machine learning

For this study, we applied four classification-based machine learning algorithms, k-NN, DT, LR, and SVM, which are widely used (Kim and Oh, 2021). The k-NN is a distance-based classification algorithm that finds the ‘k’ neighbors, which are closest to the data to be classified. The target data are allocated the same label as the closest neighbors (Mellios et al., 2020). The DT is a technique of classifying data based on the impurity of training data, such as the Gini index (Shin et al., 2017) and the entropy (Uma and

Balamurugan, 2020). The LR is a classification method that uses logit functions to predict the probability that data fall into a category between zero and one (Bourel and Segura, 2018). Multinomial logistic regression, an extended form of LR, allows multiple classes to be applied (Bourel and Segura, 2018). The SVM is a machine learning algorithm that determines the optimal hyperplane to maximize the distance between the categories. The class of new data is determined by the hyperplane (Mellios et al., 2020). These four machine learning techniques can be implemented using scikit-learn, one of Python's machine learning libraries (Raschka and Mirjalili, 2017).

3.3.4 Oversampling using SMOTE (Synthetic Minority Oversampling Technique)

As shown in Table 3.4 regarding the classification frameworks of three groups (Group1, Group2, and Group3), the dataset to be used in this study had an imbalance by class in all three groups. While the application of machine learning using such an imbalanced dataset gives rise to overfitting by excessively increasing prediction accuracy for the majority class (Ahmed et al., 2021), it may make an inaccurate prediction for the minority class (Shin et al., 2017). To overcome the problem of an imbalanced dataset, oversampling can be applied, leading to improved prediction accuracy for minority classes. It is a process of producing new data of minority classes equal to the number of data belonging to a majority class (Choi et al., 2019).

One of the widely used oversampling techniques is SMOTE (Fernandez et al., 2018). It is a method of synthesizing the interpolated points on a line connecting the adjacent groups of a minority class in a training set and labeling them as new samples of the minority class (Fernandez et al., 2018; Choi et al., 2019; Shin et al., 2017). Shin et al. (2017) and Choi et al. (2019) suggested that the cyanobacterial-related models to which SMOTE was applied outperformed those without oversampling. In this study, SMOTE was employed by using a Python library, imblearn (Raschka and Mirjalili, 2017).

3.3.5 Training, cross-validation, and test for the dataset

The dataset should be split into a training set for learning and a test set for verification of the model (Raschka and Mirjalili, 2017). In this study, the split ratio between the training set and the test set was 80%:20%. Additionally, four-fold cross-validation was performed on the training set to prevent the overfitting of the model (Arabgol et al., 2016). At the same time, the optimal parameters for each classification algorithm were found through grid search that could improve the model performance (Arabgol et al., 2016; Raschka and Mirjalili, 2017). Finally, the models built through four-fold cross-validation were evaluated using the test set. The parameters which were optimized in this study are shown in Table 3.6 (Raschka and Mirjalili, 2017).

Table 3.6. Parameters to be optimized in this study for four algorithms

Algorithm	Parameter	Description
k-NN	n_neighbors	Number of neighbors
DT	max_depth	Maximum depth of the tree
LR	C	Regularization parameter
SVM	C	Regularization parameter
	kernel	The kernel type to be used in the algorithm such as 'linear', 'poly', 'rbf', 'sigmoid', etc.

3.3.6 Model evaluation

The metrics that evaluate the performance of the classification model include Accuracy (*ACC*), Precision (*PRE*), Recall (*REC*), and F1-score (*F1*) (Raschka and Mirjalili, 2017; Ahmed et al., 2021; Mulyani et al., 2019). As shown in Figure 3.1 (Raschka and Mirjalili, 2017), we can describe each metric through a confusion matrix schematizing binary classification using True Positive (*TP*), True Negative (*TN*), False Positive (*FP*), and False Negative (*FN*). Accuracy is obtained by dividing the sum of correct predictions (*TP* + *TN*) by the total number of data. Precision represents the ratio of *TP* to the total number of samples predicted to be positive (*TP* + *FP*). Recall indicates the ratio of *TP* to the total number of samples belonging to the actual positive class (*TP* + *FN*). F1-score is expressed as a harmonic mean of Precision and Recall (Mulyani et al., 2019). Precision, Recall, and F1-score are known as the more reliable metrics for an imbalanced dataset than Accuracy (Mulyani et al., 2019). We, however, used Accuracy as a performance evaluation metric in this study. The reasons are that we developed the models using a balanced dataset through SMOTE and we needed to see their accuracy for both Negatives and Positives of the predicted classes. The four metrics are formulated as follows (Raschka and Mirjalili, 2017; Ahmed et al., 2021; Mulyani et al., 2019).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.2}$$

$$PRE = \frac{TP}{TP + FP} \tag{3.3}$$

$$REC = \frac{TP}{TP + FN} \tag{3.4}$$

$$F1 = 2\,\frac{PRE \times REC}{PRE + REC} \tag{3.5}$$

		Predicted class: *N*	Predicted class: *P*
Actual class	*N*	True Negative (*TN*)	False Positive (*FP*)
	P	False Negative (*FN*)	True Positive (*TP*)

Figure 3.1. Confusion matrix

3.3.7 Summary of the modelling procedure

We summarized the modelling procedure as follows in Figure 3.2, based on the above subsections

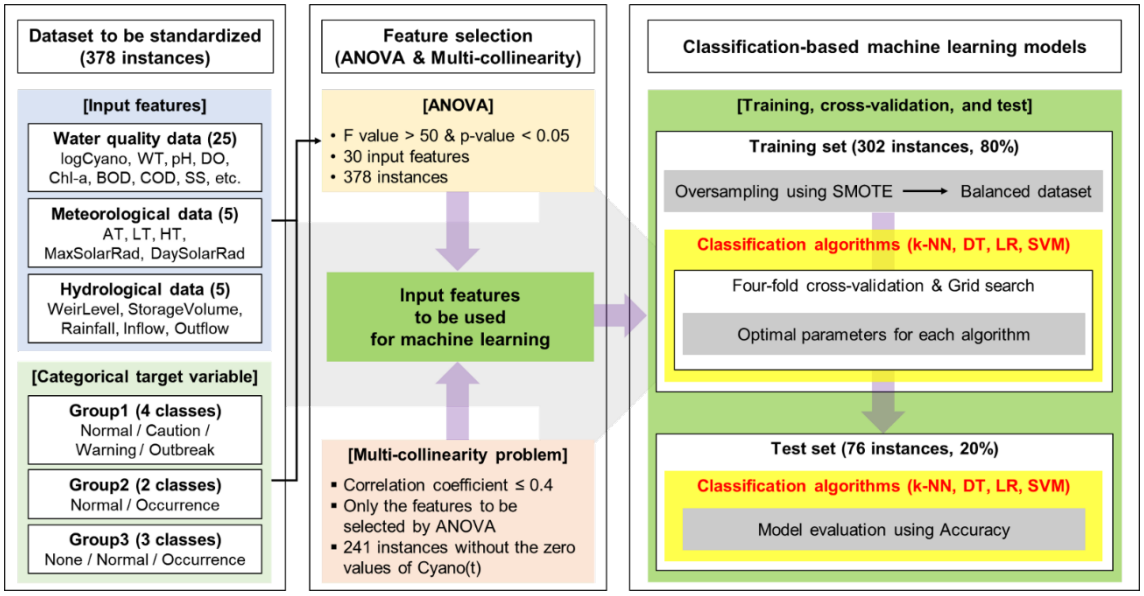


Figure 3.2. Summary of the modelling procedure

i. One-way ANOVA was carried out using 30 features including logCyano of a standardized dataset with 378 instances for three groups (Group1, Group2, and Group3). For the purpose of selecting the features having a strong correlation with the target variable, F values of more than 50 and p-values of less than 0.05 (Peng et al., 2020; Gradilla-Hernandez et al., 2020) were applied. Here, the target variable was a class based on Cyano(t+1) for each group; Normal/Caution/ Warning/Outbreak for Group1 (which was actually divided into three classes because the number of Outbreak elements was zero), Normal/Occurrence for Group2, and None/Normal/Occurrence for Group3.

ii. To address the multi-collinearity problem, a correlation analysis was performed among the features selected in the first step. As the final process for the feature selection, the paired features with low inter-correlation coefficients (0.4 or less (Patil et al., 2020; Zhang et al., 2021)) were selected. Here, Pearson's correlation analysis was performed with only 241 instances by excluding the zero values of Cyano(t) in 378 instances, as the zero values were able to distort the analysis result.

iii. The dataset consisting of the input features selected in the second step and the target variable was split into a training set and a test set by 80% and 20%. Therefore, 302 and 76 out of 378 instances were used as the training set and the test set, respectively. After that, oversampling for the training set was performed (Choi et al., 2019) by applying SMOTE. As a result of the oversampling, the number of instances by class became the same.

iv. Using the balanced datasets of three groups acquired in the third step, four classification-based machine learning algorithms including k-NN, DT, LR, and SVM, were applied. The models with optimal parameters for each machine learning method were constructed through four-fold cross-validation and grid search using the training set.

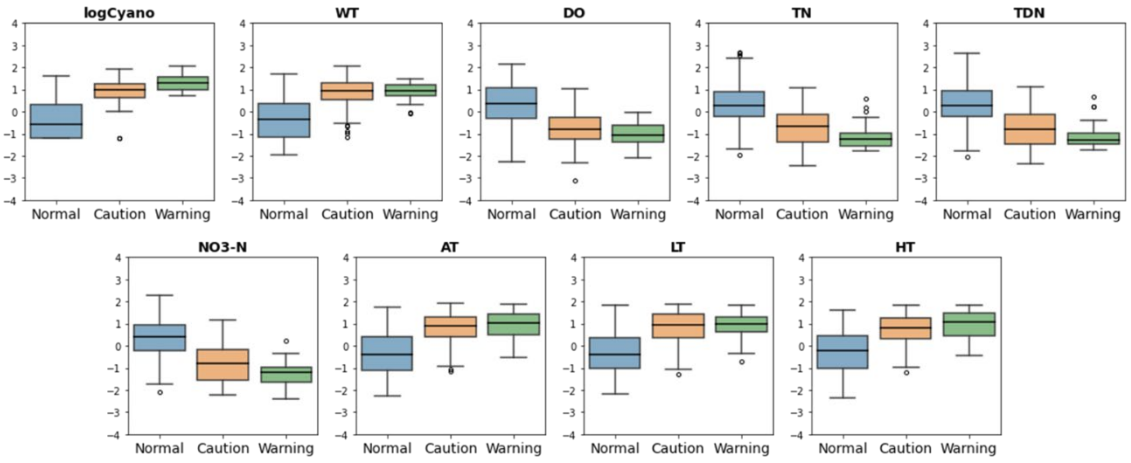
v. The optimal combination of input features and machine learning algorithms for predicting the categorical target variable was presented by evaluating the performance (Accuracy) from the test set using the models developed in the fourth step.

3.4 Results

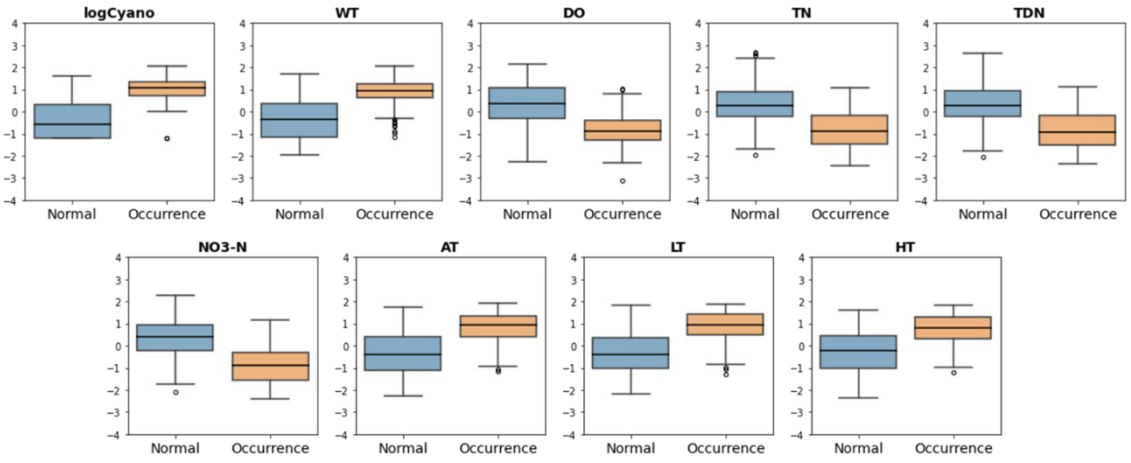
3.4.1 Determination of the modelling cases

We could determine the modelling cases based on the result of feature selection. Through one-way ANOVA, as shown in Table 3.7, nine features such as logCyano, WT, DO, TN, TDN, NO_3-N, AT, LT, and HT had significant F and p values (F value > 50, p-value < 0.05 (Peng et al., 2020; Gradilla-Hernandez et al., 2020)) for the categorical target

variable in all three groups. Figure 3.3 shows the box plots having the data distribution for each group for the selected nine features, which helps intuitively notice the differences between classes. As shown in Figure 3.3, the distinctions between classes in Group2 and Group3 are clear for the nine features, as in the results of Table 3.7. On the other hand, in Group1, the distinction between Normal and Caution is clear, but it is somewhat unclear between Caution and Warning. Nonetheless, we used Group1 to compare with the modelling results for the other two groups. This is because we applied the same classification as the Algae Alert System in Table 3.1 to Group1 and the F and p values for its three classes were significant as shown in Table 3.7.

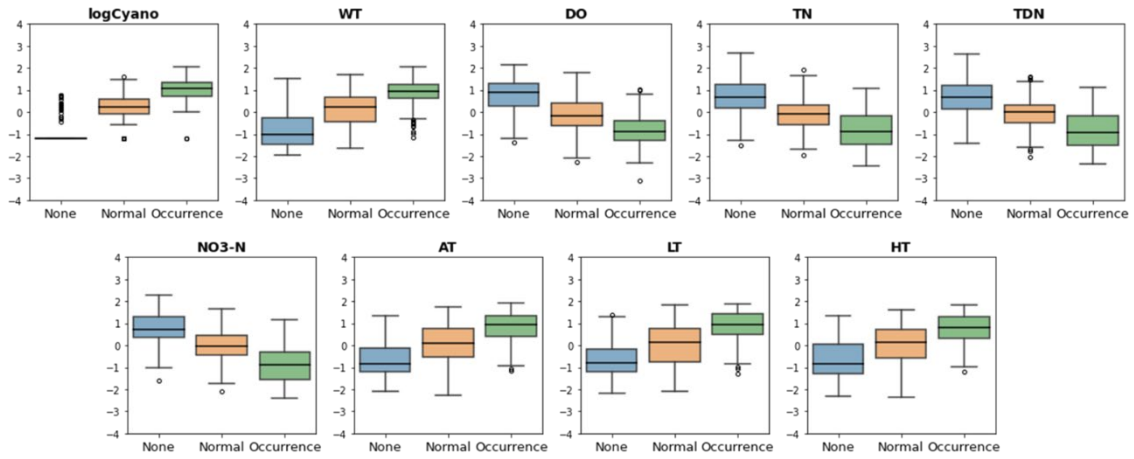


(a) Group1



(b) Group2

Figure 3.3. Cont.



(c) Group3

Figure 3.3. Box plots to show the data distribution between classes for the nine features selected by ANOVA

Table 3.7. F and p values of 30 features

(The nine features with the bold font have significant F and p values.)

Feature	Group1		Group2		Group3	
	F value	*p*-value	F value	*p*-value	F value	*p*-value
logCyano	**132.367**	**<0.001**	**256.089**	**<0.001**	**270.917**	**<0.001**
WT	**71.613**	**<0.001**	**143.214**	**<0.001**	**142.227**	**<0.001**
pH	0.545	0.580	0.313	0.576	6.180	0.002
DO	**74.182**	**<0.001**	**145.698**	**<0.001**	**131.458**	**<0.001**
Chl-a	7.137	0.001	14.118	<0.001	7.637	0.001
BOD	1.463	0.233	2.917	0.088	5.022	0.007
COD	2.599	0.076	5.186	0.023	18.898	<0.001
SS	5.244	0.006	3.924	0.048	4.928	0.008
TN	**63.964**	**<0.001**	**123.352**	**<0.001**	**108.115**	**<0.001**
TP	4.222	0.015	0.951	0.330	6.432	0.002
N/P	19.436	<0.001	38.336	<0.001	40.293	<0.001
TOC	1.499	0.225	1.456	0.228	18.843	<0.001

Table 3.7. Cont.

Feature	Group1		Group2		Group3	
	F value	***p*-value**	**F value**	***p*-value**	**F value**	***p*-value**
EC	6.176	0.002	0.170	0.680	8.701	<0.001
TotalColiform	4.703	0.010	6.984	0.009	7.137	0.001
TDN	**66.039**	**<0.001**	**128.394**	**<0.001**	**103.655**	**<0.001**
NH_3-N	5.961	0.003	11.281	0.001	6.176	0.002
NO_3-N	**85.820**	**<0.001**	**163.285**	**<0.001**	**126.452**	**<0.001**
TDP	3.874	0.022	2.428	0.120	12.020	<0.001
PO_4-P	2.922	0.055	0.594	0.441	8.241	<0.001
FecalColiform	1.754	0.175	3.176	0.076	5.414	0.005
AT	**63.407**	**<0.001**	**126.277**	**<0.001**	**98.519**	**<0.001**
LT	**66.861**	**<0.001**	**133.669**	**<0.001**	**103.961**	**<0.001**
HT	**53.737**	**<0.001**	**106.578**	**<0.001**	**83.166**	**<0.001**
MaxSolarRad	5.712	0.004	9.368	0.002	6.774	0.001
DaySolarRad	4.996	0.007	7.154	0.008	5.754	0.003
WeirLevel	3.047	0.049	4.768	0.030	9.661	<0.001
StorageVolume	2.737	0.066	4.370	0.037	9.695	<0.001
Rainfall	2.256	0.106	0.370	0.543	0.327	0.721
Inflow	3.843	0.022	0.244	0.622	6.501	0.002
Outflow	3.649	0.027	0.148	0.701	6.543	0.002

Although those nine features were highly correlated with the target variable, two features among them had a multi-collinearity problem. This could be solved by eventually selecting the features with low inter-correlation coefficients (0.4 or less (Patil et al., 2020; Zhang et al., 2021)). As shown in Figure 3.4, we could recognize that WT and DO should be eliminated because they were highly correlated with the other features. Accordingly, we were able to make eight modelling cases, which consisted of two features as shown in Table 3.8.

Figure 3.4. Pearson correlation coefficients (absolute values) among nine features selected by ANOVA

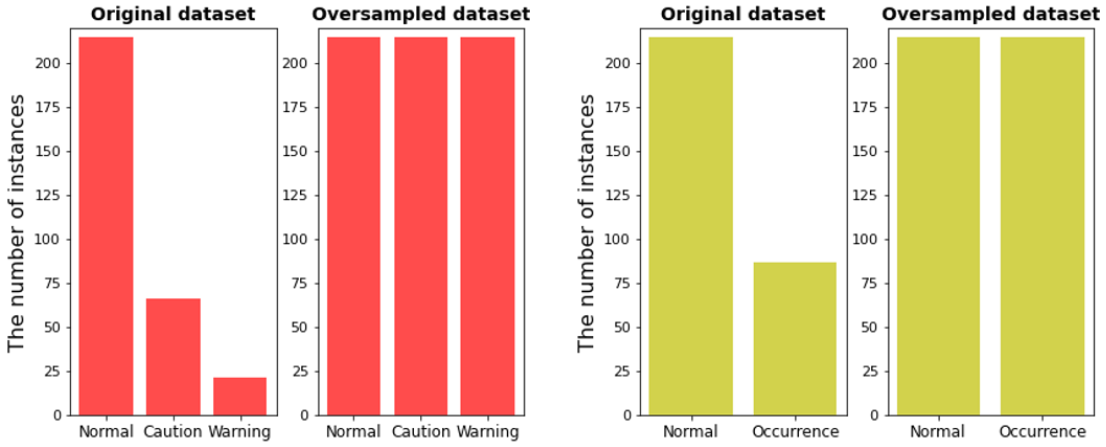
(The red circles indicate the correlation coefficients of 0.4 or less.)

Table 3.8. Modelling cases with a combination of input features

Modelling Case	Input Features
Case1	logCyano, HT
Case2	TN, AT
Case3	TN, LT
Case4	TN, HT
Case5	TDN, HT
Case6	NO_3-N, AT
Case7	NO_3-N, LT
Case8	NO_3-N, HT

3.4.2 Accuracy of the models

For eight modelling cases, the machine learning models were developed using four classification algorithms: k-NN, DT, LR, and SVM. Prior to applying those algorithms, oversampling for the training sets was implemented for the eight cases by a group as shown in Figure 3.5. As a result, a total of 96 models were built with the balanced datasets of the three groups for the eight cases using four machine learning techniques. Table 3.9 shows the parameters optimized by four-fold cross-validation and grid search of each model using the training sets. One thing we need to note in this table is that the parameter max_depth of the Case1 model using Group2 and the DT algorithm (DT-Group2-Case1) is one. Since this means that only one of the two input features was used to build the model, the need for caution is considered when using this model.



(a) Group1 *(b) Group2*

(c) Group3

Figure 3.5. Bar graphs to show oversampling for train sets of three groups

Table 3.9. Optimized parameters of four classification algorithms

Algorithm (Parameter)	Group	Case1	Case2	Case3	Case4
k-NN (n_neighbors)	Group1	3	11	12	6
	Group2	3	16	13	13
	Group3	19	16	19	17
DT (max_depth)	Group1	15	11	10	9
	Group2	1	3	4	14
	Group3	3	4	3	3
LR (C)	Group1	0.01	1	0.001	0.1
	Group2	1	1	100	1
	Group3	1	10	100	100
SVM (C/kernel)	Group1	1000/rbf	1000/rbf	1000/rbf	1000/rbf
	Group2	10/linear	1/rbf	1/rbf	0.1/rbf
	Group3	10/rbf	100/rbf	10/linear	1/linear
Algorithm (Parameter)	**Group**	**Case5**	**Case6**	**Case7**	**Case8**
k-NN (n_neighbors)	Group1	6	7	3	5
	Group2	9	5	5	3
	Group3	16	10	14	14
DT (max_depth)	Group1	15	9	12	14
	Group2	14	10	8	6
	Group3	3	4	3	2
LR (C)	Group1	1	0.01	0.01	0.001
	Group2	0.1	10	1	1
	Group3	0.1	10	1	1
SVM (C/kernel)	Group1	1000/rbf	1000/rbf	1000/rbf	1000/rbf
	Group2	1/rbf	1/rbf	100/rbf	100/rbf
	Group3	0.1/rbf	0.1/rbf	1/linear	0.1/rbf

We could verify which model was more accurate using the test sets as shown in Figure 3.6. The combination of SVM-Group2-Case1 using logCyano and HT as input features provided us with the most accurate model for predicting the cyanobacterial blooms at Chilgok Weir, which ensured the highest accuracy of 92% among the 96 models. On the other hand, the model accuracy of DT-Group3-Case3 and DT-Group3-Case5 was the lowest at 54%.

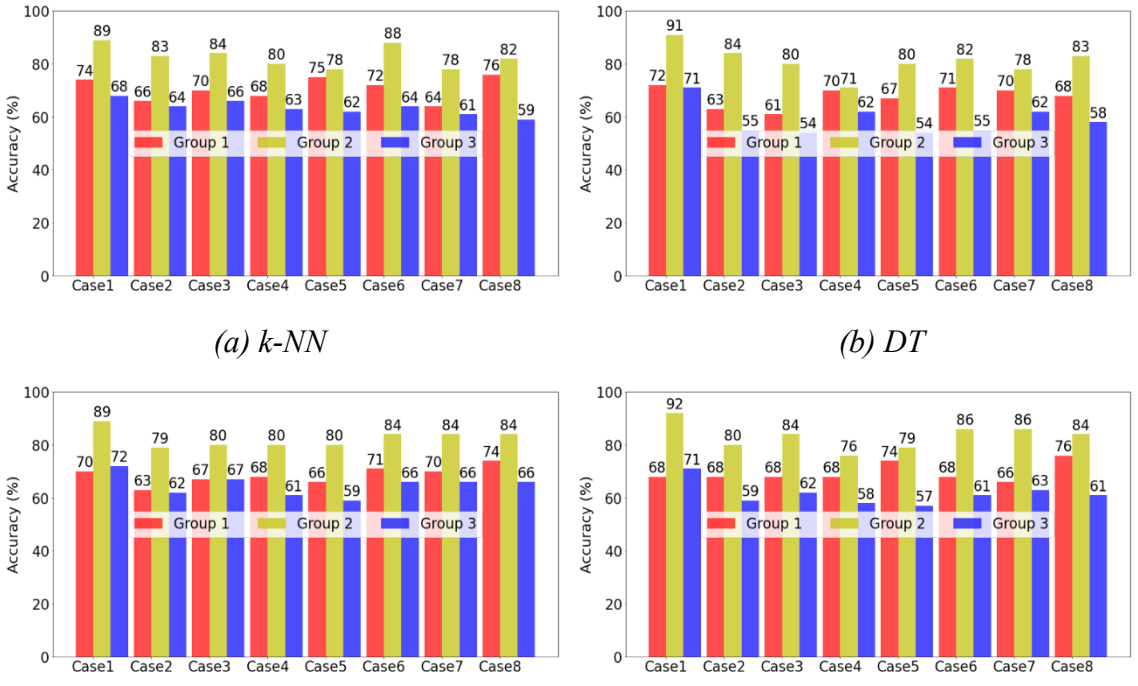
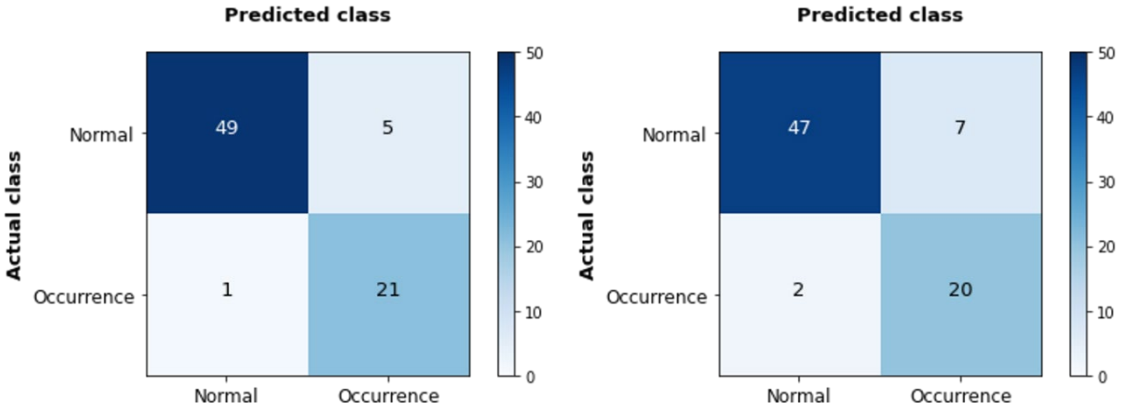


Figure 3.6. Bar graphs to show models' accuracy for test sets of three groups

Of the 96 models, 25 models with an accuracy of 80% or more came from Group2, but all 10 models with less than 60% were from Group3. The accuracy of all the models using Group2 with two classes was higher than the other groups with three classes. Moreover, when evaluating performance based on the used features, we could confirm the highest accuracy of the models using logCyano as a feature. Among the models without using logCyano, the ones with NO_3-N ensured the highest accuracy except the DT algorithm. To be specific, we obtained the highest accuracy of 88% in the models of k-NN-Group2-Case6 except for four models using logCyano. Figure 3.7 shows the confusion matrices for the two models, each with the highest accuracy when using logCyano (SVM-Group2-Case1) and when not using it (kNN-Group2-Case6). We could see that both models provided results that were not biased overall towards non-occurrence or occurrence.



(a) SVM-Group2-Case1 *(b) k-NN-Group2-Case6*

Figure 3.7. Confusion matrices

3.4.3 Summary of the modelling results

The results of the modelling study can be summarized as follows:

i. We had nine input features including logCyano, WT, DO, TN, TDN, NO_3-N, AT, LT, and HT from 30 input features by applying one-way ANOVA.

ii. Seven input features except for WT and DO were available finally for model construction due to the multi-collinearity problem.

iii. By using only two input features, we could build a model with a prediction accuracy of more than 80%.

iv. The models using Group2 with two classes surpassed the other models using Group1 and Group3 which were divided into three classes in terms of model performance.

v. The optimal combination, developing the most accurate model was SVM-Group2-Case1, whose accuracy was the highest at 92%.

vi. All the models with the highest accuracy for each of the four machine learning algorithms (k-NN, DT, LR, and SVM) included logCyano as a feature.

vii. Among the models that did not use logCyano as a feature, the ones in combination with air temperature (AT, LT, or HT) and NO_3-N enabled high predictive accuracy of more than 80%.

These results will enable the concerned decision-makers to understand how to build classification-based machine learning models for effectively and efficiently predicting the occurrences of cyanobacterial blooms (HABs). They also indicate that monitoring the

cyanobacterial cell density closely is very important for predicting HABs. For further prevention or minimization of HABs, actions could be considered controlling pollutants (e.g., NO_3-N) or water temperature in a river (e.g., by selective releases from upstream reservoirs, which do exist in the case study area considered). It should be noted, however, that the primary goal of this analysis is the effective and efficient prediction of HABs, while actions for their prevention may be quite diverse and case-study specific.

3.5 Discussion and Conclusions

We developed and evaluated the classification-based machine learning models to predict the cyanobacterial blooms after a week for the Chilgok Weir in South Korea. In order to build accurate models, we went through important processes such as feature selection, oversampling for the imbalanced dataset, and application of classification algorithms.

Through the feature selection, we could not retain features such as water temperature, total phosphorus, solar radiation, discharge, etc., which are theoretically known to affect the occurrence of cyanobacterial blooms (Park et al., 2017b; Rousso et al., 2020). However, it is noted that this was the result of the targeted data reduction, namely applying ANOVA and solving the multi-collinearity problem. At the same time, the more important thing is that we collected data on 30 features based on the theories of cyanobacterial blooms. It would be possible to develop a more accurate and efficient model when we would combine an understanding of the physical or biological processes for the target variable and a rational approach to data analysis simultaneously (Jiang et al., 2020; Moreido et al., 2021; Rousso et al., 2020).

We were able to develop classification-based machine learning models to predict cyanobacterial blooms with more than 80% accuracy using only two features. That is to say, an effective and efficient model development methodology that could increase prediction accuracy with a few features was devised. In this methodology, it was essential to select features that were involved in the target variable statistically through feature selection methods such as one-way ANOVA. Furthermore, the problems of multi-collinearity and an imbalanced dataset needed to be addressed.

We confirmed that the accuracy of the models using two classes of Group2 was overall higher than the other groups with three classes. In other words, we needed to classify the cyanobacterial density into simple two classes rather than three classes to improve the model performance. Similar to the result of this study, most multi-class classification problems are more challenging than binary ones (Chou et al., 2021). Although we concluded that the models using Group2 outperformed the others, we have to consider how to improve the performance of the models using Group1 or Group3, which can fit the real Algae Alert System through further research.

The models using algorithms other than SVM, which accuracy was highest at 92%, also made very slight differences by achieving an accuracy of 91% for DT or 89% for k-NN and LR. On the other hand, except for the highest accuracy, the algorithms' performances were different for different groups (Group1, Group2, and Group3) or the input features. Hence, we need to decide which machine learning algorithm should be employed by considering the application purpose of a model and the available datasets. This consideration is necessary because it would assist in determining how to encode the target variable (as binary, or as multi-class), and which input features should be used.

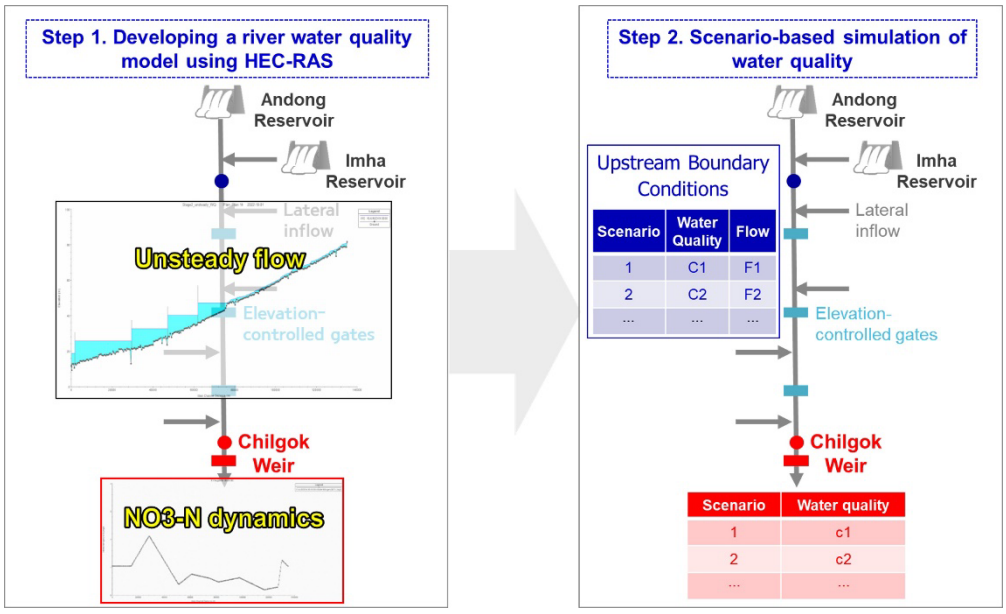
In predicting the cyanobacterial blooms, we could recognize that the current cyanobacterial density (Cyano(t)) had high auto-correlation with ones after a week (Cyano(t+1)). Even if this auto-correlation was not considered, it could be seen that some models using NO_3-N or TN as input features along with air temperature were also very accurate. From these results, we could assume that nitrogen compounds were directly or indirectly involved in cyanobacterial blooms (Park et al., 2021b; Zhao et al., 2017). Therefore, further research could identify if the control of nitrogen compounds flowing into rivers or reservoirs according to the air temperature is possible to make the cyanobacterial blooms produce or fade.

A limitation of this study relates to the fact that we did not separate the four harmful cyanobacteria genera when applying the cyanobacterial cell density (Kim et al., 2020). Referring to the previous studies (Kim et al., 2020; Mellios et al., 2020; Park et al., 2021b; Shin et al., 2017), this was our decision for acquiring a sufficient number of instances for the dataset to carry out this research. Further study will be aimed at developing models involving the cyanobacteria species by reflecting a wider spectrum of their characteristics.

4

RIVER WATER QUALITY MODELLING FOR NITRATE NITROGEN CONTROL USING HEC-RAS

This chapter focuses on developing a river water quality model and simulating the dynamics of nitrate nitrogen (NO_3-N) proven to be one of the main factors for the occurrence of CyanoHABs at Chilgok Weir of the Nakdong River, South Korea in Chapter 3.



This chapter is extracted from Kim, J., Jonoski, A., Solomatine, D. P., and Goethals, P. L. M.: Water quality modelling for nitrate nitrogen control using HEC-RAS: Case study of Nakdong River in South Korea, *Water*, 15, 247, doi:10.3390/w15020247, 2023.

ABSTRACT

The World Health Organization (WHO) and the U.S. Environmental Protection Agency (EPA) provide guidelines on the maximum levels of nitrate nitrogen (NO_3-N) contained in drinking water since excess nitrate ingestion may harm human health. Thus, monitoring and controlling the NO_3-N concentration is of paramount importance, particularly in sources of drinking water such as the Nakdong River in South Korea. This study addresses NO_3-N pollution in the Nakdong River in South Korea, where such pollution mostly comes from diffuse sources in the catchment due to the agricultural use of fertilizers. The objective of this study is to suggest guidelines for designing strategies to control NO_3-N in this river using a process-based model developed with HEC-RAS. The model was built based on water quality parameters (water temperature, dissolved oxygen, ammonia nitrogen, etc.) related to NO_3-N dynamics incorporating hydraulic and meteorological data. This model simulated NO_3-N dynamics downstream under 55 scenarios while focusing on a section near locations of drinking water intakes. The scenarios were constructed based on variations in water quantity and quality upstream. The simulation results showed that the peak concentration of NO_3-N downstream could be directly controlled by limiting the NO_3-N concentration upstream. Additionally, control of the flow rate upstream could also lead to a reduction in the overall average concentration of NO_3-N downstream, but this predominantly occurred when the NO_3-N concentration was decreasing. In conclusion, the design and implementation of strategies for the control of NO_3-N downstream should be carried out after performing a quantitative analysis of the impact of different control measures for different downstream conditions using a water quality model.

4.1 INTRODUCTION

Climate change has already negatively impacted water resources in terms of quantity and quality (UNESCO and UN-Water, 2020). This has prompted increasing interest in ways to effectively improve water quality, particularly in rivers and surface water bodies that provide water for the public water supply. A severe reduction in water quality can pose a risk to public health by increasing human exposure to contaminated water (WHO, 2022). Among the major sources of water pollution, nitrate nitrogen (NO_3-N), one of the nitrogen fractions (Celikkol et al., 2021), may cause specific cancers and adversely affect human reproduction when people take it in excess (Ward et al., 2005; Danaraj et al., 2022; Ward et al., 2018). In this regard, the maximum contaminant level (MCL) of NO_3-N has been set to 10 mg L^{-1} for drinking water by the U.S. Environmental Protection Agency (EPA). The same standard in drinking water has been applied in other countries such as South Korea (Lee et al., 2020) and Japan (Nakagawa et al., 2016). The European Nitrate Directive has required designating areas with surface water or groundwater whose nitrate

(NO_3^-) concentration has been more than 50 mg L^{-1} as Nitrate Vulnerable Zones (Musacchio et al., 2020). The 50 mg L^{-1} of NO_3^- or 11.3 mg L^{-1} (50 mg L^{-1} multiplied by 0.2258) of NO_3-N is identical to the guideline provided by the World Health Organization (WHO) (Ward et al., 2018). However, many studies showed that health risks could still be present despite nitrate ingestion below this MCL (Ward et al., 2018). Thus, the water quality of reservoirs and rivers needs to be improved by controlling the concentration of this particular pollutant to make it as low as possible since reservoirs and rivers are principal sources of drinking water.

Nitrogen fractions such as NO_3-N may flow into reservoirs or rivers due to agricultural practices such as the use of nitrogen fertilizer (Ward et al., 2018). Therefore, there is a risk of nitrate contamination in a river catchment with a lot of agricultural activities, such as the Nakdong River in South Korea (Elzain et al., 2022). Moreover, these pollutants have become water quality parameters that contribute to the complexity of water pollution (Yang and Yu, 2018). NO_3-N can be not only risky as a pollutant itself, but some studies indicated NO_3-N as one of the main drivers of Harmful Algal Blooms (HABs) (Kim et al., 2022a; Park et al., 2021b; Zhao et al., 2017). HABs have caused harm to ecology in an aquatic environment (Paerl and Otten, 2013) and have threatened public health by producing toxic substances such as microcystin (Falconer and Humpage, 2005; Ho and Goethals, 2020). This is problematic particularly in South Korea, where HABs have frequently created environmental problems with the four major rivers since 2012, when 16 weirs were constructed resulting in lentic water bodies in the rivers (Park et al., 2021a; Park et al., 2017b; Song and Lynch, 2018; Romo et al., 2013). The specific NO_3-N concentration is hardly possible to be indicated in terms of preventing or minimizing HABs because the relationship between NO_3-N and HABs depends on other factors such as the state of water flow, site-specificity, and weather. Nevertheless, if the NO_3-N concentrations are controlled when flowing into a river or reservoir, a beneficial effect can be achieved for both the aquatic environment and public health.

A water quality model can be an effective and essential tool from the perspective of Water Quality Management (WQM). A well-developed model can help decision-makers take proper precautions or emergency actions. Strategies designed with a water quality model would be more cost-effective than others, in particular, if they involve establishing new infrastructures or imposing government regulations (Aguilar et al., 2014) to control water pollution. However, success in WQM based on water quality modelling is dependent on the use of reliable data for the model setup and the high performance of the developed model.

Model selection is made with consideration of various conditions including research purposes, data collection, and the required level of model performance (Engel et al., 2007). Models (including water quality models) can be generally classified as process-based and data-driven models (Ejigu, 2021; Rousso et al., 2020). The process-based model is based

on scientific theories or knowledge, while the data-driven model uses data analytics or statistical techniques. Users must select a model that meets optimum conditions after understanding its advantages and disadvantages. To achieve the desired results by developing a process-based model, the user should fully acknowledge the fate and transport of water quality parameters (Srivastava et al., 2006; Razavi et al., 2012).

There are various modelling systems that have the capability to simulate NO_3-N dynamics in catchments and rivers—for instance, CE-QUAL-W2, SWAT, WASP7, MIKE11 (Costa et al., 2021; Alam and Dutta, 2021), and HEC-RAS (Gunawardena and Najim, 2017; Abed et al., 2020; Brunner, 2016; Teran-Velasquez et al., 2022; Taralgatti et al., 2020; Abed et al., 2021). Developing water quality models generally requires many kinds of input variables, which is challenging for model developers (Ghafoor et al., 2022). Nonetheless, HEC-RAS outweighs other one-dimensional river water quality models in terms of user interface and ease of model development, although it has not been widely used compared to the others. HEC-RAS allows users to simultaneously develop a hydraulic and a water quality model (Brunner, 2016). In addition, HEC-RAS ensures the reproduction of river flows as realistically as possible when there are inline structures such as a weir in a river. This is because it is well-equipped with various structures for geometric data and numerous boundary condition types (Brunner, 2016). Several studies on water quality have recently been conducted based on these advantages of HEC-RAS. A recent study showed tangible results for nitrogen dynamics linked to unsteady flow (Teran-Velasquez et al., 2022), while most studies on water quality models developed with HEC-RAS were limited to the analysis of steady flow (Gunawardena and Najim, 2017; Abed et al., 2020; Taralgatti et al., 2020; Abed et al., 2021).

We aim to set out the guidelines for designing strategies to control the NO_3-N concentration using a process-based model developed with HEC-RAS for the Nakdong River. This river is an important water source for many cities located in the southeastern part of South Korea. Specifically, we first produced a model of NO_3-N dynamics for the target area of the upper Nakdong River using HEC-RAS and data from 2019 to 2020. The water quality model was developed based on the hydraulic model of unsteady flow. The downstream boundary of the model was in the vicinity of the Chilgok Weir, which is 135 km away from the upstream boundary. Second, we simulated the change in NO_3-N concentration at the location of Chilgok Weir by using the model developed in the first step. For this purpose, 55 scenarios were constructed with variation in water quantity and quality at the upstream boundary. Finally, we generated guidelines for the design of strategies to control the concentration of NO_3-N at Chilgok Weir. These guidelines were based on the scenarios of the second step.

To the best of our knowledge, this is the first study for the Nakdong River designed to use HEC-RAS for the development of a river water quality model linked with unsteady flow. The novelty of this study is based on an in-depth analysis of the change in NO_3-N concentration in the lower reach of a river under controlled conditions of the upstream

boundary such as water quantity and quality. The methodology presented in this study may also be applied for controlling HABs when linked to research that suggests NO_3-N is the main driver of HABs.

4.2 MATERIALS AND METHODS

The upper reach of the Nakdong River was selected as a study area (see Figure 2.1). The study area covers 135 km in length from the confluence of the Nakdong River and the Banbyeoncheon River to the Chilgok Weir. From 2019 to 2020, the flow rate in this area varied from 5 to 4680 $m^3\ s^{-1}$ and the NO_3-N concentration varied from 0.240 to 3.099 mg L^{-1}.

4.2.1 Model description

We used HEC-RAS version 5.0.7 for this study. HEC-RAS has several capabilities such as analysis of steady flow and unsteady flow, simulation of sediment transport, and simulation of fate and transport of water quality parameters (Brunner, 2016). Of these functions, we focused on the module for the river water quality analysis, which was first added to version 4.0 in 2008. The analysis of steady or unsteady flow should precede a water quality analysis (Brunner, 2016). As we had to consider the operations of the four weirs, we performed the analysis of unsteady flow (Choi and Han, 2014) ahead of simulating the dynamics of NO_3-N, which is an output variable for this study.

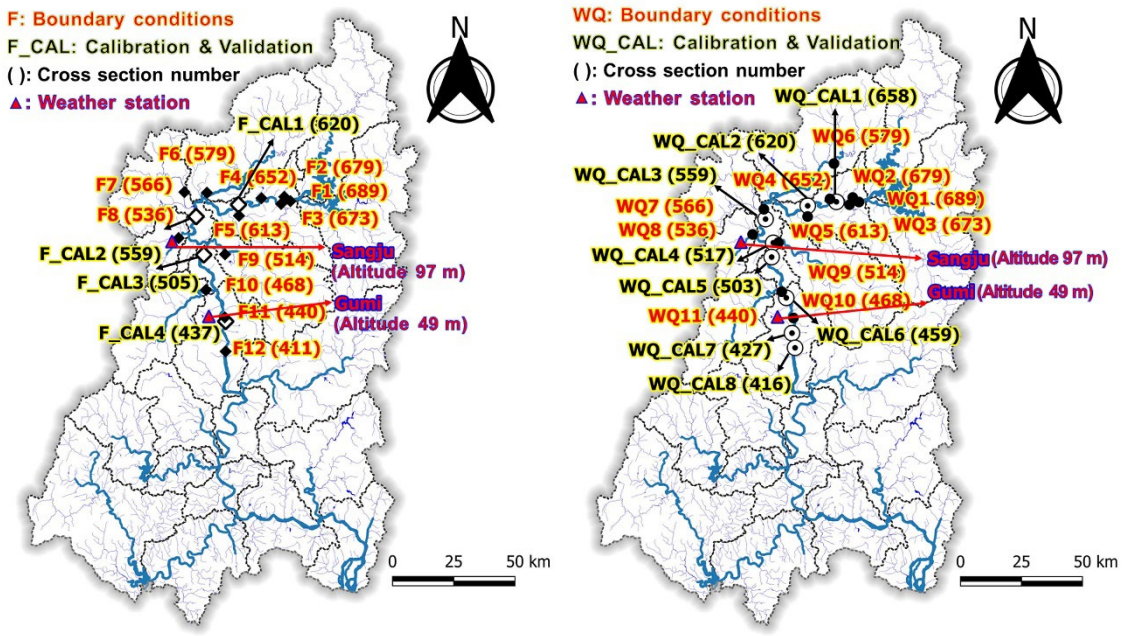
HEC-RAS allows users to build a river water quality model combined with an unsteady flow analysis with inline structures including a weir. This modelling system analyzes unsteady flow by solving the Saint-Venant equation with the implicit finite difference method. The module for analysis of unsteady flow enables the application of several boundary conditions such as stage hydrograph, flow hydrograph, lateral inflow hydrograph, elevation-controlled gates, and so forth (Brunner, 2016). These various boundary conditions help to replicate river flows as realistically as possible. HEC-RAS also solves the one-dimensional Advection–Dispersion equation for water quality analysis using an explicit numerical method called QUICKEST–ULTIMATE (Quadratic Upstream Interpolation for Convective Kinematics with Estimated Streaming Terms–Universal Limiter for Transient Interpolation Modelling of the Advective Transport Equations) (Brunner, 2016; Leonard, 1979; Leonard, 1991). The module for water quality analysis simulates the fate and transport of water temperature, dissolved oxygen (DO), carbonaceous biochemical oxygen demand (CBOD), and nutrient components such as NO_3-N (Brunner, 2016).

4.2.2 Data for HEC-RAS model

HEC-RAS requires geometric data, parameters, hydraulic data, water quality data, and meteorological data for the development of a water quality model (Brunner, 2016). The geometric data include the geometry of cross sections and the inline structures such as a weir (Brunner, 2016). Parameters for a flow model incorporate Manning's roughness coefficients of each cross section and the status of inline structures (e.g., gate conditions at weirs) (Brunner, 2016). For water quality, parameters include dispersion coefficients and different coefficients controlling the rate of change of different compounds with chemical reactions (Brunner, 2016). Furthermore, HEC-RAS needs hydraulic data such as flow rate, water quality data such as water temperature and concentrations of pollutants, and meteorological data such as atmospheric pressure (Brunner, 2016). When different nutrients are modelled (such as NO_3-N), their conversion rates (named 'pathways' in HEC-RAS) may be temperature dependent, and water temperature variations are modelled using the meteorological data (Brunner, 2016).

The geometric data were obtained from the Basic River Plan for the Nakdong River, including Manning's roughness coefficients for cross sections (numbered in HEC-RAS as 411–689, see Figure 4.1) and the inline structures. The River Act of South Korea says that institutions for river management should make a ten-year plan for river management called the Basic River Plan and confirm its validity every five years if necessary. The Basic River Plan for the Nakdong River was made in 2013.

The observational data were retrieved from 16 monitoring stations for hydraulic data, 19 monitoring stations for water quality, and two weather stations (Sangju and Gumi). The location of these stations is shown in Figure 4.1. The daily data are available for flow rate, water level, and climate, while water quality data is monitored almost weekly (48 or 36 times a year). We collected the data for model development in terms of the fate and transport of NO_3-N. The hydraulic data included flow rate and water level. The water quality data contained water temperature, chlorophyll a (Chl-a), dissolved oxygen demand (DO), total dissolved nitrogen (TDN), ammonia nitrogen (NH_3-N), and NO_3-N. Five types of meteorological data were collected, including atmospheric pressure, air temperature, relative humidity, solar radiation, and wind speed, as shown in Table 4.1. Table 4.2 shows the mean, minimum, and maximum values of the observational data of flow rate and NO_3-N in the cross sections for model calibration (2019) and validation (2020).



(a) Stations for hydraulic data *(b) Stations for water quality data*

Figure 4.1. Location of the monitoring stations including two weather stations

Table 4.1. List of the data collected for model development

Category (Source)	**Data**	**Unit**
Hydraulic data (Water Resources Management Information System)	Flow rate	$m^3\ s^{-1}$
	Water level	mamsl
Water quality data (Water Environment Information System)	Water temperature	°C
	Chlorophyll a (Chl-a)	$mg\ m^{-3}$
	Dissolved oxygen (DO)	$mg\ L^{-1}$
	Total dissolved nitrogen (TDN)	$mg\ L^{-1}$
	Ammonia nitrogen (NH_3-N)	$mg\ L^{-1}$
	Nitrate nitrogen (NO_3-N)	$mg\ L^{-1}$
Meteorological data (Open MET Data Portal)	Atmospheric pressure	hPa
	Air temperature	°C
	Relative humidity	%
	Solar radiation	$MJ\ m^{-2}$
	Wind speed	$m\ s^{-1}$

Table 4.2. Mean, minimum, and maximum values of the observational data (flow rate and NO_3-N) in the cross sections for model calibration (2019) and validation (2020)

Data (Unit)	**Cross Section Number**	**Calibration (2019)**		
		Mean	**Minimum**	**Maximum**
Flow rate (m^3 s^{-1})	620	48.85	5.06	976.45
	559	76.78	17.30	1675.61
	505	98.87	4.27	3031.83
	437	116.09	24.06	4677.58
NO_3-N (mg L^{-1})	658	1.313	0.679	3.038
	620	1.398	0.240	3.058
	559	1.750	0.807	2.872
	517	1.688	0.651	2.935
	503	1.760	0.798	2.803
	459	1.693	0.722	2.871
	427	1.886	0.624	3.099
	416	1.841	0.732	3.027
Data (Unit)	**Cross Section Number**	**Validation (2020)**		
		Mean	**Minimum**	**Maximum**
Flow rate (m^3 s^{-1})	620	94.41	10.38	1909.73
	559	173.68	18.84	2499.44
	505	212.05	23.78	3632.07
	437	270.62	21.50	4495.12
NO_3-N (mg L^{-1})	658	1.445	1.055	2.453
	620	1.547	1.095	2.512
	559	1.844	0.900	2.924
	517	1.840	0.993	2.858
	503	1.884	0.869	2.890
	459	1.917	1.179	2.957
	427	2.011	1.055	3.095
	416	2.009	1.066	2.986

4.2.3 Data preparation

We preprocessed raw data for water quality to make them suitable for model development. The reason we needed this process is that the observational data and their frequencies do not exactly correspond to those required in the modelling system. HEC-RAS requires water temperature, algae, DO, carbonaceous biochemical oxygen demand (CBOD), dissolved organic nitrogen (DON), ammonium nitrogen (NH_4-N), nitrite nitrogen (NO_2-N), and NO_3-N (Brunner, 2016) as water quality parameters related to NO_3-N dynamics, as shown in Table 4.3. To address the problem of such discrepancies between the data, we interpolated the weekly data to convert them into daily data and estimated the data which are not measured—for example, algae, CBOD, and a few nitrogen components.

Table 4.3. Components of water quality required in HEC-RAS

Data	Unit
Water temperature	°C
Algae	mg L^{-1}
Dissolved oxygen (DO)	mg L^{-1}
Carbonaceous biochemical oxygen demand (CBOD)	mg L^{-1}
Dissolved organic nitrogen (DON)	mg L^{-1}
Ammonium nitrogen (NH_4-N)	mg L^{-1}
Nitrite nitrogen (NO_2-N)	mg L^{-1}
Nitrate nitrogen (NO_3-N)	mg L^{-1}

The following are four processes we went through for data preparation. First, the weekly data for water quality were interpolated so that they were transformed into daily data, which is the same interval as the water level and flow data. We interpolated the water quality data by applying a step function to avoid distortion of the data variation (Cullinan et al., 2007). In other words, the same values as the previous observational data were placed at daily intervals until the next data were available (James, 2016; Mcintyre and Wheater, 2004).

Second, we estimated the algal biomass required as input data by using the observational data of Chl-a, which is often used as a proxy index for HABs (Kim et al., 2017; Yi et al., 2018; Zhang and Johnson, 2016). The concentration of Chl-a can be converted into the algal biomass with the stoichiometric ratio according to Equation (4.1) (Zhang and Johnson, 2016; Teran-Velasquez et al., 2022).

$$100.0\ \text{g Algae} : 40.0\ \text{g C} : 7.2\ \text{g N} : 1.0\ \text{g P}: (0.4 - 1.0)\ \text{g Chl}\text{–}a \qquad (4.1)$$

where C is carbon, N is nitrogen, and P is phosphorus.

Third, a few nitrogen fractions such as NO_2-N, NH_4-N, and DON had to be estimated because they were not monitored (Teran-Velasquez et al., 2022). NO_2-N was assumed to be zero since it hardly exists in rivers (Meybeck, 1982; Park et al., 2014; Bhuyan et al., 2020; Mihale, 2015). The concentrations of NH_3-N were determined by laboratory experiments using an ion analyzer (Park et al., 2021a) after converting ammonium ions (NH_4^+) into ammonia (NH_3) by increasing the pH of samples with sodium hydroxide (NaOH). Because NH_4^+ and NH_3 are pH-dependent, NH_3-N exists in the form of NH_4-N in most aquatic environments (Rus et al., 2012; Hem, 1985). We thus replaced the data of NH_4-N required in HEC-RAS with the available data of NH_3-N. The DON concentration was calculated by subtracting the sum of NH_3-N and NO_3-N from TDN (Celikkol et al., 2021; Park et al., 2014).

Lastly, we did not consider CBOD as an input variable because the module for water quality analysis in HEC-RAS calculates only losses due to oxidation and settling for CBOD (Brunner, 2016). We performed the sensitivity analysis on the assumption that the changes in the CBOD concentration at all the boundary conditions would not cause fluctuation in the downstream NO_3-N concentration. The result from this sensitivity analysis demonstrated that the assumption was valid, as shown in Figure 4.2.

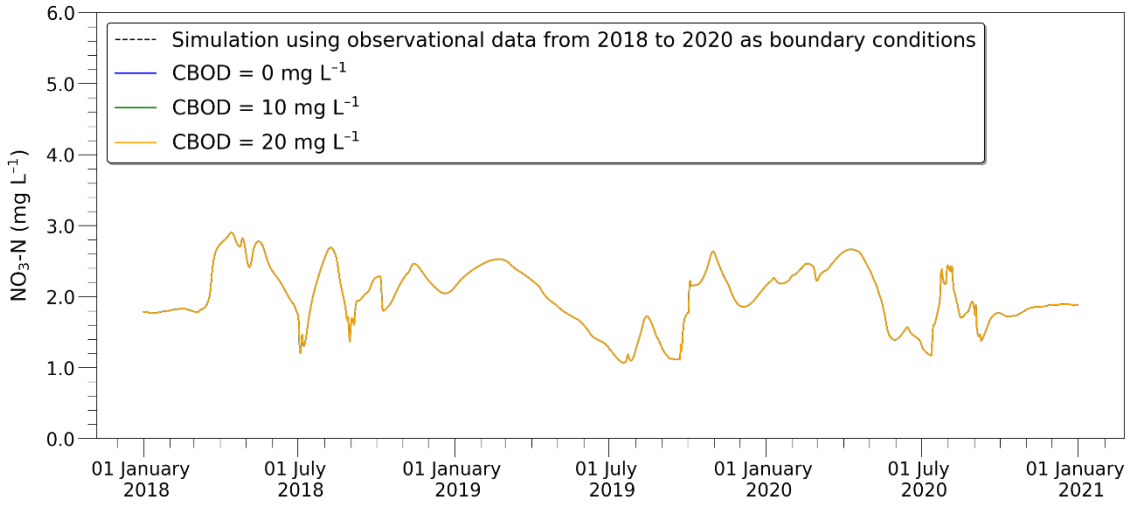


Figure 4.2. Graph showing the changes in the downstream NO_3-N concentration caused by changes in the CBOD concentration at all the boundary conditions

4.2.4 Experimental setup

To build a water quality model using HEC-RAS, we needed not only the geometric data but also the boundary conditions for modules for both unsteady flow and water quality (Brunner, 2016). We collected the geometric data by extracting the upper reach including cross sections (number 411–689) corresponding to approximately 135 km from the Basic River Plan for the Nakdong River. The daily data of flow rate were entered as boundary conditions for the cross section most upstream in addition to 10 cross sections with lateral inflows. The data of stage hydrograph was provided as a boundary condition most downstream. Regarding the four weirs included in the geometric data, we entered the data of the water levels for management as the boundary conditions of the type of elevation-controlled gate. The boundary condition of the elevation-controlled gate enables the control of the gates of the weirs in time (Brunner, 2016). This control of the gates was automatically taken into account in HEC-RAS based on each water level for the management (see Table 2.2) of the four weirs. As the boundary conditions for the water quality module, we entered the daily data interpolated from the weekly data in the cross sections where the boundary conditions for flow analysis were already given (Brunner, 2016).

We calibrated the model parameters with data from 2019 and validated the model with data from 2020. Since the peak flow in 2019 was larger than in 2020 at the monitoring station most downstream for calibration and validation, the data from 2019 were used for calibration. The warm-up period is also necessary for model development until dynamic stability is achieved for the initial conditions (Daggupati et al., 2015). Therefore, we entered the data for the warm-up period from August to December of the previous years.

The data for unsteady flow were derived from four monitoring stations for calibration and validation. For water quality analysis, we used the data from eight monitoring stations, which is twice as many stations as used for the flow analysis. The reason we used data from more stations for water quality analysis is that figuring out the fate and transport of NO_3-N is more important and complicated than flow analysis in this study. These stations were designated in consideration of the locations of the tributaries and the weirs, as shown in Figure 4.1, which illustrates the location of the monitoring stations. The main parameters related to NO_3-N dynamics are the conversion rates, shown in Table 4.4 (Brunner, 2016), and model calibration was performed based on the default values provided in HEC-RAS. Finally, for the dispersion coefficient, we used the HEC-RAS option of automatic computation based on flow data.

Table 4.4. Main parameters related to NO_3-N dynamics provided in HEC-RAS

Parameter	Description	Default Value
Beta 3	Rate constant: DON→NH_4-N	0.020
Beta 1	Rate constant: NH_4-N→NO_2-N	0.100
Beta 2	Rate constant: NO_2-N→NO_3-N	0.200
Sigma 4	Settling rate (DON)	0.001
KNR	Nitrification inhibition coefficient	0.600

We constructed 55 scenarios to understand how the concentration of NO_3-N downstream is changed by the variation in water quantity and quality at the upstream boundary. Three components such as flow rate, water temperature, and NO_3-N were related to these scenarios. Table 4.5 shows how we constructed the scenarios using these components. For example, the seventh scenario (Scenario 7) is that the flow rate of the upstream boundary increases by 50 $m^3 s^{-1}$ for 10 days from 1 January.

Table 4.5. Scenarios constructed for an understanding of NO_3-N dynamics downstream

Components *		Increment /Decrement	Period	Start Date	Scenario
Water quantity	Flow rate ($m^3 s^{-1}$)	−30	365 days	1 January	Scenario 1
		−20			Scenario 2
		−10			Scenario 3
		+50			Scenario 4
		+100			Scenario 5
		+150			Scenario 6
		+50	10 days	1 January 1 May 1 July 1 October	Scenario 7–42
		+100	20 days		
		+150	31 days		

Table 4.5. Cont.

Components *		Increment /Decrement	Period	Start Date	Scenario
Water quality	Water temperature (°C)	–20	365 days	1 January	Scenario 43
		−5			Scenario 44
		+10			Scenario 45
		Constant 0			Scenario 46
		Constant 15			Scenario 47
		Constant 30			Scenario 48
	NO_3-N (mg L^{-1})	−1.0			Scenario 49
		−0.5			Scenario 50
		+0.5			Scenario 51
		+1.0			Scenario 52
		Constant 0.0			Scenario 53
		Constant 1.5			Scenario 54
		Constant 3.0			Scenario 55

*Note(s): * The components belong to the boundary conditions at the upstream boundary.*

These scenarios were constructed under the assumption that the water quantity and quality at the upstream boundary can be controlled. In practice, controls on the water quantity and quality can be imposed by the joint operation of the Andong and Imha reservoirs and the use of SWF installed in the Imha Dam (Lee et al., 2007; Jeong et al., 2020; Park et al., 2017a). The maximum increment of flow rate, 150 $m^3 s^{-1}$, was given based on the maximum amount of water that can be released downstream via the generators of the Andong Dam and the Imha Dam. The simulations under the scenarios were carried out with data from 2018 and the developed model.

4.3 RESULTS

4.3.1 Calibration and validation for unsteady flow

We used Manning's roughness coefficients, listed in Table 4.6, for calibration of the hydraulic model. The Manning's roughness coefficient is the main parameter for calibration. We obtained the data of the coefficients from the Basic River Plan for the Nakdong River.

Table 4.6. Manning's roughness coefficients for the hydraulic unsteady model

Cross Section Number	Manning Roughness Coefficient (Unit: $m^{-1/3}$ s)
411–467	0.024
468–672	0.026
673–689	0.028

Moriasi et al. (2015) suggested the criteria of performance evaluation for watershed-scale models using Coefficient of Determination (R^2), Nash Sutcliffe Efficiency (NSE), and Percent Bias (PBIAS). According to the study, model performance for flow simulations is "Good" if $0.75 < R^2 \leq 0.85$, $0.70 < NSE \leq 0.80$, and $\pm 5\% \leq PBIAS < \pm 10\%$, while it is "Satisfactory" if $0.60 < R^2 \leq 0.75$, $0.50 < NSE \leq 0.70$, and $\pm 10\% \leq PBIAS < \pm 15\%$. These criteria are mainly applied to watershed-scale models, but they can be used for measurement of the performance of our river model built using HEC-RAS. However, we also simultaneously employed a graphical method (Moriasi et al., 2015) to assess the quality of the models. Equations (4.2)–(4.4) show R^2, NSE, and PBIAS, respectively (Moriasi et al., 2015).

$$R^2 = \left[\frac{\sum_{i=1}^{n}(O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}} \right]^2 \quad (4.2)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (4.3)$$

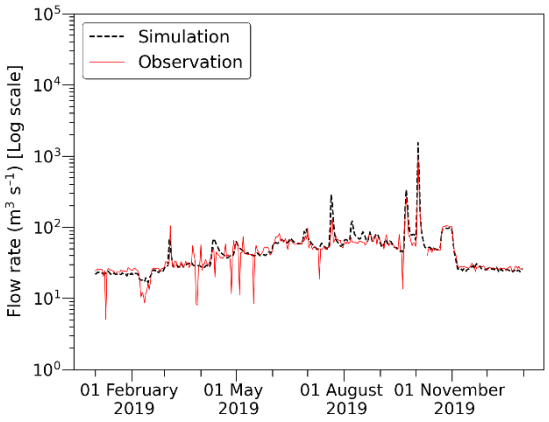
$$PBIAS = \frac{\sum_{i=1}^{n}(O_i - S_i)}{\sum_{i=1}^{n} O_i} \times 100 \quad (4.4)$$

where O is observational data and S is simulation result.

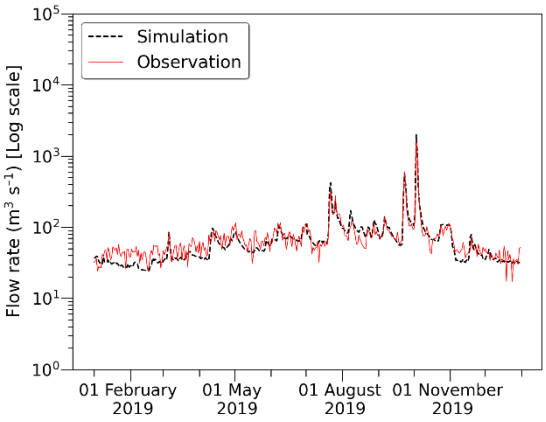
Unsteady flow was simulated using observational hydraulic data such as flow rate and water level as boundary conditions of HEC-RAS. As a result of both calibration and validation for unsteady flow, we carefully judged that the performance of our model was high overall in consideration of both the quantitative evaluation and the graphical method. The quantitatively measured model performance was more than "Satisfactory" except for one cross section (437), as shown in Table 4.7. However, the peak flows from the model simulation were not consistent with the observational data according to Figure 4.3 and Figure 4.4, so this produced an unsatisfactory outcome of PBIAS in cross section 437. Nonetheless, since the upward or downward trends in the flow rate simulated were consistent with those of the observational data, model performance was judged as high for this unsteady flow model.

Table 4.7. Hydraulic model performance for unsteady flow

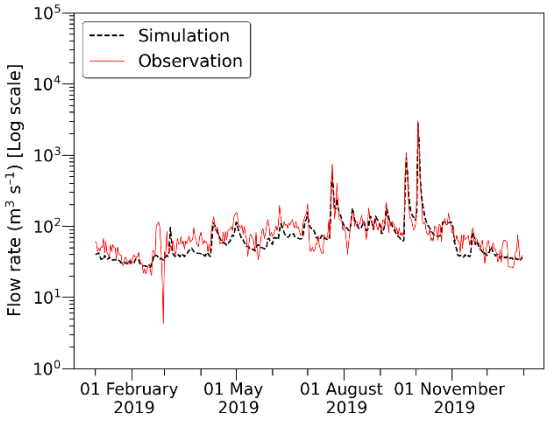
Calibration /Validation	**Cross Section Number**	**R^2**	**NSE**
Calibration	620	0.956	0.612
	559	0.975	0.945
	505	0.967	0.962
	437	0.929	0.866
Calibration /Validation	**Cross Section Number**	**PBIAS (%)**	**Performance**
Calibration	620	−10.3	Satisfactory
	559	2.0	Very Good
	505	10.5	Satisfactory
	437	11.7	Satisfactory
Calibration /Validation	**Cross Section Number**	**R^2**	**NSE**
Validation	620	0.875	0.870
	559	0.948	0.937
	505	0.952	0.918
	437	0.963	0.917
Calibration /Validation	**Cross Section Number**	**PBIAS (%)**	**Performance**
Validation	620	−9.4	Good
	559	6.5	Good
	505	9.8	Good
	437	16.7	Not Satisfactory



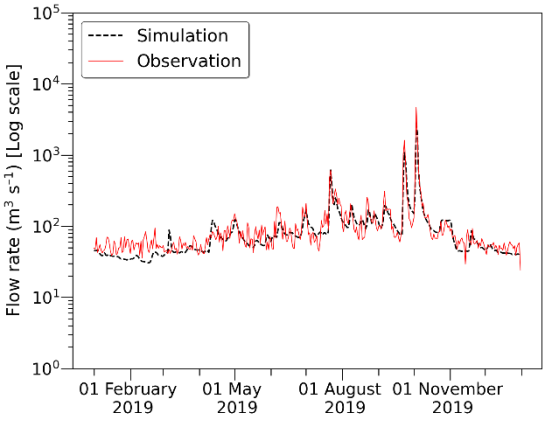
(a) Cross section 620



(b) Cross section 559

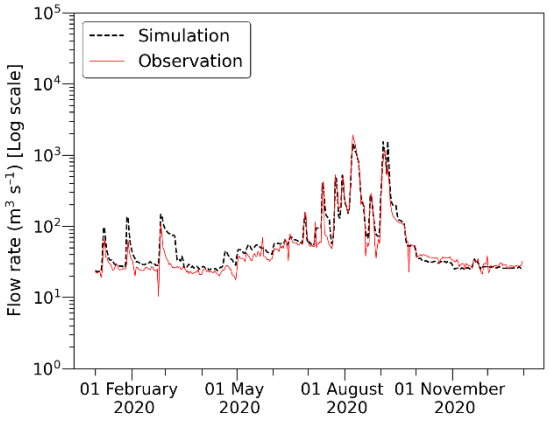


(c) Cross section 505

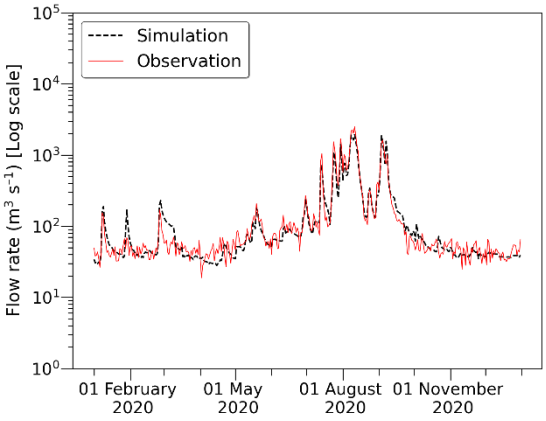


(d) Cross section 437

Figure 4.3. Hydrographs showing the difference between simulation and observation for calibration

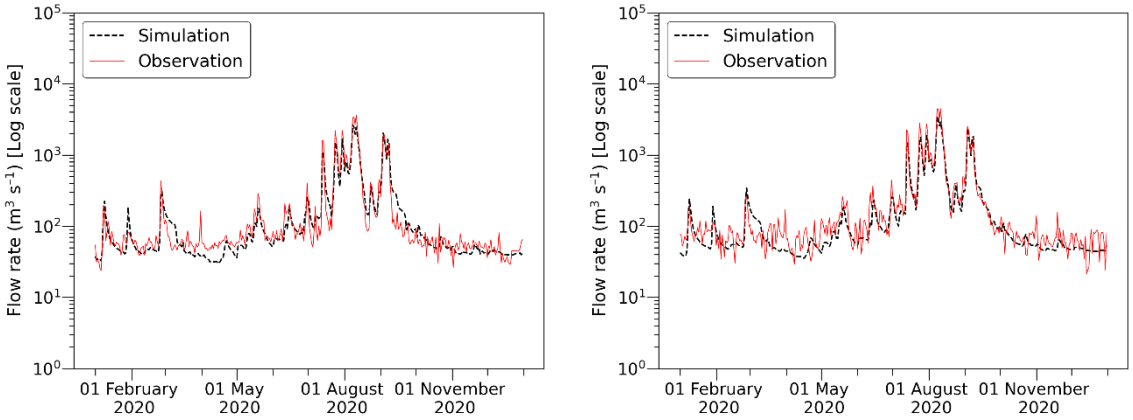


(a) Cross section 620



(b) Cross section 559

Figure 4.4 cont.



(c) Cross section 505 *(d) Cross section 437*

Figure 4.4. Hydrographs showing the difference between simulation and observation for validation

4.3.2 Calibration and validation for NO_3-N dynamics

The water quality model for NO_3-N dynamics was developed using the hydraulic model built for unsteady flow. We simulated NO_3-N dynamics using the water quality data and the meteorological data as the boundary conditions of HEC-RAS. For calibration and validation, we used the main parameters of the model related to NO_3-N dynamics (see Table 4.4). One model parameter was significantly adjusted during calibration, namely Beta 3, for which a value of 0.001 was applied, while the default values were used for the other model parameters. We simulated the water quality parameters including NO_3-N by applying these model parameters. Table 4.8 shows the mean values of both the observational data and the simulation results for the water quality parameters between 2019 (calibration) and 2020 (validation).

Table 4.8. Mean values of both the observational data and the simulation results for the water quality parameters between 2019 (calibration) and 2020 (validation)

Water Quality Parameter (Unit)		**Cross Section Number**			
		658	**620**	**559**	**517**
Water temperature (°C)	Observation	15.0	14.5	16.4	16.7
	Simulation	13.8	12.8	12.5	12.6
Water Quality Parameter (Unit)		**Cross Section Number**			
		503	**459**	**427**	**416**
Water temperature (°C)	Observation	15.7	16.2	17.4	15.7
	Simulation	12.0	12.8	12.4	12.1

Table 4.8. Cont.

Water Quality Parameter (Unit)		**Cross Section Number**			
		658	**620**	**559**	**517**
DO ($mg\ L^{-1}$)	Observation	10.6	10.5	10.6	11.0
	Simulation	10.6	10.6	10.8	10.8
Water Quality Parameter (Unit)		**Cross Section Number**			
		503	**459**	**427**	**416**
DO ($mg\ L^{-1}$)	Observation	10.9	10.4	10.8	10.3
	Simulation	11.0	10.8	10.9	11.1
Water Quality Parameter (Unit)		**Cross Section Number**			
		658	**620**	**559**	**517**
DON ($mg\ L^{-1}$)	Observation	0.483	0.424	0.418	0.428
	Simulation	0.410	0.411	0.397	0.410
Water Quality Parameter (Unit)		**Cross Section Number**			
		503	**459**	**427**	**416**
DON ($mg\ L^{-1}$)	Observation	0.359	0.375	0.425	0.379
	Simulation	0.402	0.420	0.418	0.416
Water Quality Parameter (Unit)		**Cross Section Number**			
		658	**620**	**559**	**517**
NH_4-N ($mg\ L^{-1}$)	Observation	0.062	0.048	0.055	0.045
	Simulation	0.045	0.043	0.044	0.037
Water Quality Parameter (Unit)		**Cross Section Number**			
		503	**459**	**427**	**416**
NH_4-N ($mg\ L^{-1}$)	Observation	0.053	0.050	0.077	0.091
	Simulation	0.033	0.041	0.033	0.032
Water Quality Parameter (Unit)		**Cross Section Number**			
		658	**620**	**559**	**517**
NO_3-N ($mg\ L^{-1}$)	Observation	1.379	1.473	1.798	1.765
	Simulation	1.310	1.324	1.664	1.709
Water Quality Parameter (Unit)		**Cross Section Number**			
		503	**459**	**427**	**416**
NO_3-N ($mg\ L^{-1}$)	Observation	1.822	1.810	1.949	1.925
	Simulation	1.772	1.847	1.899	1.917

We assessed the model performance for NO_3-N dynamics by adopting both the objective criteria established by Moriasi et al. (2015) and the graphical method. According to Moriasi et al., model performance for nitrogen (N) is "Good" if $0.60 < R^2 \leq 0.70$, $0.50 < NSE \leq 0.65$, and $\pm 15\% \leq PBIAS < \pm 20\%$, while it is "Satisfactory" if $0.30 < R^2 \leq 0.60$, $0.35 < NSE \leq 0.50$, and $\pm 20\% \leq PBIAS < \pm 30\%$ at the watershed scale. The gap between the watershed-scale model and our river model was closed by simultaneously employing the graphical method in the same way as when the model performance for flow simulation was assessed.

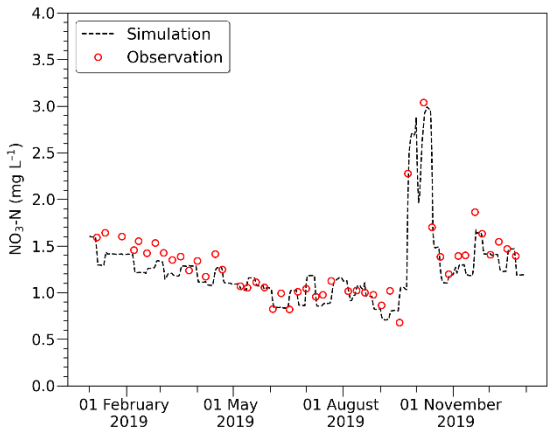
We judged that we built a robust model for NO_3-N dynamics when carefully evaluating model performance at eight monitoring stations. Model performance for NO_3-N dynamics was more than "Satisfactory" except for one cross section (620), as shown in Table 4.9. Figure 4.5 and Figure 4.6 show that NO_3-N dynamics simulated by the HEC-RAS model had a remarkably similar pattern to the observational data in eight cross sections.

Table 4.9. Model performance for NO_3-N

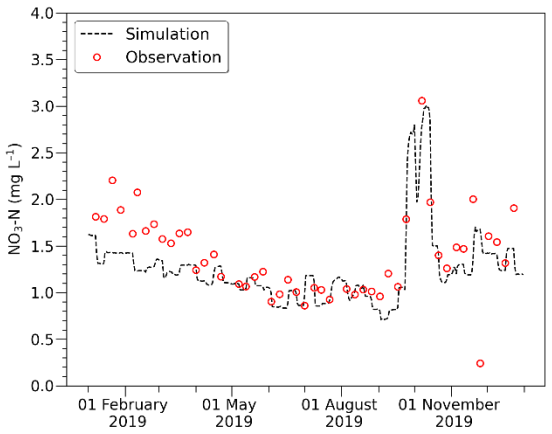
Calibration /Validation	**Cross Section Number**	**R^2**	**NSE**
Calibration	658	0.789	0.750
	620	0.438	0.301
	559	0.766	0.667
	517	0.849	0.801
	503	0.872	0.828
	459	0.895	0.803
	427	0.816	0.732
	416	0.852	0.777
Calibration /Validation	**Cross Section Number**	**PBIAS (%)**	**Performance**
Calibration	658	5.7	Very Good
	620	10.3	Not Satisfactory
	559	9.5	Very Good
	517	3.5	Very Good
	503	3.7	Very Good
	459	–5.0	Very Good
	427	0.5	Very Good
	416	–1.7	Very Good

Table 4.9. Cont.

Calibration /Validation	Cross Section Number	R^2	NSE
Validation	658	0.621	0.478
	620	0.366	–0.155
	559	0.494	0.442
	517	0.652	0.640
	503	0.611	0.605
	459	0.750	0.749
	427	0.606	0.575
	416	0.791	0.764
Calibration /Validation	**Cross Section Number**	**PBIAS (%)**	**Performance**
Validation	658	4.4	Satisfactory
	620	10.0	Not Satisfactory
	559	5.7	Satisfactory
	517	2.8	Good
	503	1.8	Good
	459	0.4	Very Good
	427	4.5	Good
	416	2.4	Very Good

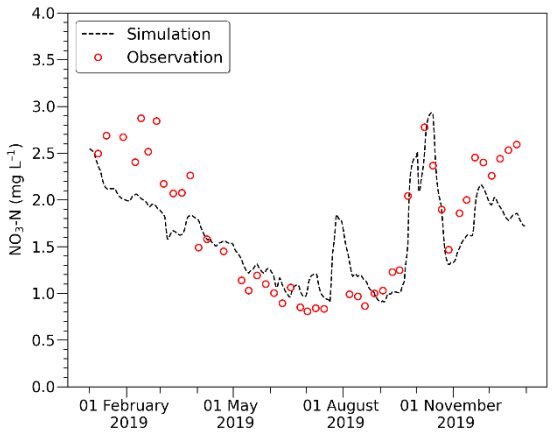


(a) Cross section 658

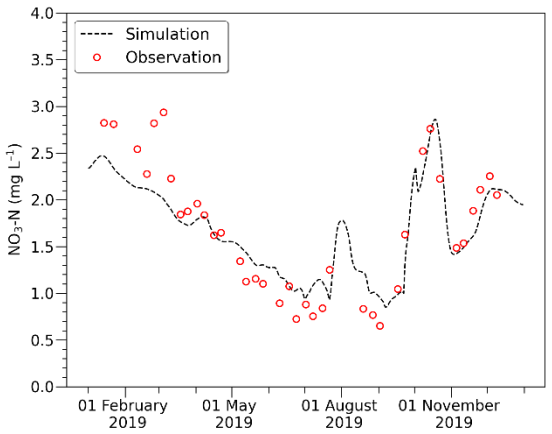


(b) Cross section 620

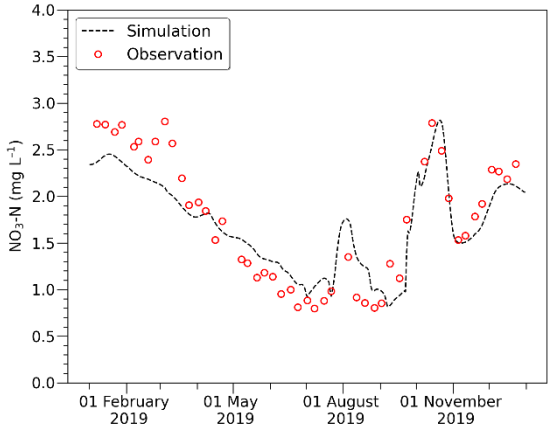
Figure 4.5. Cont.



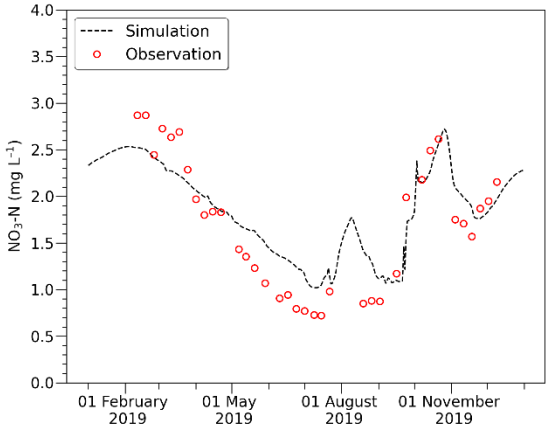
(c) Cross section 559



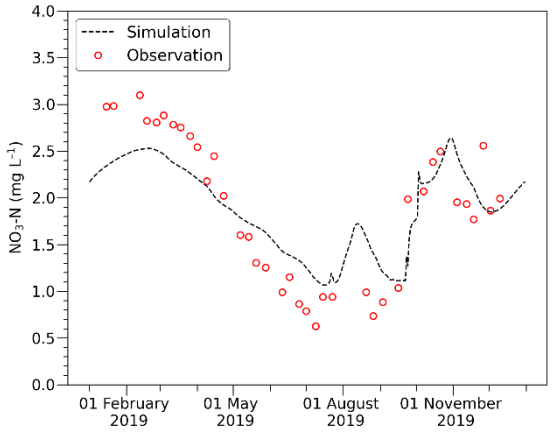
(d) Cross section 517



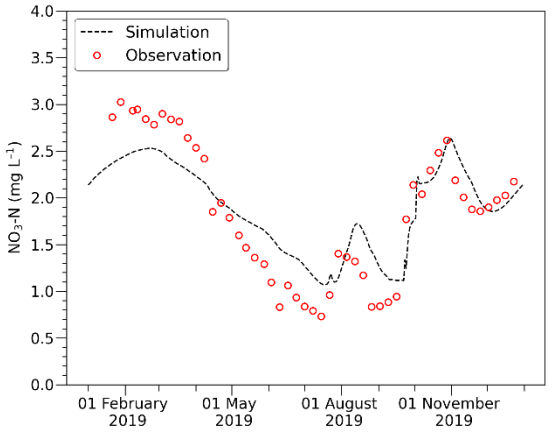
(e) Cross section 503



(f) Cross section 459

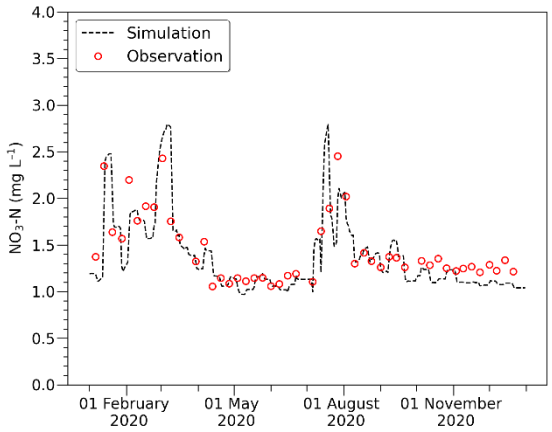


(g) Cross section 427

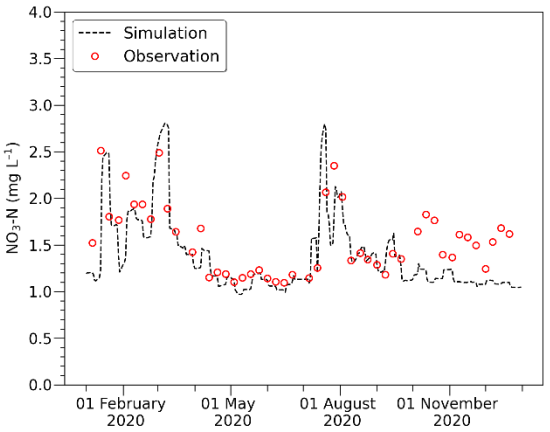


(h) Cross section 416

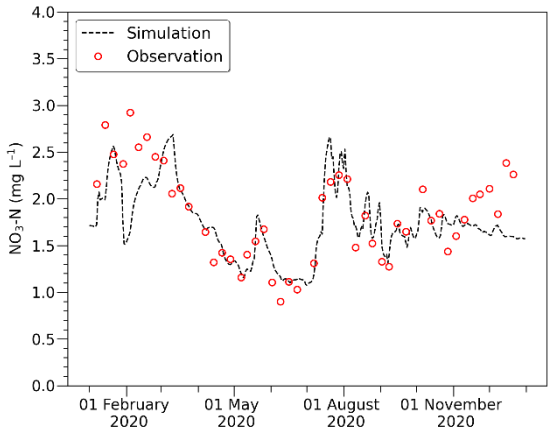
Figure 4.5. Graphs showing the difference between simulation and observation of the NO_3-N concentration for calibration



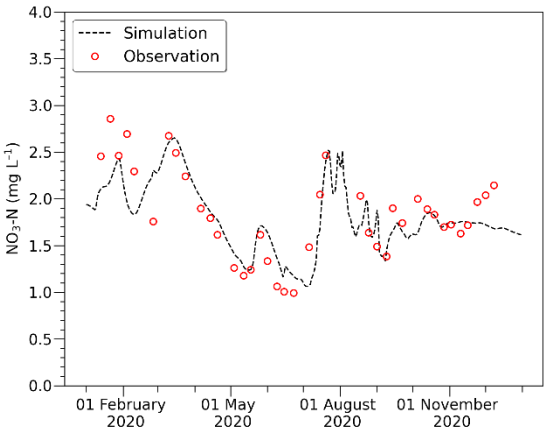
(a) Cross section 658



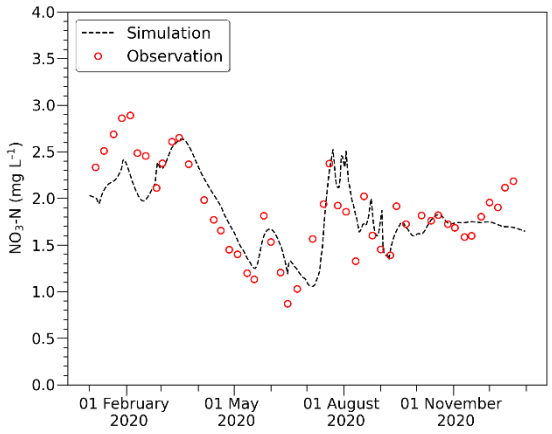
(b) Cross section 620



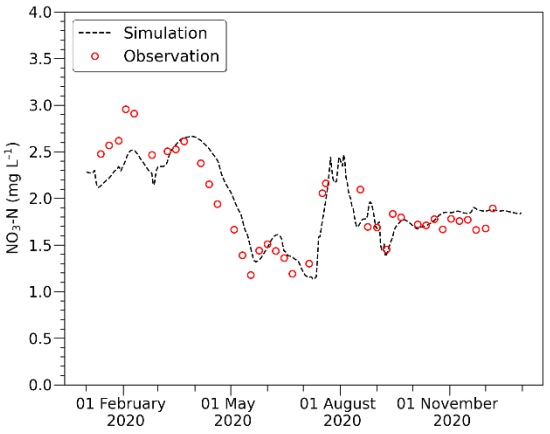
(c) Cross section 559



(d) Cross section 517

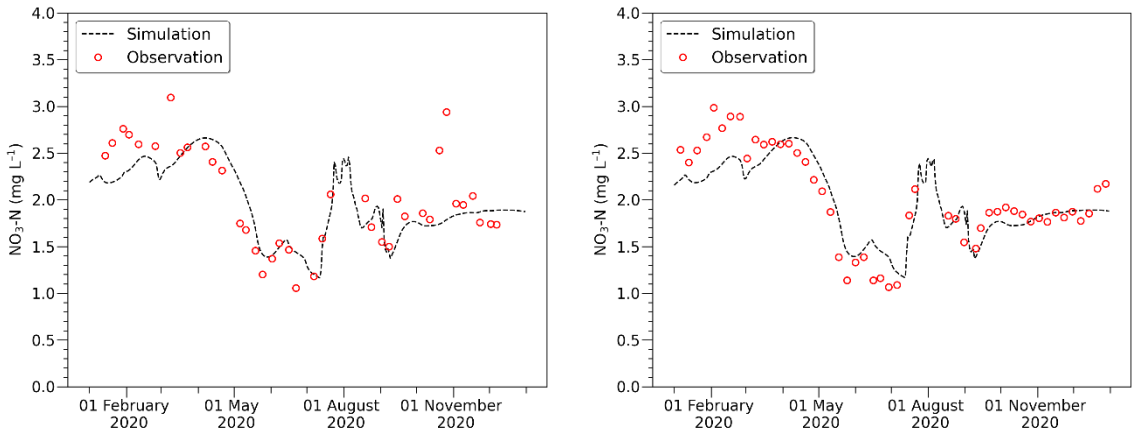


(e) Cross section 503



(f) Cross section 459

Figure 4.6. Cont.



(g) Cross section 427 *(h) Cross section 416*

Figure 4.6. Graphs showing the difference between simulation and observation of the NO_3-N concentration for validation

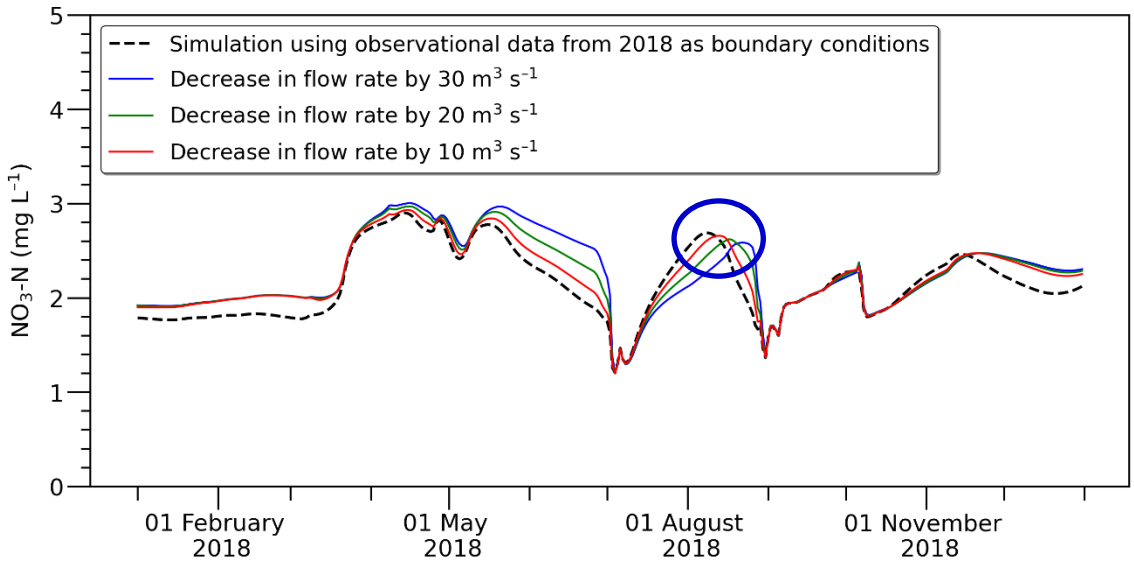
The model delivered high performance, particularly in cross section 416, which is closest to the Chilgok Weir. The station in this cross section is located most downstream among the eight monitoring stations for calibration and validation. Cross section 416 is critically important in this study because the scenarios, provided in Table 4.5, were constructed for the simulation of NO_3-N dynamics in cross section 416.

4.3.3 NO_3-N dynamics according to variation in water quantity

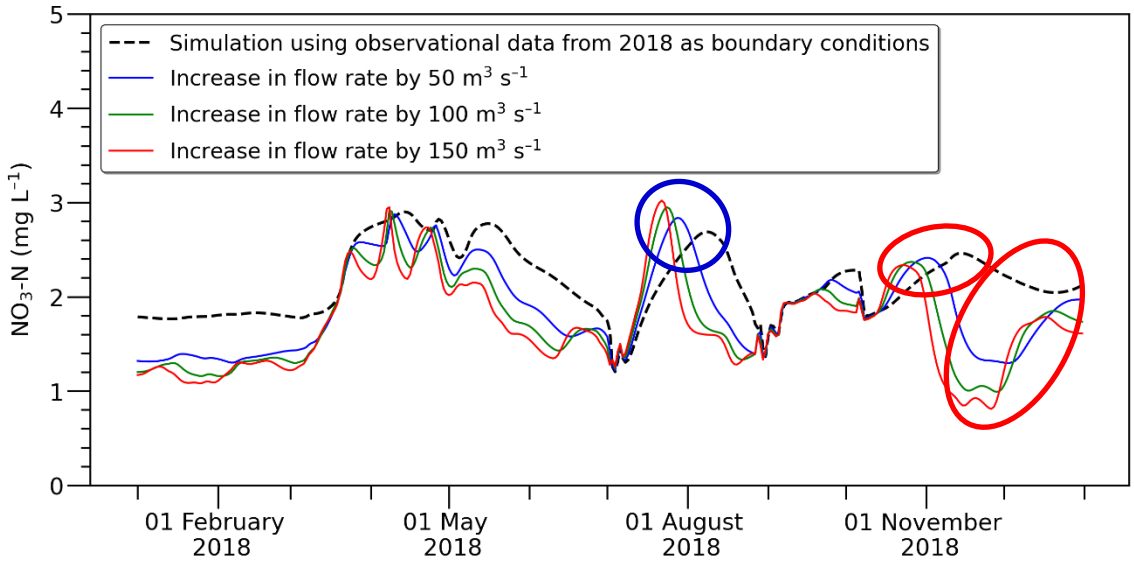
Simulations under Scenarios 1–6 indicated changes in the concentration of NO_3-N in cross section 416 caused by variations in the flow rate most upstream for the whole period (365 days), as shown in Figure 4.7. The black graph in Figure 4.7 shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions. We compared the other graphs, which are simulation results achieved by variation in flow rate, to the black graph.

The results showed that increased flow rates at the upstream boundary led to a decrease in the NO_3-N concentrations in cross section 416. However, different aspects were explored regarding the change in the NO_3-N concentration only around July and August, as indicated by the blue ellipses in Figure 4.7. In other words, the peak concentration of NO_3-N increased in the blue ellipses, although the flow rate increased at the upstream boundary. This reversal was brought about when the downstream NO_3-N concentration sharply increased in the simulation result using observational data at the boundaries (black graph). Here, the increase in flow rate seems to have accelerated the dispersion of the NO_3-N concentration downstream. The acceleration in the dispersion temporarily caused a rapid increase in the NO_3-N concentration. This hypothesis can be supported by

comparing Figure 4.7 with Figure 4.8, which shows the results simulated with the fixed dispersion coefficient of zero. In the blue ellipses of Figure 4.8, the increase in flow rate did not lead to an increase in NO_3-N concentration, unlike in Figure 4.7, which shows the results simulated with the computed dispersion coefficients.

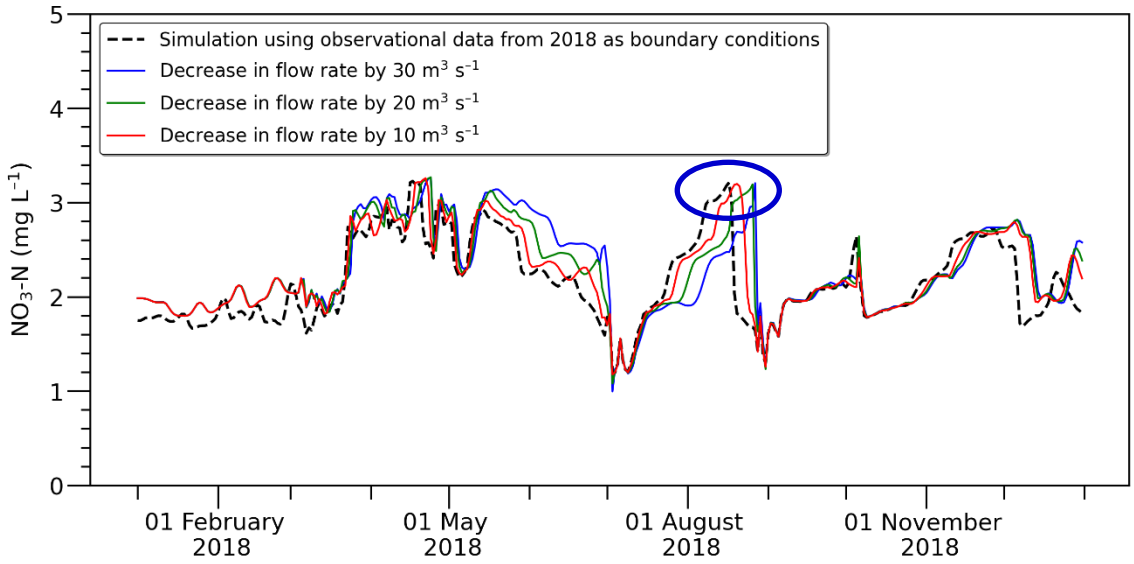


(a) Changes in the NO_3-N concentration by a decrease in flow rates (Scenarios 1–3)

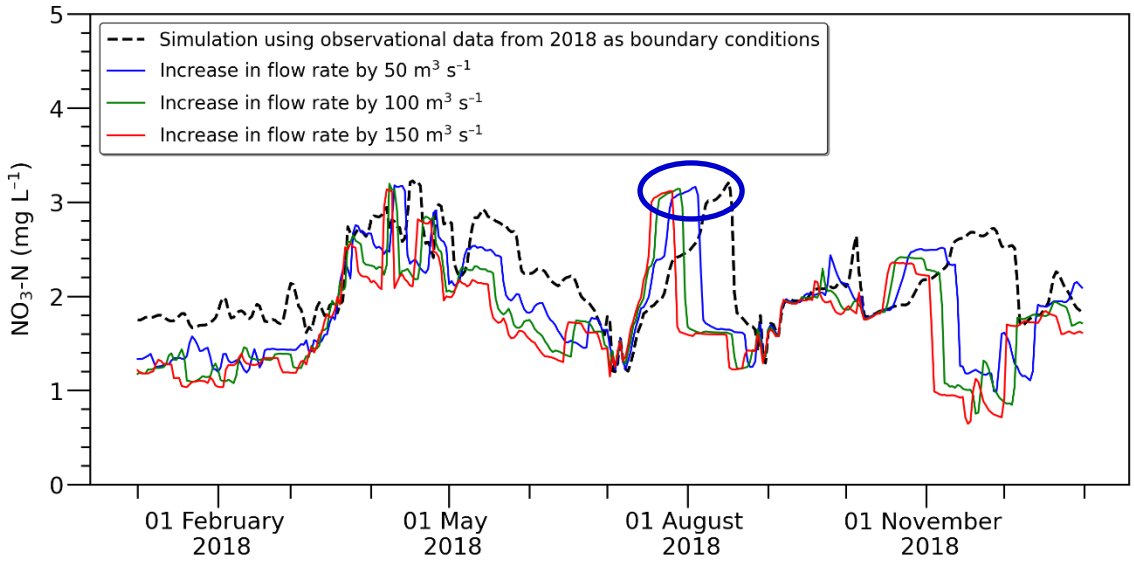


(b) Changes in the NO_3-N concentration by an increase in flow rates (Scenarios 4–6)

Figure 4.7. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate at the upstream boundary for 365 days (The black graph shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions. The dispersion coefficient was automatically computed in HEC-RAS.)



(a) Changes in the NO_3-N concentration by a decrease in flow rates (Scenarios 1–3)



(b) Changes in the NO_3-N concentration by an increase in flow rates (Scenarios 4–6)

Figure 4.8. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate at the upstream boundary for 365 days (The black graph shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions. The dispersion coefficient was set to zero.)

The effect of decreasing the NO_3-N concentration was more considerably exerted by an increase in the flow rate when the NO_3-N concentration downstream was decreasing than when it was increasing, as indicated by the red ellipses in Figure 4.7. As shown in Table

4.10, the flow rate that increased by 150 $m^3 s^{-1}$ brought about a reduction effect of only 5.1%. This effect was shown when the NO_3-N concentration was increasing. On the other hand, the rate of reduction in the NO_3-N concentration was much higher (60.3%) when the NO_3-N concentration was decreasing.

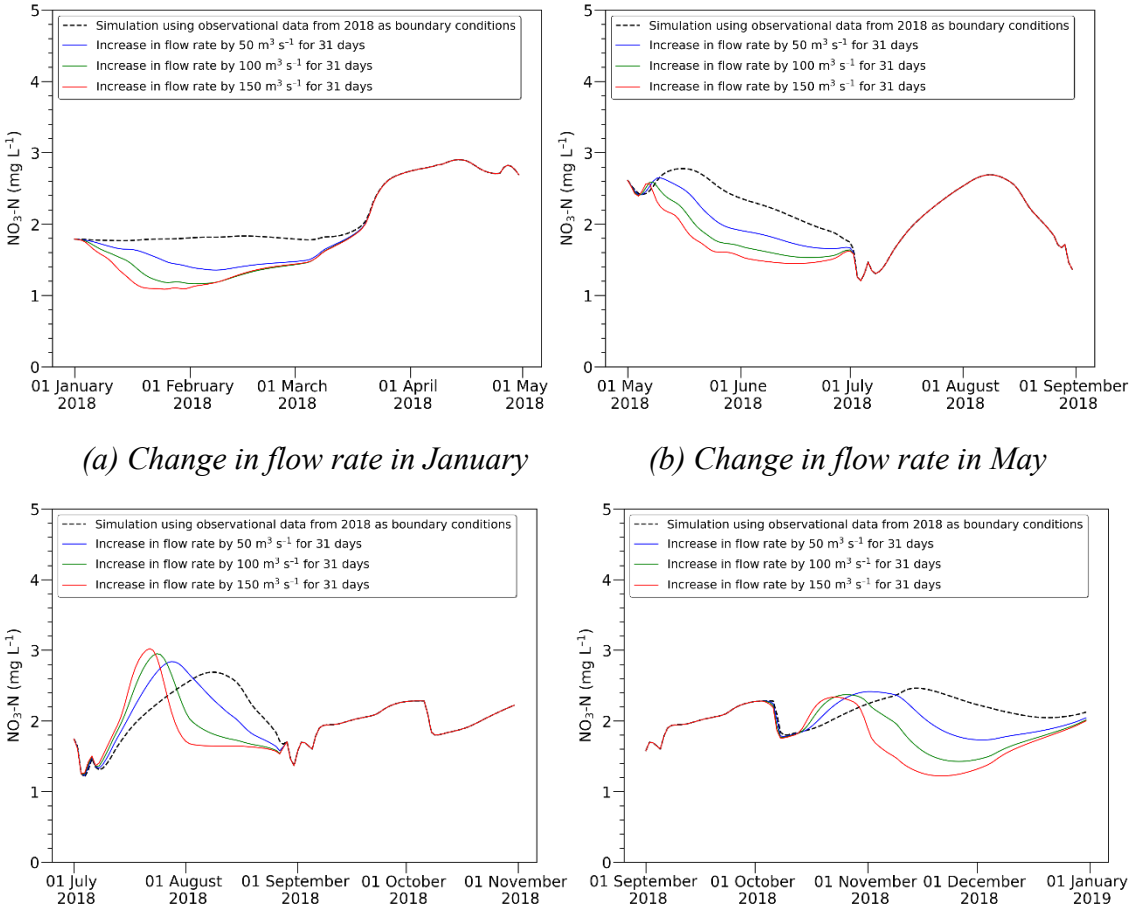
Table 4.10. Example, taken from the red ellipses in Figure 4.7b, of the change in NO_3-N concentration produced by an increase in flow rate

Flow Rate		**NO_3-N**			
Increment ($m^3 s^{-1}$)	**Rate of Increment (%)**	**Concentration (mg L^{-1})**	**Date**	**Reduction in Concentration (mg L^{-1})**	**Rate of Reduction (%)**
0	-	2.463	14 November	-	-
50	33.3	2.414	1 November	0.049	2.0
100	66.7	2.371	26 October	0.091	3.7
150	100.0	2.337	23 October	0.126	5.1
0	-	2.046	21 December	-	-
50	33.3	1.299	2 December	0.747	36.5
100	66.7	0.992	28 November	1.054	51.5
150	100.0	0.813	26 November	1.233	60.3

Interestingly, we found that a fall in the NO_3-N concentration was not proportional to a rise in the flow rate. In Figure 4.7, this point is demonstrated by the unequal changes in the NO_3-N concentration corresponding to the equal-step increase in flow rate (e.g., change in concentration is high for flow variation from 0 $m^3 s^{-1}$ to 50 $m^3 s^{-1}$, but it is insignificant for the change from 100 $m^3 s^{-1}$ to 150 $m^3 s^{-1}$). In any case, ever-increasing flow rates under Scenarios 4–6 do not match the reservoir operations in practice, because this may lead to a shortage of water supply. That is why we considered Scenarios 7–42, where the flow rates were increased at the upstream boundary temporarily instead of for the whole period (365 days).

The overall results obtained by the simulation under Scenarios 7–42 showed that the larger the flow rate, or the longer duration of the increase in flow rate, the more significant the reducing effect on the NO_3-N concentration. Nonetheless, the results showed slight

differences depending on when the flow rate started to increase. For instance, in Figure 4.9b and Figure 4.10b, it can be seen that the concentration of NO_3-N decreased compared to the black graph, depending on the amount or the duration of increased flow. On the contrary, Figure 4.9c and Figure 4.10c show opposite results to Figure 4.9b and Figure 4.10b. The only difference between these cases was the time when the flow rate started to increase. Figure 4.9b and Figure 4.10b show the results achieved under the condition where the increase in flow rate began in May, when the NO_3-N concentration was falling. On the other hand, in Figure 4.9c and Figure 4.10c, the flow rate increased at a time when the concentration of NO_3-N was markedly rising.



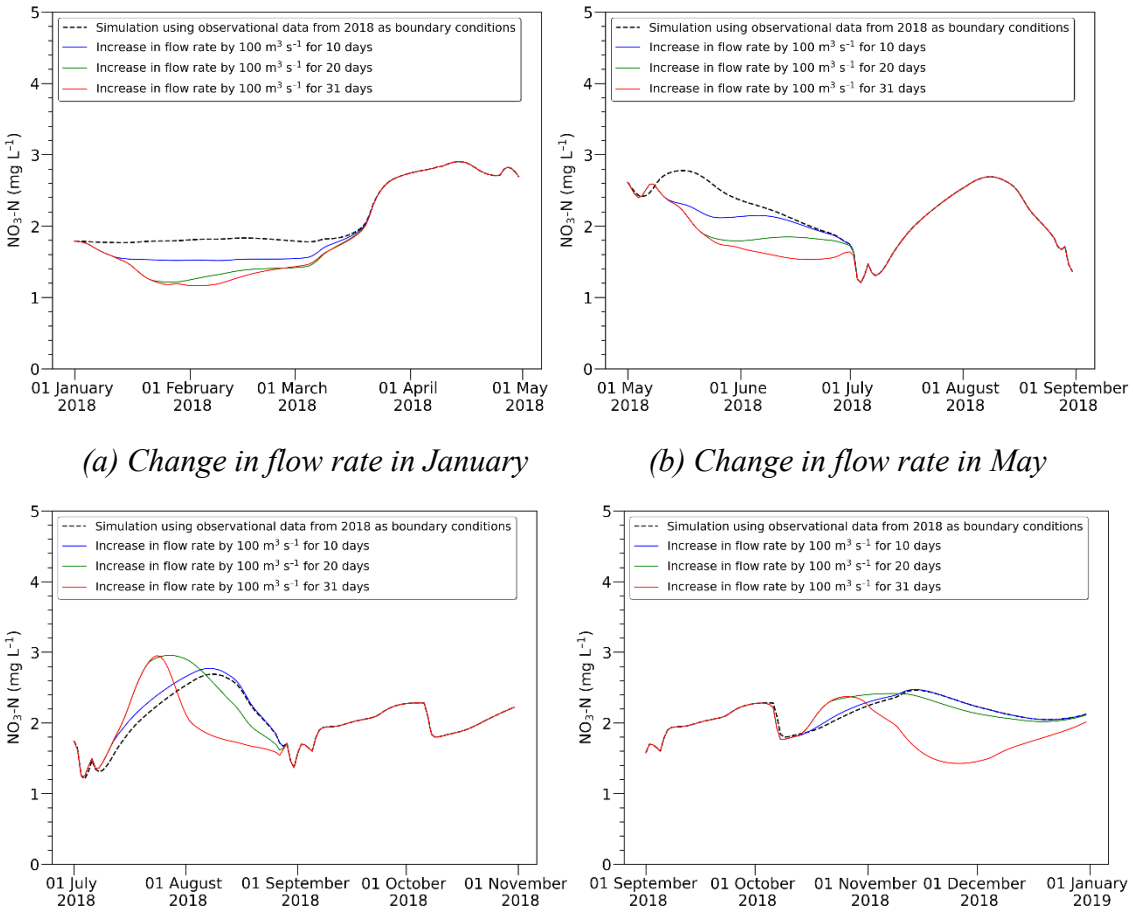
(a) Change in flow rate in January

(b) Change in flow rate in May

(c) Change in flow rate in July

(d) Change in flow rate in October

Figure 4.9. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate (50, 100, and 150 m^3 s^{-1}) at the upstream boundary for 31 days (The black graph shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions.)



(a) Change in flow rate in January *(b) Change in flow rate in May*

(c) Change in flow rate in July *(d) Change in flow rate in October*

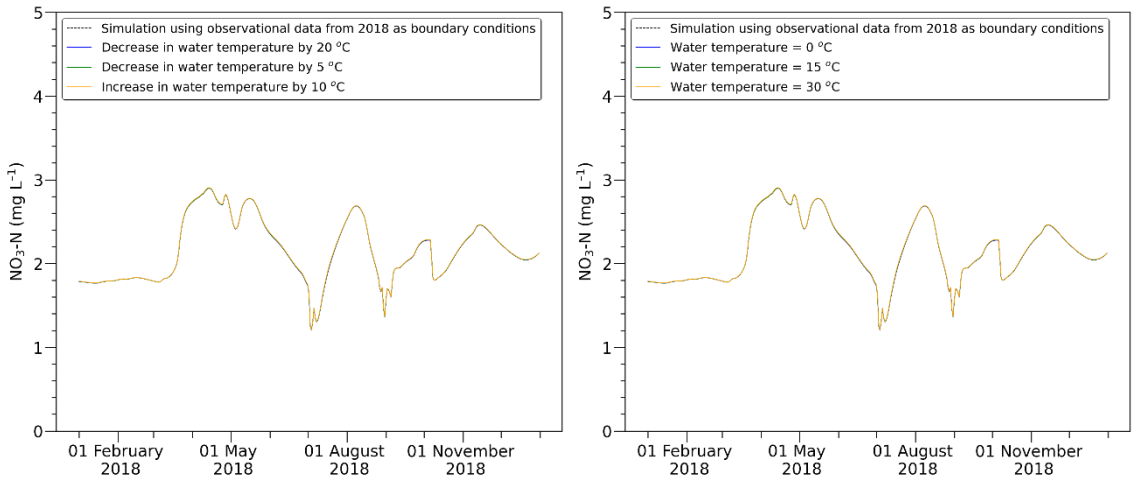
Figure 4.10. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate (100 m^3 s^{-1}) at the upstream boundary for 10, 20, and 31 days (The black graph shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions.)

In this regard, the current status of a river should be considered for decision-making related to reservoir operations in terms of WQM. Specifically, decision-makers should determine to what extent the flow rate released from a reservoir will be increased or decreased or when this action will be taken by considering the current status of the concentration of water pollutants. This will result in effective and efficient control of NO_3-N downstream.

4.3.4 NO_3-N dynamics according to variation in water quality

We learned from Scenarios 43–48 that variations in water temperature at the upstream boundary had little impact on the NO_3-N concentration in cross section 416, as shown

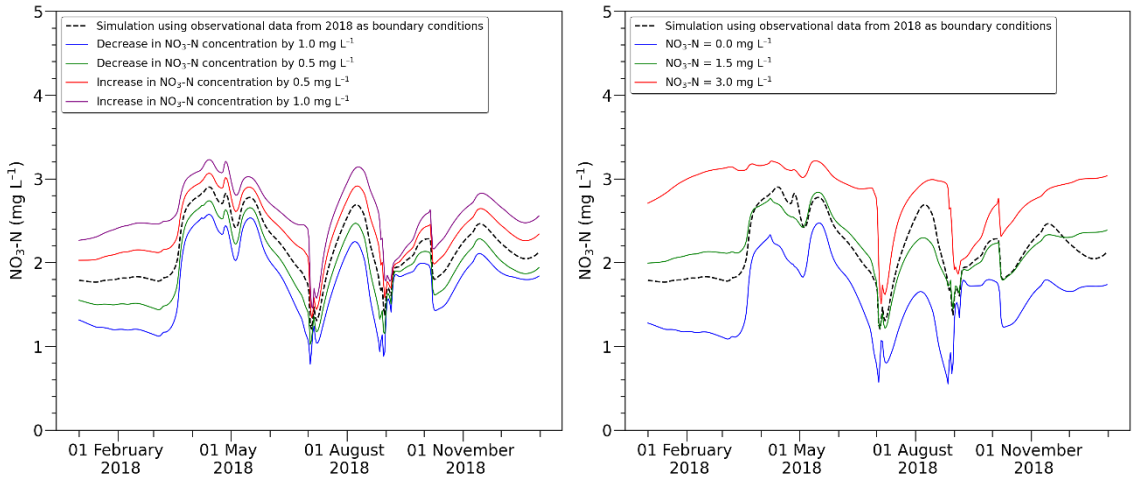
in Figure 4.11. This phenomenon seems to emerge because the water upstream is mixed with tributaries as the water flows downstream, and the water temperature of the river reaches equilibrium. This means that there is little impact on the concentration of NO_3-N downstream only with the change in water temperature at the upstream boundary.



(a) Scenarios 43–45 *(b) Scenarios 46–48*

Figure 4.11. Changes in the NO_3-N concentration in cross section 416 caused by variations in water temperature at the upstream boundary under Scenarios 43–48 (The black graphs show results simulated using the observational data from 2018 as boundary conditions.)

The simulation results under Scenarios 49–55 demonstrated that a marked variation in the NO_3-N concentration occurred downstream if the concentration of NO_3-N increased or decreased at the upstream boundary, as shown in Figure 4.12. In other words, control over the NO_3-N concentration itself in the tributaries or the upper reaches of a river would be highly effective in controlling the concentration of NO_3-N downstream. However, the amount of variation in the downstream NO_3-N concentration may increase or decrease depending not only on the change in the upstream NO_3-N concentration but also on the current status of the river, such as flow rate and water temperature. Therefore, the control method for NO_3-N should be adopted in consideration of the current status in the target area. This sufficient consideration for the downstream status enables the establishment of effective strategies for controlling the downstream NO_3-N concentration with a water quality model.



(a) Scenarios 49–52 *(b) Scenarios 52–55*

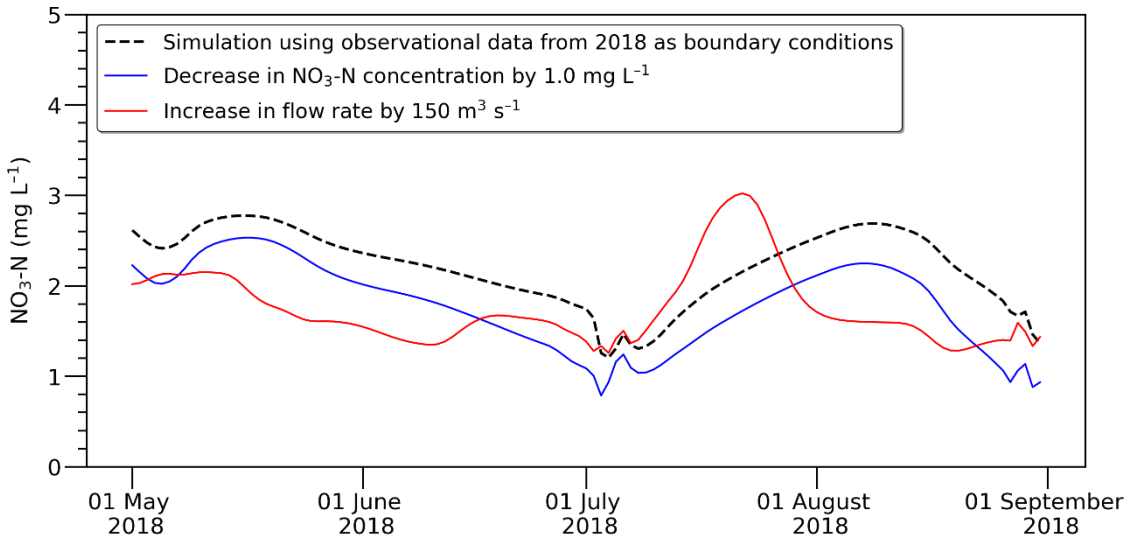
Figure 4.12. Changes in the NO_3-N concentration in cross section 416 caused by variations in the NO_3-N concentration at the upstream boundary under Scenarios 49–55 (The black graphs show results simulated using the observational data from 2018 as boundary conditions.)

4.3.5 Guidelines for design of strategies to control NO_3-N downstream

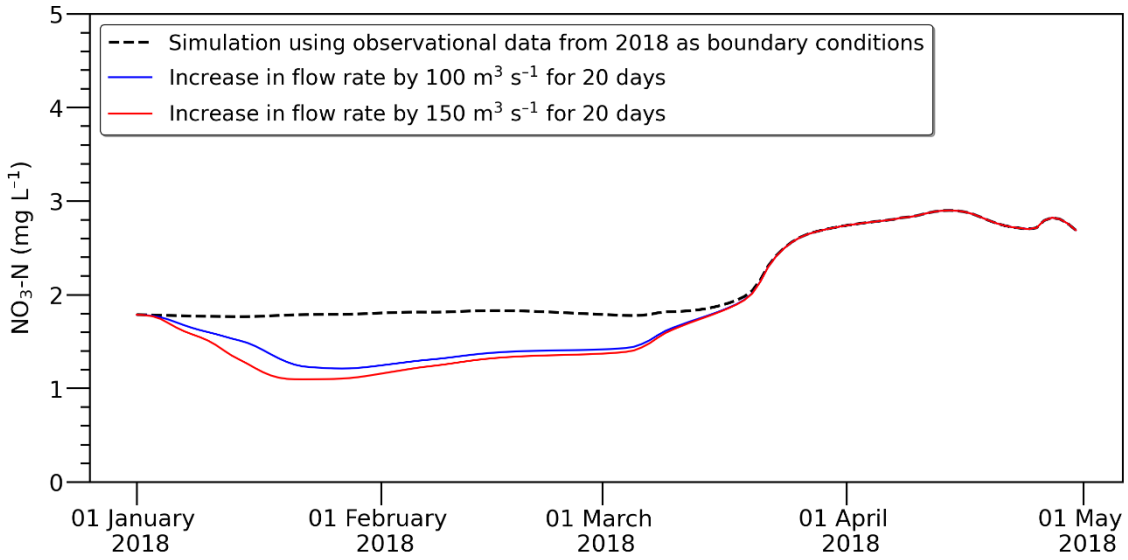
Effective strategies can be devised to control the downstream NO_3-N concentration based on the simulation results of the Scenarios of this study. Guidelines for the design of strategies can be suggested using the control methods of the flow rate or the NO_3-N concentration at the upstream boundary, which was proven effective under the Scenarios. The primary purpose of control methods should be carefully considered before employing the methods. The purpose can include control of the peak concentration or the overall average concentration of downstream NO_3-N.

Specifically, the control method of the NO_3-N concentration itself at the upstream boundary is much more practical for decreasing the highest concentration of NO_3-N downstream than a change in the flow rate at the upstream boundary. This can be demonstrated in Figure 4.13a, which shows conditions of both decreasing and increasing concentrations of NO_3-N in 2018 (black graph). The blue graph shows the variation in the NO_3-N concentration in cross section 416 when the NO_3-N concentration decreased by 1.0 mg L^{-1} at the upstream boundary. The red graph shows the simulation result achieved by a flow rate increase of 150 m^3 s^{-1} at the upstream boundary. We could clearly observe that the peak concentration in the blue graph was lower than the peak in the red graph when the NO_3-N concentration was increasing (July–August 2018). Contrastingly, when the NO_3-N concentration was decreasing (May–June 2018), we could produce the

effect of decreasing the downstream NO_3-N concentration by increasing the flow rate more than by reducing the NO_3-N concentration at the upstream boundary.



(a) Changes in the NO_3-N concentration by a decrease in the NO_3-N concentration (blue graph) or an increase in flow rate (red graph)



(b) Changes in the NO_3-N concentration by an increase in flow rate of 100 m^3 s^{-1} (blue graph) and 150 m^3 s^{-1} (red graph) for 20 days

Figure 4.13. Changes in the NO_3-N concentration in cross section 416 caused by a decrease in the NO_3-N concentration at the upstream boundary or an increase in flow rate at the upstream boundary (The black graphs show the result simulated using the observational data from 2018 as boundary conditions.)

Nonetheless, a large flow rate is not always fully effective. Figure 4.13b shows that there is a slight difference in making the downstream NO_3-N concentration decrease between an increase in flow rate of 172.8 million m^3 (100 m^3 s^{-1} for 20 days, the blue graph) and of 259.2 million m^3 (150 m^3 s^{-1} for 20 days, the red graph). The lowest concentrations of NO_3-N were 1.212 mg L^{-1} (on 27 January) and 1.097 mg L^{-1} (on 22 January) in the blue and red graphs, respectively, with a difference of only 0.115 mg L^{-1}.

4.4 DISCUSSION

The simulation results showed how the downstream NO_3-N concentration would respond depending on variation in the quantity and quality of water upstream. With these results, general guidelines for strategies to control downstream NO_3-N can be suggested with the control methods for the peak concentration and the overall average concentration of NO_3-N. The peak concentration of downstream NO_3-N can be directly controlled by limiting the concentration of NO_3-N in the tributaries or the upper reaches of a river. Control of the upstream flow rate is a viable strategy in terms of control over the overall average concentration of downstream NO_3-N when its concentration is decreasing. Notably, the strategy related to water quantity can be effectively implemented by deciding how much the flow rate should be increased after performing a quantitative analysis of the impact on the control of the downstream NO_3-N concentration. These strategies would be implemented by a combination of joint operations of the reservoirs with SWF and simulation results with the water quality model.

As mentioned earlier, the methodology presented in this study can be used in further research for the indirect regulation of HABs in rivers by controlling the NO_3-N concentration. Since HABs are produced by various factors such as climate, aquatic environments, etc., many researchers have tried to find the major drivers to predict HABs (Rousso et al., 2020). Several previous studies suggested that NO_3-N is one of the key factors underlying HABs (Kim et al., 2022a; Park et al., 2021b; Zhao et al., 2017). Accurate prediction of HABs is not easy because HABs can be produced or faded not only by chemical factors but also by biological processes (Kim et al., 2017; Reynolds, 2006; Park et al., 2017b). However, for cases when NO_3-N is determined to be a critical factor, appropriate countermeasures against HABs in a river can be introduced by predicting and controlling the NO_3-N concentration, which is relatively easier to simulate than HABs.

However, some studies have surprisingly shown that a low concentration of NO_3-N promotes HABs, although the effect could depend on the species of algae (Kim et al., 2022a; Ferber et al., 2004; Weyhenmeyer et al., 2007; Talib et al., 2008). If these findings are linked with this study, HABs could be controlled by a reduction in the flow rate released from an upstream reservoir as in Scenarios 1–3 or by an increase in the NO_3-N

concentration of the released water as in Scenarios 51 and 52 (highly unusual scenarios and hardly possible in practice). Nevertheless, since the implementation of this strategy may lead to an increase in the downstream NO_3-N concentration, an optimization process is necessary by considering an acceptable standard in the NO_3-N concentration required for drinking water sourced from the river.

All the processes for water quality modelling, such as monitoring, analyzing, predicting, and controlling water quality parameters, are closely related to human health and the stability of aquatic ecosystems (Ustaoglu et al., 2021; Forio and Goethals, 2020). This study, however, focused on the modelling process for one water quality parameter (NO_3-N). Further studies should be oriented toward sustainable development in terms of public health and ecological diversity and away from simply focusing on the water quality model. For instance, a water quality model would forecast NO_3-N concentrations in a river. The simulation result could be used for judging whether the concentrations would exceed an acceptable level regarding public health. If exceeding the acceptable level, a decision should be made in advance to reduce the NO_3-N concentrations in the river. A series of these processes would support the sustainable development of human life and aquatic ecosystems.

Moreover, we need to mention the hindrances to this study to be considered in further research. In this study, we tried to clearly understand NO_3-N dynamics depending on the changes in water quantity and quality at the upstream boundary. However, since there were limitations on available data, we needed to make some assumptions. For example, the concentrations of NH_4-N and NO_2-N required in HEC-RAS were replaced with the measured concentration of NH_3-N and zero, respectively (Rus et al., 2012; Meybeck, 1982; Park et al., 2014; Bhuyan et al., 2020; Mihale, 2015; Hem, 1985). Despite these reasonable assumptions based on observable facts, the developed model may still have uncertainty. Furthermore, the HEC-RAS model has not been widely used as a water quality model, although it has been frequently used for flow analysis. This would mean that it should be further validated as a water quality model. In this study, we attempted to develop the HEC-RAS model to simulate the NO_3-N dynamics in the Nakdong River, but its suitability for simulating other water quality parameters should be further demonstrated. Additionally, we constructed a one-dimensional model with HEC-RAS, but a multi-dimensional model would be necessary for detailed analysis of critical locations (e.g., weirs close to water supply intakes, such as the Chilgok Weir in this study). This is because the fate and transport of NO_3-N may tend to vary in a transverse or vertical direction and not only in a longitudinal direction as modelled in this study. Further studies could be conducted with consideration for adequate substitutes for the data that were not measured, the limitations of the HEC-RAS model as a water quality model, and the application of a multi-dimensional model.

4.5 CONCLUSIONS

We developed a one-dimensional process-based model to simulate the fate and transport of NO_3-N using HEC-RAS for the upper reach of the Nakdong River in South Korea. Variations in the downstream NO_3-N concentration were simulated by the developed model according to changes in water quantity and quality at the upstream boundary. For the monitoring station located near the Chilgok Weir, these simulation results were analyzed in comparison with the modelling result that was obtained using the observational data as boundary conditions without the change in water quantity and quality.

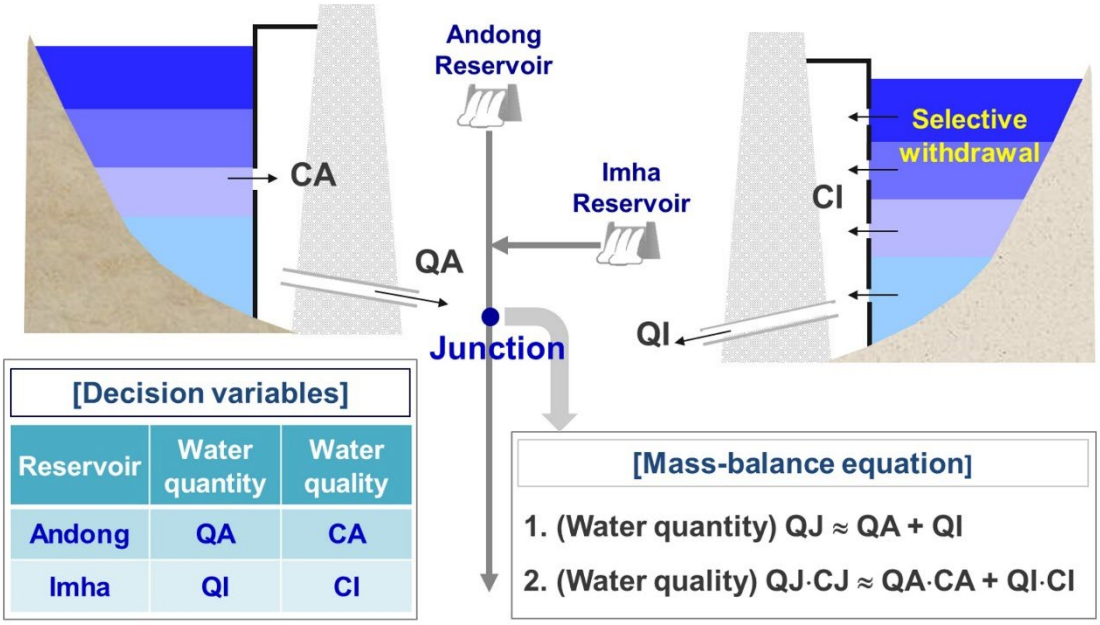
The main finding in connection with the control of water quality is that the change in the downstream NO_3-N concentration was mostly achieved by direct control of the NO_3-N concentration at the upstream boundary. In terms of the control on water quantity, we could create a growing impact on the change in the downstream NO_3-N concentration as the flow rate was increasing at the upstream boundary. However, the reducing effect on the NO_3-N concentration varied depending on how long the flow rate increased and the current status of the downstream NO_3-N concentration. Therefore, strategic decisions on WQM should be made after predicting what effect will be achieved using a water quality model.

Based on the guidelines for the design of strategies for controlling the downstream NO_3-N concentration, we learned that the unilateral decision between water quantity and quality at the upstream boundary would not be best for the improvement in downstream water quality. In this respect, further research can be conducted on the optimal operation of reservoirs in consideration of both water quantity and quality. This optimization process can be accelerated together with a surrogate model for water quality based on a broad spectrum of scenarios.

5

OPTIMAL RESERVOIR OPERATION TO MITIGATE CYANOBACTERIAL BLOOMS DOWNSTREAM

This chapter provides the optimization model for reservoir operation in terms of reducing the frequency of occurrence of CyanoHABs at Chilgok Weir and demonstrates the applicability of the framework proposed in Chapter 1.



This chapter is extracted from Kim, J., Jonoski, A., Solomatine, D. P., and Goethals, P. L. M.: Decision support framework for optimal reservoir operation to mitigate cyanobacterial blooms in rivers, submitted to *Sustainability* in 2023 (under review).

ABSTRACT

Harmful cyanobacterial blooms (CyanoHABs) can produce toxic substances which can harm public health. In South Korea, these CyanoHABs have been a problem in rivers since weirs were constructed in the middle sections of the four major rivers in 2012. To alleviate this problem, flow control has been imposed such as flushing water from reservoirs. However, this measure may cause water shortage in reservoirs because the measure may require an additional amount of water, which exceeds the water demand allocated to the reservoirs. This study aimed at establishing a practical framework for a decision support system for optimal joint operation of the upstream reservoirs (Andong and Imha) to reduce the frequency of CyanoHABs in the Nakdong River, South Korea. Methodologically, three models were introduced: 1) a machine learning model (accuracy 88%) based on the k-NN (k-Nearest Neighbor) algorithm to predict the occurrence of CyanoHABs at a selected downstream location (the Chilgok Weir located approximately 140 km downstream from the Andong Dam), 2) a multi-objective optimization model employing NSGA-II (Non-dominated Sorting Genetic Algorithm II) to determine both the quantity and quality of water released from the reservoirs, and 3) a river water quality model (R^2 0.79) using HEC-RAS to simulate the water quality parameter at Chilgok Weir according to given upstream boundary conditions. The applicability of the framework was demonstrated by simulation results using observational data from 2015 to 2019. The simulation results based on the framework confirmed that the frequency of CyanoHABs would be decreased compared to the number of days when CyanoHABs were observed at Chilgok Weir. This framework, with a combination of several models, is a novelty in terms of efficiency, and it can be a part of a solution to the problem of CyanoHABs without using an additional amount of water from a reservoir.

5.1 INTRODUCTION

Reservoir operations have been mainly focused on the management of water quantity, such as water supply, hydropower generation, and flood control (Yoo, 2009; Saadatpour et al., 2021). However, recent studies have aimed at improving water quality downstream through the efficient operation of reservoirs (Yosefipoor et al., 2022; Saadatpour et al., 2021; Saadatpour et al., 2020). Water quality downstream can be improved by discharging more clean water from a reservoir, but this reservoir operation may also cause an increase in the risk of water shortages.

Harmful cyanobacterial blooms (CyanoHABs), which negatively impact water quality, frequently occur in lentic water bodies such as lakes and reservoirs due to eutrophication (Modabberi et al., 2020; Jankowiak et al., 2019; Park et al., 2021a; Xu et al., 2015; Zhao et al., 2019; Mozafari et al., 2023). In South Korea, CyanoHABs have been an environmental problem in rivers, particularly since 2012 (Park et al., 2021a; Park et al.,

2017b; Song and Lynch, 2018), when the Korean government constructed 16 weirs in the middle of rivers within the Four Major Rivers Restoration Project (Song and Lynch, 2018). This project has raised a matter of controversy with the problem of water quality. This is because of the claim that the weirs have caused the frequent occurrence of CyanoHABs in the rivers, where the flow velocity has decreased (Song and Lynch, 2018). According to the Ministry of Environment in South Korea, BOD (Biochemical Oxygen Demand) and TP (Total Phosphorus), which are the main parameters of water quality to be managed in rivers, at major points of the four rivers as of 2021 were measured as follows: BOD 1.6 mg L^{-1} and TP 0.074 mg L^{-1} at the point Noryangjin of the Han River, BOD 2.2 mg L^{-1} and TP 0.037 mg L^{-1} at the point Waegwan of the Nakdong River, BOD 2.4 mg L^{-1} and TP 0.054 mg L^{-1} at the point Buyeo1 of the Geum River, BOD 5.3 mg L^{-1} and TP 0.169 mg L^{-1} at the point Naju of the Yeongsan River.

CyanoHABs can produce toxic substances such as microcystins (Carmichael and Boyer, 2016; Falconer and Humpage, 2005; Falconer, 2005). Human health may be damaged through the ingestion of water containing these toxic substances (Carmichael and Boyer, 2016; Falconer and Humpage, 2005; Falconer, 2005) or the inhalation of aerosolized cyanotoxins (May et al., 2018; Plaas and Paerl, 2021). Thus, the management of water quality is of paramount importance in terms of preventing the occurrence of CyanoHABs. This management of water quality requires predicting the occurrence of CyanoHABs. However, the prediction of the occurrence of CyanoHABs is challenging due to the complexity of factors involved, such as climate, water quality, flow conditions, and chemical and biological processes (Rousso et al., 2020).

In South Korea, reservoir operation for mitigating CyanoHABs has primarily focused on flow control, such as flushing water from a reservoir to a river downstream (Kim et al., 2022c; Lee and Baek, 2022). However, the use of water in reservoirs for improving water quality has not been generally factored into the design of reservoirs in South Korea (Kim et al., 2019b; Yu et al., 2017; Yoo et al., 2021). Hence, the flow control using an additional amount of water has been only temporarily taken in order to mitigate CyanoHABs (Kim et al., 2022c), since this flow control can cause water shortage. For a trade-off between the conflicting objectives such as improving water quality and alleviating water shortage, optimal operation of reservoirs must be conducted by simultaneously considering both the quantity and quality of water.

A selective withdrawal facility (SWF) is designed and used for controlling the quality of water released from a reservoir (Kim and Choi, 2021; Smith et al., 1987; Davis et al., 1987; Bohan and Grace, 1973). In South Korea, the Imha Reservoir (Lee et al., 2007; Kim et al., 2022b) and the Soyanggang Reservoir (Kim and Choi, 2021; Park et al., 2018a) are equipped with the SWF. Previous studies have demonstrated the effectiveness of the SWF regarding the exclusion of turbid water from the reservoirs, which is the primary purpose of the SWF (Lee et al., 2007; Park et al., 2018a; Kim et al., 2022b). As the SWF

enables the selection of water quality by depth, it can control not only the water quality of the reservoir upstream but also downstream. Therefore, this SWF is a critical factor for the optimal operation of reservoirs considering the improvement of water quality downstream.

However, there have been few studies on reservoir operation using the SWF in terms of addressing issues about biological parameters of water quality such as CyanoHABs in a downstream river. Previous studies (Yosefipoor et al., 2022; Saadatpour et al., 2021; Saadatpour et al., 2020) have not focused on CyanoHABs but on physical or chemical parameters such as temperature, DO (Dissolved Oxygen), and PO_4 (Phosphate). This can be because simulating biological parameters of water quality is more complex compared to physical or chemical parameters.

The main objective of this study is to demonstrate the practical framework (see Figure 1.1) for a decision support system aimed at decreasing the frequency of occurrence of CyanoHABs at Chilgok Weir, based on the optimal joint operation of two upstream reservoirs (Andong and Imha reservoirs). The two reservoirs and the Chilgok Weir are located in the Nakdong River of South Korea. Methodologically, we present three models for this framework. These models include a machine learning model based on the k-NN (k-Nearest Neighbor) algorithm for predicting the occurrence of CyanoHABs at Chilgok Weir, an optimization model employing NSGA-II (Non-dominated Sorting Genetic Algorithm II) for the joint operation of the two reservoirs considering both the quantity and quality of water, and a river water quality model using HEC-RAS to link the machine learning and optimization models. The applicability of the framework is finally demonstrated using observational data and the three models in terms of reducing the frequency of occurrence of CyanoHABs at Chilgok Weir.

5.2 Modelling Methods

5.2.1 Data preparation

The modelling process requires hydrological or hydraulic data, water quality data, and meteorological data for the study area, the upper reach of the Nakdong River (see Figure 2.1). These data can be collected from the Water Resources Management Information System, the Water Environment Information System, and the Open Met Data Portal of South Korea (Kim et al., 2021). Hydrological or hydraulic data and meteorological data are available on a daily basis, while water quality data are obtained on a weekly basis (48 or 36 times a year) for rivers and monthly basis for reservoirs at three depths (Kim et al., 2021).

All data should have the same time interval for the modelling process in this study. However, while hydraulic data and meteorological data are acquired at daily intervals,

water quality data is obtained on a weekly or monthly basis. To address the problem of the mismatch between the time steps, we transformed the weekly or monthly water quality data into daily data using a step function. The step function involves using the same value as observational data of the previous time step until the data for the next time step is available (James, 2016; Mcintyre and Wheater, 2004), maintaining consistency between water quality measurements (Cullinan et al., 2007).

5.2.2 Machine learning model

The machine learning model for Steps 1 and 5 of the framework was developed in Chapter 3 to predict the occurrence of CyanoHABs one week ahead at Chilgok Weir. The determination of the occurrence of CyanoHABs was based on cyanobacterial cell density, as specified by the Algae Alert System of South Korea, with a threshold of 1000 cells mL^{-1} (Kim et al., 2022a; Srivastava et al., 2015). If the cyanobacterial cell density was equal to or higher than this threshold, CyanoHABs were deemed to appear.

Warmer temperatures are generally favorable for CyanoHABs, as well as nutrient conditions, which was also confirmed in our model. After testing many potential influencing factors, including nutrients such as nitrogen and phosphorus, we selected nitrate nitrogen (NO_3-N) and average air temperature (AT) as the input features to build the machine learning model. The model, which was developed using these two input features and applying the k-NN algorithm, an instance-based learning classification technique, was found to ensure the best accuracy of 88%. One of the input features, NO_3-N, showed a negative correlation with the occurrence of CyanoHABs after one week at Chilgok Weir, while the other input feature, AT had a positive correlation.

5.2.3 Optimization model

We developed the multi-objective optimization model considering both the quantity and quality of water for Step 3 in the framework. The decision variables of the optimization included the amount of water supply downstream of the two reservoirs (Andong (*QA*) and Imha (*QI*)) and the quality of water (*CI*) released from the Imha Reservoir using the SWF. The water quality parameter used in this optimization process was NO_3-N, which was an input feature of the machine learning model presented in Section 5.2.2. Since the optimization problem involved the quantity and quality of water, we formulated two objective functions.

The first objective function (OF1) was related to water quantity as shown in Equation (5.1). Through the joint operation of both reservoirs, the two reservoirs can be assumed as one in terms of water supply. In this regard, we aimed at minimizing OF1 which is the difference between the sum of water quantity released from the Andong (*QA*) and Imha (*QI*) reservoirs and the sum of the water demand (QWD) to be allocated by the two reservoirs.

$$\min \quad OF1 = \left\{ \sum_{t=1}^{Days} \left[\frac{QWD_t - (QA_t + QI_t)}{QWD_t} \right]^2 \right\} \div Days \qquad (5.1)$$

where *Days* is the number of days for a simulation period.

The second objective function (OF2), to be minimized, is related to water quality as shown in Equation (5.2). It was formulated by subtracting the NO_3-N concentration (*CJ*) at the junction where the water from the two reservoirs meets from the reference concentration (CR). We formulated OF2 to maximize *CJ*, as NO_3-N concentration was negatively correlated with the occurrence of CyanoHABs after one week at Chilgok Weir (as stated in Section 5.2.2). However, this *CJ* was constrained in order not to exceed CR (Equation (5.3)). *CJ* was calculated using the quantity and quality of water from both reservoirs with the chemical mass-balance equation (Equation (5.4)) (Jain, 1996; Jha et al., 2007).

$$\min \quad OF2 = \left(\sum_{t=1}^{Days} \frac{CR - CJ_t}{CR} \right) \div Days \qquad (5.2)$$

$$CJ \leq CR \qquad (5.3)$$

$$Q_D C_D = Q_U C_U + \sum_{i=1}^{n} L_i \qquad (5.4)$$

where Q_D and C_D represent the flow rate ($m^3\ s^{-1}$) and concentration ($mg\ L^{-1}$) of a specific water quality parameter at a downstream location, respectively. At an upstream location, these values are represented by Q_U and C_U, respectively. L_i ($g\ s^{-1}$) and n denote the individual loadings and the number of inflow points, respectively, between the upstream and downstream locations.

The quality of water (*CI*) released from the Imha Reservoir was constrained in consideration of the NO_3-N concentration distributed by the depth of the reservoir. The constraint on *CI* was set between the minimum (Min. CI) and maximum (Max. CI) values of the NO_3-N concentration, as shown in Equation (5.5), based on the depth where the SWF is available. Simulation of water quality is necessary to determine the distribution

of NO_3-N concentration by the depth of a reservoir, as discussed in Step 2 of the framework. However, in this study, we demonstrated the applicability of the framework by using the observational data of the NO_3-N concentration by depth in the Imha reservoir.

$$Min.CI \leq CI \leq Max.CI \qquad (5.5)$$

Water quality in the Imha Reservoir is monitored at three stations: Imha Dam 1, Imha Dam 2, and Imha Dam 3. The daily data of Min. CI and Max. CI was retrieved from a total of four stations, including one station monitoring the quality of water downstream released from the Imha reservoir. Figure 5.1 shows Min. CI and Max. CI.

In the optimization model, we employed NSGA-II, a widely-applied genetic algorithm using a fast non-dominated sorting procedure (Deb et al., 2002). The Python library, pymoo (version 0.6.0) was used for this optimization model (Blank and Deb, 2020).

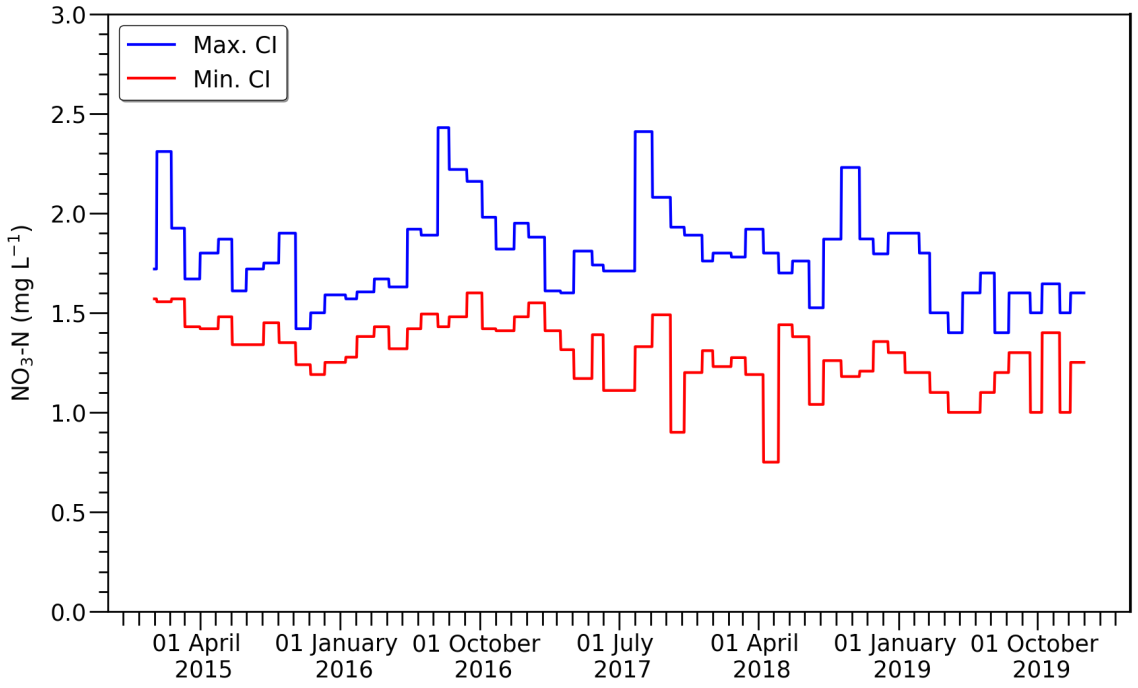


Figure 5.1. Range of NO_3-N concentrations (CI) in water released from the Imha Reservoir

5.2.4 River water quality model

In Chapter 4, we developed a river water quality model using HEC-RAS version 5.0.7 to simulate the NO_3-N dynamics at Chilgok Weir, given upstream boundary conditions. This model evaluated to have a high performance of 0.76 or higher with both R^2 (Coefficient of Determination) and NSE (Nash Sutcliffe Efficiency) was applied in Step 4 of the framework since the model covered the same area as in this study.

The one-dimensional river water quality model using HEC-RAS, a process-based modelling system, does not require a long computation time. However, the HEC-RAS model first simulates unsteady flow and then water quality using the simulation results of the unsteady flow (Brunner, 2016). In order to apply the framework, repeatedly running the HEC-RAS model in the optimization process of Step 3 of the framework is necessary, which requries significant computational time.

To address this time-consuming computational issue, we developed a surrogate model that mimics the NO_3-N dynamics simulated by the HEC-RAS model, based on an Artificial Neural Networks (ANN). Such surrogate models, based on machine learning, when trained, are much more computationally efficient compared to process-based models (Aguilar et al., 2014). The model was built based on Equation (5.6) (Yosefipoor et al., 2022; Saadatpour et al., 2021; Saadatpour et al., 2020) by using the hydraulic, water quality, and meteorological data obtained for developing the HEC-RAS model in Chapter 4. The dataset for training and testing the surrogate model consisted of 508,923 instances, which were generated using the HEC-RAS model by varying the upstream boundary conditions such as flow rate and NO_3-N concentration. This variation in the upstream boundary conditions was required to link the optimization results obtained from Step 3 in the framework with the surrogate model. The surrogate model *SM* was constructed using the Keras open-source software library in Python:

$$\hat{y}(t) = SM(x_1(t-1), x_1(t-2), \cdots, x_m(t-l)) \qquad (5.6)$$

where $\hat{y}$ is the prediction result of the NO_3-N concentration at Chilgok Weir, *SM* is the surrogate model which emulates the behavior of the HEC-RAS model, $\boldsymbol{x}$ $(x_1, x_2, \ldots x_m)$ is a vector of input features, t is the time step (day), m is the number of input features, and l is the time lag. The data used for the development of the surrogate model is the same as for the HEC-RAS model developed in Chapter 4 (see Table 4.1).

While the surrogate model has the advantage to save computation time, it may not be able to reproduce all the HEC-RAS simulation results with 100% accuracy. Hence, we compared the simulation results of the surrogate model with those of the HEC-RAS model.

5.3 Experimental Setup

5.3.1 Procedure

We established the procedure for demonstrating the applicability of the framework shown in Figure 1.1 in three stages: optimization, river water quality modelling, and simulation of the occurrence of CyanoHABs at Chilgok Weir. The first stage involves the optimization model which simulates time series data for the decision variables *QA*, *QI*,

and *CI*. These time series data are used for calculating the flow rate (*QJ*) and the NO_3-N concentration (*CJ*) at the junction where the water from the Andong and Imha reservoirs converge. *QJ* can be different from *QA*+*QI* since *QJ* includes the residual discharge for the area from the dam to the junction. In the second stage, the river water quality model runs the simulation of the NO_3-N dynamics at Chilgok Weir by using the calculated *QJ* and *CJ* as boundary conditions. Finally, the machine learning model is employed to predict the occurrence of CyanoHABs at Chilgok Weir. This model uses the NO_3-N concentration simulated in the second stage and the observational data of average air temperature as input features. Figure 5.2 shows the experimental procedure.

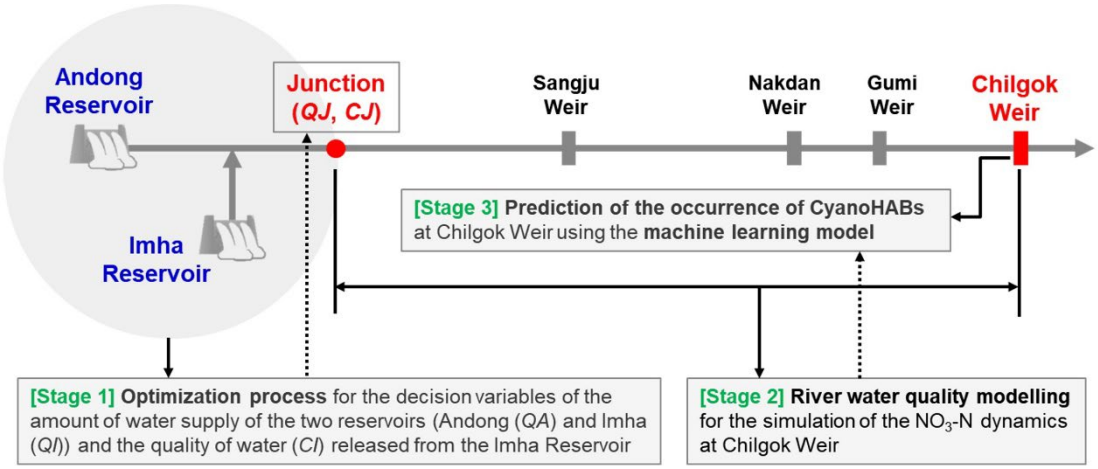


Figure 5.2. Experimental procedure

We adopted the experimental procedure based on the observational data collected over five years from 2015 to 2019. During this period there was no record of discharge via spillway for flood control from the Andong and Imha reservoirs, which was important, as this study did not consider flood routing in the reservoirs.

The observational data used in this study included the cyanobacterial cell density for 226 days, of which 72 days had a cell density of 1000 or higher. As mentioned in Section 5.2.2, we assumed that CyanoHABs occurred when the cell density was 1000 or higher. In this study, we compared the results of CyanoHABs simulated using the optimization results with the 72 days when they were observed.

Before following the experimental procedure, we took a simulation test to validate the developed models. As mentioned in Section 5.2.2, the NO_3-N concentration at Chilgok Weir showed a negative correlation with the occurrence of CyanoHABs. In consideration of this relationship, two hypothetical scenarios were constructed by increasing or decreasing the NO_3-N concentration of the upstream boundary condition by 0.50 mg L^{-1} from the values in the observational data. The river water quality model using HEC-RAS

simulated the NO_3-N dynamics at Chilgok Weir under these scenarios. The machine learning model was used to simulate the number of days with CyanoHABs at Chilgok Weir, using the NO_3-N concentration simulated and the average air temperature measured.

5.3.2 Experimental cases for optimization

We examined the effect of different constraints on the decision variables of the optimization process with nine cases, as shown in Table 5.1. These cases were based on five variables: *QA*, *QI*, *QA*+*QI*, *QJ*, and *CJ*.

QD shown in Table 5.1 is the sum of the maximum amount of water supply allocated downstream in the design stage of the Andong and Imha reservoirs, hereinafter referred to as "Design Discharge". This Design Discharge includes the water quantity for municipal and industrial use, irrigation, and environmental flow, and is presented as monthly data as shown in Table 5.2. The sum of the water demands downstream of the two reservoirs, referred to as QWD (as stated in Section 5.2.3) can also be presented as quantities by month, with slight variations from year to year. QO is the observational data for the water supply downstream of the two reservoirs. This QO was acquired on a daily basis, but it was converted into a monthly average for comparison with QD and QWD. The data for these variables (QD, QWD, and QO) are shown in Figure 5.

Table 5.1. Experimental cases based on the constraints for the optimization process

Case	Constraints for optimization	
	QA **(Unit: m^3 s^{-1})**	***QI*** **(Unit: m^3 s^{-1})**
Case1	$0 \le QA \le 161.0$	$1.0 \le QI \le 119.0$
Case2		
Case3	$0 \le QA \le 0.5QD$	$1.0 \le QI \le 0.5QD$
Case4	$0 \le QA \le 0.3QD$	$1.0 \le QI \le 0.7QD$
Case5	$0 \le QA \le 0.7QD$	$1.0 \le QI \le 0.3QD3$
Case6	$0 \le QA \le 161.0$	$0 \le QI \le 119.0$
Case7	$0.500QO \le QA \le 0.505QO$	$0.500QO \le QI \le 0.505QO$
Case8	$0.300QO \le QA \le 0.305QO$	$0.700QO \le QI \le 0.705QO$
Case9	$0.700QO \le QA \le 0.705QO$	$0.300QO \le QI \le 0.305QO$

Table 5.1. Cont.

Case	Constraints for optimization		
	QA*+*QI **(Unit: m^3 s^{-1})**	***QJ*** **(Unit: m^3 s^{-1})**	***CJ* (*CJ*≤CR)** **(Unit: mg L^{-1})**
Case1	*QA*+*QI*≤QD		*CJ*≤3.11
Case2			*CJ*≤3.50
Case3	-	*QJ*≥QWD	*CJ*≤3.11
Case4	-		*CJ*≤3.11
Case5	-		*CJ*≤3.90
Case6	*QA*+*QI*=QO		*CJ*≤4.10
Case7	-	*QJ*≥17.5	*CJ*≤3.70
Case8	-		*CJ*≤3.50
Case9	-		*CJ*≤4.50

Table 5.2. Design Discharge

Reservoir	Design Discharge by Month (Unit: m^3 s^{-1})					
	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**
Andong	19.9	19.9	19.9	20.8	33.7	49.3
Imha	13.5	13.5	13.5	13.6	14.0	14.7
Sum (QD)	33.4	33.4	33.4	34.4	47.7	64.0
Reservoir	**Design Discharge by Month (Unit: m^3 s^{-1})**					
	Jul	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Andong	40.5	50.2	36.4	22.0	19.9	19.9
Imha	14.4	14.8	14.2	13.6	13.5	13.5
Sum (QD)	54.9	65.0	50.6	35.6	33.4	33.4

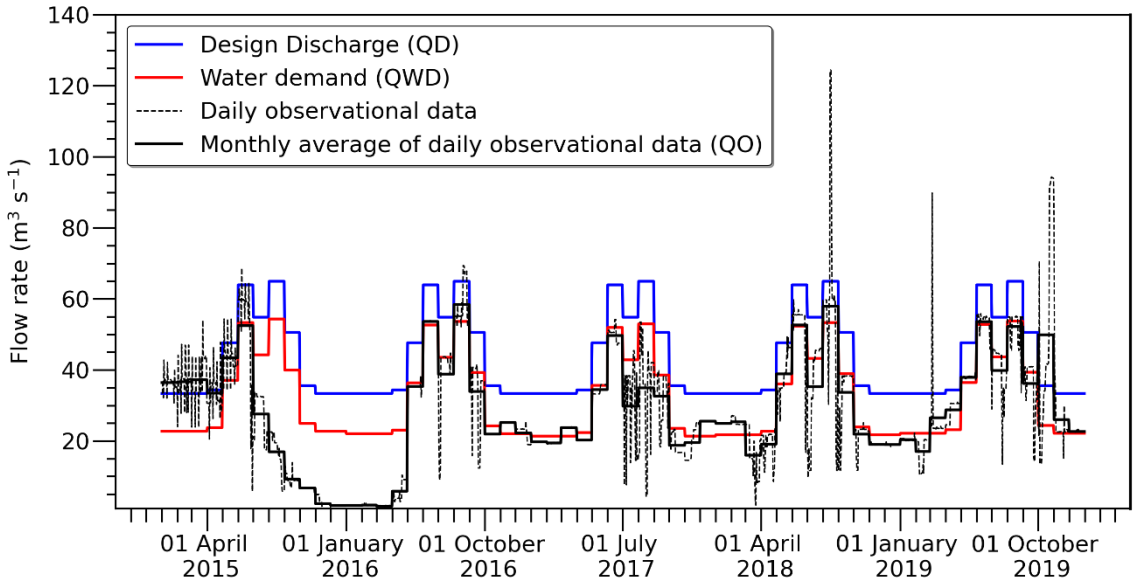


Figure 5.3. Graph showing Design Discharge, water demand, daily observational data, and monthly averages of the daily observational data for the downstream river (sum for the Andong and Imha reservoirs)

The reference concentration of NO_3-N, denoted as CR, was used as a constraint on the NO_3-N concentration (*CJ*) at the junction of the water downstream from the two reservoirs (as mentioned in Section 5.2.3). In Case 1, Case 3, and Case 4, CR was set to 3.11 mg L^{-1}, the maximum concentration of NO_3-N at the junction retrieved from the observational data between 2015 and 2019. In Case 2, CR was set to 3.50 mg L^{-1} in order to compare the results of Case 2 with those of Case 1. We made Case 2 to analyze the impact of variation in CR. For the remaining cases (Cases 5–9), CR was set to the minimum concentration of NO_3-N at which the Pareto front could be obtained during the optimization process.

In the first two cases, the maximum amounts of water were set to 161.0 m^3 s^{-1} and 119.0 m^3 s^{-1} for *QA* and *QI*, respectively. These values represent the quantities of water that can be released downstream through the hydropower generators of the two reservoirs. The minimum *QI* was set to 1.0 m^3 s^{-1}, which is a value that accounts for the water demand of a downstream river from the Imha Reservoir. This intake facility for the water demand is located between the Imha Dam and the junction where the Banbyeoncheon River, a downstream river of the Imha Reservoir, joins the Nakdong River.

We placed the constraints on the sum of water released from the two reservoirs (*QA*+*QI*) in Cases 1 and 2 to ensure that *QA*+*QI* did not exceed the Design Discharge (QD). Additionally, *QJ* was constrained to satisfy QWD, considering the joint operation of the two reservoirs. This constraint on *QJ* was also applied to Cases 3–5.

The major difference between Cases 3–5 and Cases 1–2 is in the constraints placed on *QA* and *QI*. In Cases 3–5, the maximum values of *QA* and *QI* were determined by applying to each portion of QD, where the sum of the portions equals one in each case. The goal of this approach was to reduce the number of constraints and narrow the range of constraints compared to Cases 1–2, allowing efficient optimization. To evaluate the impact of different portions on the simulation results for CyanoHABs, the values (in percentages) of 50%, 30%, and 70% were applied to Cases 3–5.

Unlike Cases 1–5, which were related to QD and QWD, Cases 6–9 were based on QO. Reservoirs are usually operated to meet water demand, but as shown in Figure 5, the amount of water released from the reservoir may be less than the water demand due to drought or the status of the flow rate downstream. To assess whether the frequency of occurrence of CyanoHABs downstream could be reduced by releasing a similar amount of water to the observed, the constraint on *QA*+*QI* in Case 6 was set to be equivalent to QO. In Cases 6–9, the constraint of *QJ* was set to exceed 17.5 m^3 s^{-1}, considering the stability of the water level calculation in the HEC-RAS model used after the optimization process. When values of water level for unsteady flow in each cross section are calculated using HEC-RAS, a dry condition for a cross section makes an error (Brunner, 2016). To avoid this error, a minimum flow rate is required and the value of 17.5 m^3 s^{-1} was used as the minimum flow rate for the HEC-RAS model developed in Chapter 4.

The difference between Case 6 and Cases 7–9 was in the constraints applied to *QA* and *QI*. In Cases 7–9, we applied the earlier defined portions to QO for the constraints on *QA* and *QI*, which narrowed the range of *QA* and *QI*, enabling efficient optimization.

We carried out the optimization for these nine cases and then simulated the number of days with CyanoHABs at Chilgok Weir using the river water quality model and the machine learning model. The simulation results allowed us to evaluate the impact of changing the optimization constraints on the number of days with CyanoHABs.

5.4 Results and Discussion

5.4.1 Simulation test

We generated hypothetical data by modifying the observational data of the NO_3-N concentrations by increments or decrements of 0.50 mg L^{-1} for the simulation test, as stated in Section 5.3.1. These data were used as the upstream boundary condition for the HEC-RAS model. The other boundary conditions and model data remained unchanged. Figure 5.4 shows the results of the simulation test in terms of NO_3-N concentration at Chilgok Weir. Table 5.3 shows the number of days with CyanoHABs after one week at Chilgok Weir, obtained after applying the machine learning model based on the NO_3-N concentration shown in Figure 5.4.

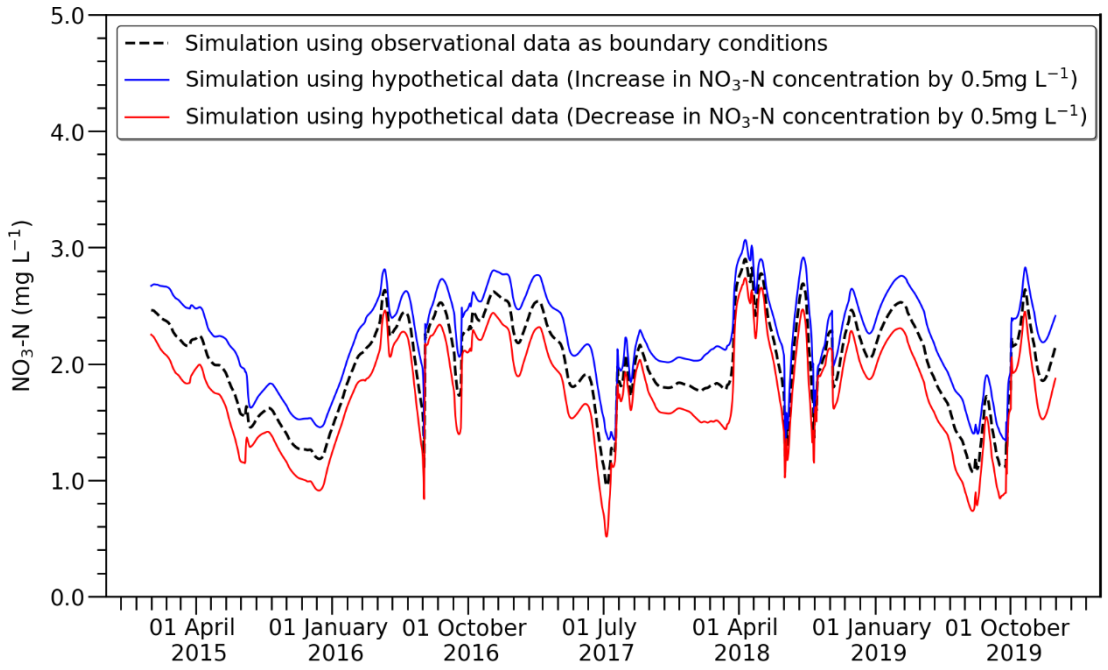


Figure 5.4. Results of the simulation test (the NO_3-N concentrations at Chilgok Weir using the hypothetical data and the HEC-RAS model)

Table 5.3. Results of the simulation test (the number of days with CyanoHABs after one week at Chilgok Weir using the NO_3-N concentrations of Figure 5.4)

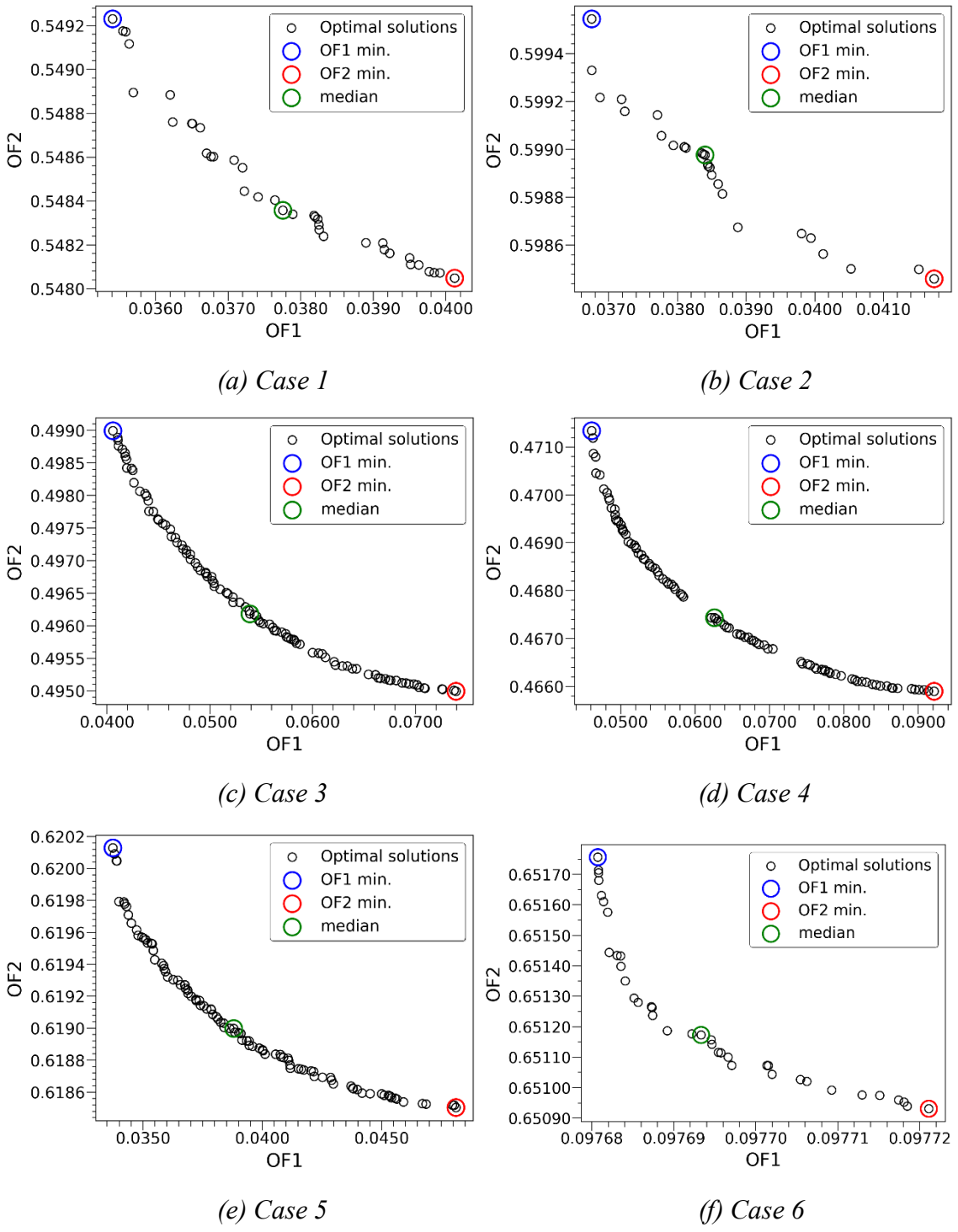
CyanoHABs	**Observation**	**Increment of 0.50 mg L^{-1}**	**Decrement of 0.50 mg L^{-1}**
Occurrence	72 days	66 days	103 days
Nonoccurrence	154 days	160 days	123 days
Sum	226 days	226 days	226 days

The results of the simulation test showed a negative correlation between the NO_3-N concentration upstream and the number of days with CyanoHABs, as shown in Table 5.3. These findings indicated that an increase in the NO_3-N concentration upstream resulted in a decrease in the number of days with CyanoHABs at Chilgok Weir, as expected. Hence, the simulation results demonstrated the validity of both the machine learning model and the HEC-RAS model.

5.4.2 Optimization results

We simulated the time series data for the decision variables, *QA*, *QI*, and *CI*, corresponding to a period of 1826 days over five years from 2015 to 2019 by applying the two objective functions presented in Section 5.2.3. The optimization process involved

the application of constraints for the nine cases outlined in Section 5.3.2. The optimization results were obtained as shown in Figure 5.5, which presents the Pareto front for each case.

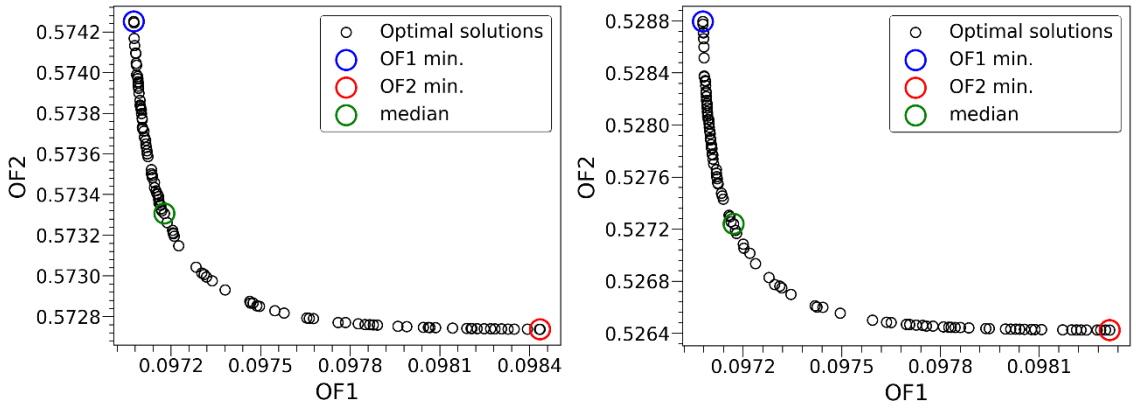


(a) Case 1 *(b) Case 2*

(c) Case 3 *(d) Case 4*

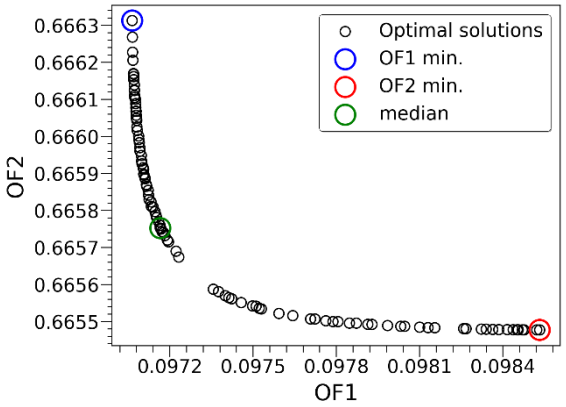
(e) Case 5 *(f) Case 6*

Figure 5.5. Cont.



(g) Case 7

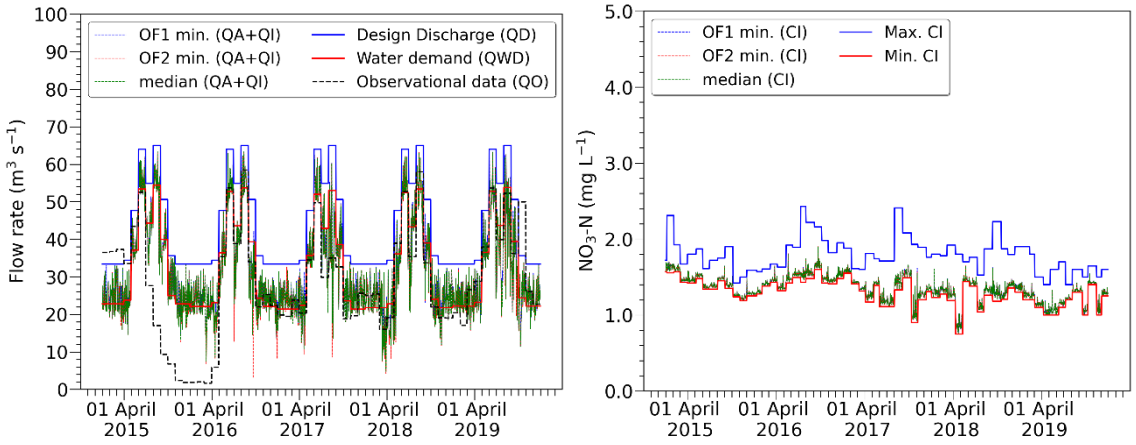
(h) Case 8



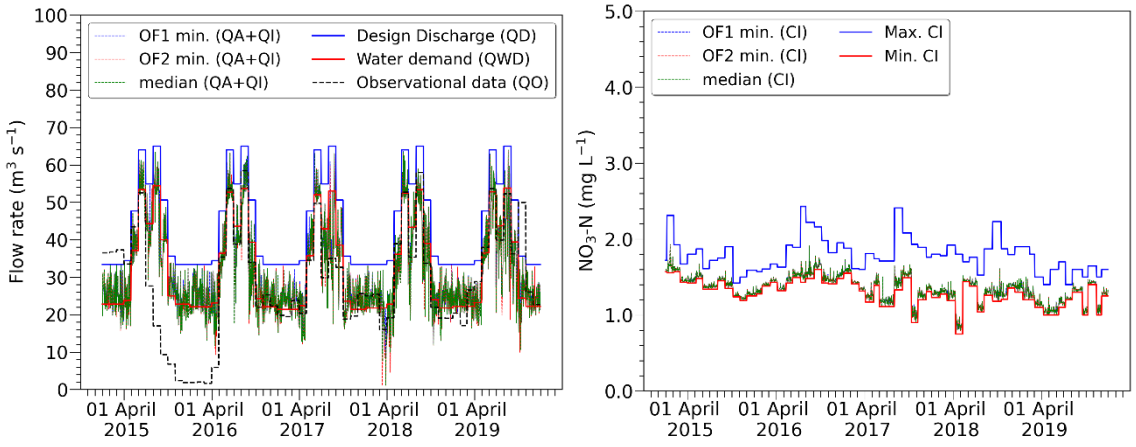
(i) Case 9

Figure 5.5. Pareto front (OF1 and OF2 are the objective functions related to water quantity and water quality, respectively.)

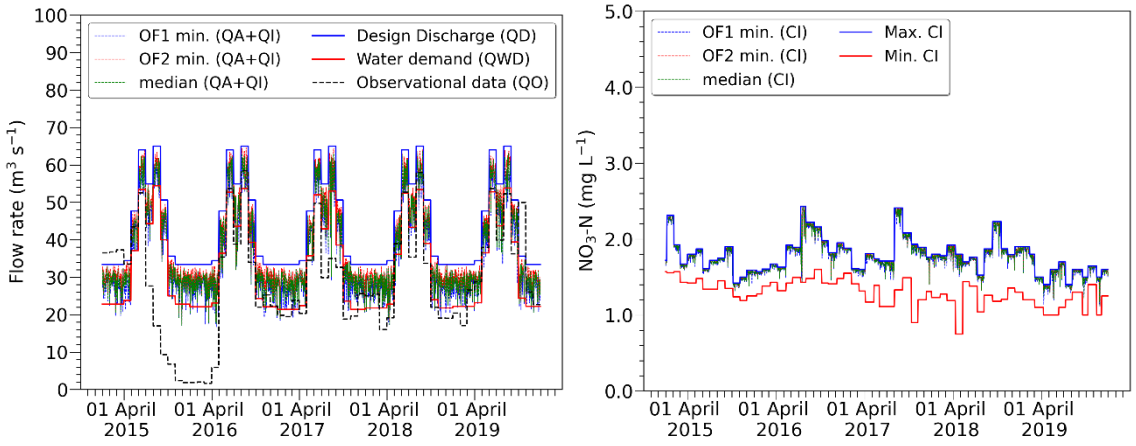
The differences were minor between the optimal solutions, as shown through the range of the x-axis (OF1) and y-axis (OF2) in Figure 5.5. This would lead to only small variations between the decision variables for the solutions. To further examine the relationship between the decision variables and the optimal solutions, three optimal solutions were selected for each case, as shown in Figure 5.5. These included a solution of most minimizing the first objective function (OF1 min.), a solution of most minimizing the second objective function (OF2 min.), and a median of the optimal solutions. The simulation results, as shown in Figure 5.6, confirmed slight differences between the decision variables (*QA+QI* and *CI*) corresponding to the three selected solutions in the nine cases.



(a) QA+QI for Case 1 *(b) CI for Case 1*

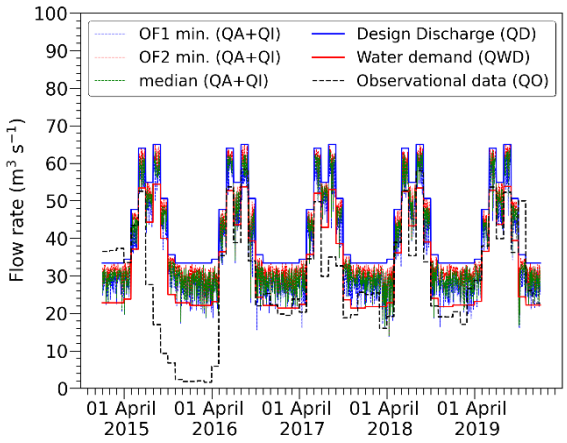


(c) QA+QI for Case 2 *(d) CI for Case 2*

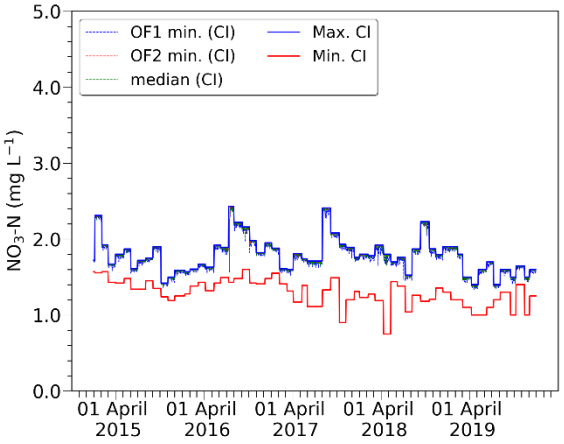


(e) QA+QI for Case 3 *(f) CI for Case 3*

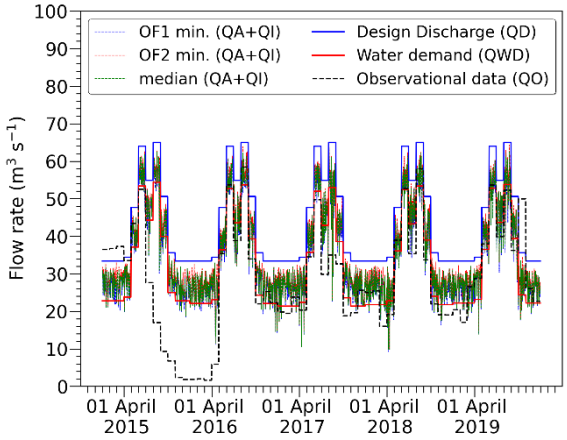
Figure 5.6. Cont.



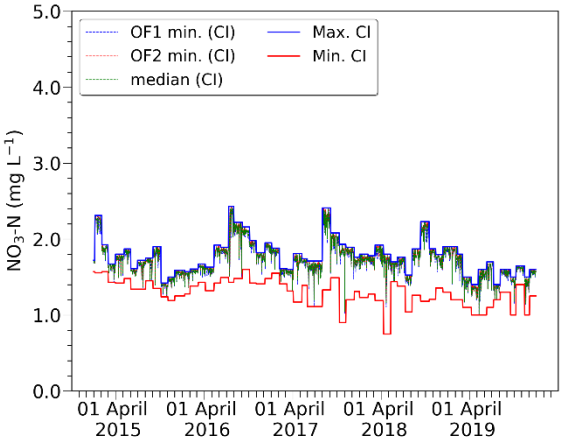
(g) QA+QI for Case 4



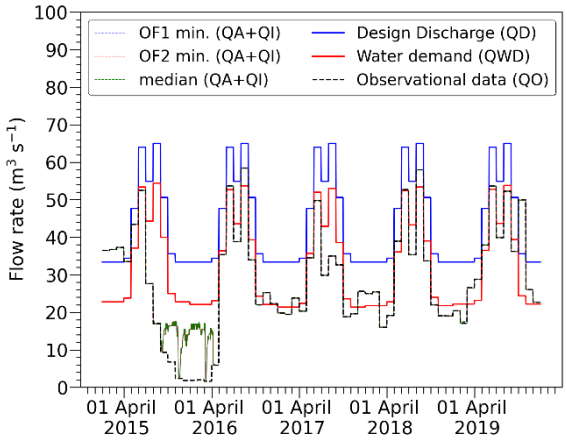
(h) CI for Case 4



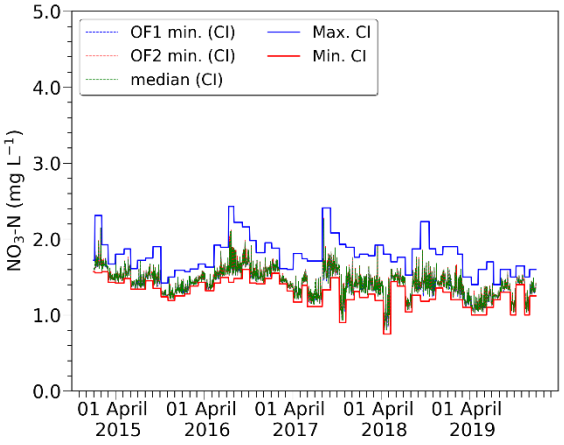
(i) QA+QI for Case 5



(j) CI for Case 5

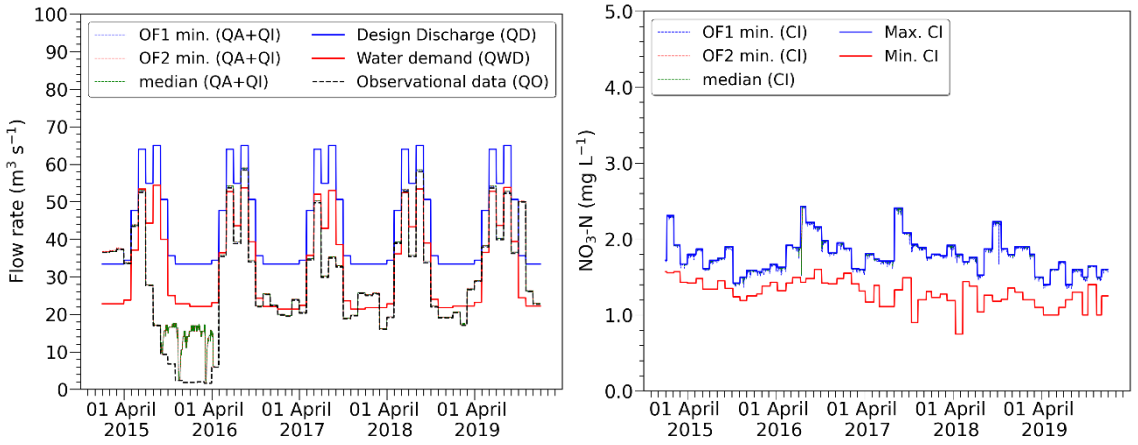


(k) QA+QI for Case 6

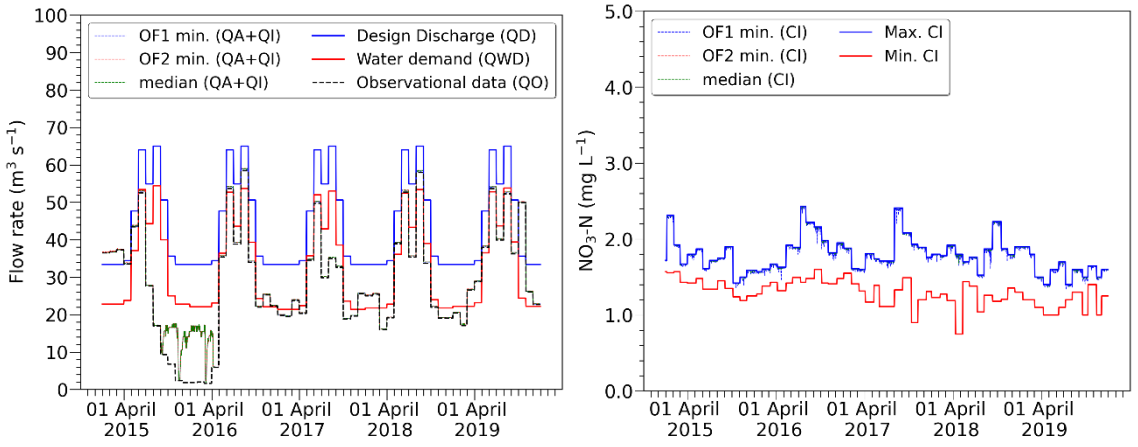


(l) CI for Case 6

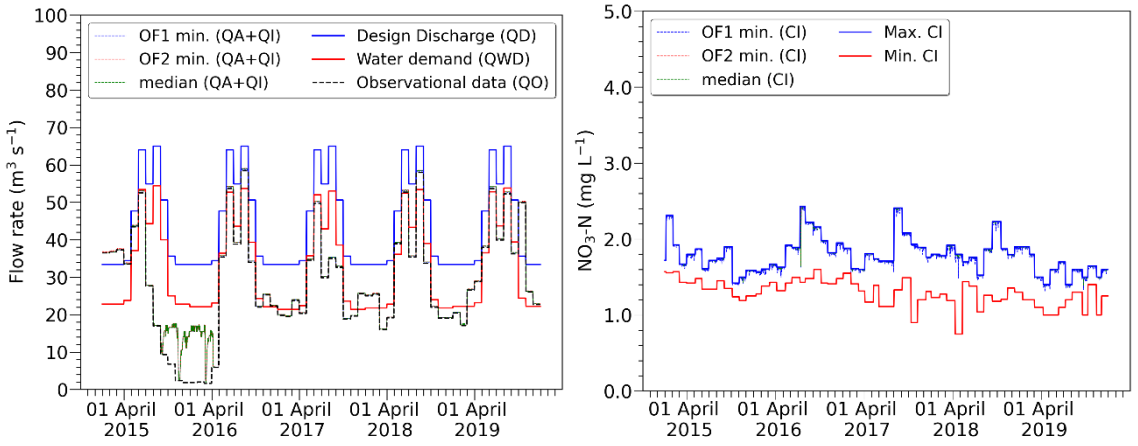
Figure 5.6. Cont.



(m) QA+QI for Case 7 *(n) CI for Case 7*



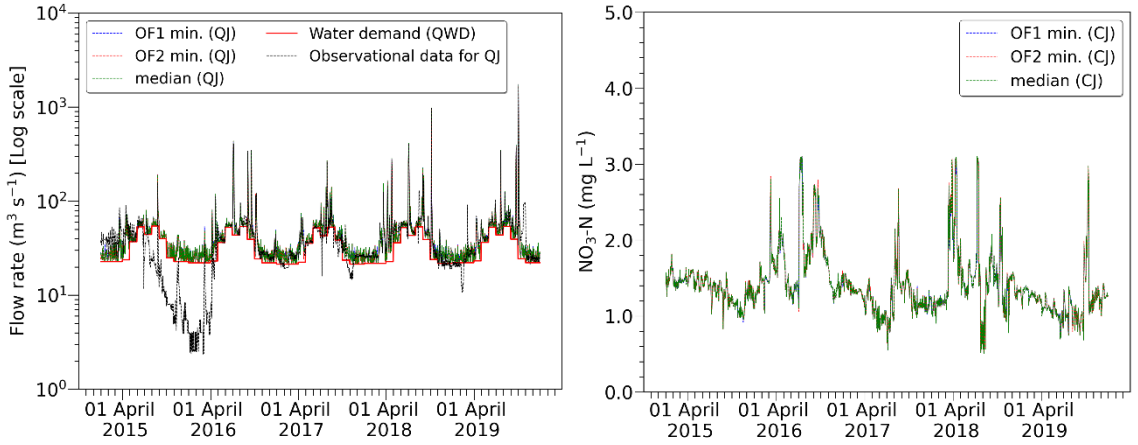
(o) QA+QI for Case 8 *(p) CI for Case 8*



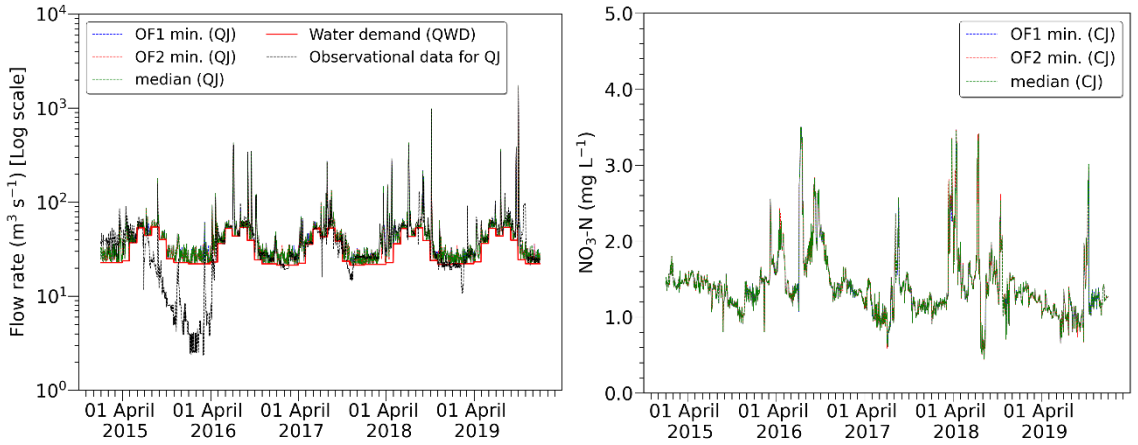
(q) QA+QI for Case 9 *(r) CI for Case 9*

Figure 5.6. Optimization results (QA+QI and CI) for three optimal solutions (OF1 min., OF2 min., median)

Figure 5.7 shows the results of *QJ* and *CJ* simulated by the constraints specified in the nine cases. These *QJ* and *CJ* were used as the upstream boundary conditions for the river water quality model. In Cases 1–5, the simulation results indicated that the values of *QJ* were consistently greater than or equal to QWD throughout the entire period. On the other hand, the simulation results of Cases 6–9 showed that the values of *QJ* might be lower than QWD. This is because Cases 6–9 used the observation data (QO) as the constraint on *QA*+*QI*, as shown in Table 5.1. Furthermore, the simulation results confirmed that *CJ* did not exceed the CR specified in each case.

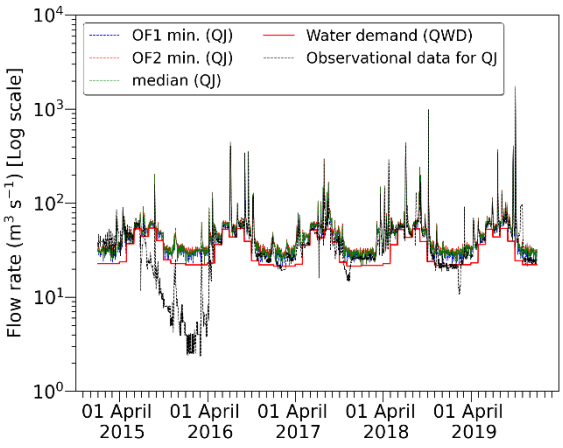


(a) QJ for Case 1 *(b) CJ for Case 1*

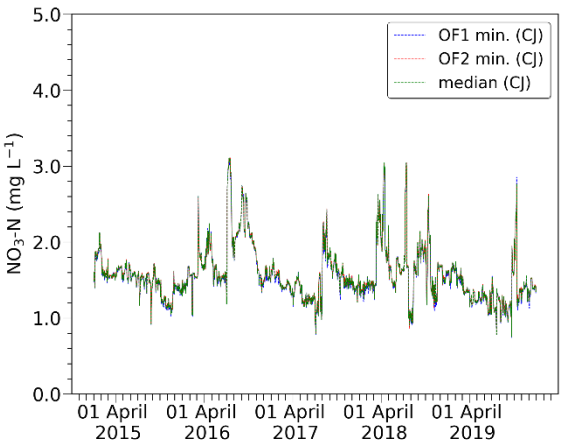


(c) QJ for Case 2 *(d) CJ for Case 2*

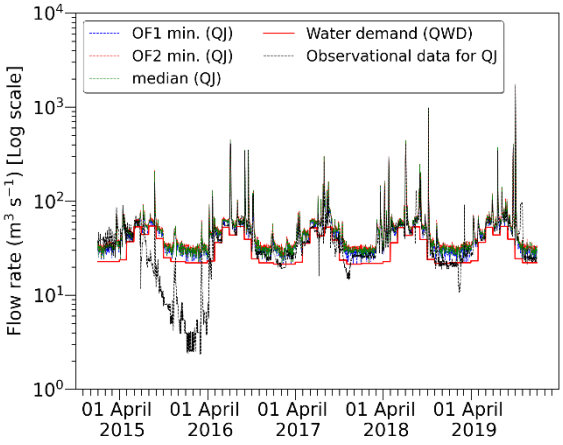
Figure 5.7. Cont.



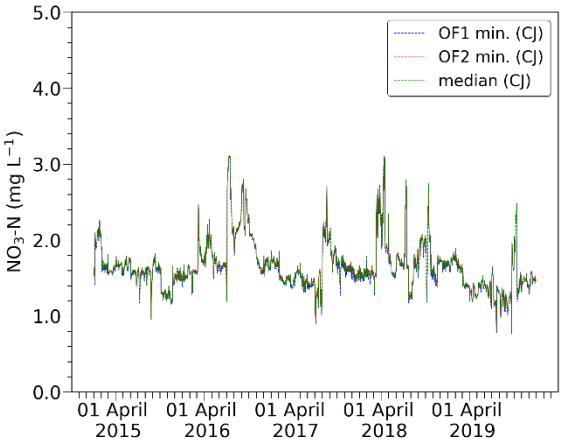
(e) QJ for Case 3



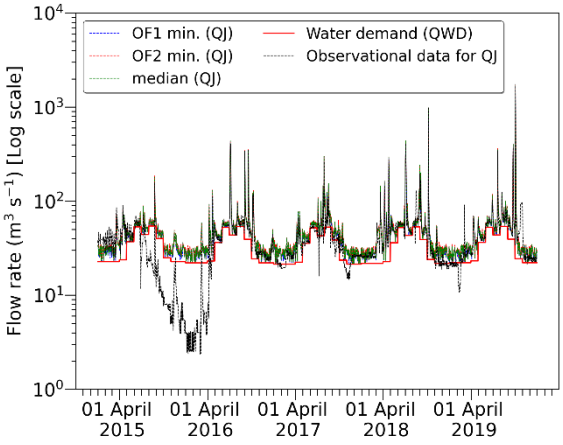
(f) CJ for Case 3



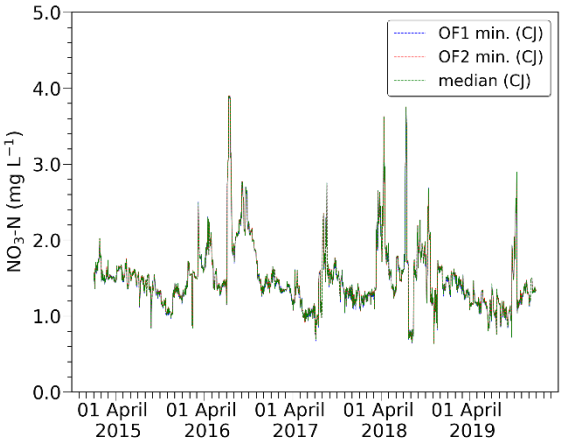
(g) QJ for Case 4



(h) CJ for Case 4

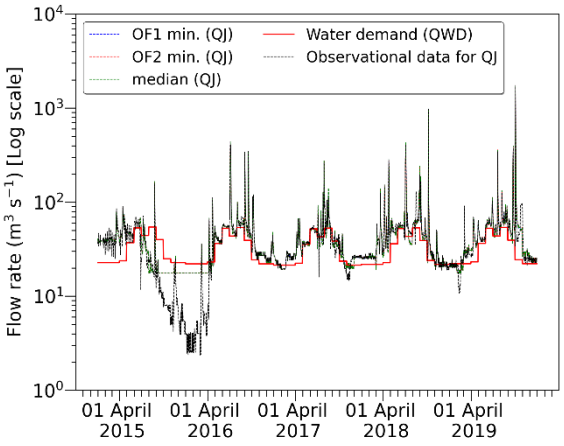


(i) QJ for Case 5

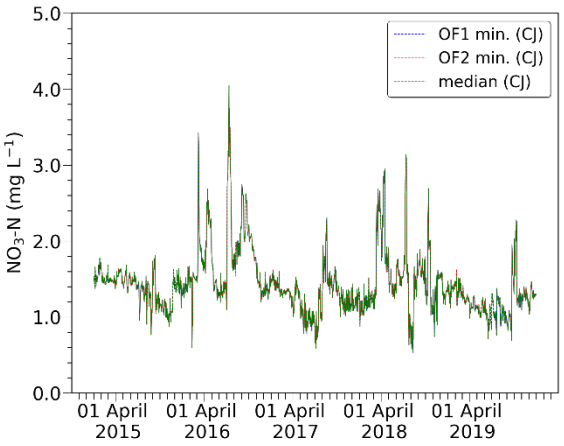


(j) CJ for Case 5

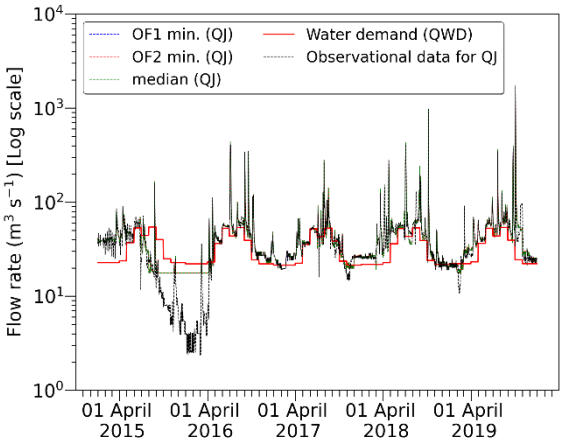
Figure 5.7. Cont.



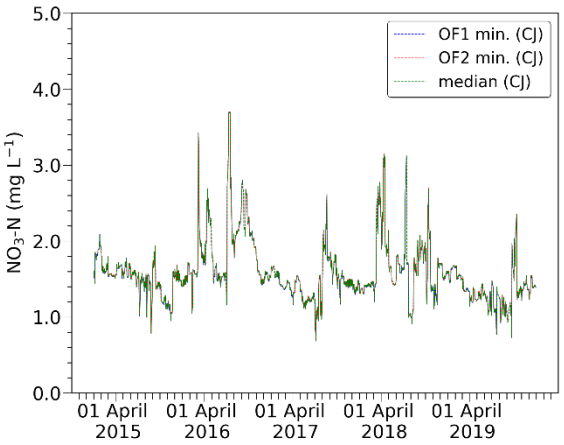
(k) QJ for Case 6



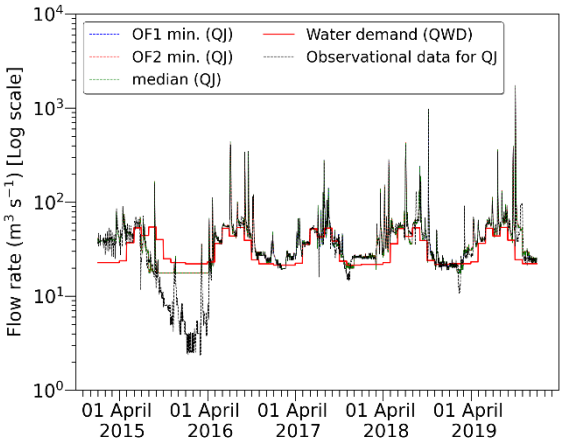
(l) CJ for Case 6



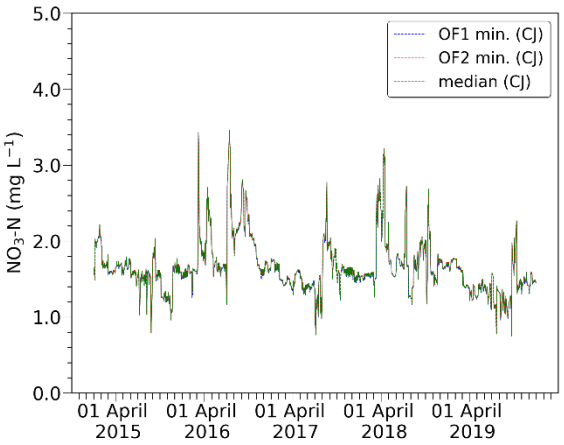
(m) QJ for Case 7



(n) CJ for Case 7

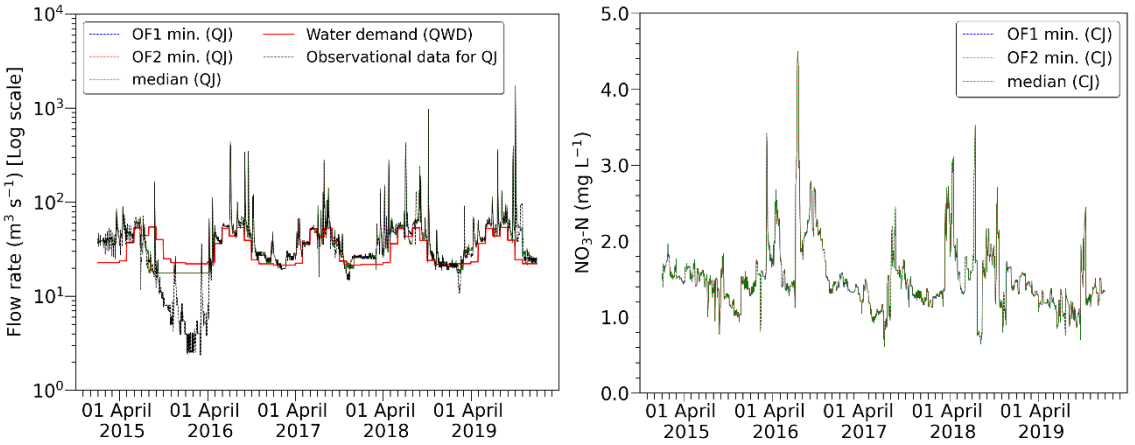


(o) QJ for Case 8



(p) CJ for Case 8

Figure 5.7. Cont.



(q) QJ for Case 9 *(r) CJ for Case 9*

Figure 5.7. Optimization results (QJ and CJ) for three optimal solutions (OF1 min., OF2 min., median)

5.4.3 NO_3-N dynamics at Chilgok Weir

The NO_3-N dynamics at Chilgok Weir were simulated by using the *QJ* and *CJ* values for the three optimal solutions (OF1 min., OF2 min., and median) as the upstream boundary conditions for the river water quality model. We used two models for the simulation: the HEC-RAS model and the surrogate model. The surrogate model was developed to replicate the NO_3-N dynamics simulated by the HEC-RAS model, as mentioned in Section 5.2.4.

We assessed the performance of the surrogate model by using both the graphical method and the traditional performance indices such as R^2, NSE, and RMSE (Root Mean Square Error). Figure 5.8 shows the simulation results of the NO_3-N concentrations at Chilgok Weir for Case 1. Although the simulation results of the two models were not 100% identical, the trend of the NO_3-N concentration, whether increasing or decreasing, was almost the same. The performance indices for the surrogate model are shown in Table 5.4. R^2 and NSE were higher than 0.90 and 0.80, respectively. In addition, RMSE was less than 0.20 for all cases. These performance indices are represented by Equations (4.2), (4.3), and (5.7), respectively (Moriasi et al., 2015). Based on these performance measures, the performance of the surrogate model can be judged as high.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - S_i)^2} \qquad (5.7)$$

where O is observational data and S is simulation result.

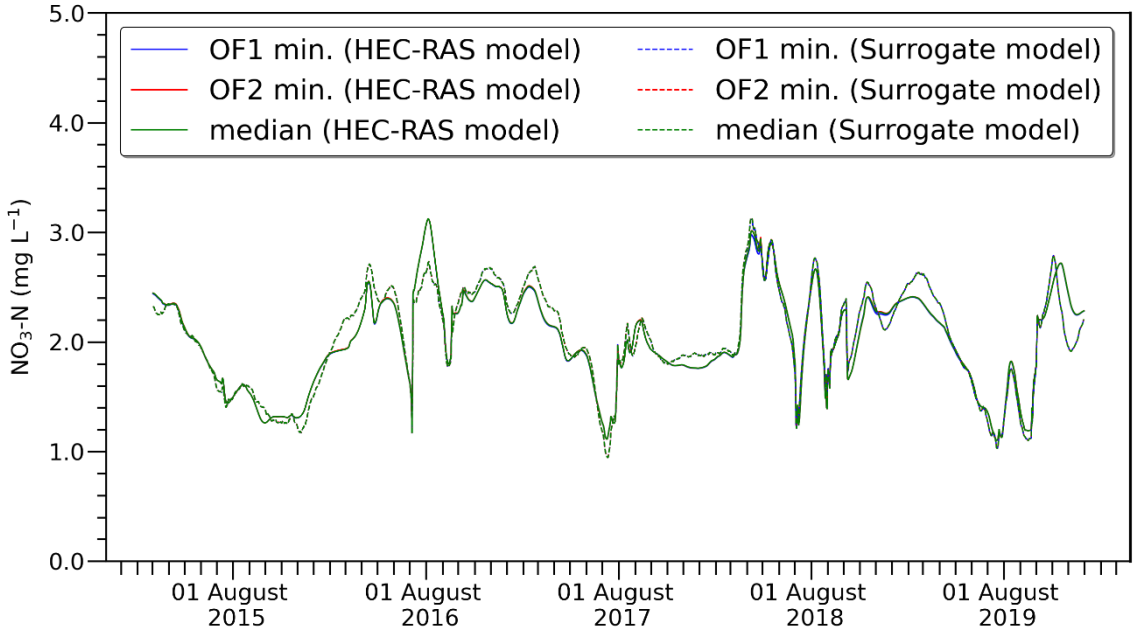


Figure 5.8. Simulation results of the NO_3-N concentrations at Chilgok Weir using the HEC-RAS model and the surrogate model for three optimal solutions

Table 5.4. Performance of the surrogate model for three optimal solutions

Case	R^2		
	OF1 min.	**OF2 min.**	**median**
Case1	0.921	0.921	0.920
Case2	0.918	0.917	0.918
Case3	0.915	0.912	0.914
Case4	0.905	0.902	0.905
Case5	0.907	0.906	0.907
Case6	0.954	0.954	0.954
Case7	0.943	0.943	0.943
Case8	0.932	0.932	0.932
Case9	0.943	0.943	0.943

Table 5.4. Cont.

Case	NSE		
	OF1 min.	**OF2 min.**	**median**
Case1	0.910	0.912	0.911
Case2	0.908	0.908	0.908
Case3	0.887	0.880	0.884
Case4	0.846	0.837	0.842
Case5	0.896	0.893	0.895
Case6	0.948	0.947	0.947
Case7	0.920	0.919	0.919
Case8	0.884	0.881	0.882
Case9	0.935	0.934	0.934
Case	**RMSE**		
	OF1 min.	**OF2 min.**	**median**
Case1	0.129	0.129	0.129
Case2	0.131	0.132	0.131
Case3	0.136	0.138	0.137
Case4	0.154	0.155	0.155
Case5	0.138	0.139	0.138
Case6	0.098	0.098	0.098
Case7	0.118	0.119	0.119
Case8	0.140	0.142	0.141
Case9	0.110	0.110	0.110

Apart from the high performance, the surrogate model had the advantage of saving computation time by approximately 1/3 or 1/4 compared to the HEC-RAS model. Despite these advantages, we needed to evaluate the final outcome of this study using the NO_3-N concentrations simulated by the surrogate model. The final outcome is the number of days with CyanoHABs after one week at Chilgok Weir.

5.4.4 Prediction of the occurrence of CyanoHABs at Chilgok Weir

We simulated the number of days with CyanoHABs after one week at Chilgok Weir. The simulation results were obtained by using the NO_3-N concentrations simulated by the two river water quality models (HEC-RAS and surrogate models). As shown in Table 5.5, the prediction results based on the surrogate model were significantly different from those based on the HEC-RAS model despite the high performance (see Table 5.4) of the surrogate model itself. Given the importance of ensuring the accuracy of the river water quality model in this study, we analyzed the simulation results based on the HEC-RAS model.

Table 5.5. Simulation results of the number of days with CyanoHABs after one week at Chilgok Weir for three optimal solutions

Case	HEC-RAS Model (Unit: days)		
	OF1 min.	**OF2 min.**	**median**
Case1	71	72	70
Case2	68	69	69
Case3	67	70	68
Case4	62	61	62
Case5	71	73	72
Case6	75	74	75
Case7	69	69	69
Case8	70	73	72
Case9	70	70	70
Case	**Surrogate Model (Unit: days)**		
	OF1 min.	**OF2 min.**	**median**
Case1	76	76	76
Case2	77	77	77
Case3	77	77	77
Case4	76	76	76
Case5	77	77	77
Case6	73	73	73
Case7	74	74	74
Case8	73	73	73
Case9	73	74	74

By comparing the observational data of 72 days with a cyanobacterial cell density of 1000 or higher from 2015 to 2019, we assessed the effect of reducing the number of days with CyanoHABs for nine cases. The results showed that, except for Case 6, the number of days was 72 or less in eight cases. Therefore, the applicability of the framework for the optimal operation of reservoirs was demonstrated in terms of reducing the frequency of CyanoHABs at Chilgok Weir.

Case 4 had the most noticeable effect in reducing the number of days with CyanoHABs among the nine cases. This simulation result of Case 4 would appear to be related to the constraint on the maximum amount of water from the Imha reservoir, which accounted for 70% of the Design Discharge (QD). This is because increasing the amount of water from the Imha reservoir would raise the likelihood of increasing the pollution load of NO_3-N downstream, as water quality could be regulated from the reservoir using the SWF. On the other hand, Case 6 showed the simulation result exceeding 72 days despite the fact that Case 6 had the objective functions and constraints which were formulated for the same purpose of reducing the number of days with CyanoHABs as the other cases.

As shown in Figures 5.6k and 5.6l for Case 6, *CI* for the optimal solutions varied near Min. CI while *QA+QI* was constrained on QO. This means that the variation in *CI* more greatly affected not only *CJ* but also the NO_3-N concentrations at Chilgok Weir than *QA+QI*. Comparing Case 6 with Cases 7–9, the values of *CI* for Cases 7–9 varied near Max. CI as shown in Figures 5.6m–5.6r even if the simulation results of water quantity for Cases 7–9 were similar to those for Case 6. As a result, the number of days with CyanoHABs at Chilgok Weir increased since *CI* (NO_3-N concentration) for Case 6 was low.

Interestingly, the simulation results in Cases 1 and 2 revealed that the effect of reducing the number of days with CyanoHABs would be produced with only the constraint on the water quality. The performance of Case 2 outweighed that of Case 1. In Cases 1 and 2, the reference concentrations (CR) were 3.11 mg L^{-1} and 3.50 mg L^{-1}, respectively. The other constraints in the two cases were the same. This result showed insight into how to reduce the frequency of occurrence of CyanoHABs downstream by controlling the quality of water from a reservoir by using the SWF, even with the same amount of water.

Among the simulation results in Cases 7–9, the reducing effect of the number of days with CyanoHABs was most significant in Case 7, followed by Case 9 and then Case 8. Cases 7–9 imposed constraints on the quantity of water based on the observational data (QO). These simulation results indicated that the optimal joint operation of the two reservoirs would lead to a reduction in the number of days with CyanoHABs using the same amount of water from the two reservoirs as the observational data. Nonetheless, the effect of Case 7 (69 days) was not as remarkable as that of Case 4 (61–62 days), where the constraint on the water quantity was set based on the water demand (QWD) and QD.

These simulation results suggest that the quantity of water from the reservoirs can have an impact on the NO_3-N loadings downstream.

The simulation results for the given cases confirmed that optimal operation of reservoirs, which simultaneously consider both the quantity and quality of water, would effectively decrease the frequency of CyanoHABs downstream. However, this method may not always be practical, as in Case 6, where the simulation result showed the number of days exceeding 72. This outcome was likely due to the indirect use of optimization results (*QJ* and *CJ*), which served as upstream boundary conditions of the river water quality model, for predicting the occurrence of CyanoHABs at Chilgok Weir. Thus, to achieve tangible results in reducing the frequency of CyanoHABs downstream, the series of processes outlined in the framework should take place in an orderly and systemic manner.

5.5 CONCLUSIONS

In this study, we demonstrated the applicability of the framework (see Figure 1.1) for a decision support system aimed at reducing the frequency of CyanoHABs at Chilgok Weir of the Nakdong River in South Korea, based on multi-objective optimization of the joint operation of the Andong and Imha reservoirs. This framework was designed by using three models: the machine learning model, the optimization model, and the river water quality model. In order to resolve the computational constraints on the optimization loop, a surrogate machine learning model which could replace the process-based model was developed. To demonstrate the applicability of this framework, these models were applied to the observational data from 2015 to 2019. The simulation results showed that the implementation of this framework would reduce the incidence of CyanoHABs downstream. Accordingly, this new approach to reservoir operation considering both the quantity and quality of water had applicability in terms of mitigating CyanoHABs downstream. Additionally, the framework is a novelty in terms of efficiency as it can be a part of a solution to the problem of CyanoHABs without using an additional amount of water from a reservoir.

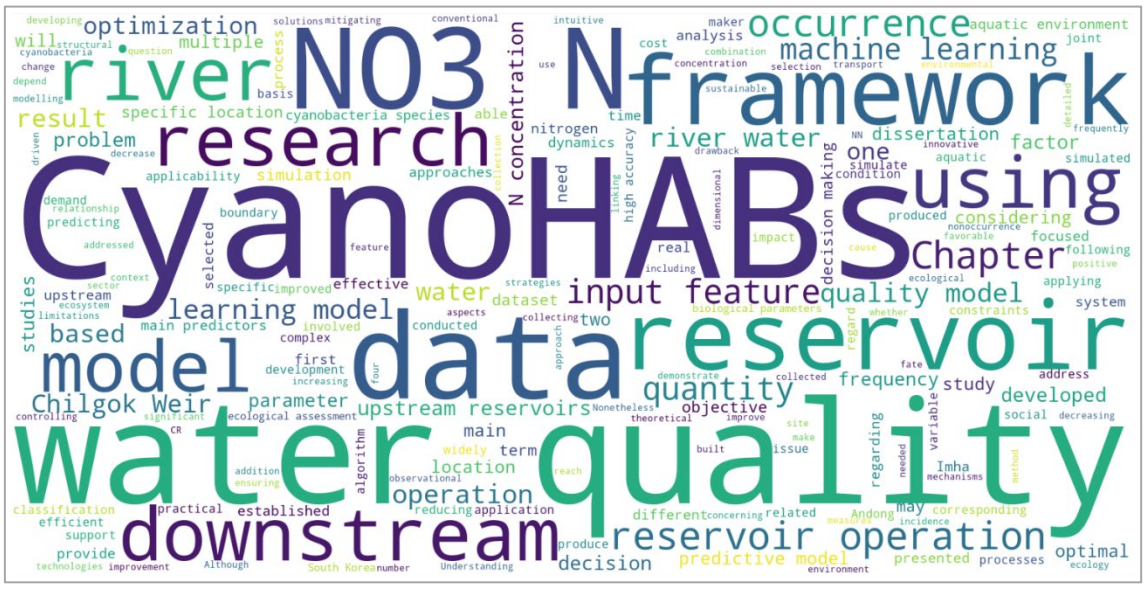
The framework established in this study offered a methodology not to prevent the occurrence of CyanoHABs but to reduce their frequency. Nevertheless, the framework has advantages in terms of efficiency for two reasons. First, it can decrease the frequency of CyanoHABs in rivers without incurring any costs, unlike the current technologies for algae removal (Yang et al., 2023). Second, the optimal operation of reservoirs does not require an additional amount of water, unlike the reservoir operation which has been involved in flushing water to reduce the frequency of CyanoHABs in a river downstream. Therefore, this framework related to the management of water quality can efficiently support sustainable development in terms of human health and the management of aquatic ecosystems (Ustaoglu et al., 2021; Forio and Goethals, 2020).

However, this study also has limitations. First, this study did not consider flood routing in the reservoirs. In this regard, we used observational data from 2015 to 2019 that did not have a record of discharge through the spillway for flood control of the Andong and Imha reservoirs. Secondly, errors in the river water quality model, particularly in the surrogate model, had an apparently negative impact on the simulation results of the occurrence of CyanoHABs. Further studies are needed to transcend these limitations by dealing with the optimization process considering flood routing in reservoirs and the performance improvement of river water quality models.

6

CONCLUSIONS

This chapter concludes this dissertation with reflections on the research questions raised in Chapter 1, research outcomes, research limitations, and recommendations for further studies.



6.1 Reflections on Research Questions

The main objective of this dissertation is to demonstrate the applicability of the framework presented in Figure 1.1. This framework for optimal reservoir operation was established to decrease the frequency of occurrence of CyanoHABs at a specific location downstream. Throughout this dissertation, the research questions raised in Chapter 1 have been addressed in order to accomplish the main objective as follows:

i. What is an effective and efficient way to predict the occurrence of CyanoHABs at a specific location downstream in terms of linking with the operation of upstream reservoirs?

In Chapter 3, the machine learning models ensuring high accuracy of more than 80% were developed using only two input features based on four classification algorithms such as k-Nearest Neighbor (k-NN), Decision Tree (DT), Logistic Regression (LR), and Support Vector Machine (SVM). To build an effective model for predicting the occurrence or nonoccurrence of CyanoHABs with high accuracy, input features were first selected by applying ANOVA (Analysis of Variance) and solving a multi-collinearity problem. Next, an oversampling method of SMOTE was adopted to overcome the problem of having an imbalanced dataset on CyanoHABs.

As a result, a model applying the k-NN algorithm ensured high accuracy in predicting the occurrence or nonoccurrence of CyanoHABs at Chilgok Weir. This efficient model was developed by using only two input features: average air temperature (AT) and nitrate nitrogen (NO_3-N). Understanding the input features affecting CyanoHABs downstream enables the development of viable strategies for reservoir operations. This operation of upstream reservoirs can be conducted for controlling the specific water quality parameter which is one of the input features of the machine learning model. The input feature to be controlled by reservoir operation is NO_3-N concentration in this research.

ii. How can a river water quality model be developed to simulate the fate and transport of water quality parameters involved in CyanoHABs to cover a river reach between upstream reservoirs and a specific location downstream?

In Chapter 4, a river water quality model using HEC-RAS was built to simulate the dynamics of NO_3-N concentration which was identified in Chapter 3 as one of the main factors for the occurrence of CyanoHABs at Chilgok Weir. This river water quality model covered the river reach between the confluence of the water from the upstream reservoirs (Andong and Imha) and the Chilgok Weir.

By applying this river water quality model, the fate and transport of NO_3-N were understood under the scenarios based on variations in the quantity and quality of water at the upstream boundary. The simulation results showed how different aspects of the NO_3-N dynamics downstream were depending on flow rate and NO_3-N concentration upstream.

These changes in the quantity and quality of water at the upstream boundary can be artificially produced by the actual operation of the upstream reservoirs.

iii. What optimization process for the operation of upstream reservoirs should be set up, for simultaneously considering both the quantity and quality of water downstream?

In Chapter 5, the optimization model for joint operation of the upstream reservoirs (Andong and Imha) was produced to satisfy the demand for both the quantity and quality downstream. The decision variables for water quantity were constrained based on the optimal joint operation of the two reservoirs. The decision variable for water quality was the NO_3-N concentration, which is the main factor for the occurrence of CyanoHABs at Chilgok Weir, by considering the use of a selective withdrawal facility of the Imha Reservoir.

iv. How can the optimal operation of upstream reservoirs be coupled to a predictive model for CyanoHABs and a river water quality model?

The three models developed in each chapter were connected within the framework. The optimization model first simulated the decision variables for the quantity and quality (NO_3-N) of water released from the two reservoirs (Andong and Imha). By using the optimization results as the upstream boundary conditions, the river water quality model simulated the dynamics of NO_3-N at Chilgok Weir. Finally, the machine learning model predicted the occurrence of CyanoHABs at Chilgok Weir by using the NO_3-N concentrations simulated by the river water quality model as one of the input features.

In Chapter 5, the applicability of the framework was demonstrated by simulations using observational data for the study area. The simulation results based on the framework confirmed that the frequency of CyanoHABs would be decreased compared to the number of days when CyanoHABs were observed at Chilgok Weir.

6.2 RESEARCH OUTCOMES

In this dissertation, a practical framework has been presented by combining different types of water quality models with optimal operation of upstream reservoirs for decreasing the frequency of CyanoHABs in a downstream river. The research of this dissertation has produced results that conform not only to the innovative approaches from the scientific perspective distinguished from the conventional reservoir operation, but also to the social interest through the improvement of the aquatic environment downstream, as presented in the following sections.

6.2.1 Scientific perspective

This research provides the following scientific perspective:

i. There is a novelty in the approach for the optimal operation of reservoirs to address the issues concerning CyanoHABs downstream. Although there have been studies on optimization for reservoir operation considering the demands on both the quantity and quality of water in a downstream river, these studies have focused not on biological parameters of water quality such as CyanoHABs, but on physical or chemical parameters (Yosefipoor et al., 2022; Saadatpour et al., 2021; Saadatpour et al., 2020). The framework established in this research will be able to contribute to the decision-making of reservoir operation in practice to create a favorable aquatic environment downstream by reducing the incidence of CyanoHABs related to complex processes of biological parameters (Kim et al., 2017; Park et al., 2017b).

ii. The predictive model for CyanoHABs developed in Chapter 2 can help decision-makers in the water sector to formulate effective strategies for preventing the occurrence of CyanoHABs at a specific location. If many factors were involved in decision-making, a decision-maker could struggle to implement a strategy. This means that a decision-maker can need a predictive model which produces intuitive results only with a few factors involved in issues to be addressed. In this regard, the machine learning model based on a classification algorithm using two input features can be effective in terms of decision-making.

iii. The combination of the optimization model for reservoir operation (Chapter 5) with the one-dimensional river water quality model (Chapter 4) and the machine learning model (Chapter 3) can facilitate improved integration of the reservoir-river system considering the quantity and quality of water. This comprehensive framework offers the optimal operation of the reservoir away from conventional approaches to reservoir operation which have only focused on water quantity. Additionally, efficiency in computing time can be improved by employing the one-dimensional model and data-driven model. In particular, this framework is anticipated to be practically used in South Korea, which introduced a policy on the integrated management of reservoirs and rivers in 2022.

6.2.2 Environmental and social impact

The results from this research are expected to have the following environmental and social impacts:

i. The application of the framework can not only improve the aquatic environment in a river but also produce positive effects on public health by decreasing the frequency of CyanoHABs which produce toxic substances.

ii. Beneficial impacts on the aquatic ecosystem in rivers can be created owing to the improvement of the aquatic environment. Although this research did not carry out the ecological assessment, mitigating CyanoHABs can provide fish and other organisms with favorable environments.

iii. A reduction in the cost of producing drinking water can be achieved as the quality of raw water in rivers can be improved without using an additional amount of water from reservoirs.

iv. Practical applications of this research method to South Korea can provide an opportunity to settle the social conflict between different stakeholders regarding the question of whether the river water quality has become low due to the Four Major Rivers Restoration Project or not by reducing the frequency of CyanoHABs in rivers.

v. This research can make a significant contribution towards expanding the scope of the role of reservoirs from only ensuring water supply to improving the water environment of rivers. Furthermore, the research context of this dissertation can ultimately lead to a paradigm shift in reservoir operation as there is increasing demand for clean water. However, for the paradigm shift in reservoir operation, reservoirs should be equipped with SWFs or joint operation of reservoirs must be carried out. Agencies in charge of reservoir operation need to recognize the necessity for these structural or non-structural measures including costs to be paid. These costs may seem to be a drawback in the short term, but the advantages in the long term outweigh the drawback. This is because these measures will be one of the solutions to problems with water quality caused by climate change and thus will support sustainable development in the water sector.

6.3 LIMITATIONS AND RECOMMENDATIONS

6.3.1 Research Limitations

This research focused on controlling the NO_3-N concentration in the river to reduce the frequency of occurrence of CyanoHABs, since NO_3-N was selected as main input feature of the machine learning model. Choosing NO_3-N as main input feature was the result of the feature selection based on not only the theoretical knowledge of CyanoHABs but also a rational approach to data analysis, as mentioned in Chapter 3. Nonetheless, this study may have limitations regarding ecological parameters considered, because factors such as TN (Total Nitrogen) and TP that are widely recognized as main predictors for CyanoHABs (Rousso et al., 2020) were not incorporated in the input features of the machine learning model. Accordingly, in the process of selecting input features (water quality data) related to the occurrence of CyanoHABs, a more theoretical understanding

of the complex mechanisms of algal growth would be needed, and detailed consideration related to the water quality processes may be required.

CyanoHABs are widely known to frequently occur in a condition of nutrient over-enrichment such as nitrogen and phosphorous (Paerl and Otten, 2013; Paerl et al., 2001; Noori et al., 2021). Nevertheless, NO_3-N was negatively correlated with the occurrence of CyanoHABs at Chilgok Weir as shown in Chapter 3. Interestingly, while a positive correlation between nitrogen compounds and CyanoHABs is widely accepted, the relationship between them can depend on site-specific factors and cyanobacteria species present (Park et al., 2021a; Zakova et al., 1993; Harrow-Lyle and Kirkwood, 2020; Deng et al., 2007; Jahan et al., 2010; Rousso et al., 2020). This highlights the need for context-specific approaches to managing freshwater systems, which take into account local conditions and the ecology of the system.

Despite the importance of identifying cyanobacteria species, the dataset on cyanobacterial cell density was not split into the groups of harmful cyanobacteria genera when developing the machine learning model. This was because analysis considering cyanobacteria species could cause the problem of an insufficient number of instances for the dataset. Hence, this machine learning model has limitations, since it did not involve the distinction of different cyanobacteria species.

In addition, because of this relationship between NO_3-N and CyanoHABs at Chilgok Weir, releasing water with a high concentration of NO_3-N from the reservoirs can decrease the frequency of CyanoHABs. However, this reservoir operation may cause the problem of increasing the NO_3-N concentration in a downstream river. Therefore, constraints on NO_3-N for the optimization process should be carefully formulated to address this issue. In this regard, we imposed constraints on the NO_3-N concentration downstream using the reference concentration (CR). These constraints enabled the NO_3-N concentration downstream to be maintained below CR.

6.3.2 Recommendations for further studies

One of the most significant factors in developing a robust predictive model is data availability (Goethals and Forio, 2018; Ho and Goethals, 2022). Fortunately, data on CyanoHABs including cyanobacterial cell density have been regularly examined and managed by the environmental authority in South Korea. However, obtaining water quality data is more difficult than collecting data on water quantity. This is because most water quality data are generally collected in situ and obtained in laboratory experiments. In this regard, these water quality data are available on a weekly or monthly basis. Water quality modelling requires hydrological or hydraulic data and meteorological data collected at least on a daily basis as well as water quality data. The problem of differences in the time interval between these data may affect the performance of the predictive model. To resolve this problem, the need for frequently collecting water quality has arisen.

Emerging technologies such as real-time and on-site data collection based on remote sensing and the Internet of Things (Chowdury et al., 2019; Ho and Goethals, 2022) can be pragmatic solutions to the problem. Further research will be able to demonstrate the effectiveness of linking these technologies regarding real-time data collection with the framework established in this study.

This research did not incorporate the simulation of water quality in reservoirs corresponding to Step 2 of the framework. This decision was made based on the objective of this dissertation, which was aimed at demonstrating the applicability of the framework using observational data on water quality in the two reservoirs. Further studies will be able to make a practical application of the framework through the use of the modelling systems, such as CE-QUAL-W2, EELCOM-CAEDYM, and EFDC, for the simulation of water quality in reservoirs (Gao and Li, 2014).

This study can offer scalability for enhancing sustainable development when linked to ecological assessments in a river. The impact of mitigating CyanoHABs on the aquatic ecosystem can be first analyzed and the analysis result can be included as a factor of reservoir operation (Forio and Goethals, 2020). Further studies on the reservoir operation considering ecological assessments will be able to suggest innovative approaches in terms of the diversity of aquatic ecology in addition to the quantity and quality of water.

As the framework established in this study focused on only one specific location, the Chilgok Weir, further studies can be aimed at reducing the incidence of CyanoHABs at multiple locations. If the main predictors for CyanoHABs depend on a location, machine learning models, river water quality models, and optimization models should be developed corresponding to each location. Machine learning models first need to be developed corresponding to multiple locations, and the main predictors of the occurrence of CyanoHABs should be selected for each location. River water quality models can be built for each location to simulate the dynamics of the multiple parameters of water quality selected as the main predictors. An optimization model can be designed to address objective functions regarding water quantity and the multiple parameters of water quality. To simplify the objective functions, a water quality index can incorporate the multiple parameters of water quality (Yosefipoor et al., 2022; Saadatpour et al., 2021; Saadatpour et al., 2020).

Concerning decision-making, the predictive models using nominal or ordinal data can be more efficient than real-valued data as the results using the former type of data are much more intuitive (Kim et al., 2020). Nonetheless, building a model based on real-valued data, rather than on a classification basis, could support more detailed decision-making. Further research would be needed on whether the feature selection process presented in this study can improve the accuracy of such a model.

Since cyanobacteria are one of the biological parameters of water quality with complex mechanisms, more studies recently have been conducted on predicting the occurrence of

CyanoHABs using data-driven models (Rousso et al., 2020), which are developed with historical datasets. As these datasets are outcomes of complicated processes for CyanoHABs, further studies can be conducted by using machine learning models with multiple combinations of input features, combined with Interpretable Artificial Intelligence (IAI) and eXplainable Artificial Intelligence (XAI) that provide post hoc explanations based on ecological aspects (Başağaoğlu et al., 2022).

REFERENCES

Korea legislation research institute. The statutes of the Republic of Korea home page: Act on the investigation, planning, and management of water resources: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=55436&lang=ENG, last access: 28 September 2022.

Korea legislation research institute. The statutes of the Republic of Korea home page: Water environment conservation act: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=54838&lang=ENG, last access: 28 September 2022.

Korea legislation research institute. The statutes of the Republic of Korea home page: Weather act: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=54691&lang=ENG, last access: 28 September 2022.

National institute of environmental research. Water environment information system: https://water.nier.go.kr, last access: 24 April 2023.

Korea legislation research institute. The statutes of the Republic of Korea home page: River act: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=57588&lang=ENG, last access: 28 September 2022.

Abed, B. S., Daham, M. H., and Al-Thamiry, H. A.: Assessment and modelling of water quality along al-gharraf river (iraq), J. Green Eng., 10, 13565–13579, 2020.

Abed, B. S., Daham, M. H., and Ismail, A. H.: Water quality modelling and management of Diyala River and its impact on Tigris River, J. Eng. Sci. Technol., 16, 122–135, 2021.

Aguilar, J., Van Andel, S.-J., Werner, M., and Solomatine, D. P.: Hydrodynamic and water quality surrogate modeling for reservoir operation, 11th International Conference on Hydroinformatics, New York, NY, USA, 2014.

Ahmed, M., Mumtaz, R., and Mohammad, S.: Analysis of water quality indices and machine learning techniques for rating water pollution: A case study of Rawal Dam, Pakistan, Water Supply, 21, 3225–3250, doi:10.2166/ws.2021.082, 2021.

Ahn, J. M., Kim, J., Park, L. J., Jeon, J., Jong, J., Min, J.-H., and Kang, T.: Predicting cyanobacterial harmful algal blooms (CyanoHABs) in a regulated river using a revised EFDC model, Water, 13, 439, doi:10.3390/w13040439, 2021.

Al-Abadi, A. M., Handhal, A. M., and Al-Ginamy, M. A.: Evaluating the Dibdibba aquifer productivity at the Karbala-Najaf plateau (Central Iraq) using GIS-based tree machine learning algorithms, Nat. Resour. Res., 29, 1989–2009, doi:10.1007/s11053-019-09561-x, 2020.

Alam, M. J. and Dutta, D.: Modelling of nutrient pollution dynamics in river basins: A review with a perspective of a distributed modelling approach, Geosciences, 11, 369, doi:10.3390/geosciences11090369, 2021.

Arabgol, R., Sartaj, M., and Asghari, K.: Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines (SVMs) model, Environ. Model. Assess., 21, 71–82, doi:10.1007/s10666-015-9468-0, 2016.

Bae, S. and Seo, D.: Changes in algal bloom dynamics in a regulated large river in response to eutrophic status, Ecol. Model., 454, 109590, doi:10.1016/j.ecolmodel.2021.109590, 2021.

Başağaoğlu, H., Chakraborty, D., Lago, C. D., Gutierrez, L., Şahinli, M. A., Giacomoni, M., Furl, C., Mirchi, A., Moriasi, D., and Şengör, S. S.: A review on interpretable and explainable artificial intelligence in hydroclimatic applications, Water, 14, 1230, doi:10.3390/w14081230, 2022.

Bhuyan, K. J., Saharia, P. K., Sarma, D., Bhagawati, K., Baishya, S., and Hazarika, S. R.: Effect of periphyton (Streblus asper lour.) assemblage on water quality parameters and growth perfomance of Jayanti Rohu and Amur Common carp in the aquaculture system, J. Krishi Vigyan, 9, 74–80, doi:10.5958/2349-4433.2020.00084.7, 2020.

Blank, J. and Deb, K.: Pymoo: Multi-objective optimization in python, Ieee Access, 8, 89497–89509, doi:10.1109/Access.2020.2990567, 2020.

Bohan, J. P. and Grace, J. L.: Selective withdrawal from man-made lakes, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, USA, 1973.

Bourel, M. and Segura, A. M.: Multiclass classification methods in ecology, Ecol. Indic., 85, 1012–1021, doi:10.1016/j.ecolind.2017.11.031, 2018.

Brunner, G. W.: HEC-RAS river analysis system user's manual version 5.0, U.S. Army Corps of Engineers, Institute for Water Resources, Hydrologic Engineering Center, Davis, CA, USA, 2016.

Carmichael, W. W. and Boyer, G. L.: Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes, Harmful Algae, 54, 194–212, doi:10.1016/j.hal.2016.02.002, 2016.

Celikkol, S., Fortin, N., Tromas, N., Andriananjamanantsoa, H., and Greer, C. W.: Bioavailable nutrients (N and P) and precipitation patterns drive cyanobacterial blooms in Missisquoi Bay, Lake Champlain, Microorganisms, 9, 2097, doi:10.3390/microorganisms9102097, 2021.

Choi, H.-G. and Han, K.-Y.: Development and applicability assessment of 1-D water quality model in Nakdong River, KSCE J. Civ. Eng., 18, 2234–2243, doi:10.1007/s12205-014-0457-7, 2014.

Choi, J.-H., Kim, J., Won, J., and Min, O.: Modelling chlorophyll-a concentration using deep neural networks considering extreme data imbalance and skewness, 2019 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea,, 631–634, doi:doi.org/10.23919/ICACT.2019.8702027, 2019.

Chou, J. S., Pham, T. T. P., and Ho, C. C.: Metaheuristic optimized multi-level classification learning system for engineering management, Appl. Sci., 11, 5533, doi:10.3390/app11125533, 2021.

Chowdury, M. S. U., Bin Emran, T., Ghosh, S., Pathak, A., Alam, M. M., Absar, N., Andersson, K., and Hossain, M. S.: IoT based real-time river water quality monitoring system, Procedia Comput. Sci., 155, 161–168, doi:10.1016/j.procs.2019.08.025, 2019.

Costa, C. M. d. B., Leite, I. R., Almeida, A. K., and de Almeida, I. K.: Choosing an appropriate water quality model—A review, Environ. Monit. Assess., 193, 38, doi:10.1007/s10661-020-08786-1, 2021.

Cullinan, V. I., May, C. W., Brandenberger, J. M., Judd, C., and Johnston, R. K.: Development of an empirical water quality model for stormwater based on watershed land use in Puget Sound, Space and Naval Warfare Systems Center, Marine Environmental Support Office, Bremerton, WA, USA, 2007.

Daggupati, P., Pai, N., Ale, S., Douglas-Mankin, K. R., Zeckoski, R. W., Jeong, J., Parajuli, P. B., Saraswat, D., and Youssef, M. A.: A recommended calibration and validation strategy for hydrologic and water quality models, T. Asabe 58, 1705–1719, doi:10.13031/trans.58.10712, 2015.

Danaraj, J., Ushani, U., Packiavathy, S. V., Dharmadhas, J. S., Karuppiah, T., Anandha Kumar, S., and Aooj, E. S.: Climate change impacts of nitrate contamination on human health, in: Climate change impact on groundwater resources: Human health risk assessment in arid and semi-arid regions, edited by: Panneerselvam, B., Pande, C. B., Muniraj, K., Balasubramanian, A., and Ravichandran, N., Springer International Publishing, Cham, Switzerland, 257–278, 10.1007/978-3-031-04707-7_14, 2022.

Davis, J. E., Holland, J. P., Schneider, M. L., and Wilhelms, S. C.: Select: A numerical, one-dimensional model for selective withdrawal, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, USA, 1987.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga-ii, IEEE Trans. Evol. Comput., 6, 182–197, doi:10.1109/4235.996017, 2002.

Deng, D. G., Xie, P., Zhou, Q., Yang, H., and Guo, L. G.: Studies on temporal and spatial variations of phytoplankton in Lake Chaohu, J. Integr. Plant Biol., 49, 409–418, doi:10.1111/j.1744-7909.2007.00390.x, 2007.

Ejigu, M. T.: Overview of water quality modeling, Cogent Eng., 8, 1891711, doi:10.1080/23311916.2021.1891711, 2021.

Elzain, H. E., Chung, S. Y., Senapathi, V., Sekar, S., Lee, S. Y., Priyadarsi, R. D., Hassan, A., and Sabarathinam, C.: Comparative study of machine learning models for evaluating groundwater vulnerability to nitrate contamination, Ecotoxicol. Environ. Saf., 229, 113061, doi:10.1016/j.ecoenv.2021.113061, 2022.

Engel, B., Storm, D., White, M., Arnold, J., and Arabi, M.: A hydrologic/water quality model application protocol, J. Am. Water Resour. Assoc., 43, 1223–1236, doi:10.1111/j.1752-1688.2007.00105.x, 2007.

Falconer, I. R.: Is there a human health hazard from microcystins in the drinking water supply?, Acta hydrochim. hydrobiol., 33, 64–71, doi:10.1002/aheh.200300551, 2005.

Falconer, I. R. and Humpage, A. R.: Health risk assessment of cyanobacterial (blue-green algal) toxins in drinking water, Int. J. Environ. Res. Public Health, 2, 43–50, doi:10.3390/ijerph2005010043, 2005.

Ferber, L. R., Levine, S. N., Lini, A., and Livingston, G. P.: Do cyanobacteria dominate in eutrophic lakes because they fix atmospheric nitrogen?, Freshw. Biol., 49, 690–708, doi:10.1111/j.1365-2427.2004.01218.x, 2004.

Fernandez, A., Garcia, S., Herrera, F., and Chawla, N. V.: SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary, J. Artif. Intell. Res., 61, 863–905, doi:10.1613/jair.1.11192, 2018.

Forio, M. A. E. and Goethals, P. L. M.: An integrated approach of multi-community monitoring and assessment of aquatic ecosystems to support sustainable development, Sustainability 12, 5603, doi:10.3390/su12145603, 2020.

Gao, L. L. and Li, D. L.: A review of hydrological/water-quality models, Front. Agric. Sci. Eng., 1, 267–276, doi:10.15302/J-Fase-2014041, 2014.

Ghafoor, J., Forio, M. A. E., and Goethals, P. L. M.: Spatially explicit river basin models for cost-benefit analyses to optimize land use, Sustainability 14, 8953, doi:10.3390/su14148953, 2022.

Gnana, D. A. A., Balamurugan, S. A. A., and Leavline, E. J.: Literature review on feature selection methods for high-dimensional data, Int. J. Comput. Appl., 136, 9–17, 2016.

Gobler, C. J.: Climate change and harmful algal blooms: Insights and perspective, Harmful Algae, 91, 101731, doi:10.1016/j.hal.2019.101731, 2020.

Goethals, P. L. M. and Forio, M. A. E.: Advances in ecological water system modeling: Integration and leanification as a basis for application in environmental management, Water, 10, 1216, doi:10.3390/w10091216, 2018.

Gradilla-Hernandez, M. S., de Anda, J., Garcia-Gonzalez, A., Meza-Rodriguez, D., Montes, C. Y., and Perfecto-Avalos, Y.: Multivariate water quality analysis of Lake Cajititlan, Mexico, Environ. Monit. Assess., 192, 5, doi:10.1007/s10661-019-7972-4, 2020.

Gunawardena, M. P. and Najim, M. M. M.: Adapting Sri Lanka to climate change: Approaches to water modelling in the upper Mahaweli catchment area, in: Climate change research at universities: Addressing the mitigation and adaptation challenges, edited by: Leal Filho, W., Springer International Publishing, Cham, Switzerland, 95–115, 10.1007/978-3-319-58214-6_6, 2017.

Harrow-Lyle, T. and Kirkwood, A. E.: The invasive macrophyte Nitellopsis obtusa may facilitate the invasive mussel Dreissena polymorpha and Microcystis blooms in a large, shallow lake, Can. J. Fish. Aquat., 77, 1201–1208, doi:10.1139/cjfas-2019-0337, 2020.

Hem, J. D.: Study and interpretation of the chemical characteristics of natural water, U.S. Department of the Interior, U.S. Geological Survey, Alexandria, VA, USA, 1985.

Ho, L. and Goethals, P.: Research hotspots and current challenges of lakes and reservoirs: A bibliometric analysis, Scientometrics, 124, 603–631, doi:10.1007/s11192-020-03453-1, 2020.

Ho, L. and Goethals, P.: Machine learning applications in river research: Trends, opportunities and challenges, Methods Ecol. Evol., 13, 2603–2621, doi:10.1111/2041-210x.13992, 2022.

Jahan, R., Khan, S., Haque, M. M., and Choi, J. K.: Study of harmful algal blooms in a eutrophic pond, Bangladesh, Environ. Monit. Assess., 170, 7–21, doi:10.1007/s10661-009-1210-4, 2010.

Jain, C. K.: Application of chemical mass balance to upstream downstream river monitoring data, J. Hydrol., 182, 105–115, doi:10.1016/0022-1694(95)02932-X, 1996.

James, R. T.: Recalibration of the Lake Okeechobee Water Quality Model (LOWQM) to extreme hydro-meteorological events, Ecol. Model., 325, 71–83, doi:10.1016/j.ecolmodel.2016.01.007, 2016.

Jankowiak, J., Hattenrath-Lehmann, T., Kramer, B. J., Ladds, M., and Gobler, C. J.: Deciphering the effects of nitrogen, phosphorus, and temperature on cyanobacterial bloom intensification, diversity, and toxicity in western Lake Erie, Limnol. Oceanogr., 64, 1347–1370, doi:10.1002/lno.11120, 2019.

Jeong, J.-W., Kim, Y.-O., and Seo, S. B.: Evaluating joint operation rules for connecting tunnels between two multipurpose dams, Hydrol. Res., 51, 392–405, doi:10.2166/nh.2020.053, 2020.

Jha, R., Ojha, C. S. P., and Bhatia, K. K. S.: Non-point source pollution estimation using a modified approach, Hydrol. Process., 21, 1098–1105, doi:10.1002/hyp.6291, 2007.

Jiang, S. J., Zheng, Y., and Solomatine, D.: Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning, Geophys. Res. Lett., 47, e2020GL088229, doi:10.1029/2020GL088229, 2020.

Jo, C. D., Lee, C. G., and Kwon, H. G.: Effects of multifunctional weir construction on key water quality indicators: A case study in Nakdong River, Korea, Int. J. Environ. Sci. Technol., 19, 11843–11856, doi:10.1007/s13762-022-03973-8, 2022.

Kim, D.-H., Lee, S.-K., Chun, B.-Y., Lee, D.-H., Hong, S.-C., and Jang, B.-K.: Illness associated with contamination of drinking water supplies with phenol, J. Korean Med. Sci., 9, 218–223, doi:10.3346/jkms.1994.9.3.218, 1994.

Kim, D. and Shin, C.: Algal boom characteristics of Yeongsan River based on weir and estuary dam operating conditions using EFDC-NIER model, Water, 13, 2295, doi:10.3390/w13162295, 2021.

Kim, J., Jonoski, A., and Solomatine, D. P.: A classification-based machine learning approach to the prediction of cyanobacterial blooms in Chilgok Weir, South Korea, Water, 14, 542, doi:10.3390/w14040542, 2022a.

Kim, J., Lee, T., and Seo, D.: Algal bloom prediction of the lower Han River, Korea using the EFDC hydrodynamic and water quality model, Ecol. Model., 366, 27–36, doi:10.1016/j.ecolmodel.2017.10.015, 2017.

Kim, J., Seo, D., Jang, M., and Kim, J.: Augmentation of limited input data using an artificial neural network method to improve the accuracy of water quality modeling in a large lake, J. Hydrol., 602, 126817, doi:10.1016/j.jhydrol.2021.126817, 2021.

Kim, J., Kim, H. Y., Choi, H. G., Jeong, S., and Lee, Y.: The innovative operation of Imha Reservoir, E3S Web of Conferences, 01029, doi:10.1051/e3sconf/202234601029, 2022.

Kim, J., Kwak, J., Ahn, J. M., Kim, H., Jeon, J., and Kim, K.: Oscillation flow dam operation method for algal bloom mitigation, Water, 14, 1315, doi:10.3390/w14081315, 2022c.

Kim, S., Kim, S., Mehrotra, R., and Sharma, A.: Predicting cyanobacteria occurrence using climatological and environmental controls, Water Res., 175, 115639, doi:10.1016/j.watres.2020.115639, 2020.

Kim, S., Chung, S., Park, H., Cho, Y., and Lee, H.: Analysis of environmental factors associated with cyanobacterial dominance after river weir installation, Water, 11, 1163, doi:10.3390/w11061163, 2019a.

Kim, S. K. and Choi, S.-U.: Assessment of the impact of selective withdrawal on downstream fish habitats using a coupled hydrodynamic and habitat modeling, J. Hydrol., 593, 125665, doi:10.1016/j.jhydrol.2020.125665, 2021.

Kim, W., Lee, J., Kim, J., and Kim, S.: Assessment of water supply stability for drought-vulnerable boryeong multipurpose dam in South Korea using future dry climate change scenarios, Water, 11, 2403, doi:10.3390/w11112403, 2019b.

Kim, Y. and Oh, S.: Machine-learning insights into nitrate-reducing communities in a full-scale municipal wastewater treatment plant, J. Environ. Manag., 300, 113795, doi:10.1016/j.jenvman.2021.113795, 2021.

Korean Government: The 1st master plan for national water management (2021–2030), Korean Government, Sejong-si, Republic of Korea, 2021.

Kosten, S., Huszar, V. L., Bécares, E., Costa, L. S., van Donk, E., Hansson, L. A., Jeppesen, E., Kruk, C., Lacerot, G., and Mazzeo, N.: Warmer climates boost cyanobacterial dominance in shallow lakes, Glob. Chang. Bio., 18, 118–126, doi:10.1111/j.1365-2486.2011.02488.x, 2012.

Lee, C.-M., Hamm, S.-Y., Cheong, J.-Y., Kim, K., Yoon, H., Kim, M., and Kim, J.: Contribution of nitrate-nitrogen concentration in groundwater to stream water in an agricultural head watershed, Environ. Res., 184, 109313, doi:10.1016/j.envres.2020.109313, 2020.

Lee, D. Y. and Baek, K. O.: Study of the mitigation of algae in Lake Uiam according to the operation of the Chuncheon Dam and the Soyang Dam, KSCE Journal of Civil and Environmental Engineering Research, 42, 171–179, doi:10.12652/Ksce.2022.42.2.0171, 2022.

Lee, H.-J., Park, H.-K., and Cheon, S.-U.: Effects of weir construction on phytoplankton assemblages and water quality in a large river system, Int. J. Environ. Res. Public Health, 15, 2348, doi:10.3390/ijerph15112348, 2018.

Lee, J., Bae, S., Lee, D.-R., and Seo, D.: Transportation modeling of conservative pollutant in a river with weirs-the Nakdong River case, J. Korean Soc. Environ. Eng., 36, 821–827, doi:10.4491/KSEE.2014.36.12.821, 2014.

Lee, N.: Water policy and institutions in the Republic of Korea, Asian Development Bank Institute, Tokyo, Japan, 2019.

Lee, S., Kim, J., Noh, J., and Ko, I. H.: Assessment of selective withdrawal facility in the Imha Reservoir using CE-QUAL-W2 model, J. Korean Soc. Water Environ., 23, 228–235, 2007.

Lee, S., Kim, J., Choi, B., Kim, G., and Lee, J.: Harmful algal blooms and liver diseases: Focusing on the areas near the four major rivers in South Korea, J. Environ. Sci. Health C, 37, 356–370, doi:10.1080/10590501.2019.1674600, 2019.

Leonard, B. P.: A stable and accurate convective modelling procedure based on quadratic upstream interpolation, Comput. Methods Appl. Mech. Eng., 19, 59–98, doi:10.1016/0045-7825(79)90034-3, 1979.

Leonard, B. P.: The ULTIMATE conservative difference scheme applied to unsteady one-dimensional advection, Comput. Methods Appl. Mech. Eng., 88, 17–74, doi:10.1016/0045-7825(91)90232-U, 1991.

May, N. W., Olson, N. E., Panas, M., Axson, J. L., Tirella, P. S., Kirpes, R. M., Craig, R. L., Gunsch, M. J., China, S., Laskin, A., Ault, A. P., and Pratt, K. A.: Aerosol emissions from great lakes harmful algal blooms, Environ. Sci. Technol., 52, 397–405, doi:10.1021/acs.est.7b03609, 2018.

McIntyre, N. R. and Wheater, H. S.: A tool for risk-based management of surface water quality, Environ. Model. Softw., 19, 1131–1140, doi:10.1016/j.envsoft.2003.12.003, 2004.

Mellios, N., Moe, S. J., and Laspidou, C.: Machine learning approaches for predicting health risk of cyanobacterial blooms in Northern European lakes, Water, 12, 1191, doi:10.3390/w12041191, 2020.

Meybeck, M.: Carbon, nitrogen, and phosphorus transport by world rivers, Am. J. Sci., 282, 401–450, 1982.

Mihale, M. J.: Nitrogen and phosphorus dynamics in the waters of the Great Ruaha River, Tanzania, J. Water Resour. Ocean Sci., 4, 59–71, doi:10.11648/j.wros.20150405.11, 2015.

Modabberi, A., Noori, R., Madani, K., Ehsani, A. H., Mehr, A. D., Hooshyaripor, F., and Klove, B.: Caspian Sea is eutrophying: The alarming message of satellite data, Environ. Res. Lett., 15, 124047, doi:10.1088/1748-9326/abc6d3, 2020.

Moreido, V., Gartsman, B., Solomatine, D. P., and Suchilina, Z.: How well can machine learning models perform without hydrologists? Application of rational feature selection to improve hydrological forecasting, Water, 13, 1696, doi:10.3390/w13121696, 2021.

Moriasi, D. N., Gitau, M. W., Pai, N., and Daggupati, P.: Hydrologic and water quality models: Performance measures and evaluation criteria, T. Asabe 58, 1763–1785, doi:10.13031/trans.58.10715, 2015.

Mozafari, Z., Noori, R., Siadatmousavi, S. M., Afzalimehr, H., and Azizpour, J.: Satellite-based monitoring of eutrophication in the earth's largest transboundary lake, Geohealth, 7, e2022GH000770, doi:10.1029/2022GH000770, 2023.

Mulyani, E., Hidayah, I., and Fauziati, S.: Dropout prediction optimization through SMOTE and ensemble learning, 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 516–521, doi:10.1109/ISRITI48646.2019.9034673, 2019.

Musacchio, A., Re, V., Mas-Pla, J., and Sacchi, E.: EU Nitrates Directive, from theory to practice: Environmental effectiveness and influence of regional governance on its performance, Ambio, 49, 504–516, doi:10.1007/s13280-019-01197-8, 2020.

Nagawa, K., Suzuki, M., Yamamoto, Y., Inoue, K., Kozawa, E., Mimura, T., Nakamura, K., Nagata, M., and Niitsu, M.: Texture analysis of muscle mri: Machine learning-based classifications in idiopathic inflammatory myopathies, Sci. Rep., 11, 9821, doi:10.1038/s41598-021-89311-3, 2021.

Nakagawa, K., Amano, H., Asakura, H., and Berndtsson, R.: Spatial trends of nitrate pollution and groundwater chemistry in Shimabara, Nagasaki, Japan, Environ. Earth Sci., 75, 234, doi:10.1007/s12665-015-4971-9, 2016.

Noori, R., Ansari, E., Jeong, Y. W., Aradpour, S., Maghrebi, M., Hosseinzadeh, M., and Bateni, S. M.: Hyper-nutrient enrichment status in the Sabalan Lake, Iran, Water, 13, 2874, doi:10.3390/w13202874, 2021.

Omer, N. H.: Water quality parameters, in: Water quality-science, assessments and policy, IntechOpen, London, UK, 3–20, 2020.

Paerl, H. W.: Controlling cyanobacterial harmful blooms in freshwater ecosystems, Microb. Biotechnol., 10, 1106–1110, doi:10.1111/1751-7915.12725, 2017.

Paerl, H. W. and Huisman, J.: Climate change: A catalyst for global expansion of harmful cyanobacterial blooms, Environ. Microbiol. Rep., 1, 27–37, doi:10.1111/j.1758-2229. 2008.00004.x, 2009.

Paerl, H. W. and Otten, T. G.: Harmful cyanobacterial blooms: Causes, consequences, and controls, Microb. Ecol., 65, 995–1010, doi:10.1007/s00248-012-0159-y, 2013.

Paerl, H. W. and Scott, J. T.: Throwing fuel on the fire: Synergistic effects of excessive nitrogen inputs and global warming on harmful algal blooms, Environ. Sci. Technol., 44, 7756–7758, doi:10.1021/es102665e, 2010.

Paerl, H. W., Fulton, R. S., Moisander, P. H., and Dyble, J.: Harmful freshwater algal blooms, with an emphasis on cyanobacteria, TheScientificWorldJournal, 1, 76–113, doi:10.1100/tsw.2001.16, 2001.

Park, H.-K., Lee, H.-J., Heo, J., Yun, J.-H., Kim, Y.-J., Kim, H.-M., Hong, D.-G., and Lee, I.-J.: Deciphering the key factors determining spatio-temporal heterogeneity of cyanobacterial bloom dynamics in the Nakdong River with consecutive large weirs, Sci. Total Environ., 755, 143079, doi:10.1016/j.scitotenv.2020.143079, 2021a.

Park, H.-S. and Chung, S.-W.: Water transportation and stratification modification in the Andong-Imha linked reservoirs system, J. Korean Soc. Water Environ., 30, 31–43, doi:10.15681/KSWE.2014.30.1.031, 2014.

Park, H., Chung, S., Cho, E., and Lim, K.: Impact of climate change on the persistent turbidity issue of a large dam reservoir in the temperate monsoon region, Climatic Change, 151, 365–378, doi:10.1007/s10584-018-2322-z, 2018a.

Park, J., Sohn, S., and Kim, Y.: Distribution characteristics of total nitrogen components in streams by watershed characteristics, J. Korean Soc. Water Environ., 30, 503–511, doi:10.15681/KSWE.2014.30.5.503, 2014.

Park, J., Wang, D., and Lee, W. H.: Evaluation of weir construction on water quality related to algal blooms in the Nakdong River, Environ. Earth Sci., 77, 408, doi:10. 1007/s12665-018-7590-4, 2018b.

Park, J. C., Um, M.-J., Song, Y.-I., Hwang, H.-D., Kim, M. M., and Park, D.: Modeling of turbidity variation in two reservoirs connected by a water transfer tunnel in South Korea, Sustainability 9, 993, doi:10.3390/su9060993, 2017a.

Park, Y., Pyo, J., Kwon, Y. S., Cha, Y., Lee, H., Kang, T., and Cho, K. H.: Evaluating physico-chemical influences on cyanobacterial blooms using hyperspectral images in inland water, Korea, Water Res., 126, 319–328, doi:10.1016/j.watres.2017.09.026, 2017b.

Park, Y., Lee, H. K., Shin, J.-K., Chon, K., Kim, S., Cho, K. H., Kim, J. H., and Baek, S.-S.: A machine learning approach for early warning of cyanobacterial bloom outbreaks in a freshwater reservoir, J. Environ. Manag., 288, 112415, doi:10.1016/j.jenvman. 2021.112415, 2021b.

Patil, V. B., Pinto, S. M., Govindaraju, T., Hebbalu, V. S., Bhat, V., and Kannanur, L. N.: Multivariate statistics and water quality index (WQI) approach for geochemical assessment of groundwater quality—a case study of Kanavi Halla Sub-Basin, Belagavi, India, Environ. Geochem. Health, 42, 2667–2684, doi:10.1007/s10653-019-00500-6, 2020.

Peng, Y. P., Khaled, U., Al-Rashed, A. A. A. A., Meer, R., Goodarzi, M., and Sarafraz, M. M.: Potential application of Response Surface Methodology (RSM) for the prediction and optimization of thermal conductivity of aqueous CuO (II) nanofluid: A statistical approach and experimental validation, Physica A, 554, 124353, doi:10.1016/ j.physa.2020.124353, 2020.

Plaas, H. E. and Paerl, H. W.: Toxic cyanobacteria: A growing threat to water and air quality, Environ. Sci. Technol., 55, 44–64, doi:10.1021/acs.est.0c06653, 2021.

Pyo, J., Cho, K. H., Kim, K., Baek, S.-S., Nam, G., and Park, S.: Cyanobacteria cell prediction using interpretable deep learning model with observed, numerical, and sensing data assemblage, Water Res., 203, 117483, doi:10.1016/j.watres.2021.117483, 2021.

Raschka, S. and Mirjalili, V.: Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2, third edition, Int. J. Knowl.-Based Organ., 10, 3175783, 2017.

Razavi, S., Tolson, B. A., and Burn, D. H.: Review of surrogate modeling in water resources, Water Resour. Res., 48, W07401, doi:10.1029/2011wr011527, 2012.

Reynolds, C. S.: The ecology of phytoplankton, Cambridge University Press, New York, NY, USA, 2006.

Romo, S., Soria, J., Fernandez, F., Ouahid, Y., and Baron-Sola, A.: Water residence time and the dynamics of toxic cyanobacteria, Freshw. Biol., 58, 513–522, doi:10.1111/j.1365-2427.2012.02734.x, 2013.

Rousso, B. Z., Bertone, E., Stewart, R., and Hamilton, D. P.: A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes, Water Res., 182, 115959, doi:10.1016/j.watres.2020.115959, 2020.

Rus, D. L., Patton, C. J., Mueller, D. K., and Crawford, C. G.: Assessing total nitrogen in surface-water samples: Precision and bias of analytical and computational methods, U.S. Department of the Interior, U.S. Geological Survey, Reston, VA, USA, 2012.

Saadatpour, M., Afshar, A., and Solis, S. S.: Surrogate-based multiperiod, multiobjective reservoir operation optimization for quality and quantity management, J. Water Resour. Plann. Manag., 146, doi:10.1061/(Asce)Wr.1943-5452.0001252, 2020.

Saadatpour, M., Javaheri, S., Afshar, A., and Solis, S. S.: Optimization of selective withdrawal systems in hydropower reservoir considering water quality and quantity aspects, Expert Syst. Appl., 184, 115474, doi:10.1016/j.eswa.2021.115474, 2021.

Shin, J., Yoon, S., and Cha, Y.: Prediction of cyanobacteria blooms in the lower Han River (South Korea) using ensemble learning algorithms, Desalin. Water Treat., 84, 31–39, doi:10.5004/dwt.2017.20986, 2017.

Smith, D. R., Wilhelms, S. C., Holland, J. P., Dortch, M. S., and Davis, J. E.: Improved description of selective withdrawal through point sinks, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, USA, 1987.

Smith, G. J. and Daniels, V.: Algal blooms of the 18th and 19th centuries, Toxicon, 142, 42–44, doi:10.1016/j.toxicon.2017.12.049, 2018.

Song, H. and Lynch, M. J.: Restoration of nature or special interests? A political economy analysis of the Four Major Rivers Restoration Project in South Korea, Crit. Criminol., 26, 251–270, doi:10.1007/s10612-018-9384-0, 2018.

Srivastava, A., Ahn, C.-Y., Asthana, R. K., Lee, H.-G., and Oh, H.-M.: Status, alert system, and prediction of cyanobacterial bloom in South Korea, Biomed Res. Int., 2015, 584696, doi:10.1155/2015/584696, 2015.

Srivastava, P., McVair, J. N., and Johnson, T. E.: Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed, J. Am. Water Resour. Assoc., 42, 545–563, doi:10.1111/j.1752-1688.2006.tb04475.x, 2006.

Talib, A., Recknagel, F., Cao, H., van der Molen, D., and Abu Hasan, Y.: Use of hybrid EA models for the prediction of chlorophyll-a and phytoplankton functional groups abundance in two shallow lakes, Malays. J. Math. Sci., 2, 11–28, 2008.

Taralgatti, P. D., Pawar, R. S., Pawar, G. S., Nomaan, M. H., and Limkar, C. R.: Water quality modeling of Bhima River by using HEC-RAS software, Int. J. Eng. Adv. Technol., 9, 2886–2889, doi:10.35940/ijeat.B3481.029320, 2020.

Teran-Velasquez, G., Helm, B., and Krebs, P.: Longitudinal river monitoring and modelling substantiate the impact of weirs on nitrogen dynamics, Water, 14, 189, doi:10.3390/w14020189, 2022.

Tong, Y. D., Xu, X. W., Qi, M., Sun, J. J., Zhang, Y. Y., Zhang, W., Wang, M. Z., Wang, X. J., and Zhang, Y.: Lake warming intensifies the seasonal pattern of internal nutrient cycling in the eutrophic lake and potential impacts on algal blooms, Water Res., 188, 116570, doi:10.1016/j.watres.2020.116570, 2021.

Tousi, E. G., Duan, J. N. G., Gundy, P. M., Bright, K. R., and Gerba, C. P.: Evaluation of E. coli in sediment for assessing irrigation water quality using machine learning, Sci. Total Environ., 799, 149286, doi:10.1016/j.scitotenv.2021.149286, 2021.

Uma, K. V. and Balamurugan, S. A. A.: C5.0 decision tree model using Tsallis entropy and association function for general and medical dataset, Intell. Autom. Soft Comput., 26, 61–70, doi:10.31209/2019.100000153, 2020.

UNESCO and UN-Water: United Nations world water development report 2020: Water and climate change, UNESCO, Paris, France, 236 pp., 2020.

Ustaoglu, F., Tas, B., Tepe, Y., and Topaldemir, H.: Comprehensive assessment of water quality and associated health risk by using physicochemical quality indices and multivariate analysis in Terme River, Turkey, Environ. Sci. Pollut. Res., 28, 62736–62754, doi:10.1007/s11356-021-15135-3, 2021.

Vien, B. S., Wong, L., Kuen, T., Rose, L. F., and Chiu, W. K.: A machine learning approach for anaerobic reactor performance prediction using long short-term memory recurrent neural network, Struct. Health Monit. 8apwshm, 18, 61, doi:10.21741/9781644901311-8, 2021.

Ward, M. H., deKok, T. M., Levallois, P., Brender, J., Gulis, G., Nolan, B. T., and VanDerslice, J.: Workgroup report: Drinking-water nitrate and health—recent findings and research needs, Environ. Health Perspect., 113, 1607–1614, doi:10.1289/ehp.8043, 2005.

Ward, M. H., Jones, R. R., Brender, J. D., de Kok, T. M., Weyer, P. J., Nolan, B. T., Villanueva, C. M., and van Breda, S. G.: Drinking water nitrate and human health: An updated review, Int. J. Environ. Res. Public Health, 15, 1557, doi:10.3390/ijerph 15071557, 2018.

Weyhenmeyer, G. A., Jeppesen, E., Adrian, R., Arvola, L., Blenckner, T., Jankowski, T., Jennings, E., Nõges, P., Nõges, T., and Straile, D.: Nitrate-depleted conditions on the increase in shallow northern European lakes, Limnol. Oceanogr., 52, 1346–1353, doi:10.4319/lo.2007.52.4.1346, 2007.

WHO: Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda, World Health Organization, Geneva, Switzerland, 2022.

Wu, S. S., Hu, X. L., Zheng, W. B., He, C. C., Zhang, G. C., Zhang, H., and Wang, X.: Effects of reservoir water level fluctuations and rainfall on a landslide by two-way ANOVA and K-means clustering, B Eng. Geol. Environ., 80, 5405–5421, doi:10.1007/ s10064-021-02273-8, 2021.

Wurtsbaugh, W. A., Paerl, H. W., and Dodds, W. K.: Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum, Wiley Interdiscip. Rev. Water, 6, e1373, doi:10.1002/wat2.1373, 2019.

Xu, H., Paerl, H. W., Qin, B., Zhu, G., Hall, N. S., and Wu, Y.: Determining critical nutrient thresholds needed to control harmful cyanobacterial blooms in eutrophic Lake Taihu, China, Environ. Sci. Technol., 49, 1051–1059, doi:10.1021/es503744q, 2015.

Xu, X. D., Lin, H., Liu, Z. H., Ye, Z. L., Li, X. Y., and Long, J. P.: A combined strategy of improved variable selection and ensemble algorithm to map the growing stem volume of planted coniferous forest, Remote Sens., 13, 4631, doi:10.3390/rs13224631, 2021.

Yajima, H. and Derot, J.: Application of the Random Forest model for chlorophyll-a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases, J. Hydroinform., 20, 206–220, doi:10.2166/hydro.2017.010, 2018.

Yang, L. and Yu, X.: A review of development and application on river comprehensive water quality model QUAL2k, in: Iop conference series: Earth and environmental science, IOP Publishing, Bristol, UK, 10.1088/1755-1315/189/2/022034, 2018.

Yang, Y., Chen, H., and Lu, J.: Inactivation of algae by visible-light-driven modified photocatalysts: A review, Sci. Total Environ., 858, 159640, doi:10.1016/j.scitotenv. 2022.159640, 2023.

Yang, Y., Huang, T. T., Shi, Y. Z., Wendroth, O., and Liu, B. Y.: Comparing the performance of an autoregressive state-space approach to the linear regression and artificial neural network for streamflow estimation, J. Environ. Inf., 37, 36–48, doi:10. 3808/jei.202000440, 2021.

Yi, H.-S., Park, S., An, K.-G., and Kwak, K.-C.: Algal bloom prediction using extreme learning machine models at artificial weirs in the Nakdong River, Korea, Int. J. Environ. Res. Public Health, 15, 2078, doi:10.3390/ijerph15102078, 2018.

Yoo, C. and Cho, E.: Effect of multicollinearity on the bivariate frequency analysis of annual maximum rainfall events, Water, 11, 905, doi:10.3390/w11050905, 2019.

Yoo, C., Jun, C., Zhu, J., and Na, W.: Evaluation of dam water-supply capacity in Korea using the water-shortage index, Water, 13, 956, doi:10.3390/w13070956, 2021.

Yoo, J.-H.: Maximization of hydropower generation through the application of a linear programming model, J. Hydrol., 376, 182–187, doi:10.1016/j.jhydrol.2009.07.026, 2009.

Yosefipoor, P., Saadatpour, M., Solis, S. S., and Afshar, A.: An adaptive surrogate-based, multi-pollutant, and multi-objective optimization for river-reservoir system management, Ecol. Eng., 175, 106487, doi:10.1016/j.ecoleng.2021.106487, 2022.

Yu, J. S., Shin, J. Y., Kwon, M., and Kim, T.-W.: Bivariate drought frequency analysis to evaluate water supply capacity of multi-purpose dams, KSCE Journal of Civil and Environmental Engineering Research, 37, 231–238, doi:10.12652/Ksce.2017.37.1.0231, 2017.

Zakova, Z., Berankova, D., Kockova, E., and Kriz, P.: Influence of diffuse pollution on the eutrophication and water-quality of reservoirs in the Morava river basin, Water Sci. Technol., 28, 79–90, doi:10.2166/wst.1993.0406, 1993.

Zeng, Q. H., Liu, Y., Zhao, H. T., Sun, M. D., and Li, X. Y.: Comparison of models for predicting the changes in phytoplankton community composition in the receiving water system of an inter basin water transfer project, Environ. Pollut., 223, 676–684, doi:10.1016/j.envpol.2017.02.001, 2017.

Zhang, Y. P., Yao, X. Y., Wu, Q., Huang, Y. B., Zhou, Z. X., Yang, J., and Liu, X. W.: Turbidity prediction of lake-type raw water using random forest model based on meteorological data: A case study of Tai lake, China, J. Environ. Manag., 290, 112657, doi:10.1016/j.jenvman.2021.112657, 2021.

Zhang, Z. and Johnson, B. E.: Aquatic nutrient simulation modules (NSMs) developed for hydrologic and hydraulic models, U.S. Army Engineer Research and Development Center, Environmental Laboratory, Vicksburg, MS, USA, 2016.

Zhao, C. S., Shao, N. F., Yang, S. T., Ren, H., Ge, Y. R., Feng, P., Dong, B. E., and Zhao, Y.: Predicting cyanobacteria bloom occurrence in lakes and reservoirs before blooms occur, Sci. Total Environ., 670, 837–848, doi:10.1016/j.scitotenv.2019.03.161, 2019.

Zhao, W. X., Li, Y. Y., Jiao, Y. J., Zhou, B., Vogt, R. D., Liu, H. L., Ji, M., Ma, Z., Li, A. D., Zhou, B. H., and Xu, Y. P.: Spatial and temporal variations in environmental

variables in relation to phytoplankton community structure in a eutrophic river-type reservoir, Water, 9, 754, doi:10.3390/w9100754, 2017.

Zhou, Q., Zhang, X. D., Yu, L. F., Ren, L. L., and Luo, Y. Q.: Combining WV-2 images and tree physiological factors to detect damage stages of populus gansuensis by Asian longhorned beetle (Anoplophora glabripennis) at the tree level, Ecosyst, 8, 35, doi:10.1186/s40663-021-00314-y, 2021.

LIST OF ACRONYMS

ACC	Accuracy
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
AT	Average air Temperature
BOD	Biochemical Oxygen Demand
C	Carbon
CAEDYM	Computational Aquatic Ecosystem Dynamics Model
CBOD	Carbonaceous Biochemical Oxygen Demand
C_D	Concentration of a specific water quality parameter at a Downstream location
Chl-a	Chlorophyll a
CI	Quality of water released from the Imha Reservoir
CJ	NO_3-N concentration at the Junction where the water from the two reservoirs (Andong and Imha) meets
COD	Chemical Oxygen Demand
CR	Reference Concentration
C_U	Concentration of a specific water quality parameter at an Upstream location
Cyano	Cyanobacterial cell density
CyanoHABs	Harmful Cyanobacterial Blooms
DaySolarRad	Total amount of Solar Radiation
DO	Dissolved Oxygen

DON	Dissolved Organic Nitrogen
DT	Decision Tree
DYRESM	DYnamics REservoir Simulation Model
EC	Electrical Conductivity
EFDC	Environmental Fluid Dynamics Code
ELCOM	Estuary and Lake COmputer Model
ELM	Extreme Learning Machine
EPA	United States Environmental Protection Agency
F1	F1-score
Fe	Iron
FecalColiform	Fecal Coliforms
FN	False Negative
FP	False Positive
HABs	Harmful Algal Blooms
HEC-RAS	Hydrologic Engineering Center's River Analysis System
HT	Highest air Temperature
Inflow	Weir Inflow
KMA	Korea Meteorological Administration
k-NN	k-Nearest Neighbor
K-water	Korea water resources corporation
L_i	Individual Loadings of Inflow points between upstream and downstream locations

logCyano	Logarithm of Cyanobacterial cell density
LR	Logistic Regression
LT	Lowest air Temperature
mamsl	Meters Above Mean Sea Level
MaxSolarRad	Maximum amount of Solar Radiation for one hour
MCL	Maximum Contaminant Level
Min.	Minimum
Max.	Maximum
N	Nitrogen
N/P	TN/TP ratio
NH_3-N	Ammonia Nitrogen
NH_4-N	Ammonium Nitrogen
NIER	National Institute of Environmental Research
NO_2-N	Nitrite Nitrogen
NO_3-N	Nitrate Nitrogen
NSE	Nash Sutcliffe Efficiency
NSGA-II	Non-dominated Sorting Genetic Algorithm II
OF1	Objective Function 1
OF2	Objective Function 2
Outflow	Weir Outflow
P	Phosphorus
PBIAS	Percent Bias

pH	Hydrogen ion concentration
PO_4	Phosphate
PO_4-P	Phosphate Phosphorus
PRE	Precision
QA	Amount of water supply downstream of the Andong Reservoir
Q_D	Flow rate at a Downstream location
QD	Design Discharge
QI	Amount of water supply downstream of the Imha Reservoir
QJ	Flow rate at the Junction where the water from the two reservoirs (Andong and Imha) meets
QO	Observational data for the water supply downstream of the two reservoirs (Andong and Imha)
Q_U	Flow rate at an Upstream location
QUICKEST–ULTIMATE	Quadratic Upstream Interpolation for Convective Kinematics with Estimated Streaming Terms–Universal Limiter for Transient Interpolation Modelling of the Advective Transport Equations
QWD	Sum of the Water Demand allocated to the two reservoirs (Andong and Imha)
R^2	Coefficient of Determination
Rainfall	Rainfall in weir catchment area
REC	Recall
RF	Random Forest
RMSE	Root Mean Square Error
SM	Surrogate Model

SMOTE	Synthetic Minority Oversampling Technique
SS	Suspended Solids
StorageVolume	Storage Volume of weir
SVM	Support Vector Machine
SWAT	Soil and Water Assessment Tool
SWF	Selective Withdrawal Facility
TDN	Total Dissolved Nitrogen
TDP	Total Dissolved Phosphorus
TN	Total Nitrogen
TN	True Negative
TOC	Total Organic Carbon
TotalColiform	Total Coliforms
TP	Total Phosphorus
TP	True Positive
UN-Water	United Nations Water
WASP	Water Quality Analysis Simulation Program
WeirLevel	Water Level of Weir
WHO	World Health Organization
WQM	Water Quality Management
WT	Water temperature

LIST OF TABLES

Table 2.1. Details about Andong and Imha reservoirs 12

Table 2.2. Details about Sangju, Nakdan, Gumi, and Chilgok weirs 13

Table 3.1. Criteria for algae alert in South Korea 19

Table 3.2. List of features 20

Table 3.3. Nine-year mean, minimum, median, and maximum values for each feature in the raw dataset (378 instances) 21

Table 3.4. Classification framework for each group 23

Table 3.5. Input features (including the mean, minimum, median, and maximum values for each feature) after preprocessing the dataset using logarithmic transformation and standardization, together with the classified target variable 24

Table 3.6. Parameters to be optimized in this study for four algorithms 27

Table 3.7. F and p values of 30 features 31

Table 3.8. Modelling cases with a combination of input features 33

Table 3.9. Optimized parameters of four classification algorithms 35

Table 4.1. List of the data collected for model development 47

Table 4.2. Mean, minimum, and maximum values of the observational data (flow rate and NO_3-N) in the cross sections for model calibration (2019) and validation (2020) 48

Table 4.3. Components of water quality required in HEC-RAS 49

Table 4.4. Main parameters related to NO_3-N dynamics provided in HEC-RAS 52

Table 4.5. Scenarios constructed for an understanding of NO_3-N dynamics downstream 52

Table 4.6. Manning's roughness coefficients for the hydraulic unsteady model 54

Table 4.7. Hydraulic model performance for unsteady flow 55

Table 4.8. Mean values of both the observational data and the simulation results for the water quality parameters between 2019 (calibration) and 2020 (validation) 57

Table 4.9. Model performance for NO_3-N 59

Table 4.10. Example, taken from the red ellipses in Figure 4.7b, of the change in NO_3-N concentration produced by an increase in flow rate 66

Table 5.1. Experimental cases based on the constraints for the optimization process 84

Table 5.2. Design Discharge 85

Table 5.3. Results of the simulation test (the number of days with CyanoHABs after one week at Chilgok Weir using the NO_3-N concentrations of Figure 5.4)..... 88

Table 5.4. Performance of the surrogate model for three optimal solutions 98

Table 5.5. Simulation results of the number of days with CyanoHABs after one week at Chilgok Weir for three optimal solutions..... 100

LIST OF FIGURES

Figure 1.1. Framework for optimal reservoir operation to reduce the frequency of CyanoHABs at a specific location downstream 6

Figure 1.2. Outline of Chapters 3 to 5 7

Figure 2.1. Location and schematization of the study area 11

Figure 2.2. Water transfer tunnel and Selective Withdrawal Facility (SWF) 12

Figure 3.1. Confusion matrix..... 28

Figure 3.2. Summary of the modelling procedure..... 28

Figure 3.3. Box plots to show the data distribution between classes for the nine features selected by ANOVA..... 31

Figure 3.4. Pearson correlation coefficients (absolute values) among nine features selected by ANOVA..... 33

Figure 3.5. Bar graphs to show oversampling for train sets of three groups..... 34

Figure 3.6. Bar graphs to show models' accuracy for test sets of three groups 36

Figure 3.7. Confusion matrices 37

Figure 4.1. Location of the monitoring stations including two weather stations 47

Figure 4.2. Graph showing the changes in the downstream NO_3-N concentration caused by changes in the CBOD concentration at all the boundary conditions..... 50

Figure 4.3. Hydrographs showing the difference between simulation and observation for calibration 56

Figure 4.4. Hydrographs showing the difference between simulation and observation for validation 57

Figure 4.5. Graphs showing the difference between simulation and observation of the NO_3-N concentration for calibration 61

Figure 4.6. Graphs showing the difference between simulation and observation of the NO_3-N concentration for validation 63

Figure 4.7. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate at the upstream boundary for 365 days (The black graph shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions. The dispersion coefficient was automatically computed in HEC-RAS.) 64

Figure 4.8. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate at the upstream boundary for 365 days (The black graph shows the NO_3-N

concentration simulated using the observational data from 2018 as boundary conditions. The dispersion coefficient was set to zero.) 65

Figure 4.9. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate (50, 100, and 150 $m^3 s^{-1}$) at the upstream boundary for 31 days (The black graph shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions.)..... 67

Figure 4.10. Changes in the NO_3-N concentration in cross section 416 caused by variations in flow rate (100 $m^3 s^{-1}$) at the upstream boundary for 10, 20, and 31 days (The black graph shows the NO_3-N concentration simulated using the observational data from 2018 as boundary conditions.)..... 68

Figure 4.11. Changes in the NO_3-N concentration in cross section 416 caused by variations in water temperature at the upstream boundary under Scenarios 43–48 (The black graphs show results simulated using the observational data from 2018 as boundary conditions.) 69

Figure 4.12. Changes in the NO_3-N concentration in cross section 416 caused by variations in the NO_3-N concentration at the upstream boundary under Scenarios 49–55 (The black graphs show results simulated using the observational data from 2018 as boundary conditions.) 70

Figure 4.13. Changes in the NO_3-N concentration in cross section 416 caused by a decrease in the NO_3-N concentration at the upstream boundary or an increase in flow rate at the upstream boundary (The black graphs show the result simulated using the observational data from 2018 as boundary conditions.)..... 71

Figure 5.1. Range of NO_3-N concentrations (CI) in water released from the Imha Reservoir..... 81

Figure 5.2. Experimental procedure 83

Figure 5.3. Graph showing Design Discharge, water demand, daily observational data, and monthly averages of the daily observational data for the downstream river (sum for the Andong and Imha reservoirs) 86

Figure 5.4. Results of the simulation test (the NO_3-N concentrations at Chilgok Weir using the hypothetical data and the HEC-RAS model) 88

Figure 5.5. Pareto front (OF1 and OF2 are the objective functions related to water quantity and water quality, respectively.)..... 90

Figure 5.6. Optimization results (QA+QI and CI) for three optimal solutions (OF1 min., OF2 min., median)..... 93

Figure 5.7. Optimization results (QJ and CJ) for three optimal solutions (OF1 min., OF2 min., median)..... 97

Figure 5.8. Simulation results of the NO_3-N concentrations at Chilgok Weir using the HEC-RAS model and the surrogate model for three optimal solutions 98

ABOUT THE AUTHOR

Jongchan Kim carried out his PhD research at the Hydroinformatics and Socio-Technical Innovation department of IHE Delft Institute for Water Education thanks to financial support from K-water (Korea Water Resources Corporation). His PhD diploma has been awarded by TU Delft (Delft University of Technology) in a joint degree program. Before joining IHE Delft, he has worked for K-water (since 2005), which is a governmental corporation in South Korea in charge of water resources management such as operating multi-purpose reservoirs and multi-regional water supply facilities. He has built his career in the water sector in fields of hydrologic investigation and multi-purpose reservoir operation.

Journals publications

- ✓ Kim, J., Jonoski, A., and Solomatine, D. P.: A classification-based machine learning approach to the prediction of cyanobacterial blooms in Chilgok Weir, South Korea, *Water*, 14, 542, doi:10.3390/w14040542, 2022.
- ✓ Kim, J., Jonoski, A., Solomatine, D. P., and Goethals, P. L. M.: Water quality modelling for nitrate nitrogen control using hec-ras: Case study of nakdong river in south korea, *Water*, 15, 247, doi:10.3390/w15020247, 2023.
- ✓ Kim, J., Jonoski, A., Solomatine, D. P., and Goethals, P. L. M.: Decision support framework for optimal reservoir operation to mitigate cyanobacterial blooms in rivers, submitted to *Sustainability* in 2023 (under review).

Conference proceedings

- ✓ Kim, J., Kim, H. Y., Choi, H. G., Jeong, S., and Lee, Y.: The innovative operation of Imha reservoir, E3S Web of Conferences, 01029, doi:10.1051/e3sconf/ 202234601029, 2022.

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Jongchan Kim

born on the 22nd of November 1978 in Ulsan, South Korea

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Delft, 14 September 2023

Chair of the SENSE board

Prof. dr. Martin Wassen

The SENSE Director

Prof. Philipp Pattberg

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The SENSE Research School declares that **Jongchan Kim** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 43.1 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2020)
- Research in context activity: Organizing a session in EKC (Europe-Korea Conference on science and technology) (2021)

Other PhD and Advanced MSc Courses

- Hydraulic Modelling Using HEC-RAS, IHE Delft (2020)
- QGIS for Hydrological Applications, IHE Delft (2020)
- Using Creativity to Maximize Productivity and Innovation in your PhD, TU Delft (2020)
- Engineering Ethics, TU Delft (2020)
- Analytical Storytelling, TU Delft (2020)
- The Informed Researcher: Information and Data Skills, TU Delft (2020)
- Foundations of Educational Design, TU Delft (2020)
- Cross Cultural Communication Skills in Academia, TU Delft (2020)
- Effective Negotiation: Win-Win Communication, TU Delft (2020)
- Problem-solving and Decision-making – The Covid-19 edition, TU Delft (2020)
- Brain Management, TU Delft (2020)
- Career Development - Exploring a research career outside academia, TU Delft (2020)
- Information Technology and Software Engineering, IHE Delft (2021)
- Scientific Text Processing with LaTeX, TU Delft (2022)

Management and Didactic Skills Training

- Supervising MSc student with the thesis entitled 'Andong Dam & Imha Dam Inflow Prediction Using Artificial Neural Network Model' (2021)

Oral Presentations

- *Approach to IWRM-based smart technology for reservoir operations.* EKC: Europe-Korea Conference on Science and Technology, 31st of October 2021, Paris, France
- *The innovative operation of Imha reservoir.* ICOLD 2021, 16th of November 2021, Online

SENSE coordinator PhD education

Dr. ir. Peter Vermeulen

IHE DELFT
Institute for Water Education
under the auspices of UNESCO

TU Delft Delft University of Technology

In the water sector, issues concerning the aquatic environment have been extensively discussed due to climate change. Particularly, water quality problems such as harmful cyanobacterial blooms (CyanoHABs) can be threats to the water environment while harming human health and aquatic ecosystems. This study focused on establishing a practical framework for the optimal operation of upstream reservoirs to address the problem of CyanoHABs in a downstream river. Furthermore, the applicability of this framework was demonstrated using observational data related to the quantity and quality of the upstream reservoirs in the study area, the Nakdong River of South Korea. Methodologically, three models were incorporated: a machine learning model to predict the occurrence of CyanoHABs, a river water quality model to simulate a water quality parameter influencing CyanoHABs, and an optimization model for the joint operation of upstream reservoirs. The research findings can support the decision-making of reservoir operation to create a favorable aquatic environment in a downstream river by reducing the frequency of CyanoHABs downstream. In particular, the framework to be established in this research can be a novelty in terms of efficiency since it can be a solution to the problem of CyanoHABs without using an additional amount of water from an upstream reservoir.

This book is printed on paper from sustainably managed forests and controlled sources

9 789073 445529