Optimizing River Discharge Forecasts with Machine Learning for Japanese Public Dams Operation

Hascoet Tristan
Yoshimi, Keisuke
Dossa, Rousslan
Takiguchi, Tetsuya

(Citation)
国民経済雑誌,227(2):45-64

(Issue Date)
2023-02-10

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100479010

(URL)
https://hdl.handle.net/20.500.14094/0100479010

ISSN 0387-3129

国民経済雑誌

Optimizing River Discharge Forecasts with Machine Learning for Japanese Public Dams Operation

Tristan Hascoet
Keisuke Yoshimi
Rousslan Dossa
Tetsuya Takiguchi

The Kokumin-Keizai Zasshi（*Journal of Economics & Business Administration*）
Vol. 227, No. 2（February, 2023）

The Kokumin-Keizai Zasshi (*Journal of Economics & Business Administration*)
Vol. 227, No. 2 (February, 2023), 45-64

Optimizing River Discharge Forecasts with Machine Learning for Japanese Public Dams Operation

Tristan Hascoet[a]
Keisuke Yoshimi[b]
Rousslan Dossa[c]
Tetsuya Takiguchi[d]

Despite a climate and topology favorable to hydropower (HP) generation, HP only accounts for 4¥% of today's Japanese primary energy consumption mix. In recent years, calls for improving the efficiency of Japanese HP have emerged from prominent voices in the Ministry of Land, Infrastructure, Transport and Tourism. Among potential optimizations, data-driven dam operation policies using accurate river discharge forecasts have been advocated for. Meanwhile, Machine Learning (ML) has recently made important strides in hydrological modeling, with forecast accuracy improvements demonstrated on both precipitation nowcasting and river discharge prediction. We are interested in the convergence of these societal and technological contexts: our goal is to provide scientific evidence and actionable insights towards the implementation of more efficient dam operation policies using ML-based river discharge forecasts on a national scale. Towards this goal, this work presents a framework that simulates dam operation using uncertain river discharge forecasts. This framework aims to jointly quantify river discharge forecast accuracy, and the impact of forecast errors on dam operation efficiency. We conduct a preliminary study of ML-based discharge forecast on a dataset of 127 Japanese public dams we have assembled, and attempt to quantify the impact of different forecast error components on operation efficiency.

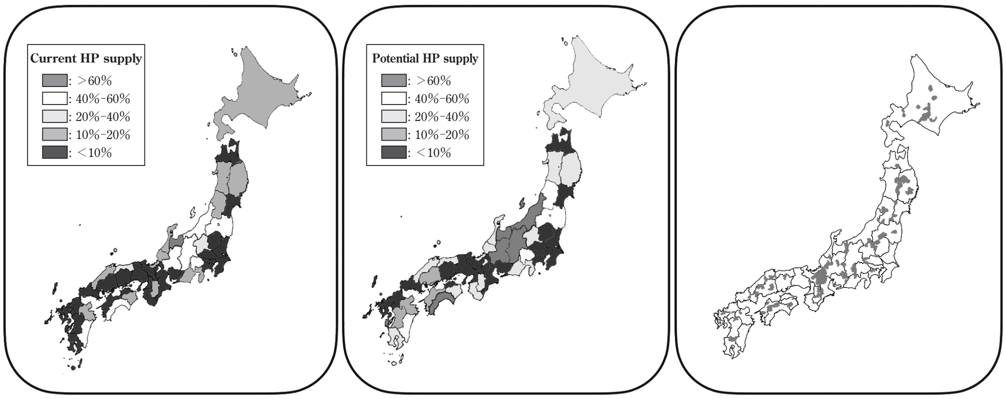
Keywords Machine Learning, Hydrology, River Discharge Forecast, Dam Operation

a Graduate School of Business Administration Kobe Univeristy, tristan@people.kobe-u.ac.jp
b Graduate School of System Informatics Kobe Univeristy, yoshimi@stu.kobe-u.ac.jp
c Graduate School of System Informatics Kobe Univeristy, doss@ai.cs.kobe-u.ac.jp
d Graduate School of System Informatics Kobe Univeristy, takigu@kobe-u.ac.jp

1 Introduction

Dam operation is a problem of control under uncertainty, in which dam operators aim to maximize multiple objectives (flood protection, HP generation, etc.) given uncertain forecasts of river discharge flowing into dam reservoirs. The more accurate river discharge forecasts are, the more efficiently (in terms of both flood protection and HP generation) dams can be operated. The uncertainty of discharge forecasts can be attributed to two main factors: Uncertainty in precipitation forecast (how much rain will fall) and uncertainty in hydrological modeling (how much of the fallen rain will flow into rivers). High levels of uncertainty and abundant alternative energy sources have lead Japanese public dam operators to adopt conservative operation strategies. However, three factors may come to challenge this status quo: First, social and environmental pressures on fossil fuels and nuclear energy production, combined with the rapid development of intermittent renewable energy sources, are foreseen to increase the value of HP generation. Second, climate change is expected to have a deep impact on Japanese surface water distribution (Synthesis Report, 2018), which is set to challenge current water supply operations. Third, improvement of both physical and statistical inference models are foreseen to increase the accuracy of forecast, allowing for better-informed dam operation policies. Combined, these three factors may come to challenge the current risk-benefit analysis towards more efficient operation policies leveraging accurate river forecast.

Figure 1



(Left) Current energy supply rates (Energy consumed / HP production) coming from HP (Center) Potential HP supply rate. Both figures were taken and translated from 角 et al. 2019 (Right): Illustration of our dataset's dam locations. Many collected dams are located in prefectures with high potential for optimization.

In this context, we aim to provide the scientific evidence and actionable insights for dam infrastructure managers and policy makers to implement energy-efficient and flood-resistant dam operation policies leveraging Machine Learning (ML)-based discharge forecast on a national scale. To do so, we have implemented the conceptual framework illustrated in Figure 2, which we detail in Section 3. This framework simulates Japanese public dam operations on a national scale so as to characterize the uncertainty of river discharge forecast, and to quantify the impact of these uncertainties on the efficiency of dam operation policies. Using this framework, we aim to provide answers the following questions:

- To what accuracy can ML models forecast river discharge?
  - What is the impact of different error components (e.g.; atmospheric or hydrological uncertainties)?
  - How does accuracy evolve with horizon times?
  - Can reliable uncertainty estimates on the forecasted discharge be achieved?
- How does forecast accuracy impact operation efficiency?
  - What is the impact of different error components?
  - On what time horizons is forecast accuracy the most impacting?
- What is the impact of dam dimensioning on operation efficiency?
  - In particular, what potential efficiency gains can be expected from dam heightening?

The answer to these questions represent the scientific evidence and actionable insights we aim to provide for dam infrastructure managers and policy makers. This work presents the results of our initial efforts towards achieving this goal, focused on analysing the accuracy of ML river discharge forecast, and on evaluating the impact of forecast accuracy on dam operation efficiency. On the river discharge forecast side, we report positive results, showing that advanced Deep Learning (DL) models tend to outperform both global hydrological models and linear baselines. We find uncertainty on precipitation observations and forecasts to be the most impacting factor for river discharge forecast accuracy, and that for high level of precipitation uncertainties, non-linear models perform on-par with linear models. In addition, we find that forecast accuracy of ML models improves when models are fitted to noisy precipitation forecast inputs, suggesting that these models can skillfully make use of the patterns of precipitation uncertainties to improve their predictions. On the problem of evaluating the impact of forecast accuracy on dam operation efficiency, we report a negative result: We propose a Reinforcement Learning (RL) formulation to the problem of dam operation, and find that our approach does not manage to deal with extreme weather events (e.g.; typhoons), which induces random variations in our results. These random variations are due to the difficulty of the model to handle rare statistical extremes, rather than the expression of a meaningful trend, which constitutes our negative result. Nevertheless, this experiment provides us with future research directions: we will focus on integrating explicit modeling of heavy rain events to our framework. Section 6 discusses the limitations of our current approach and lays out a path towards providing more definitive answers to the questions outlined above. The remainder of this paper is organised as follows: We present our proposed framework in Section 3. Section 4 focuses on the analysis of ML-driven discharge forecast, and Section 5 on evaluating the impact of forecast errors on dam operation efficiency. We start by presenting the dataset we have collected in Section 2, and the remainder of this section further motivates our study with additional context.

1. 1 Societal Context

A mountainous topology and a heavy rain climate lend Japan a high potential for HP generation. Historically, Japan has extensively relied on HP generation during the first half of the 20 th century, favoring HP over fire-based energy for its base load supply, a policy known as 「水主火徒」. As the post-war period of great economic development called for increased energy consumption, fossil fuel plants were preferred to HP for their ability to quickly and efficiently answer the rapid increase in demand. Later, the oil shock has seen Japan strategically develop nuclear power generation to ensure its energy independence. Due to longer infrastructure development times, HP lost its competitiveness in times of rapidly increasingly energy needs, so that its operating infrastructure has been comparatively little optimized (角 et al. 2019). Today, nuclear incidents and international pledges to reduce carbon emissions have come to threaten the long term viability of Japan's current energy mix. While solar and wind power generation are being intensively developed, their intermittent nature does not allow them to cover for the base load and demand response capacity provided by fossil fuel plants. In this context HP generation is seen as a valuable low-carbon alternative to fossil fuels for both base load and demand response needs to complement the development of intermittent renewable energy sources. For all its benefits, several voices from the MLIT have been advocating for a more efficient use of Japanese water resources towards HP generation (角 et al. 2019, 竹村 2016). Figure 1, drawn from a 2019 report on the state and future of Japanese HP (角 et al. 2019), shows the current rate of energy demand supplied by HP per prefecture, and contrasts it to potentially achievable supply rates, illustrating large potential gains. Among the potential optimization, the implementation of power efficient dam operation policies using accurate river discharge forecasts has been identified. Furthermore, climate change is expected to have a deep impact on surface water distribution in Japan (Synthesis Report, 2018), with impacting local disparities including faster snow melt in the northern and Japan sea regions, increased drought periods in the south, and increased flooding risks due to heavy rain events across the country. Both the destabilization of surface water distribution and the need for sustainable energy supply call for better forecasting abilities to optimize water resource management operations.

1. 2 Technological Context

Dam operation is the problem of satisfying two opposing objectives. On the one hand, HP production, crop irrigation, household and industrial consumption benefit from keeping high water volumes in dam reservoirs. On the other hand, flood control benefits from low water levels as empty reservoirs can better buffer strong river discharges so as to prevent flooding of downstream settlements. Optimal operation policies aim to maximize HP generation and other water-supplied services while minimizing the risk of flooding by keeping appropriate levels of water in the reservoir at all time. If water levels are kept high to maximize HP generation, operators must be able to preemptively lower water levels in order to accommodate for high river discharge following heavy rain events. Failing to do so comes with dire consequences: disastrous flooding of downstream urban settlements, and even possible dam failures. Thus, the implementation of efficient dam operation policies requires a precise knowledge of incoming discharges ahead of time, i.e.; accurate river discharge forecast. In the present study, we focus on river discharge forecast horizons of up to 3 days, which we estimated as the time needed for most small to medium size dams to preemptively empty their reservoirs so as to buffer heavy rain event discharges.

Historically, both atmospheric and hydrological modeling have been mostly addressed by physical simulation models. However, following the success of DL approaches across an increasing array of science and engineering fields, recent years have seen an increasing interest in applying DL methodology to river discharge modeling, hydrology and the earth sciences at large. In a series of recent works, ML-based river discharge models have been shown to outperform traditional methods on several benchmarks (Kratzert et al. 2018), with notable voices advocating for further development and wider applications of statistical approaches over physical models (Nearing et al. 2021). In the meantime, another line of work has shown ML-based precipitation nowcasting to outperform state-of-the art ensemble physical atmosphere simulations (Ravuri et al. 2021, Espeholt et al. 2021). These recent successes are emblematic of a wider trend that sees statistical approaches increasingly impacting the earth sciences, further exemplified by leading institutions integrating statistical approaches at the heart of their development strategy (ECMWF, 2021). Together, these trends beg the question of whether ML can provide river discharge forecast accurate enough to empower efficient dam operation policy implementations in Japan, as called for by prominent policy makers (角 et al. 2019, 竹村 2016). This work aims to layout the foundation to answer this question.

2 Dataset

We have assembled a dataset covering 127 public dams across Japan, whose locations are illustrated in Figure 1. For each dam, we have collected historical hourly reservoir inflow discharges provided by the MLIT, spanning from the year 1980 to 2020, as well as the dam dimensions (wall height, reservoir volume, control door average and potential discharge and HP turbine power). Atmospheric observations (precipitation, temperature, wind, etc.), and forecasts for the same period were collected from different sources, interpolated to each dam's drainage area (the area from which rain flows into the dam), and aligned to the in-situ river discharges observations.

As precipitation forecasts, we have collected historical data from different physical simulations provided by the Japanese Meteorological Association (JMA) on different spatio-temporal scales, including the Global Spectral Model (GSM), Meso-Scale Model (MSM) and Local Forecast Model (LFM). As atmospheric observations, we used assimilated data provided by the MSM model for precipitation, wind and temperature. Additionally, we have collected remote sensing precipitation estimates (Kubota et al. 2020), and in-situ precipitation measurements provided by the MLIT. We also collected snow melt data from the Today Earth simulations.

Table 1 summarizes the different kind of variables collected for each dam. This dataset will

Table 1: Summary of the variables in our dataset

Data Source	Variable	Type	Unit
MLIT	Discharge	In-Situ	m^3/s
JMA	Precipitation	Forecast	mm
MLIT	Precipitation	In-Situ	mm
GSMap	Precipitation	Remote Sensing	mm
JMA	Precipitation	Assimilated Model	mm
JMA	Temperature	Assimilated Model	degrees
JMA	Wind	Assimilated Model	m/s
TE [9]	Snow melt	Model	m^3/s

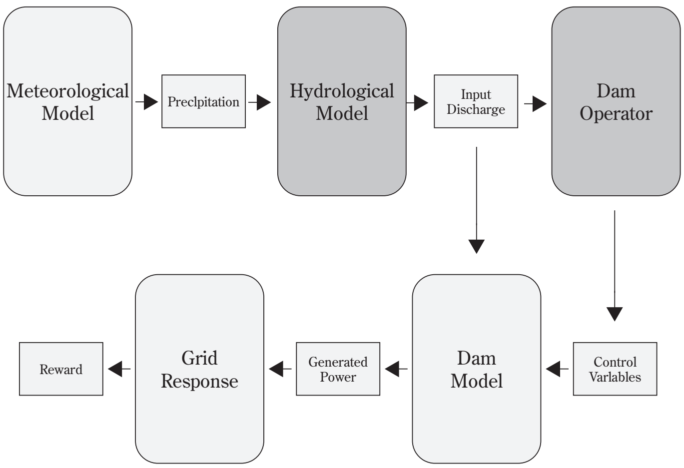
be released, along additional Japanese river discharge measurements, following existing hydrological standard in an upcoming paper. Beyond the present study, our goal is to foster future ML research efforts on improving discharge forecast of Japanese rivers by providing this data as a standard benchmark. Similar efforts have been undertaken to provide such benchmark in different countries (e.g.; Addor et al. 2017), and to homogenize datasets from different countries (Kratzert et al. 2022). We hope that this data will enable the inclusion of Japan to this international effort for future research to benefit Japan as well.

Our study is currently limited to using the JMA's GMS precipitation forecast; evaluating the impact of more local simulations (i.e.; MSM and LFM) is left for future work.

3 Framework

Figure 3 illustrates our proposed framework. Components for which we have optimized ML models are shown in darker color. The Meteorological Model component provides atmospheric forecast and observations including precipitation, surface wind and temperatures. These atmospheric variables are given as inputs to the Hydrological Model, which computes river discharge forecasts. River discharge forecasts are used by the Dam Operator component, which computes control variables acting on the Dam Model. The Dam Model keeps track and up-dates the state of the dam reservoir given incoming river discharges (in-situ observations provided by the MLIT), and the outflow discharge controlled by the operator. A reward is computed from the Dam Model operation: A positive reward is attributed by the Grid Response component, quantifying the benefit gained from HP, and a negative reward is attributed that quantifies the flood

Figure 2: Schematic illustration of our proposed framework.



risk resulting from dangerous operations. We further describe each component below:

Meteorological Model: In the present study, we use JMA's GSM model to provide atmospheric forecast. In some ablation study, we also simulate ideal forecast using actual observation data so as to isolate the impact of atmospheric forecast errors on the discharge forecast accuracy. In future work, we also plan to integrate both finer-grain JMA simulations (MSM, LFM) and recent Deep Learning based precipitation nowcasting models (Ravuri et al. 2021, Espeholt et al. 2021). Our implementation is modular so as to allow the integration of new models in the future.

The **Hydrological Model** implements a function

$$F_t = f(X_t) \tag{1}$$

$$f : \mathbb{R}^D \rightarrow \mathbb{R}^{72} \tag{2}$$

where X_t represents D-dimensional feature vector representing atmospheric forecast and observations provided by the Meteorological Model (cf. Figure 3), and $F_t \in \mathbb{R}^{72}$ represents hourly river discharge forecast for the following three days. We have experimented with different hydrological models including global hydrological modes, linear baselines non-linear ML models and recent hydrological DL models. Section 4 details our analysis.

Dam model: The dam model keeps track of water levels throughout the simulation. Our dam model features two doors: the flood control door releases a large volume o^F of the reservoir's water. The HP door releases a smaller volume o^{HP} which generates HP. Given a certain discharge d_t incoming at time t, the dam simulator updates the reservoir's current volume following the $v_t = v_{t-1} + d_t - o_t$ rule. Here, the amount of water released is denoted as o_t, and can be either o^F or o^{HP}, depending on which door the dam operator opens. We run hourly simulations, so discharge variables represent volumes exchanged during one hour of operation. A function g characterizing the reservoir geometry translates water volume to water height $h_t = g(v_t)$, which is used for the reward computation. We modeled dam reservoir geometry as a truncated cylinder, parameterized by their actual height and heuristically defined radius. When the HP door is activated by the operator, the amount of generated power P_t is proportional to the current water level: $P_t = C \times h_t$.

Dam Operator: The dam operator implements a control function

$$a_t = c(v_t, F_t) \tag{3}$$

$$c : \mathbb{R} \times \mathbb{R}^{72} \rightarrow \mathcal{A} \tag{4}$$

which outputs a discrete control variables $a_t \in \mathcal{A}$, given two inputs: the current dam reservoir state v_t and the forecasted river discharges F_t. $\mathcal{A} = \{a_{HP}, a_F, a_{idle}\}$ represents a set of three discrete actions: either open the HP door, the flood control door, or remain idle. In Section 5, we present our early attempt to learn an optimal control function c using RL.

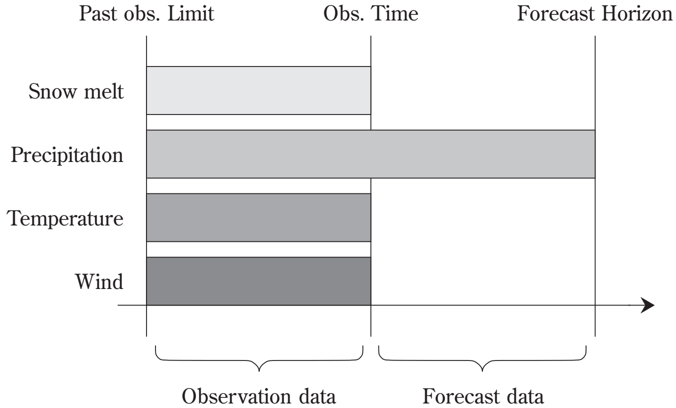
Operation Reward: At each step, a reward $r_t = r_t^{HP} + r_t^F$ is computed to quantify the efficiency of current dam operation. This reward is the sum of two components: A positive reward is given for HP generation. In this work, we use a constant reward for HP: $r_t^{HP} = K \times P_t$, with K a constant. In future work, we may implement a more complex reward system using a grid response simulation that would account for HP value fluctuations. The r_t^{flood} component represents a negative reward related to flood risk. In this work, we represent flood risk as a function of reservoir water levels, which can be seen as quantifying the risk of dam failures as water levels reach high values. In the future, we may instead quantify flood risk as cumulative outflow, which is the subject of regulations to prevent downstream flooding in practical operations.

4 River Discharge Forecast

In this section, we aim to maximize discharge forecast accuracy and to characterize forecast errors by quantitatively answering the questions listed below. To do so, we use a Machine Learning methodology: we fit possibly non-linear parameterized functions to regress future river discharge on a set of training data, and evaluate the ability of the fitted functions to estimate river discharge on a held-out set of test data. For each dam, we split our dataset into a training set ranging up to January 2018, a validation set consisting in data between January and December 2018 and a test set made of data from January to December 2019. The validation set is used to calibrate the model regularization to avoid overfitting the training data. We report our results in terms of forecast accuracy on the test set.

We use the in-situ measurements provided by the MLIT as river discharge ground-truth for both training (i.e.; fitting the parameterized function to the training set) and testing (evaluating forecast accuracy on the test set). Figure 3 illustrates the inputs X_t, at a given time step t, used to regress river discharges d_{t+T} at a given forecast horizon T. For a given horizon time T ranging from 1 to 72 hours ahead, X_t includes past hourly observations of different atmospheric variables over the past 85 hours, as well as forecasted precipitation data up to the target horizon. We use the Mean Squared Error (MSE) as a loss function to train the models. Models are functions f_θ parameterized by a set of parameters $\theta \in \Theta$. Training a model consists in fitting the parameterized function to the training data by minimizing the the loss function over the parameters on the training set Tr.

Figure 3: Schematic illustration of the input data X_t to our model



We use past observations of both atmospheric variables and discharge up to the current observation time t. In addition, we use forecasted precipitations up to the horizon time T.

$$\mathcal{L}(\theta)=\sum_{t\in T_r}(f_\theta(X_t)-d_{t+T})^2 \tag{5}$$

$$\theta^*=min_{\theta\in\Theta}\mathcal{L}(\theta) \tag{6}$$

We evaluate different classes of model f_θ: A linear model, a standard Multi-Layer Perceptron (MLP), gradient boosting models (LGB, XGB), and the deep learning model (LSTM) presented in (Kratzert et al. 2018). Each model is fitted and evaluated individually per dam and per horizon time. We report the average accuracy over the full test set (averaged across dams). As large dams show significantly higher river discharges, the average MSE across the dataset is dominated by the error on large dams. In order to report results that more evenly represent the average performace across dams, we report our results in terms of the Nash-Sutcliffe model efficiency coefficient (NSE), which is common practice in hydrology. Given an observed discharge d_{t+T} and a predicted discharge $f(X_t)$, NSE is defined as below:

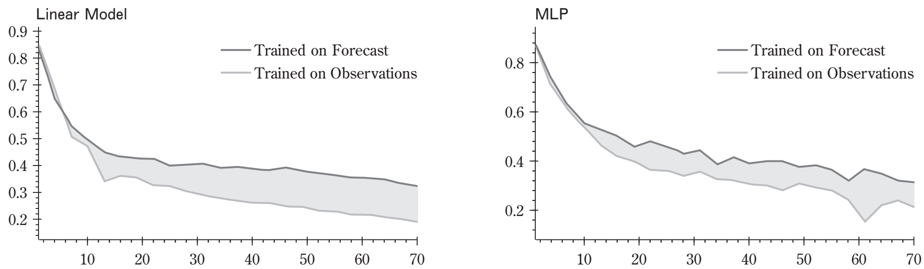
$$NSE(d_{t+T},\ f_\theta(X_t))=1-\frac{\Sigma_t(d_{t+T}-f_\theta(X_t))^2}{\Sigma_t(d_{t+T}-\hat{d})^2} \tag{7}$$

$$E(T)=\frac{1}{|Te|}\times\sum_{t\in Te}NSE(f_\theta(X_t)-d_{t+T}) \tag{8}$$

where $\hat{d}$ denotes the average observed discharge over the observation period. NSE can be roughly understood as the ratio of variance in future discharge explained by the model, with a score of 1 representing perfect forecast ability and a score of zero representing an accuracy

accuracy similar to a constant prediction corresponding to the observed average. To answer the below questions, we repeat the training and evaluation procedure described above, varying either the model f_θ, the variables X_t used as input, or the horizon time T.

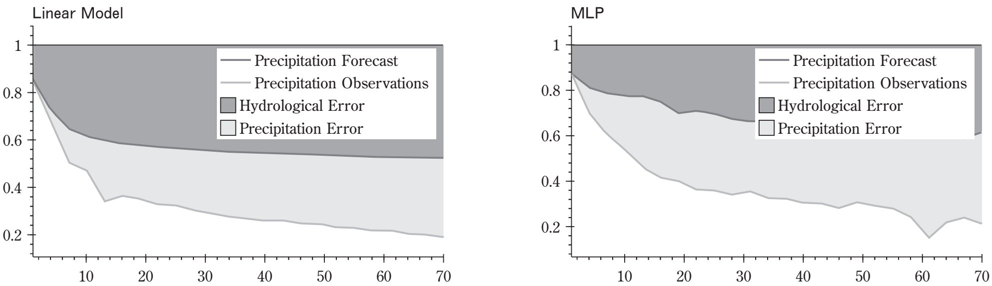
Figure 4



(a) Linear Model (b) MLP

Accuracy of models trained on either forecasted precipitations or observed precipitation (oracle forecast). The difference between both curves represent improvement brought by knowledge of the precipitation forecast uncertainties.

Figure 5



(a) Linear Model (b) MLP

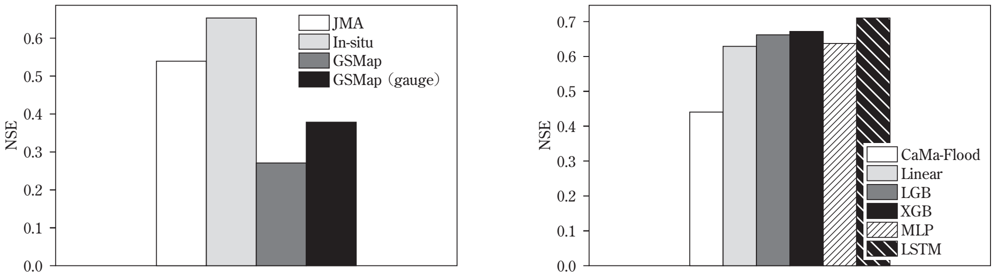
Accuracy of models trained and evaluated on either forecasted precipitations or observed precipitation (oracle forecast). The upper area can be seen as quantifying the hydrological error component, while the lower curve aims to to quantify the precipitation forecast component.

What practical forecast accuracy can be achieved on different forecast horizons? The upper curve in Figure 4 presents the evolution of accuracy with forecast horizons obtained by our best effort models in practical use-case situations: in-situ measurements of past discharge and atmospheric variables with GSM precipitation forecast were used as inputs to a linear and MLP model. A sharp decrease in accuracy can be observed within the first day time horizons. The below experiments further analyse the nature of these errors.

Are discharge forecast errors most impacted by precipitation forecast errors (how much rain will fall) or hydrological errors (how much of the fallen rain will flow into the

dam)? To answer this question, we ran the following experiment: In a first run, we train and evaluate models using oracle precipitation forecasts (i.e., we simulate perfectly accurate precipitation forecast using future in-situ observations of precipitations). In a second run, we train evaluate the learned models using the actual uncertain GSM forecast as precipitation forecast. This allows us to isolate discharge forecast error components caused by hydrological and precipitation forecast errors respectively: We consider the errors of the first run to stem from hydrological uncertainties and the difference in accuracy between both runs to be the impact of GSM forecast errors. Figure 5 shows the result of our experiments for two models, a linear model and a higher capacity MLP. The sharp increase in errors induced by short-term precipitation forecast motivates us to integrate higher precision short-term forecasts to maintain high accuracy on these horizon times. We are currently considering either JMA's physical simulations or precipitation nowcasting DL models.

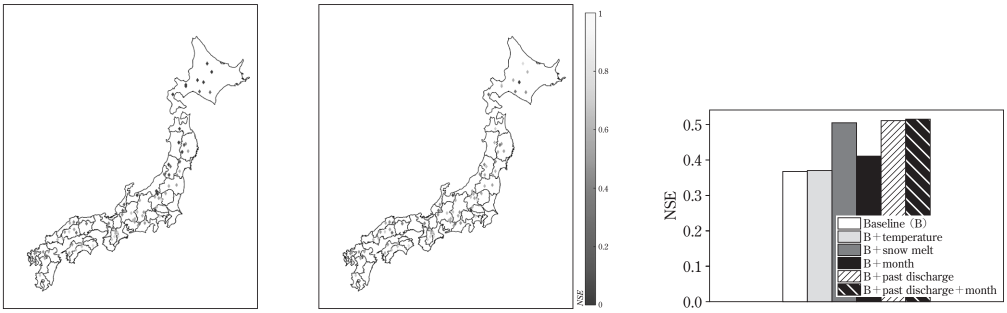
Figure 6: Impact of different modeling on discharge forecast accuracy



(a) Impact of precipitation data (b) Impact of hydrological model

What variables are most predictive of river discharge? Figure 6(a) highlights the importance of quality precipitation estimates, by showing results on a 24 hours horizon forecast using different precipitation observations. In-situ observations provide large improvements over both remote sensing estimates (GSMap) and assimilated model simulations (JMA). Temperature seems to provide little predictive power, as shown in Figure 7(c). Snow melt was also found to have an important impact for dams in the north and along the Japan sea. We study the impact of providing snow melt information (simulation data provided by the Today Earth service) as input to the model and compare forecast accuracy with and without snow melt data in Figure 7. Large accuracy gains can be observed in heavy snowfall regions (North and Japan sea). We also find that using past discharge observations with conditioning of the model on the current month allows to recover similar accuracy, which suggests that snow melt-induced discharges may be smooth enough to be estimated from past discharge observations and season-

Figure 7: Illustration of the impact of snow melt modeling on discharge forecast accuracy



(a) Without snow melt (b) With snow melt (c) Impact of additional data

ality only. It remains to be seen whether this strategy may work for longer horizon times.

What hydrological models are most accurate for river Japanese dams discharge modeling? We compare the accuracy of different ML models to that of a global hydrology model (Yamazaki et al. 2011) on a one day horizon forecast, and show that ML models tend to outperform the hydrology model. This may be due to ML model relying extensively on high-precision local data, while the global model does not. In addition, it can be seen that more expressive models outperform the baseline linear models. This trend was only observed for high precision precipitation estimates, while the gap between ML models and the linear baseline diminishes as the uncertainty in precipitation estimates increases. Indeed, Figure 4 and 5 both show that the MLP accuracy tends towards the linear baseline accuracy for longer horizon times when using uncertain forecasts.

Can ML models leverage knowledge of the uncertainty in precipitation forecasts to improve river discharge forecast? To answer this question, we run the following experiment: In a first run, we train ML models to forecast river discharge given oracle precipitation forecast. We both control overfitting and evaluate the accuracy of trained models on actual precipitation forecast. In that setting, the models are trained to model accurate hydrological phenomena, i.e., to output actual river discharges corresponding to actual precipitations. In a second run, we train models to regress river discharges on the GSM precipitation forecast instead of the oracle. This way, the model has knowledge of the given precipitation forecast error distribution. If the precipitation forecast shows systematic biases, the model shall thus fit these biases and reproduce them on the test set. Figure 5 shows the results of this experiment, with the lower curve representing oracle precipitation forecast training and the upper curve showing results for the GSM forecast training. Non-negligible improvements can be observed for the latter. At this point of the study, we have not yet elucidated the reasons behind this improvement. Future analysis will investigate if structural biases are indeed identified in the precipitation forecast and what these biases are. Nevertheless, knowledge of precipitation forecast errors seems to improve forecast ability. Another possible line of improvement would be to provide explicit precipitation error estimates as input to the model, using ensemble simulations provided by the JMA.

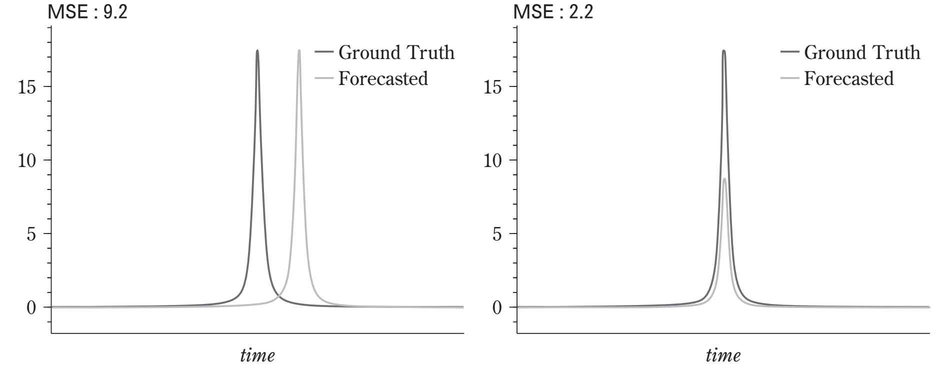
5 Impact on Dam Operation

In the previous section, we have analysed the impact of different modeling components on river discharge forecast errors. Maybe the most fundamental question remaining towards addressing our final goal is: how do forecast errors impact the efficiency of dam operations, and what forecast accuracy is required to enable data-driven energy-efficient policies? Answering these questions would provide precious insights: From an academic perspective, it would provide directions for future research on river discharge to focus on minimizing the most impactful error components. From a more practical perspective, it would provide actionable insights for dam operators to implement data-driven operation policies. In this section, we thus focus on quantifying the impact of different forecast error patterns on dam operation efficiency and propose a RL mehtodology to do so. Unfortunately, at this point of our study, difficulties in calibrating the model to deal with rare and extreme rain event prevent us from drawing tangible conclusions. Nevertheless, we present our methodology and discuss ways forward to address our current difficulties.

5.1 Motivation

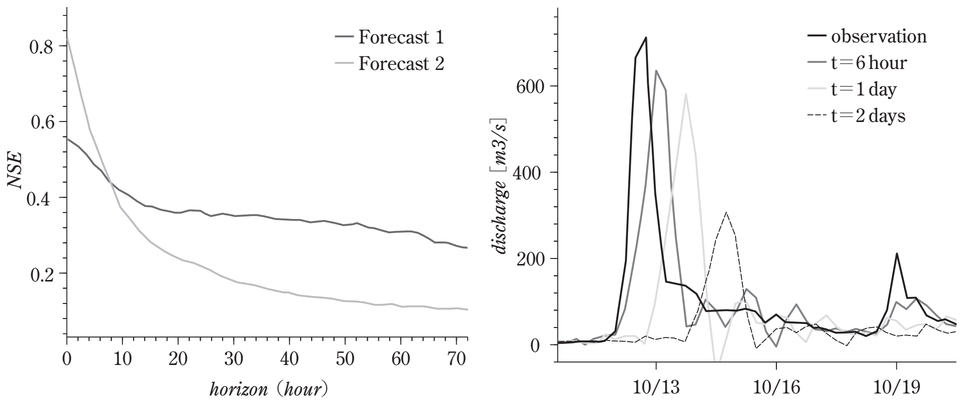
Evaluating the impact of forecast errors on dam operation is not as straightforward as it first seems. Indeed, forecast errors stem from a variety of causes: some errors are due to uncertainty in precipitation forecasts, while some errors come from inaccurate measurements of precipitations, or from hydrological errors caused by unexpected runoff coefficient (i.e.; how much of the fallen rain ends up flowing into the river). These different causes of uncertainty lend river discharge forecast errors a multidimensional structure, manifesting themselves in various forms. Figure 8 schematically illustrates two different error components: peak discharge time delay and peak discharge underestimates. Using a classical MSE metric, the time delay component seems the most severe. However, when it comes to dam operation, an operator given the delayed forecast would only empty the dam reservoir slightly too early, which would have little impact on the operation outcome. On the other hand, peak underestimate could lead the dam operator to not empty the reservoir enough preemptively, which could lead

Figure 8: Illustration of different forecast error components



(Left) Peak estimation delay (Right) Peak underestimate

Figure 9: Illustration of practical forecasts using our best effort model



(Left) Forecast accuracy decreases in time differently for different models. (Right) Illustration of discharge forecasted from different time horizons.

to a disastrous dam failure. This example illustrates the fact that straightforward accuracy metrics do not necessarily correlate with practical concerns of operation efficiency. But the complexity of estimating the impact of different error components does not end there. Preemptively keeping low reservoir levels due to an overestimation of forecasted peak discharge would not have the same impact during dry and wet seasons. During dry seasons, low precipitation levels would lead to long periods of low water levels, incurring more efficiency loss than during rainy seasons in which reservoirs would be filled faster.

Figure 9 illustrates another form of complexity in evaluating the quality of discharge forecasts. Forecasts are made for different time horizons, with accuracy typically decreasing with time horizon. Is a high short-term accuracy (as illustrated with the red curve in Figure 2

(Left)) more beneficial to dam operation efficiency, or is longer forecast horizon time accuracy (red curve) more impactful? In the face of this complexity, we propose a fully data-driven approach to forecast quality assessment. We propose to use the reward of an RL agent, trained to optimize dam operation efficiency given an uncertain river discharge forecast, as a metric assessing the value of this forecast to dam operation optimization. We compare the relative efficiency loss incurred by agents using uncertain forecasts relatively to that of an agent operating with perfect oracle forecast.

5. 2 Model and experiment

The dam operator is faced with a problem of control under uncertainty: it seeks an optimal operation policies, in terms of average reward, given uncertain discharge forecasts. In our case, the operator policy is implemented through the function c in Equation 3. We propose to address this problem through the framework of RL: we parameterize function c as a MLP, which we refer to as the RL agent. In RL terminology, the remaining modules of our proposed framework (Figure 2), constitutes the RL environment, a program that provides the RL agent with its inputs (the current dam state, and the discharge forecast) and its output (the reward resulting from its actions). The agent is trained to maximize the average reward $R=\sum_t r_t$ of his operation over the course of the evaluation period. We use the Deep Q-Network framework for training of the agent. For brevity, we omit the mathematical definition of this model and refer interested readers to the original paper for further details (Mnih et al. 2013).

Because the agent learns its policy from uncertain forecast, we expect the efficiency of the learned policy to reflect the forecast uncertainty: the higher the level of discharge forecast uncertainty is, the more conservative the learned policy should be so as to avoid high negative rewards stemming from unexpected future discharges, which would simultaneously decrease the positive rewards. Our goal is to quantify the impact of forecast errors on dam operation efficiency. We start by defining a baseline agent trained on perfect river discharge oracle forecast. Abusing notation for the sake of brevity, we denote the oracle forecast by $\hat{F}$ and denote by $R(\hat{F})$ the average reward of this agent over the course of the simulation. Given an uncertain river discharge forecasts F over the same period of time, we similarly train an RL agent and evaluate its average reward $R(F)$. The metric we use in our experiments to evaluate the efficiency loss incurred by the uncertainty in a given forecast F is the ratio:

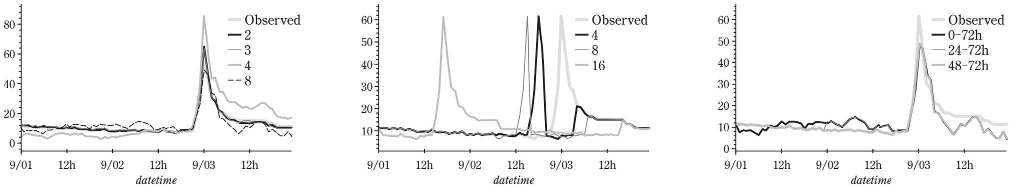
$$E(F)=\frac{R(F)}{R(\hat{F})} \tag{9}$$

The more impactful forecast error components are, the lower the above efficiency ratio is expected to tend. The least impactful forecast errors are, the closer to one this ratio should be. Our final goal is to evaluate the impact of practical forecast modeling designs on operation efficiency, but we start by calibrating our RL methodology on artificial forecast error patterns for which we know the expected trend. We simulate uncertain forecast by applying synthetic noise models to the oracle forecast. We use the following noise models:

Multiplicative Noise Model: We apply a multiplicative noise model to simulate peak underestimate and overestimate error components (illustrated in Figure 8 (Right)). Forecasts with different noise levels were generated by sampling a multiplicative noise coefficient for each time step of the simulation from a Gaussian distribution $\mathcal{N}(1, \sigma)$. We then smoothed the noise coefficient in time using a Gaussian filter with a time window of 48 hours, to simulate temporal consistency of forecast errors. Figure 10(a) illustrates the resulting artificial forecasts. We expect the agent efficiency loss to gradually decrease with the noise σ.

Peak Time Delay Simulation: We simulate discharge peak time estimation delays (illustrated in Figure 8 (Right)) by segmenting the 98th top percentile of river discharge. We randomly shift 10 hours windows of river discharge centered on these peaks with time delays drawn from a uniform distribution in the $[-N, N]$ hours range, for different values of N. We expect peak delays with relatively low amplitude (within the one day range) to have little effect on the efficiency as the model should learn to anticipate these delays.

Figure 10: Illustration of forecasts generated with different noise models



(a) Multiplicative Noise Model for different σ

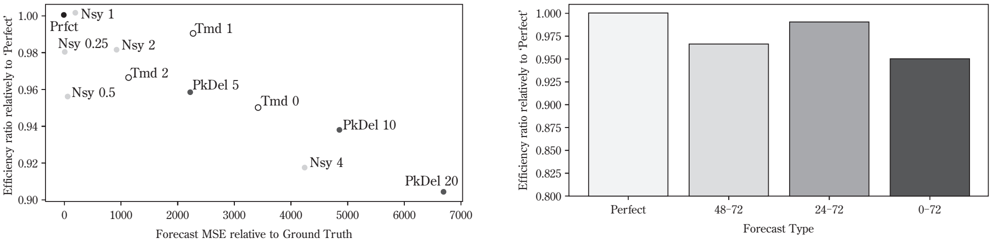
(b) Peak Time Delay Model for different N

(c) Time-shifted Multiplicative Model using different horizons

Time-shifted Multiplicative Noise Model: Finally, to measure the impact of noise on different forecast horizon, we selectively apply the multiplicative noise model to gradually closer time horizons, as illustrated in Figure 10(c). We expect efficiency to decrease as we apply the noise to shorter-term horizons.

We train one DQN agent to maximize the expected reward on each of the noisy forecast illustrated in Figure 10, and evaluate the efficiency of the learned policy in terms of the ratio E de-

Figure 11: Efficiency ratios og DQN agent trained with different forecast error components



(a) Relationship between efficiency loss and MSE for different noise models

(b) Efficiency loss evolution with the Time-shifted Multiplicative noise Model applied on different forecast horizons

fined in Equation 10. Due to heavy computational loads, in this preliminary study, we only evaluate the policy for a single dam. A DQN agent is trained over four different seeds for each of the forecast type, i.e., results are averaged over four runs with different random initial conditions. We discuss the result of our experiment, shown in Figure 11, through the lense of the following questions we aim to answer:

How do forecast errors at different time horizon impact operation efficiency? In Figure 10(b), noise is applied on gradually shorter-term horizons. The x axis represents the time horizon (in hours) on which the noise was applied. As we apply noise on shorter forecast horizons, the overall trend confirms that efficiency decreases, stressing the importance of short-term forecast accuracy. However, one outlier stands out when applying noise on the 24th to 72 nd-hour horizon of the forecast. This highlights the random variation we still have in our results, despite our best efforts to calibrate the model, which we further describe below.

How does dam operation efficiency correlate with standard metrics for different error components? Overall, dam operation efficiency tends to follow the MSE trends. The results of the multiplicative noise model (shown as Nsy in Figure 11(a)) suggest that deviation from the overall trend may be due to random variations in our experiments rather than meaningful phenomena.

How do different error components impact dam operation efficiency? The results presented in Figure 11(a) results contradict our expectations: Surprisingly, we found peak estimation delays to have an important impact on dam efficiency, while we expected peak discharge estimation errors (as modeled by the multiplicative noise model Nsy) to be the most impactful. The Multiplicative noise model results show important random variations, with higher noise level forecasts occasionally reaching higher efficiency than lower ones (e.g., $\sigma=2$ efficiency is higher than $\sigma=0.5$). This is especially notable for the results of the multiplicative noise model with $\sigma = 1$, for which the agent efficiency outperforms that of the agent using the oracle forecast despite non-negligible noise levels. Similar random variations can be observed for the time-shifted multiplicative noise models, as previously mentioned.

Upon closer inspection, we found that the errors causing these random variations concentrate on a few rare extreme precipitation events, for which the agent does not manage to preemptively empty the reservoir enough. Despite our best efforts to calibrate the model, we have not yet managed to stabilize these variations, which prevents us from drawing tangible conclusions for the moment. Taming these random variations by better addressing these rare extreme events will be the topic of future work.

6 Limitations, Future Work and Conclusion

The current work has presented our preliminary efforts in applying ML to optimize and evaluate the impact of river discharge forecast on Japanese public dam operation efficiency. Our initial efforts have been focused on collecting the data, proposing a methodology, and analysing forecast errors. Despite encouraging first results, much remains to be done towards our final goal.

In the short term, the instability of the dam operator model will be addressed, either by using alternative formulations of the agent state, or by introducing more structured modeling that explicitly addresses extreme rain events. We may also integrate JMA's tropical typhoon information tracking data to our modeling for explicit heavy-rain modeling. We will keep optimizing river discharge forecast by both discharge model improvements and integration of finer-grained precipitation forecast.

We plan to open our data and simulation code with the following goals: to foster future ML research efforts on improving discharge forecast of Japanese rivers by providing a standard benchmark, and to provide a foundation into which improved precipitation and discharge forecast models derived by future research can be integrated so as to analyze the impact of their improvements on Japanese public dam operation efficiency. Despite its current limitations, we believe that our framework can bring valuable insights to assist dam infrastructure and policy maker in water resource management. Beyond the questions addressed in this study, our framework may also be used to study the impact of dam parameter on their operation. In particular, dam heightening, a process in which the dam wall height is elevated to increase reservoir capacity has been advocated for to increase operation efficiency of some Japanese dams (角 et al. 2019). Once operational, our framework would allow to quantify the potential benefits of such operation by experimenting with different reservoir geometry parameters.

This work was supported in part by JSPS KAKENHI (Grant No. JP20K19823).

References

Addor, Nans, et al. “The CAMELS data set: catchment attributes and meteorology for large-sample studies.” Hydrology and Earth System Sciences 21.10 (2017): 5293-5313.

ECMWF STRATEGY 2021-2030 report, ECMWF (2021)

Espeholt, Lasse, et al. “Skillful twelve hour precipitation forecasts using large context neural networks.” arXiv preprint arXiv: 2111.07470 (2021).

Kratzert, Frederik, et al. “Rainfall-runoff modelling using long short-term memory (LSTM) networks.” Hydrology and Earth System Sciences 22.11 (2018): 6005-6022.

Kratzert, F., Nearing, et al. (2022). Caravan-A global community dataset for large-sample hydrology.

Kubota, Takuji, et al. “Global Satellite Mapping of Precipitation (GSMaP) products in the GPM era.” Satellite precipitation measurement. Springer, Cham (2020): 355-373.

Mnih, Volodymyr, et al. “Playing Atari with Deep Reinforcement Learning”, arXiv preprint arXiv: 1312.5602186 (2013).

Nearing, Grey S., et al. “What role does hydrological science play in the age of machine learning?” Water Resources Research 57.3 (2021): e2020WR028091.

Ravuri, Suman, et al. “Skilful precipitation nowcasting using deep generative models of radar.” Nature 597.7878 (2021): 672-677.

Synthesis Report on Observations, Projections and Impact Assessments of Climate Change, 2018, Ministry of the Environment Ministry of Education, Culture, Sports, Science and Technology Ministry of Agriculture, Forestry and Fisheries Ministry of Land, Infrastructure, Transport and Tourism Japan Meteorological Agency.

Today’s Earth, https://www.eorc.jaxa.jp/water/ (accessed 2022-08-01).

Yamazaki, Dai, et al. “A physically based description of floodplain inundation dynamics in a global river routing model.” Water Resources Research 47.4 (2011).

国土文化研究所編，角哲也，井上素行，池田駿介，上阪恒雄（監修）『今こそ問う 水力発電の価値—その恵みを未来に生かすために』技報堂出版，2019年

竹村公太郎 『水力発電が日本を救う—今あるダムで年間 2 兆円超の電力を増やせる』東洋経済新報社，2016年