Modeling Earth Systems and Environment (2025) 11:102
https://doi.org/10.1007/s40808-024-02265-z

ORIGINAL ARTICLE

Deep reinforcement learning for multiple reservoir operation planning in the Chao Phraya River Basin

Yutthana Phankamolsil[1] · Areeya Rittima[2] · Wudhichart Sawangphol[3] · Jidapa Kraisangka[3] · Allan Sriratana Tabucanon[4] · Yutthana Talaluxmana[5] · Varawoot Vudhivanich[6]

Received: 28 September 2024 / Accepted: 19 December 2024

Abstract

This study demonstrates application of Deep Deterministic Policy Gradient (DDPG)-based algorithm to provide comprehensive and flexible plans for reservoir operation planning of the multiple reservoir system in the Chao Phraya River Basin (CPYRB), Thailand aiming to mitigate flood and drought risks in the region. The multi-agent-based Deep Reinforcement Learning (DRL) model is accordingly constructed considering 7-D predicted inflow, reservoir water released from adjacent reservoir, downstream flow condition, and changes in reservoir water storage, as state variables. The desired goal is to increase water storage levels in all reservoirs by 10–15% to ensure higher potential in supplying water for crop cultivation over the dry seasons and preventing flood occurrences during wet season. Simulation results from 2009 to 2022 indicate that DRL–DDPG-based algorithm can perform well in solving sequential decision problems for optimal operation of multiple reservoir system to achieve the desired water storage goal. It can offer realistic simulation results of seasonal and annual release schemes and reservoir release ratios among reservoirs in the system compared to actual operation and Fmincon and ANFIS optimizations. Importantly, DRL model demonstrates a significant advantage in view of increasing the long-term water storage levels in all reservoirs as targeted in the modelling process while maintaining the similar and consistent release schemes in the reservoir system. For the multipurpose multiple reservoir system operation, adjusting the dynamic desired goals within multi-agent-based RL model is advisable to attain the specific desired outcomes and address various water scenarios.

Keywords Deep Reinforcement Learning (DRL) · Deep Deterministic Policy Gradient (DDPG) algorithm · Artificial Intelligence (AI) · Reservoir operation planning · Chao Phraya River Basin

✉ Areeya Rittima
areeya.rit@mahidol.ac.th

Yutthana Phankamolsil
yutthana.pha@mahidol.ac.th

Wudhichart Sawangphol
wudhichart.saw@mahidol.ac.th

Jidapa Kraisangka
jidapa.kra@mahidol.ac.th

Allan Sriratana Tabucanon
allansriratana.tab@mahidol.ac.th

Yutthana Talaluxmana
fengynt@ku.ac.th

Varawoot Vudhivanich
fengvwv@ku.ac.th

[1] Environmental Engineering and Disaster Management Program, Mahidol University, Kanchanaburi Campus, Nakhon Pathom, Thailand

[2] Faculty of Engineering, Mahidol University, Phuttamonthon, Nakhon Pathom, Thailand

[3] Faculty of Information and Communication Technology, Mahidol University, Phuttamonthon, Nakhon Pathom, Thailand

[4] Faculty of Environment and Resource Studies, Mahidol University, Phuttamonthon, Nakhon Pathom, Thailand

[5] Department of Water Resources Engineering, Faculty of Engineering, Kasetsart University, Bangkok, Thailand

[6] Department of Irrigation Engineering, Faculty of Engineering at Kamphaeng Saen, Kasetsart University, Bangkok, Thailand

Published online: 23 January 2025

Introduction

Reservoir operation planning is a crucial task in water resources management involving the strategic determination of the optimal release and water storage from reservoir system to meet water demand sectors at all possible time steps. It plays essential role in addressing reservoir management strategy and establishing water allocation plans specifically for flood and drought risk mitigation driven by considerable change in climate variability. Reservoir operation planning studies for both single and multiple reservoir systems are primarily based on analyzing past reservoir data including water inflow and water outflow, estimating current and future water demand data, and modelling the reservoir operation system using the modern computer-based technology. The guidelines for releasing reservoir water incorporated with the recommended release scheme and established water allocation plan are expected to achieve as the intended goals of reservoir operation planning task. As the conventional decision-making process for reservoir operation scheme relies on the historical past data and traditionally uses predetermined rule curve as guideline, this requires extensive calculations particularly for large-scale multi-reservoir operation system (Oliveira and Loucks 1997). Additionally, the results are based on the subjective judgement by dam operators which may not capture its reality well. Moreover, accounting for non-linear relationships among relevant reservoir management factors are hardly performed. Due to these limitations, the superior techniques like Artificial Intelligence (AI) and simulation-based optimization have been progressively developed to enhance capabilities of learning, reasoning, problem-solving, and decision making of the complex reservoir operation system (Fayaed et al. 2013; Zhang et al. 2018; Seifollahi-Aghmiuni and Bozorg-Haddad 2019; Lai et al. 2022).

In recent years, integration of Artificial Intelligence (AI) technologies has driven a paradigm shift for reservoir operation and its adaptability (Yadav et al. 2023). AI is a field of computer science focusing on the simulation of human cognitive abilities by computer intelligent machine that are programmed to think and act rationally like human to solve complicated problems (Verma 2018). It is proven that AI is a powerful toolset in optimizing current reservoir operations, delivering improved decision-making competence, and allocating water resources more effectively (Yadav et al. 2023). In contrast to the physical-based models, AI-based models can acquire the various reservoir operation rules from hydrological big data and real-time operation data (Zhang et al. 2018). AI incorporates a broad range of techniques, algorithms, and approaches such as Machine Learning (ML), Reinforcement Learning (RL), Deep Reinforcement Learning (DRL), Fuzzy Logic (FL), Evolutionary Algorithm (EA), Optimization Algorithm (OA), and Hybrid Models (HM) (Yadav et al. 2023) which benefits for the specific applications in reservoir management. AI can potentially offer the improved decision-making capabilities in various means such as enhanced data processing for hydrological time-series prediction (Tounsi et al. 2022; Dastour and Hassan 2023), augmented precision for flood prediction (Hu et al. 2019), and adaptability for water resource management (Belayneh et al. 2016).

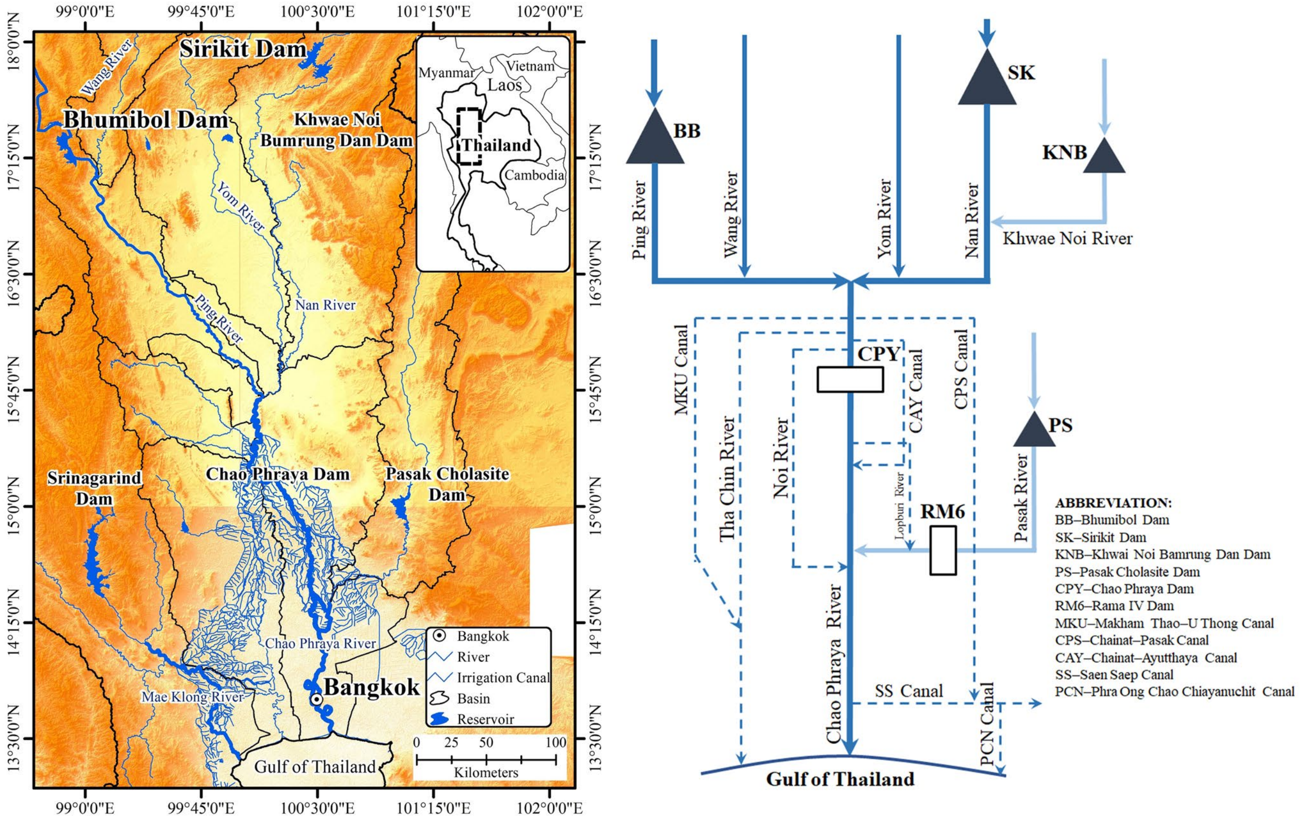
Reinforcement Learning (RL) is a sub-field of Machine Learning (ML), which is a branch of modern Artificial Intelligent (AI). Its background of RL is primarily rooted in Dynamic Programming (DP) and Markov Decision Processes (MDPs) in solving sequential decision problems (Wiering and van Otterlo 2012; Tabas 2020). Both DP and MDPs have the similar mathematical foundation used to describe the sequential decision-making problems (Wenwu et al. 2018). MDPs have been basically used to address most of the RL problems as it can model the environments with a finite set of environmental states, actions, transition probabilities and reward functions. RL has been progressively developed to leverage the learning of dynamic system behaviors by reward-driven trial and error process (Kaelbling et al. 1996). In recent years, RL has been driven by the AI research advancement in computer science which has yielded transformative and paradigm-shifting technologies. It has been found that RL algorithms have been widely applied in various fields including optimal operation of reservoir systems (Castelletti et al. 2001, 2010; Mahootchi et al. 2007; Madani and Hooshyar 2014; Dariane and Moradi 2016; Wenwu et al. 2018; Hu et al. 2022a, b). The key benefits of RL focusing on the long-term goal and uncertain environment have been proven through many applications for reservoir management (Mahootchi et al. 2007; Wang et al. 2020) and water resource system management (Hung and Yang 2021). The superior performance of RL-based reservoir operating policy has been proven to significantly outperform than those policy designed by human (Wang et al. 2020). In addition, RL has been applied for water resource scheduling of multi-reservoir system which exhibits better performance than traditional dynamic programming (Lee and Labadie 2007). Importantly, RL technique enables to adjust itself to learn the dynamic environment and create the proper response and reactions to these changes effectively (Mahootchi et al. 2007).

The core elements of RL model basically include: (1) environment, which is the genuine physical system that the agent works or simulated environment, (2) state, which is current situation of the environment, (3) agent, which is the system component that receive the states to take action, (4) reward, which is the response of environment due to the agent's action, (5) policy, which is mapping procedure of the agent's state to the action, and (6) value, which is the future reward that the system agent would gain by taking the action in a specific state.

As the goal of RL is to maximize the cumulative reward over time, therefore, designing a proper reward function is the most essential task for state–agent–action interaction (Bhattacharya et al. 2003). Establishing RL model can be started with defining the RL problem which includes agent's goal, state space, action space, and reward function. Agent's goal is what the agent needs to solve, so formulating explicit and measurable RL goal is definitely significant. The set of possible situations that the agent can encounter is defined as state space. The action space is the set of possible actions that agent can take. Determining the size of state and action spaces are based on the characteristic and relevant information of the system. The reward function defines the goal in the RL problems by mapping each perceived state of the environment to an assigned number called "reward". It influences the behavior and learning of RL model for successful or unsuccessful outcomes. Implementation of RL problem can be manipulated by selecting RL learning algorithm which can be classified into two categories: (1) Value-based and (2) Policy-based. The value-based RL learns a value function (Q-value) to guide a specific action for a next sequential state which the present and expected future rewards are maximized. However, it can be suitable and more efficient for some environments with small state and action spaces (Andriotis and Papakonstantinou 2019). The policy-based RL directly learns to take action by mapping states to actions into policy. It is regarded as more adaptable for the environment system with continuous action spaces (Nguyen et al. 2020). In the learning process of RL, maximizing the collective rewards of the dynamic environment system for taking a particular action in a given state (Q-value) is intended to achieve by the value-based and policy-based RL. Additionally, setting up appropriate hyperparameters, such as learning rate, discount factor, and exploration rate, is made in the RL implementation process. Learning process of RL models can be implemented by direct interaction between the agent and the states in environment. The latest updated value functions of the states can be used to update for each iteration. The agent selects the best admissible action that provides the best value function.

Q-Learning or fitted Q-iteration, a classical value-based RL algorithm, is widely used as learning method suitable for an environment with small state-action spaces. It uses tabular representation to collect the Q-value indicating state–action relation. Since this decision table approach of the classical RL cannot handle well with the large number of state–action combinations resulting in the curse of dimensionality problem, Deep Q-Networks (DQNs) which is the value-based Deep Reinforcement Learning (DRL) algorithm, was developed to take the discrete actions (Xu et al. 2021). To solve issues in continuous spaces and high dimensional states and actions, DRL was initially developed by combining the classical RL with deep neural networks representation (François-Lavet et al. 2018; Xu et al. 2021; Jiang et al. 2024). The enhanced learning capability of DRL in complex environments has been proven and its application to various fields has been extensively promoted (Mnih et al. 2015). DRL has been increasingly applied for reservoir system management (Rieker and Labadie 2012; Wang et al. 2020), optimal operation of multipurpose reservoir systems (Peacock and Labadie 2018), optimal hydropower reservoir operation (Xu et al. 2020, 2021; Wu et al. 2024), water division optimization (Jiang et al. 2024), and real-time control of stormwater systems (Mullapudi et al. 2020). In addition, to advance the performance of RL dealing with high-dimensional state spaces and continuous actions, Deep Deterministic Policy Gradient (DDPG) which is a sort of DRL algorithm, is newly developed for decision making process in the complicated environment. It combines elements of the value-based and policy-based RL in accordance with actor-critic networks (Alturkistani and El-Affendi 2022). This allows DDPG to learn both the value function and policy and take the optimal actions in a large and complex environment. DDPG has been proven in term of capability to successfully solve the problems with large model parameters and non-linear dynamics (Sumiea et al. 2023). Additionally, due to stability and convergence properties of DDPG algorithm, it has been applied in a broad range of challenging tasks including robotics, simulation-based issues, energy management, and reservoir operation decision and control (Tabas and Samadi 2024).

In this study, DRL modelling-based design approach focusing for multiple reservoir operation planning was demonstrated and applied for the Chao Phraya River Basin (CPYRB). Due to unbalancing between the water availability and water demand in this region, the reservoir operation management and planning plays crucial role in driving the water resources management policy for implementation against flood and drought problems. CPYRB is the largest basin in Central Thailand occupying drainage area of approximately 160,000 km^2 or nearly 30% of the country area. CPYRB has shifted from the uncontrolled basin to the highly developed basin with multipurpose storage dams, extensive canal infrastructures serving more than 10 million rai (16,000 km^2) of irrigated land (Kyaw et al. 2024), and expansion of industrial and urban area since 2000s (Molle 2002). There are four main storage dams in CPYRB: Bhumibol (BB), Sirikit (SK), Khwae Noi Bunrung (KNB), and Pasak Cholasite dams, which built across Ping, Nan, Khwae Noi, and Pasak rivers, respectively as illustrated the basin map in Fig. 1a and river schematic diagram in Fig. 1b. Their reservoir capacities are 13,462, 9,510, 939,



(a) Topographic map

(b) River schematic diagram

Fig. 1 The Chao Phraya River Basin in the Central Thailand

and 960 million cubic meters (MCM). These dams supply reservoir water for both local demand and joint demand in the wide floodplain area along the Chao Phraya rivers. The Chao Phraya (CPY) diversion dam is acted to re-regulate the downstream flow released from BB, SK, and KNB before distributing into canal irrigation system in the Greater Chao Phraya Irrigation Scheme (GCPYIS) and downstream river reach where PS dam joins. More than 70% of the water allocated from main dams has been supplied for agricultural purpose in GCPYIS. The remaining has been utilized for non-agricultural water needs including municipal and industrial uses, and ecological conservation along the tributaries and main rivers as well as hydropower production. The current reservoir operation from 2000 to 2022 in CPYRB reveals that the average release portions of all dams are 0.3352:0.3391:0.1015:0.1643 for BB, SK, KNB, and PS dams, respectively. However, these water allocation ratios have been considerably altered corresponding to the water availability and water demand situations within the basin.

Managing risks of flooding and drought events driven by climate variability and economic development acceleration in CPYRB have become critical priority in the context of water resource management. In 2011, the worst flooding triggered by the tropical monsoon storms was occurred and sparsely spread in the northern, northeastern, and central Thailand creating huge damages and economic losses in CPYRB and neighboring basins. Since 2011, CPYRB has frequently experienced a sudden increase in monsoon flooding particularly at the end of wet season (September–October) which may continue into November in flood prone area along the lower reach of Chao Phraya and Pasak rivers. Moreover, CPYRB has suffered the consecutive and prolonged droughts during dry season (November–April) arising more frequently from 2016 to 2018. This highlights the necessity of establishing suitable water allocation plan along with generating proper reservoir release scheme to effectively handle flood and drought risks for both short-term and long-term operations in this region. Consequently, this study aims to investigate the capability of Deep Reinforcement Learning (DRL) for multiple reservoir operation planning in CPYRB. A multi-agent system for multiple reservoir operation is implemented using the DRL–DDPG algorithm to determine the water releases from all reservoirs corresponding to the targeted water storage levels. Making decision on the water release by multi-agent-based DRL relies on keeping higher storage levels of all reservoirs up to 10–15% above the long-term average as established to ensure effective reservoir management planning and mitigate flood and drought risks in this region.

Methods

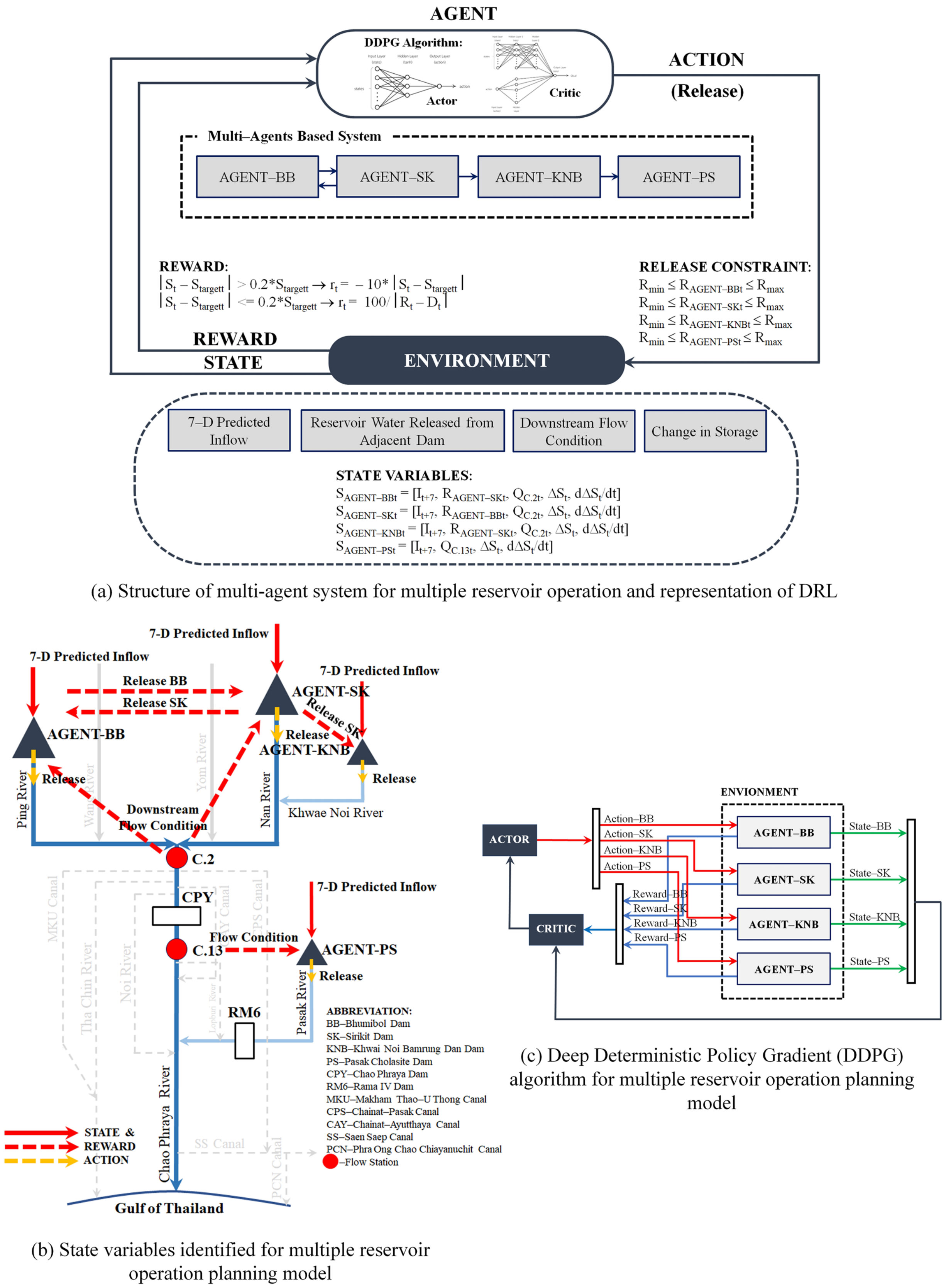
Development of multiple reservoir operation planning model by deep reinforcement learning

To develop the daily multiple reservoir operation model by the multi-agent-based DRL for CPYRB, the operation of each reservoir is represented as agent including: (1) Agent-BB, (2) Agent-SK, (3) Agent-KNB, and (4) Agent-PS. The determination of daily reservoir release of each dam is defined as the agent's action that receives the state variables from the environment system. The state variables of Agent-BB identified for this study consist of 7-day predicted inflow (I_{t+7}) (Kraisangka et al. 2022), observed reservoir release of SK dam at time t ($R_{agent\text{-}SKt}$), key downstream flow condition at C.2 station at time t ($Q_{C.2t}$), change in water storage at time t (ΔS_t) and derivative of the change in water storage with respect to time t ($d\Delta S_t/dt$). As the water storage levels of all dams in CPYRB are aimed to increase by 10–15% compared to the long-term average to moderate water scarcity during dry season, therefore, the targeted water storage levels of four main agents are accordingly generated. To achieve this, two key state variables; ΔS_t and $d\Delta S_t/dt$ are also incorporated into the reservoir operation planning model by DRL. The change in storage (ΔS_t) is subtraction term of the targeted water storage volume ($S_{targett}$) and the water storage volume generated by DRL model (S_t).

Similarly, the state variables of Agent-SK are 7-day predicted inflow (I_{t+7}), observed reservoir release of BB dam at time t ($R_{agent\text{-}BBt}$), key downstream flow condition at C.2 station at time t ($Q_{C.2t}$), change in water storage at time t (ΔS_t) and derivative of the change in water storage with respect to time t ($d\Delta S_t/dt$). As the KNB dam which was built across the Khwae Noi river, a major tributary of the Nan river, supplies water to the central region downstream of SK dam, therefore, defining the state variable of KNB-agent also incorporates the observed reservoir water released from SK dam ($R_{agent\text{-}SKt}$). In addition, 7-day predicted inflow (I_{t+7}), downstream flow condition at C.2 station at time t ($Q_{C.2t}$), change in water storage at time t (ΔS_t) and derivative of the change in water storage with respect to time t ($d\Delta S_t/dt$) are determined as key state variables for Agent-KNB to potentially satisfy the joint water demand for CPYRB. The key gauged flow at C.13 station at time t ($Q_{C.13t}$) located downstream of the Chao Phraya diversion dam is considered as the major state variable of Agent-PS together with 7-day predicted inflow (I_{t+7}), change in water storage at time t (ΔS_t) and derivative of the change in water storage with respect to time t ($d\Delta S_t/dt$) as illustrated in Fig. 2a and b.

Identifying the state variables of all agents refers to the physically-connected reservoir system and joint operation among all reservoirs in CPYRB, together with upstream and downstream influencing factors on reservoir operation. The anticipated inflow data at 7-day lead time and current and desired reservoir storage status of each single reservoir is considered as one of the important factors to make reservoir operation response corresponding to the changing water conditions. Furthermore, the downstream flow conditions at key selected stations (C.2 and C.13) and the release schemes of adjacent reservoirs impact the release decision and response for a reservoir to aid multiple reservoir operation and jointly serve the downstream demand in the lower basin. Additionally, considering initial downstream flow conditions at these key stations can prevent simultaneous dam releases from different tributaries of CPY river that may coincide and potentially lead to severe flooding in the downstream economical areas. In other words, downstream flow conditions are determined as state variable in the DRL model to identify the potential flood risks and constraints.

In the decision-making process, the agent (representing reservoir operation system of each dam) utilizes DRL to take an action (representing water release from each reservoir) through the process of trial and error driven by the assigned rewards. Each action of determining the amount of released water is referred to an "episode" which consists of numerous simulation iterations. Deep Deterministic Policy Gradient (DDPG) which is a reinforcement learning algorithm, is used for multiple reservoir operation modelling in CPYRB. DDPG employs an actor-critic approach combining value-based (Q-value) and policy-based (policy gradient) techniques that can implement large state spaces in the environment to take indiscrete action. In DDPG algorithm, the agent takes the action corresponding to the maximum Q-value from the current state. The Q-value signifies the expected future reward for taking a certain action in a given state. In other words, by doing this, the agent aims to maximize its expected future reward. To learn from the past experience and improve future decisions, each action made


AGENT
DDPG Algorithm:
Actor
Critic
ACTION
(Release)
Multi–Agents Based System
AGENT–BB
AGENT–SK
AGENT–KNB
AGENT–PS
REWARD:
$|S_t - S_{targett}| > 0.2{*}S_{targett} \rightarrow r_t = -10{*}|S_t - S_{targett}|$
$|S_t - S_{targett}| <= 0.2{*}S_{targett} \rightarrow r_t = 100/|R_t - D_t|$
RELEASE CONSTRAINT:
$R_{min} \le R_{AGENT-BBt} \le R_{max}$
$R_{min} \le R_{AGENT-SKt} \le R_{max}$
$R_{min} \le R_{AGENT-KNBt} \le R_{max}$
$R_{min} \le R_{AGENT-PSt} \le R_{max}$
REWARD
STATE
ENVIRONMENT
7–D Predicted Inflow
Reservoir Water Released from Adjacent Dam
Downstream Flow Condition
Change in Storage
STATE VARIABLES:
$S_{AGENT-BBt} = [I_{t+7}, R_{AGENT-SKt}, Q_{C.2t}, \Delta S_t, d\Delta S_t/dt]$
$S_{AGENT-SKt} = [I_{t+7}, R_{AGENT-BBt}, Q_{C.2t}, \Delta S_t, d\Delta S_t/dt]$
$S_{AGENT-KNBt} = [I_{t+7}, R_{AGENT-SKt}, Q_{C.2t}, \Delta S_t, d\Delta S_t/dt]$
$S_{AGENT-PSt} = [I_{t+7}, Q_{C.13t}, \Delta S_t, d\Delta S_t/dt]$
7-D Predicted Inflow
Release BB
Release SK
AGENT-SK
AGENT-BB
Release SK
Release
AGENT-KNB
Release
Release
Ping River
Wang River
Yom River
Nan River
Khwae Noi River
Downstream Flow Condition
C.2
CPY
MKU Canal
CAY Canal
CPS Canal
Tha Chin River
Noi River
Lopburi River
Flow Condition
C.13
AGENT-PS
Release
Pasak River
RM6
Chao Phraya River
SS Canal
PCN Canal
Gulf of Thailand
STATE &
REWARD
ACTION
ABBREVIATION:
BB–Bhumibol Dam
SK–Sirikit Dam
KNB–Khwai Noi Bamrung Dan Dam
PS–Pasak Cholasite Dam
CPY–Chao Phraya Dam
RM6–Rama IV Dam
MKU–Makham Thao–U Thong Canal
CPS–Chainat–Pasak Canal
CAY–Chainat–Ayutthaya Canal
SS–Saen Saep Canal
PCN–Phra Ong Chao Chiayanuchit Canal
–Flow Station
ENVIONMENT
ACTOR
CRITIC
Action–BB
Action–SK
Action–KNB
Action–PS
Reward–BB
Reward–SK
Reward–KNB
Reward–PS
State–BB
State–SK
State–KNB
State–PS

(a) Structure of multi-agent system for multiple reservoir operation and representation of DRL

(b) State variables identified for multiple reservoir operation planning model

(c) Deep Deterministic Policy Gradient (DDPG) algorithm for multiple reservoir operation planning model

Fig. 2 Conceptual idea for the development of multiple reservoir operation model by DRL for CPYRB

by the agent is recorded as behavior in an Artificial Neural Network (ANN) structure. The action with the highest Q-value is selected and the policy gradient in the ANNs is accordingly adjusted as shown in Fig. 2c.

The modelling process by DRL–DDPG begins with the actor, a component of agent, receiving a state value and sending an action to the environment. Subsequently, the environment sends a reward and a state value to the critic neural network to update the Q-value. The critic neural network then sends the Q-value back to the actor neural network for gradient adjustment. In the other words, the critic neural network acts as a guideline, advising the actor neural network on which action will yield the highest Q-value.

In this study, the reward function is termed as a function of water storage maintaining to reach the established target levels and satisfaction of releasing reservoir water to meet the water demand for each dam. Consequently, calculating the rewards is subject to two conditions as expressed in the following equations.

$$\left|S_t^i - S_{\text{targett}}^i\right| > 0.2 * S_{\text{targett}}^i \rightarrow r_t^i = -10 * \left|S_t^i - S_{\text{targett}}^i\right|; \quad \forall i = 1,\ldots.,N,\ \&t = 1,\ldots,T \tag{1}$$

$$\left|S_t^i - S_{\text{targett}}^i\right| \le 0.2 * S_{\text{targett}}^i \rightarrow r_t^i = 100/\left|R_t^i - D_{\text{t}}^i\right|; \quad \forall i = 1,\ldots.,N,\ \&t = 1,\ldots,T \tag{2}$$

where R^{i}_{t} is the reservoir release (outflow) of reservoir i at time step t (or DRL release), S^{i}_{t} and $S^{i}_{targett}$ are the reservoir water storage and targeted water storage of reservoir i at time step t, respectively, and D^{i}_{t} the reservoir water demand of reservoir i at time step t, including both local and joint water demands. In this study, the local and joint water demands in the basin are calculated considering both agricultural and non-agriculture requirements and encompassed into the DRL-based reservoir operation model. The calculation of local water demand for each of four main storage dams are based on the agricultural water need for small-scale irrigation scheme, and non-agricultural water demand including municipal, industrial, and ecological needs to represent its local demand supplied by the adjacent dam and incorporate environmental challenges in each CPY tributaries. For joint water demand, agricultural water requirement over the planting seasons in GCPYIS which is the largest irrigation scheme in the central region, is

Table 1 Targeted water storages of all dams specified for multiple reservoir operating planning

Dam Month		BB		SK		KNB		PS	
		Avg. storage	Targeted storage	Avg. storage	Targeted storage	Avg. storage	Targeted storage	Avg. storage	Targeted storage
Unit	–	MCM	MCM	MCM	MCM	MCM	MCM	MCM	MCM
Jan	DS	9,043	10,493	6,820	7,819	548	683	661	757
Feb		8,534	9,983	6,350	7,349	451	586	539	635
Mar		7,865	9,314	5,786	6,785	362	498	419	514
Apr		7,177	8,627	5,206	6,205	287	422	308	404
May	WS	6,681	8,130	4,753	5,752	239	375	235	331
Jun		6,517	7,966	4,591	5,590	223	358	204	300
Jul		6,412	7,862	4,714	5,713	226	362	182	278
Aug		6,677	8,126	5,491	6,490	326	462	193	288
Sep		7,618	9,067	6,567	7,566	525	660	429	525
Oct		8,828	10,277	7,183	8,182	706	842	808	904
Nov	DS	9,361	10,810	7,245	8,244	738	873	843	939
Dec		9,342	10,791	7,052	8,051	684	820	776	871
Initial Water Storage Increased in WS[1/]			+1,449 (+15%)		+999 (+15%)		+136 (+15%)		+96 (+10%)
Initial Water Storage Increased in DS[2/]			+1,450 (+15%)		+999 (+15%)		+135 (+15%)		+96 (+10%)
MPL			3,800		2,850		43		3
NPL			13,462		9,510		939		960

[1/] The difference between the targeted and average water storage levels of each dam in April

[2/] The difference between the targeted and average water storage levels of each dam in October accordingly estimated. Estimating municipal and industrial water demand in the lower CPYRB, and ecological needs along downstream reach of CPYR to prevent seawater intrusion is also conducted to account for all water demand sectors. Since supplying water to joint demand in CPYRB is proportionally shared by four dams in the basin, therefore, the average release portions of 0.3352:0.3391:0.1015:0.1643 for BB:SK:KNB:PS dams, are accordingly used to determine the individual water demand.

To maintain the targeted storage levels of all agents in the reservoir system, the negative reward, $-10 * \left|S_t^i - S_{\text{targett}}^i\right|$ is given for each step of agent's action when the difference in DRL water storages and targeted storage level of reservoir i is greater than 20% of targeted storage levels or $\left|S_t^i - S_{\text{targett}}^i\right| > 0.2 * S_{\text{targett}}^i$. In contrast, the positive reward, $100/\left|R_t^i - D_t^i\right|$ is given for each step of agent's action when the DRL water storage and targeted level of reservoir i is in a range of less than 20% of targeted storage levels or $\left|S_t^i - S_{\text{targett}}^i\right| \leq 0.2 * S_{\text{targett}}^i$.

It is noticeable that when the storage difference term or $\left|S_t^i - S_{\text{targett}}^i\right|$ is large *($>0.2{*}S^i_{targett}$)*, the considerable negative reward *($<<-10$)* is given to the DRL model by multiplying the storage difference term with –10. In other words, a large discrepancy of DRL and targeted water storage levels yields a considerable negative penalty reward for the DRL model. When the storage difference term or $\left|S_t^i - S_{\text{targett}}^i\right|$ is small *($\leq 0.2{*}S^i_{targett}$)* indicating that the DRL model can achieve well with the targeted storage levels, the significant positive reward is given to the DRL model as an inverse function of water deficit term, $\left|R_t^i - D_t^i\right|$ by dividing 100 with water deficit term. A maximum positive reward of + 100 is given to the model when there is no water deficit.

DRL–DDPG Algorithm:

The computational process of DRL–DDPG formulated in this study is presented in the following;

DRL–DDPG Algorithm:

```
Initialize a neural network with random weight; $\theta^Q$ in critic network $Q(s, a/\theta^Q)$ and $\theta^u$ in actor network $\mu(s/\theta^u)$
Initialize target network $\theta'$ and $\mu'$ with weight $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer $R$
for episode = 1, $M$ do
  Initialize a random process $N$ for action exploration
  Receive initial observation state $s_1$
  for t=1, $T$ do
    Select action $a_t = \mu(s|\theta^\mu) + N_t$ according to the current policy and exploration noise
    Execute action $a_t$ and observe reward $r_t$ and observe new state $s_{t+1}$
    Store transition $(s_t, a_t, r_t, s_{t+1})$ in $R$
    Randomly sample a minibatch of $N$ transitions of $N$ transition $(s_t, a_t, r_t, s_{t+1})$ from $R$
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
    Update critic value by minimizing the loss: $L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
    Update the actor policy using the sample policy gradient:
        $\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}$
    Update the target networks:
        $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
        $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
  end for
end for
```

Where s_t and s_{t+1}: state at time t and time t+1(new state) in which s∈S where S is a finite set containing all possible states. a_t: action at time t in which a∈A where A is a finite set containing all possible actions. r_t: reward at time t due to the action a_t and state s_t. γ: reward discount factor to control the importance of future reward ranging between 0–1. τ: changing rate for target network weight. R: experience replay memory. J: policy score function to calculate the expected reward of policy. N: exploration noise. T: number of time steps. M: number of episodes. L: loss function to quantify discrepancy between the agent's predicted values and the actual values. θ, α, ∇, and y: weighted value, learning rate, gradient value, and target value of neural network, respectively. Q or $Q(s, a)$: Q–function (DQN) which is reward summation in the present and future times. $Q(s, a/\theta^Q)$: $Q(s, a)$ estimated by neural network θ. Q' or $Q'(s, a)$: target Q–function. μ or $\mu(s)$: deterministic policy function. μ' or $\mu'(s)$: target deterministic policy function.

The multiple reservoir operation system in CPYRB is primarily governed by the principle of mass balance as expressed in the following equation;

$$S^i_{t+1} = S^i_t + I^i_t - R^i_t - E^i_t - Spill^i_t; \forall i = 1, \ldots, N, \& t = 1, \ldots, T \quad (3)$$

where S^i_t and S^i_{t+1} are the reservoir water storages of reservoir i at time step t and t + 1, respectively, I^i_t is the reservoir inflow of reservoir i at time step t, R^i_t is the release of reservoir i at time step t achieved by DRL–DDPG algorithm, E^i_t is the evaporation losses of reservoir i at time step t, and $Spill^i_t$ is the spilled water from the reservoir i at time step t. The decision on the DRL release of each reservoir is constrained by the minimum release, R^i_{min} and maximum release, R^i_{max} to ensure the minimum environmental flow requirement and maximum safe channel capacity of each dam. In addition, the available water storages after releasing reservoir water by DRL model should lie between minimum water storage, S^i_{min} and maximum water storage, S^i_{max} of each reservoir.

$$R^i_{\min} \leq R^i_t \leq R^i_{\max} \quad (4)$$

$$S^i_{\min} \leq S^i_t \leq S^i_{\max} \quad (5)$$

Setting up the targeted reservoir water storage for reservoir operation planning

The reservoir operation planning is served as the fundamental undertaking for the strategic reservoir management to achieve the specific purpose. This enables the reservoir planners to better understand and establish strategic operation policy for sustainable water security. The main objective of this study is to demonstrate the deep reinforcement learning technique to recommend the release scheme for multiple reservoir operation planning task. Consequently, the increased levels of targeted storages of four main dams in CPYRB by 10–15% compared to the long-term average, is generated as expressed in Table 1 and Fig. 3. This leads to the enhanced potential to intensively supply water not only for irrigation over the crop cultivation periods but also the downstream water needs. However, to protect dams from downstream flooding as a result of reservoir operation, the percentage increase of water storage levels in the reservoir system is determined lying between the Normal Pool Level (NPL) and Minimum Pool Level (MPL) of each reservoir. According to the increased water storages in reservoirs during dry season (Nov.–Apr.), the extra amount of water storage of + 1,450, + 999, + 135, and + 96 MCM for BB, SK, KNB, and PS dams can be increased in October before the subsequent crop planting season begins. Similarly, during wet season (May.–Oct.), the additional water storages in May of + 1,499, + 999, + 136, and + 96 MCM for these dams can be enhanced to meet agricultural and non-agricultural water demands throughout the wet season. Not only the increased water storage in all dams is utilized for the downstream water conservation purpose over the crop cultivation periods, but also it is beneficial for hydropower production. Based on this, optimal daily reservoir release scheme from 1/11/2009 to 31/12/2022 is accordingly accomplished using DRL approach.

Evaluating the effectiveness of DRL model in reservoir management for CPYRB

The daily long-term simulation from 1/11/2009 to 31/12/2022 was conducted using the DRL–DDPG-based operation model for CPYRB. The simulated reservoir water storages, reservoir releases, and release ratios of four main dams; BB, SK, KNB, and PS, were accordingly evaluated to explore the short-term and long-term operational capabilities of the DRL model in comparison to actual operation. In the last step, the comparative analysis of the DRL model against two optimization approaches previously studied for reservoir management in CPYRB focusing on BB and SK dams was benchmarked; (1) non-linear optimization programming using Fmincon function (Kyaw et al. 2022) and (2) Adaptive Neuro Fuzzy Inference System (ANFIS) (Kyaw et al. 2024).

Result and discussion

The simulated reservoir operation accomplished by DRL model for CPYRB

The followings are the daily simulated reservoir operations from 1/11/2009 to 31/12/2022 for BB, SK, KNB, and PS dams under the alternative reservoir operation schemes generated by the DRL model. The water storage levels simulated by DRL model are compared to both the targeted and observed water storage levels as shown in Fig. 4. It is revealed that DRL model recommends to release the optimal volume of water from all reservoirs to reach the increased water storage levels as determined as a desired goal. This substantially results in lowering the considerable fluctuations of water storage levels in all reservoirs. The reservoir releases at the current time step of all dams performed by DRL–DDPG-based algorithm is accomplished by the refinement process to get the maximum reward values which learns from the current and next future time steps to find the optimal action using actor-critic neural networks. Therefore, the optimal daily releases to achieve the targeted water storage levels, which will be used to establish the seasonal and

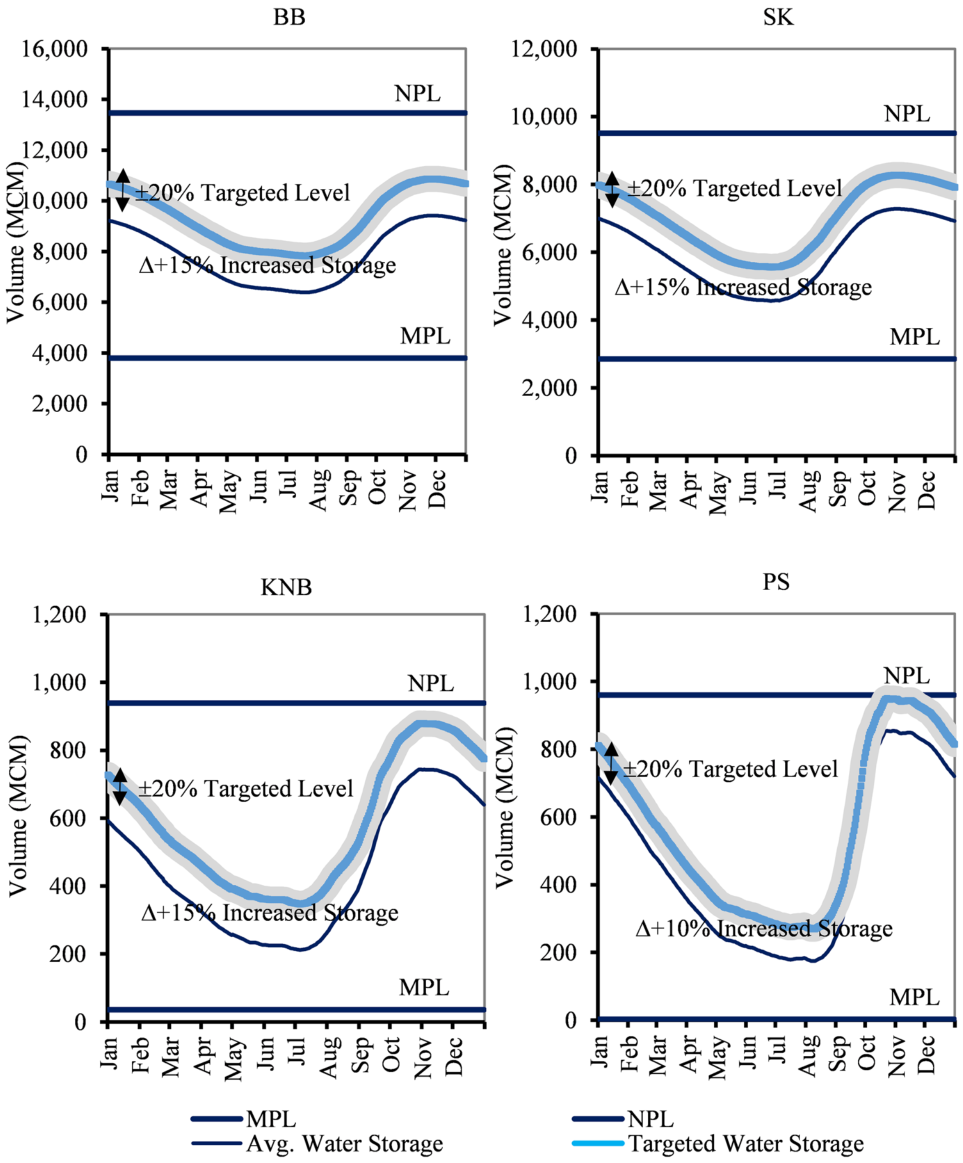


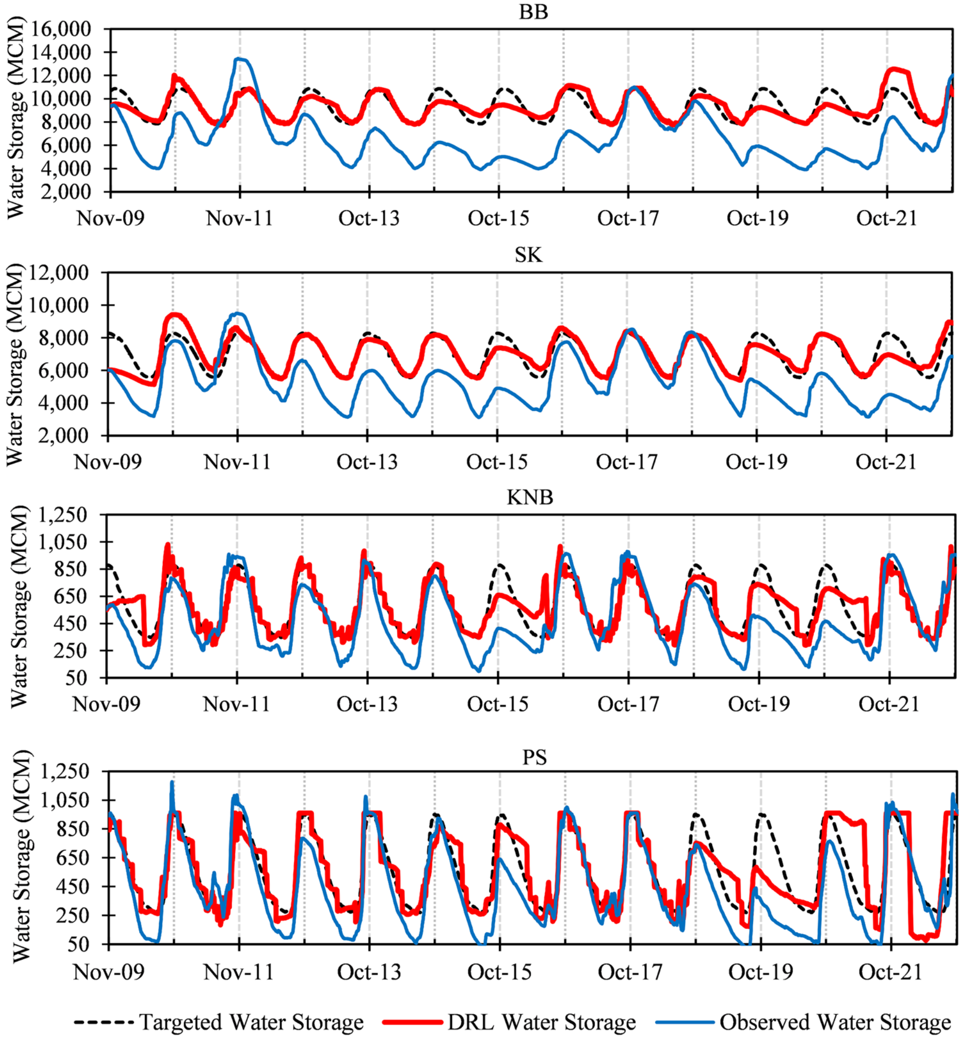
Fig. 3 Targeted water storage levels specified for multiple reservoir operation planning model by DRL

annual water allocation plans for CPYRB, are accordingly generated.

Effectiveness of DRL model in reservoir management for CPYRB

As DRL model performs well to conform the targeted water storage levels of all dams constructed as annual plan for reservoir operation, the effectiveness of DRL model in reservoir management for CPYRB is accordingly assessed. The total amount of annual water releases performed by DRL model from 2010 to 2022 are compared to the observed annual releases as shown the results in Fig. 5. In addition, the water release ratio among reservoirs in the CPYRB system are also calculated and presented. It is found that DRL generates the different annual release patterns according to the various water circumstances compared to the observed releases. In wet year of 2011, DRL model recommends releasing larger volume of released water from all dams to deplete reservoir storage levels and keep as the targeted water storages. As a result, critical flood risks for the subsequent time periods in 2012 can be certainly moderated. In critical dry years of 2014, 2015, 2019, and 2020, the total annual releases implemented by DRL model are likely to be the lowest compared to those in wet and normal years. However, adjusting the dynamic

Fig. 4 Comparison of reservoir water storages obtained from DRL model and observed water storages



targeted water storage levels to suit to water supply and water demand conditions in the reservoir system is highly recommended to reduce flood and drought risks.

The simulated results over the period 2010–2022 also indicate that average values of annual release achieved by DRL model are slightly higher than those observed releases by + 2.93% and + 0.16% for BB and KNB dams while slightly lower by – 2.91% and – 9.51% for SK and PS dams, respectively. While DRL model determines reservoir releases annually based on targeted water storage levels and state variables in the environment system, the average total release across all dams in CPYRB closely aligns with observed release. Consequently, small percentage difference of – 1.71% is found as indicated in Table 2.

Based on the seasonal analysis of reservoir release, it is found that during crop cultivation periods in dry season, DRL model recommends increasing additional water release of PS dam by + 46.01% while lowering reservoir water from BB, SK, and KNB by – 31.75%, – 16.91%, and – 17.92%, respectively. In contrast, DRL model achieves targeted water storage levels by suggesting to increase the water releases during crop cultivation periods in wet season from BB, SK, and KNB dams by + 66.28%, + 16.86%, and + 14.84%, respectively and lowering the release water from PS dams by – 34.27%.

Corresponding to the simulated results of reservoir operation aiming to keep the increased levels of water storages of all reservoirs up to 10–15% of the average, the release ratios of all reservoirs are considerably assessed and compared to the actual operation in the multiple reservoir system. Table 3 and Fig. 6 presents the reservoir release ratios in dry years, normal years, and wet years for short-term multiple operation, as well as the reservoir release ratios for long-term multiple operation which are obtained from the multiple reservoir operation planning model accomplished by DRL model. It is illustrated that DRL model suggests adjusting reservoir water allocation schemes among reservoirs in the system for both short-term and long-term operations to increase

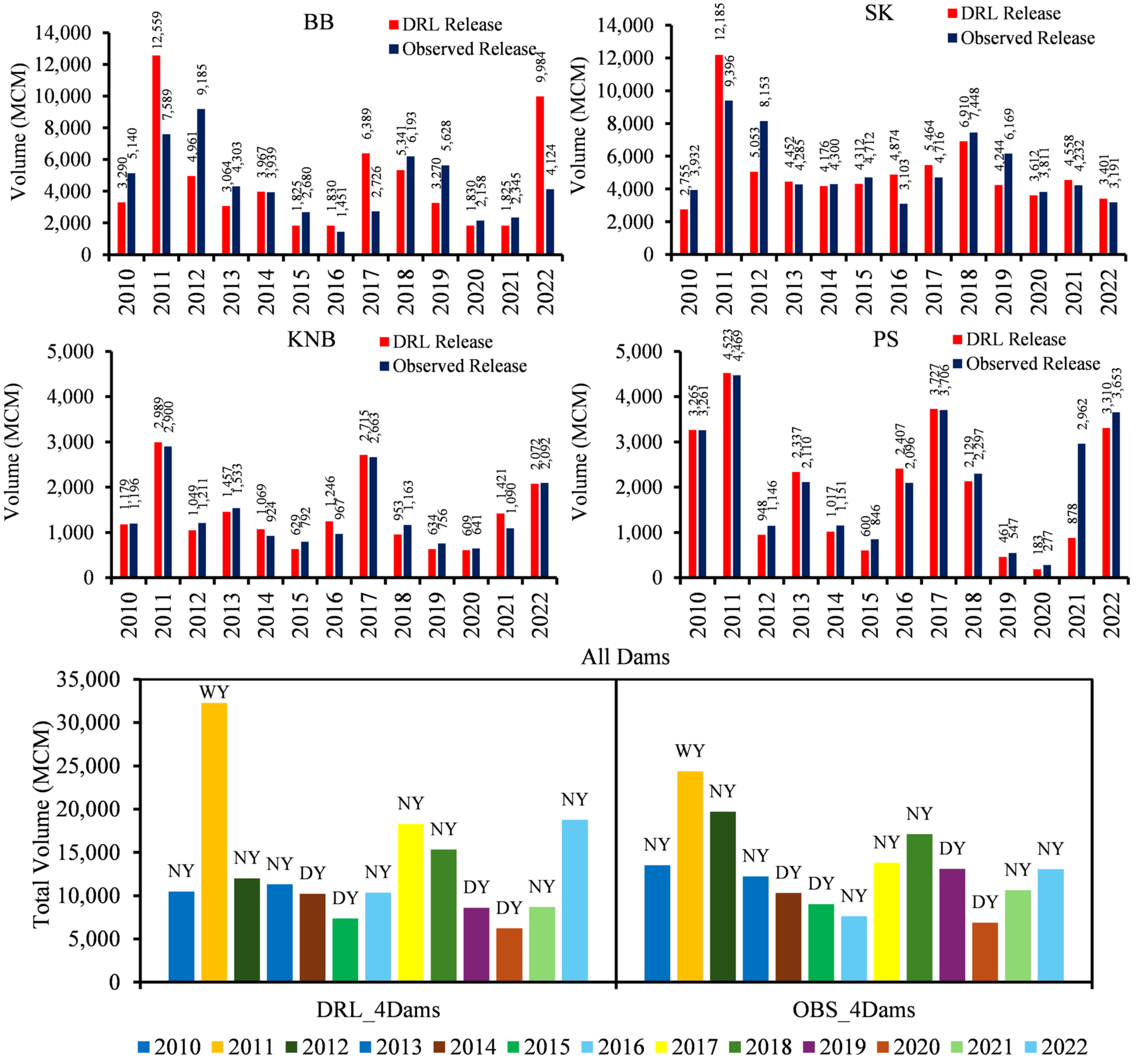


Fig. 5 Comparison of reservoir releases achieved by DRL model and observed releases

the water availability in all reservoirs. During dry years, DRL model proposes increasing higher release ratios for SK and KNB dams and slightly lower ratios for BB and PS dams to moderate drought risk. DRL model suggests raising the reservoir release from BB, SK, and KNB dams during normal years. Additionally, higher release from BB dam is recommended to mitigate flood risk during wet years. These results signify the benefit of DRL modeling for a typically successive reservoir operation planning task in establishing the annual water allocation plan based on specific storage level targets identified as an example in this study. However, in the operational practice, these targeted levels of water storages in the multiple reservoir system of CPYRB can be dynamically adjusted by policy makers to suit with the water circumstances and perspectives on achieving a sustainable reservoir management of reservoir system.

Comparative analysis of DRL model with other optimization approaches in reservoir management for CPYRB

To explore capability of proposed DRL–DDPG based operation model for multiple reservoir operation planning, the comparative analysis comparing the simulated results of the DRL model for BB and SK dams with other state-of-the-art reservoir operation optimization techniques previously

Table 2 Summary of average reservoir releases of all dams simulated from 2010 to 2022

Dam	Annual			Dry season (Nov.–Apr.)			Wet season (May.–Oct.)		
	DRL release[1/]	OBS release[1/]	ΔDIFF[3/]	DRL release[2/]	OBS release[2/]	ΔDIFF[3/]	DRL release[2/]	OBS release[2/]	ΔDIFF[3/]
Unit	MCM	MCM	%	MCM	MCM	%	MCM	MCM	%
BB	4,318	4,195	+2.93%	1,850	2,711	– 31.75%	2,657	1,598	+66.28%
SK	4,744	4,886	– 2.91%	2,374	2,857	– 16.91%	2,553	2,184	+16.86%
KNB	1,292	1,290	+0.16%	475	578	– 17.92%	880	766	+14.84%
PS	1,864	2,060	– 9.51%	928	635	+46.01%	1,009	1,534	– 34.27%
Total	12,218	12,431	– 1.71%	5,627	6,782	– 17.03%	7,098	6,083	+16.69%

[1/] Evaluated using the data from 1/1/2010 to 31/12/2022

[2/] Evaluated using the data from 1/11/2009 to 31/12/2022

[3/] ΔDIFF–Percentage difference, OBS–Observed data

studied in CPYRB, was also conducted. Two main techniques are non-linear optimization programming (Kyaw et al. 2022) and Adaptive Neuro Fuzzy Inference System (ANFIS) applied for reservoir optimization in CPYRB (Kyaw et al. 2024). The optimization-based solution technique using non-linear programming solver (Fmincon function) was developed for BB and SK reservoir operation system aiming to address water scarcity in the region while flooding conditions due to the dam releases was constrained. Consequently, setting up the objective function for multi-reservoir operation model was referred to the minimization of the water scarcity indicating inability to satisfy the joint water demand in CPYRB. The hybrid neuro-fuzzy-based reservoir operation model for BB and SK dams was also developed by aiming to aid the reservoir operation system in alleviating water scarcity and moderating floods by optimizing operational rules using ANFIS technique. To formulate the ANFIS model structures, three main variables, namely reservoir inflow, reservoir water storage, and targeted water demand, were determined as input variables and current dam release was specified as the output variable. The model was trained and tested using 80% and 20% of dataset, respectively, by doing this, the optimal reservoir operational releases of BB and SK dams were solved and presented. As the local and joint water demand data, reservoir data, and reservoir system constraints used in these two previous studies are consistent with this study, consequently, a comparative analysis in reservoir management for CPYRB

Table 3 Reservoir release ratio accomplished by DRL model for short-term and long-term operations

Operation	Avg. release ratio	DRL	OBS
Short-term: DY[1/]	BB:SK:KNB:PS	0.3273:0.5165:0.0903:0.0659	0.3554:0.4908:0.0820: 0.0718
Short-term: NY[1/]	BB:SK:KNB:PS	0.3268:0.3754:0.1167:0.1812	0.3106:0.3602:0.1157:0.2135
Short-term: WY[1/]	BB:SK:KNB:PS	0.3894:0.3778:0.0927:0.1402	0.3116:0.3858:0.1191:0.1835
Long-term: LT[1/]	BB:SK:KNB:PS	0.3280:0.4163:0.1030:0.1528	0.3352:0.3991:0.1015:0.1643

[1/]DY–Dry Year, NY–Normal Year, WY–Wet Year, LT–Long-term Operation from 2010 to 2022

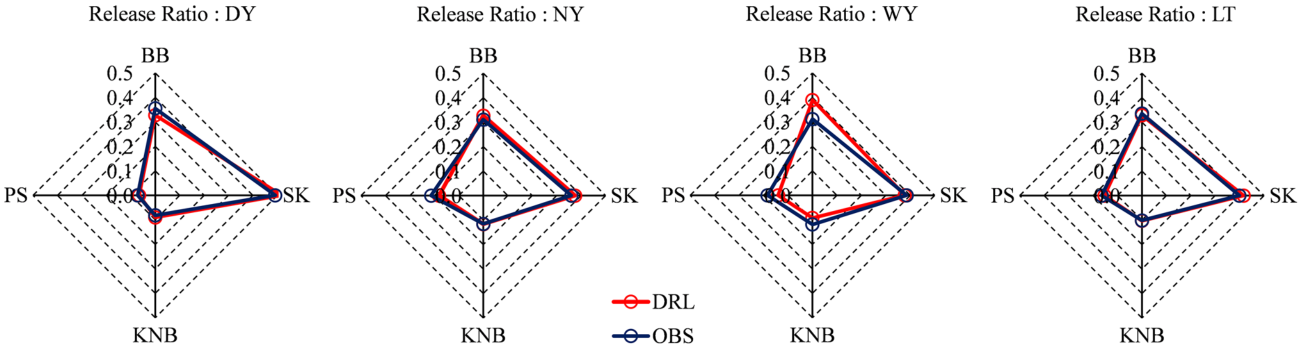


Fig. 6 Reservoir release ratio accomplished by DRL model. (Note: DY–dry year, NY–normal year, WY–wet year, LT–long-term operation, OBS–observed data)

could be conducted to compare the relative performances of these models. The comparison of simulated annual reservoir releases for BB and SK dams from 2010 to 2022, accomplished by the DRL, Fmincon, and ANFIS, models is shown in Fig. 7. These long-term simulations incorporate the impact of climate variability, including the severe flood of 2011 and consecutive prolonged droughts from 2016 to 2018.

Compared to actual operation and Fmincon and ANFIS optimization models, it is illustrated that DRL model generated the similar annual release patterns of two main storge dams with potentially different release volumes across various water conditions. The statistical correlations between the DRL model and actual operation, Fmincon, and ANFIS optimizations were 0.6744, 0.7525, and 0.6690, respectively. During the consecutive critical drought years apparently occurred from 2016 to 2018, the DRL model recommended to increase the release volume supplied from BB and SK dams significantly to moderate water scarcity in the basin. During the severe flood year in 2011, the DRL model attempted to deplete the reservoir water storage to align with the targeted storage lines resulting in substantially larger releases from these two dams compared to actual operation and two optimization models. Consequently, the smaller releases accomplished by DRL model was found in subsequent year, 2012 due to lower flood risks. However, the average long-term releases of BB and SK dams performed by DRL models are slightly different compared to actual operation, Fmincon, and ANFIS optimization techniques. It is estimated that the percentage difference in reservoir releases of DRL, Fmincon, and ANFIS are – 4.19%, + 2.58%, + 2.15%, respectively in comparison with the actual operation. Importantly, the DRL model demonstrated a significant advantage over these two optimization models in view of increasing the long-term water storages lying approximately + 15% for BB and SK dams as targeted in the DRL modelling process. Whereas, the Fmincon and ANFIS models could only increase water storages by + 12.48% (BB) and + 5.23% (SK) (Kyaw et al. 2022) and + 6.94% (BB) and + 1.62% (SK) (Kyaw et al. 2024), respectively. While the developed DRL model inherently incorporated release and storage constraints to regulate downstream flooding and maintain reservoir levels within safe limits, the continuous larger releases from two main storage dams particularly in 2011 led to the higher operational flood risk and damage compared to the other two optimization techniques. To address this limitation, it is advisable to include flood control information into the reward penalty function to improve the DRL model's capability.

DRL–DDPG hyperparameter tuning

For DRL–DDPG hyperparameter tuning, this study adopted a trial-and-error process and employed the optimal values obtained from recent research work (Tabas and Samadi 2024) as guideline to fine-tune optimal hyperparameter values as summarized in Table 4. It is emphasized that the reward discount factor (γ) of 0.90 is a substantial hyperparameter of DRL model that significantly controls the future rewards implication in the learning process of the model's agent. Additionally, a fine-tune value of changing rate for target network weight (τ) of 1×10^{-3} can help improve the stability of the training process to characterize the dynamic behaviors of reservoir operation system and targeted storage levels. The exploration noise (N) is specified to 0.30 encouraging agents to explore a broad range of release actions. The critic and actor learning rates (α) are set to 1×10^{-2} and 1×10^{-3}, respectively that influences the speed of learning process of model's agent. In this study, the initial weighted values (θ) are randomly initialized and updated using gradient value (∇) to guide the learning process of direction improvement. It is found that a key advantage of DRL–DDPG based

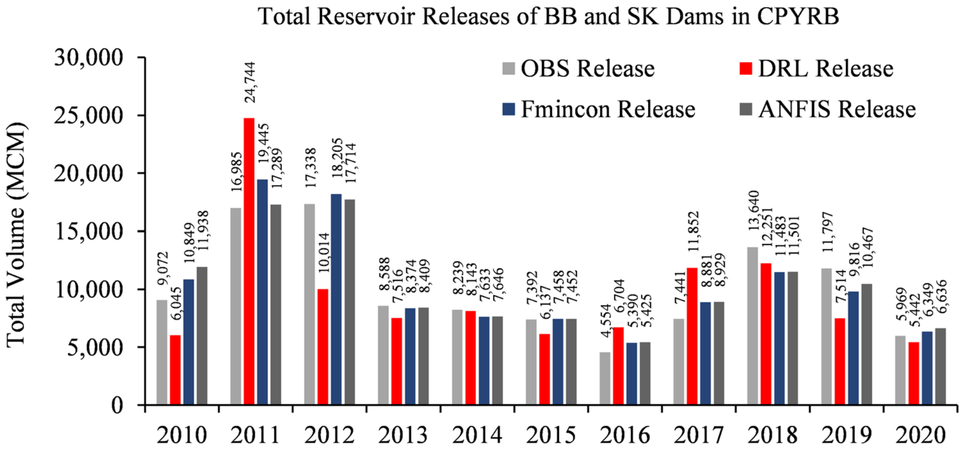


Fig. 7 Comparison of total reservoir releases of BB and SK dams accomplished by DRL–Fmincon–ANFIS models

operation model is its capability to significantly shorten computational time allowing to generate reservoir release schemes with a limited number of episode simulations. However, the speed and competence of the agent's convergence to that desired goal is significantly subject to assigned reward function.

Conclusions

The challenge in optimal operation of multiple reservoir system involves making complicated decisions on time-varying state variables like available water storage, future reservoir inflows, water demand requirements, existing release scheme of adjacent dam in the system, and downstream flow conditions. Consequently, establishing robust water allocation plan to ensure the efficient and sustainable system operation and mitigate flood and drought risks is definitely important for reservoir operation planning task. This study demonstrates application of DRL–DDPG-based algorithm to provide comprehensive and flexible plans for reservoir operation planning of the multiple reservoir system in CPYRB. The multi-agent-based DRL model is constructed considering 7-D predicted inflow, reservoir water released from adjacent reservoirs, downstream flow conditions, and changes in water storage, as state variables. The goal of multiple reservoir system operation is set by increasing 10–15% of water storage levels to all reservoirs in CPYRB and ensuring higher potential in supplying water for crop cultivation over the dry seasons. At the same time, these increased storage lines are determined not exceeding the normal pool levels to avoid flooding occurrences. In other words, the reservoir operators are assumed to set operational targeted goal to allow DRL–DDPG-based algorithm help recommend the proper seasonal and annual release schemes. Simulation results indicate that DRL–DDPG-based algorithm can perform well in solving sequential decision problems for optimal operation of multiple reservoir system to achieve the desired goal in CPYRB. It can provide reasonable and realistic simulated results in terms of seasonal and annual release schemes and reservoir release ratios among reservoirs in the system in comparison to observed operation and Fmincon and ANFIS optimization techniques. Importantly, it can shorten computational time significantly to gain reservoir release schemes with small number of episode simulation. However, the speed and competence of the agent's convergence to that desired goal is significantly subject to reward function design. For model utilization, reservoir planners can simply adjust targeted storage levels for each dam in the DRL model to meet desired goals in the multiple reservoir system by considering current and future water supply and demand conditions. The trade-off between flood control and drought mitigation measures to set up the optimal levels of targeted reservoir storages is definitely recommended to ensure the successful DRL-based reservoir operation. By considering this, the model enables the establishment of comprehensive and flexible water allocation plans and release guideline trajectory for sustainable reservoir operation planning in CPYRB.

Recommendation

The design of reward function is one of the critical aspects of DRL applications for optimal operation of reservoir systems. To fully capture the complex dynamics and trade-offs in reservoir operation, it is recommended to explicitly incorporate water deficit and flood control measures, and other relevant factors such as impact of reservoir releases on downstream ecosystems, cost of water supply, long-term sustainability of water resources, and intended hydropower production, etc. into the reward function for the achievement of specific objectives of reservoir optimization. In addition, to balance exploration and exploitation of the DRL model's agent exploring the new actions and reducing the undesired actions for determining reservoir releases, the magnitude of the penalty reward should be specified carefully. Moreover, this study utilizes long-term historical data from 2010 to 2022 for DRL-based simulation to incorporate the impact of climate variability like the severe flood of 2011 and prolonged droughts from 2016 to 2018. However, conducting scenario-based simulations to reflect the future climate variability and water demand conditions are needed for the further investigation of model's ability to adapt to changing circumstances. Furthermore, adjusting the dynamic targeted

Table 4 Optimal values of DRL–DDPG hyperparameters identified based on a trial-and-error process

Agent	Reward discount factor, γ	Changing rate for target network weight, τ	Exploration noise, N	Buffer size	Critic learning rate	Actor learning rate	Batch size
BB	0.90	1×10^{-3}	0.30	1×10^{6}	1×10^{-2}	1×10^{-3}	64
SK	0.90	1×10^{-3}	0.30	1×10^{6}	1×10^{-2}	1×10^{-3}	64
KNB	0.90	1×10^{-3}	0.30	1×10^{6}	1×10^{-2}	1×10^{-3}	64
PS	0.90	1×10^{-3}	0.30	1×10^{6}	1×10^{-2}	1×10^{-3}	64

levels of water storages in the multiple reservoir system of CPYRB is strongly recommended to suit with the water circumstances and perspectives on achieving a sustainable reservoir management and flood and drought risks mitigation.

Acknowledgements This research was funded by the Office of the National Research Council of Thailand (NRCT). The authors are grateful to the Royal Irrigation Department (RID), Electricity Generating Authority of Thailand (EGAT) for providing research data.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by A Rittima and Y Phankamolsil. The first draft of the manuscript was written and reviewed by A Rittima and Y Phankamolsil and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Mahidol University. National Research Council of Thailand (NRCT), SIP6230022.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Consent to participate The authors declare that they are aware and consent their participation in this paper.

Consent to publish The authors declare that they consent the publication of this paper.

References

Alturkistani H, El-Affendi MA (2022) Optimizing cybersecurity incident response decisions using deep reinforcement learning. Int J Electr Comput Eng Syst 12(6):6768–6776. https://doi.org/10.11591/ijece.v12i6.pp6768-6776

Andriotis CP, Papakonstantinou KG (2019) Managing engineering systems with large state and action spaces through deep reinforcement learning. Reliab Eng Syst Saf 191:106483. https://doi.org/10.1016/j.ress.2019.04.036

Belayneh A, Adamowski J, Khalil B (2016) Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustain Water Resour Manag 2(1):87–101. https://doi.org/10.1007/s40899-015-0040-5

Bhattacharya B, Lobbrecht AH, Solomatine DP (2003) Neural networks and reinforcement learning in control of water systems. J Water Resour Plan Manag 129(6):58–465. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:6(458)

Castelletti A, Corani G, Rizzoli AE, Soncini-Sessa R, Weber E (2001) A reinforcement learning approach for the operational management of a water system. Elsevier Yokohama J 2001:22–23

Castelletti A, Galelli S, Restelli M, Soncini-Sessa R (2010) Tree-based reinforcement learning for optimal water reservoir operation. Water Resour Res. https://doi.org/10.1029/2009WR008898

Dariane AB, Moradi AM (2016) Comparative analysis of evolving artificial neural network and reinforcement learning in stochastic optimization of multireservoir systems. Hydrolog Sci J 61(6):1141–1156. https://doi.org/10.1080/02626667.2014.986485

Dastour H, Hassan QK (2023) A machine-learning framework for modeling and predicting monthly streamflow time series. Hydrology 10(4):95. https://www.mdpi.com/2306%E2%80%935338/10/4/95

Fayaed SS, El-Shafie A, Jaafar O (2013) Reservoir-system simulation and optimization techniques. Stoch Environ Res Risk Assess 27(7):1751–1772. https://doi.org/10.1007/s00477-013-0711-4

François-Lavet V, Henderson P, Islam R, Bellemare MG, Pineau J (2018) An introduction to deep reinforcement learning. Found Trends Mach Le 11(3–4):219–354. https://doi.org/10.1561/2200000071

Hu R, Fang F, Pain CC, Navon IM (2019) Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. J Hydrol 575:911–920. https://doi.org/10.1016/j.jhydrol.2019.05.087

Hu H, Quan Z, Hu Q, Zhang Y (2022a) Reservoir optimal operation based on reinforcement learning. J Phys Conf Ser 2400:012039. https://doi.org/10.1088/1742-6596/2400/1/012039

Hu J, Wang Y, Pang Y, Liu Y (2022b) Optimal maintenance scheduling under uncertainties using linear programming-enhanced reinforcement learning. Eng Appl Artif Intell 109:104655. https://doi.org/10.1016/j.engappai.2021.104655

Hung F, Yang YCE (2021) Assessing adaptive irrigation impacts on water scarcity in nonstationary environments–A multi-agent reinforcement learning approach. Water Resour Res 57:e2020WR029262. https://doi.org/10.1029/2020WR029262

Jiang Q, Li J, Sun Y, Huang J, Zou R, Ma W, Guo H, Wang Z, Liu Y (2024) Deep-reinforcement-learning-based water diversion strategy. Environ Sci Ecotechnol 17:100298. https://doi.org/10.1016/j.ese.2023.100298

Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. J Artif Intell Res 4:237–285

Kraisangka J, Rittima A, Sawangphol W, Phankamolsil Y, Tabucanon AS, Talaluxmana Y, Vudhivanich V (2022) Application of machine learning in daily reservoir inflow prediction of the Bhumibol Dam, Thailand. 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI–CON), Prachuap Khiri Khan, Thailand 2022, pp. 1–4

Kyaw KM, Rittima A, Phankamolsil Y, Tabucanon AS, Sawangphol W, Kraisangka J, Talaluxmana Y, Vudhivanich V (2022) Optimization-based solution for reducing water scarcity in the Greater Chao Phraya River Basin, Thailand: through re-operating the Bhumibol and Sirikit reservoirs using non-linear programming solver. Eng J 26(10):39

Kyaw KM, Rittima A, Phankamolsil Y, Tabucanon AS, Sawangphol W, Kraisangka J, Talaluxmana Y, Vudhivanich V (2024) Re-operating the Bhumibol and Sirikit Dams using hybrid neuro-fuzzy technique to solve the water scarcity and flooding problems in the Chao Phraya River Basin. App Envi Res 46(1):009

Lai V, Huang YF, Koo CH, Ahmed AN, El-Shafie A (2022) A review of reservoir operation optimisations: from traditional models to metaheuristic algorithms. Arch Comput Methods Eng 29:3435–3457. https://doi.org/10.1007/s11831-021-09701-8

Lee J-H, Labadie JW (2007) Stochastic optimization of multireservoir systems via reinforcement learning. J Water Resour Res. https://doi.org/10.1029/2006WR005627

Madani K, Hooshyar M (2014) A game theory–reinforcement learning (GT–RL) method to develop optimal operation policies for multi-operator reservoir system. J Hydrol 519:732–742

Mahootchi M, Tizhoosh HR, Ponnambalam KP (2007) Reservoir operation optimization by reinforcement learning. J Water Manag Model. https://doi.org/10.14796/JWMM.R227-08

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D (2015) Human-level control through deep reinforcement learning. Nature 518:529–533

Molle F (2002) The closure of the Chao Phraya River Basin in Thailand: its causes, consequences and policy implications. Conference on Asian Irrigation in Transition-Responding to the Challenges Ahead, 22–23 April 2002 Workshop, Asian Institute of Technology, Bangkok, Thailand

Mullapudi A, Lewis MJ, Gruden CL, Kerkez B (2020) Deep reinforcement learning for the real time control of stormwater systems. Adv Water Resour. https://doi.org/10.1016/j.advwatres.2020.103600

Nguyen TT, Nguyen ND, Nahavandi S (2020) Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications. IEEE Trans Cybern 50(9):3826–3839. https://doi.org/10.1109/TCYB.2020.2977374

Oliveira R, Loucks DP (1997) Operating rules for multireservoir systems. Water Resour Res 33(4):839–852. https://doi.org/10.1029/96WR03745

Peacock ME, Labadie JW (2018) Deep reinforcement learning for optimal operation of multipurpose reservoir systems. 9th International Congress on Environmental Modelling and Software in Ft. Collins, Colorado, USA, June 24–28, 2018

Rieker JD, Labadie JW (2012) An intelligent agent for optimal river-reservoir system management. Water Resour Res. https://doi.org/10.1029/2012wr011958

Seifollahi-Aghmiuni S, Bozorg-Haddad O (2019) Simulation–optimization tool for multiattribute reservoir systems. J Hydrol Eng 24(9):04019028. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001817

Sumiea EHH, Abdulkadir SJ, Ragab MG, Al-Selwi SM, Fati SM, AlQushaibi A, Alhussian H (2023) Enhanced deep deterministic policy gradient algorithm using grey wolf optimizer for continuous control tasks. IEEE Access 11:139771–139784. https://doi.org/10.1109/ACCESS.2023.3341507

Tabas SS, Samadi V (2024) Fill-and-spill: deep reinforcement learning policy gradient methods for reservoir operation decision and control. J Water Resour Plan Manag 150(7):04024022. https://doi.org/10.1061/JWRMD5.WRENG-6089

Tabas SS (2020) Reinforcement learning policy gradient methods for reservoir operation management and control. Dissertation, Clemson University

Tounsi A, Temimi M, Gourley JJ (2022) On the use of machine learning to account for reservoir management rules and predict streamflow. Neural Comput Appl 34(21):18917–18931. https://doi.org/10.1007/s00521-022-07500-1

Verma M (2018) Artificial intelligence and its scope in different areas with special reference to the field of education. Int J Adv Educ 3(1):5–10

Wang X, Nair T, Li H, Wong YSR, Kelkar N (2020) Efficient reservoir management through deep reinforcement learning. AI for Earth Science Workshop at NeurIPS 2020.

Wenwu L, Mbanze D, Xueying Z (2018) Model dependent reinforcement learning algorithm for reservoir operation stochastic optimization. Int J Hydro 2(5):579–585. https://doi.org/10.15406/ijh.2018.02.00129

Wiering M, Van Otterlo M (2012) Reinforcement learning: state of the art. Springer. https://doi.org/10.1007/978-3-642-27645-3

Wu R, Wang R, Hao J, Wu Q, Wang P (2024) Multiobjective multihydropower reservoir operation optimization with transformer-based deep reinforcement learning. J Hydrol 632:130904

Xu W, Zhang X, Peng A, Liang Y (2020) Deep reinforcement learning for cascaded hydropower reservoirs considering inflow forecasts. Water Resour Manage 34(9):3003–3018. https://doi.org/10.1007/s11269-020-02600-w

Xu W, Meng F, Guo W, Li X, Fu G (2021) Deep reinforcement learning for optimal hydropower reservoir operation. J Water Resour Plan Manag 147(8):04021045. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001409

Yadav A, Minocha VK, Kumar R (2023) Optimizing reservoir operation with artificial intelligence: a state-of-the-art review. Int J Eng Technol Manage Sci 7(5):368–373

Zhang D, Lin J, Peng Q, Wang D, Yang T, Sorooshian S, Liu X, Zhuang J (2018) Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. J Hydrol 565:720–736. https://doi.org/10.1016/j.jhydrol.2018.08.050