RESEARCH ON DAM INFLOW PREDICTION DURING SEVERE FLOOD USING MACHINE LEARNING METHODS

MAKOTO NAKATSUGAWA
Muroran Institute of Technology, Muroran, Japan, mnakatsu@mmm.muroran-it.ac.jp

RIKO SAKAMOTO
Muroran Institute of Technology, Muroran, Japan, 18041031@mmm.muroran-it.ac.jp

YOSUKE KOBAYASHI
Muroran Institute of Technology, Muroran, Japan, ykobayashi@csse.muroran-it.ac.jp

ABSTRACT

We examined machine learning methods to determine which one would be optimal for predicting inflow during severe flood to a dam and reservoir water level. Predictions using machine learning were done for Kanayama Dam, at the upper reaches of the Sorachi River, which is a tributary of the Ishikari River, and for Satsunaigawa Dam, at the upper reaches of the Satsunai River, which is a tributary of the Tokachi River. The predictions were done by using hydrological information for the basins of these two rivers collected at the heavy rainfall disaster of August 2016. As machine learning based prediction methods, RF (Random Forest), FCNN (Fully Connected Neural Network), RNN (Recurrent Neural Network) and regression analysis by Elastic Net, which is a sort of sparse modeling, were examined. As a result, FCNN and Elastic Net demonstrated roughly the same accuracy, with Nash-Sutcliffe (NS) coefficients of 0.7 or greater. With Elastic Net, for cases other than those whose predicted rainfall had indeterminacy, the results were the most accurate and then the NS coefficients were 0.7 or greater. We are sure that obtained results give a promise to improve dam operation for disaster mitigation due to flood.

Keywords: dam inflow prediction, severe flood, machine learning, Elastic Net, Neural Network

1. INTRODUCTION

In recent years, heavy rainfall disasters have occurred throughout Japan. In October 2019, heavy rainfall caused by a huge typhoon caused severe disasters in eastern Japan, and in July 2018, heavy rainfall disasters were caused by rain fronts in western Japan. In Hokkaido, four typhoons approached or passed over area successively in August 2016, causing severe flood damage. During these heavy rainfalls, it was estimated that Kanayama Dam and Satsunaigawa Dam would exceed their flood control capacities, and “disaster prevention operation during an abnormal flood”, in which discharging an amount of water equal to the inflow, was implemented at both dams. Such unusual dam operations have been increasing in recent years because of heavy rainfalls that are considered to be a consequence of climate change. Under such circumstances, there is desire to enhance the functions of dams, and studies are under way on enhancing flood control functionality by performing discharge in advance based on the predicted inflow.

Currently, prediction method mainly employed by dam operators is an inflow calculation method using a model that physically and conceptually describes runoff phenomenon, such as a storage function model that uses predicted rainfall as an input condition. However, results depend on accuracy of model parameters and predicted rainfall, and in many cases, good results cannot be obtained. In recent years, prediction methods using machine learning algorithms have been proposed. Sunohara et al. (2006) estimated the inflow at Shiroyama Dam in Kanagawa Prefecture by using a neural network (NN). As a result, it was shown that predictions with up to three hours lead time were possible. Tamura et al. (2018) predicted dam inflow with 48 hours lead time by applying a deep neural network (DNN), which has two or more intermediate layers in the NN, and they used the rainfall data from Hitokura Dam in Hyogo Prefecture. These results show that if predicted rainfall is accurate, it is possible to predict the inflow with accuracy sufficient for practical use. However, when using an NN, it is necessary to prepare a highly accurate predicted rainfall. There are many issues in using NN, including that relationship between cause and result (prediction) is unknown and that it is necessary to invest a great deal of effort in optimizing the many parameters that govern results of NN.

The authors (Sakamoto et al. (2019)) compared the prediction accuracy of several machine learning methods and demonstrated that elastic net (EN) algorithm, a commonly used method of sparse modeling, is useful for dam inflow prediction. Sparse modeling can reduce a large amount of complicated data into a model with an easy-to-understand causal relationship, such as a regression equation. In addition, sparse modeling is distinguished by its ability to analyze the given data by eliminating unnecessary variables which cause over-learning and to complement the missing (sparse) information. In recent years, attention has been paid to successful imaging of black holes based on telescope data from in various parts of world (The Event Horizon Telescope Collaboration, (2019)). This is an application example of sparse modeling. There have been many studies using sparse modeling from 2013 to 2017 under a Japanese national project to promote the use of such modeling in various fields.

The current study focuses on advantages of above-mentioned EN algorithms and aims to apply them to dam inflow prediction. The authors examined practical application of this method by investigating, in particular, time lag between peak of predict value and the peak of observed value and by comparing prediction accuracy of 6-hour cumulative value to that of 1-hour value, which were issues identified in our previous study (Sakamoto et al. (2019)).

2. METHODS

2.1 Target locations and cases

Our study subjects are Kanayama Dam and Satsunaigawa Dam, managed by the Hokkaido Regional Development Bureau of the MLIT (Ministry of Land, Infrastructure, Transport and Tourism). This paper predicts dam inflow based on two types of machine learning algorithms that were calculated for floods that occurred from August to September 2016, and the results are assessed. Figure 1 shows the catchment basin of each dam and basic data on the dams are shown in Table 1.

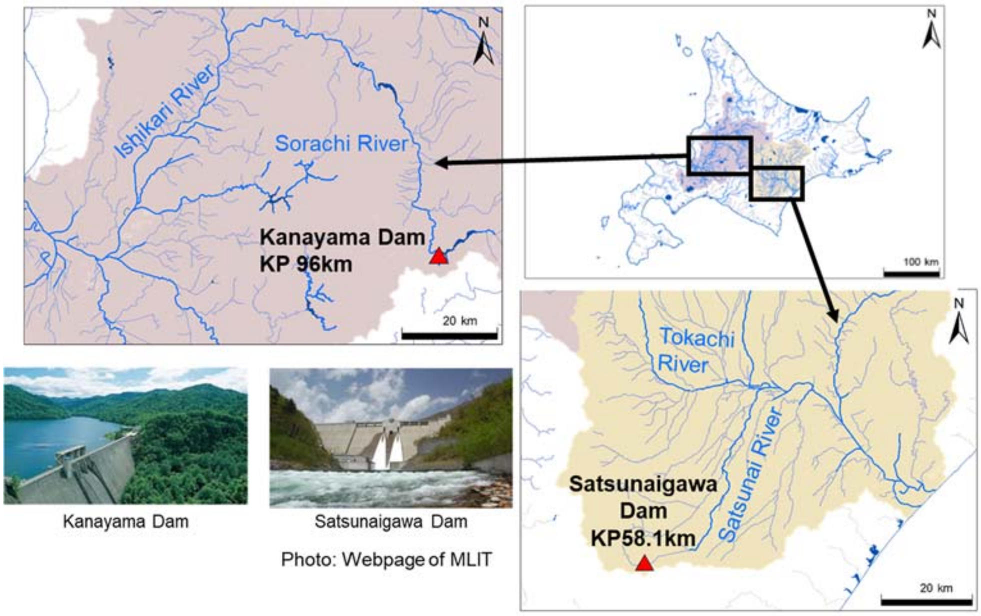

Figure 1. Map of dam locations.

Table 1. Various data of the dams.

	Kanayama Dam	Satsunaigawa Dam
Dam height (m)	57.3	114.0
Crest length (m)	288.5	300.0
Gross capacity of reservoir (m^3)	150,450,000	54,000,000
Active storage capacity (m^3)	130,420,000	42,000,000
Highest reservoir level at flooding (m)	345.0	484.0
Water level for starting disaster prevention operation for abnormal floods (m)	343.7	474.0
Flood storage preparation level (m)	338.5	466.0
Lowest water level (m)	320.0	447.5

2.1.1 Kanayama Dam

Kanayama Dam is a hollow gravity dam at the upper reaches of the Sorachi River in the Ishikari River System. The maximum inflow at the flood of August 2016 was 1,556 m3/s, which exceeded the design flow of 1,000

m^3/s. The water level at the start of disaster prevention operation during an abnormal flood is elevation level (EL) 343.7 m. Since this water level was exceeded at that time, disaster-prevention operations for an abnormal flood were conducted.

2.1.2 Satsunaigawa Dam

Satsunaigawa Dam is a concrete gravity dam at the upper reaches of the Satsunai River in the Tokachi River System. The maximum inflow at the flood of August 2016 was 714 m^3/s, which exceeded the design flow of 700 m^3/s. The water level at the start of disaster prevention operation during an abnormal flood is EL 474.0m. Since this water level was exceeded at that time, disaster-prevention operations for an abnormal flood were conducted.

2.2 Objective variable and explanatory variables used for prediction

Hourly inflow prediction is important for on-site dam management operations. However, this prediction is difficult, because the prediction accuracy for rainfall that changes greatly every hour is low. Accordingly, we used the inflow of 6-hour accumulated value as an objective variable. In this paper, we started by calculating an inflow prediction model that predicts hourly 6-hour accumulated values. Finally, we predict hourly inflow values using a machine learning algorithm with highly accurate results of 6-hour accumulated value prediction.

The explanatory variables (i.e., the variables which are input in the model) shown in Table 2 are the hourly values for inflow and soil water index, and the other rainfall data are 6-hour accumulated values. As values for these variables, we used the average rainfall for the basin and the point rainfall of each dam obtained from the Hydrology and Water Quality Database of the MLIT. The rainfall for each location is the rainfall observed at each dam site. For the assumed predicted rainfall, the measured average rainfall for the catchment basin is used. The soil water index is an index that reflects the accumulated effect of rainfall on the soil moisture condition, and it is calculated by using the average rainfall for the basin; therefore, hourly values are used. We collected these data from the Hydrology and Water Quality Database. The data collection period differs between the dams for all types of data collected. The observation data for Kanayama Dam were collected from January 2007 to December 31, 2018, and the observation data f Satsunaigawa Dam was collected from June 2002 to December 31, 2018.

Figure 2 shows the hydrographs of hourly 6-hour accumulated inflow for both dams. Red curves show the training data for the prediction model from the first day of data acquisition until December 31, 2012. Blue curves show the validation data, from January 1, 2013 to December 31, 2015, used for optimizing the hyper parameters that determine the model architecture of the learning algorithms. Green curves show the open test data, from January 1, 2016 to December 31, 2018, which were not involved in the learning process of the model and were prepared for use as “unknown” test data. The test data case for August 2016, which had the greatest observed inflow during the subject period, is reported in this paper. We conducted a prediction calculation for two cases of prediction lead time (LT): 6-hour LT and 12-hour LT.

Table 2. Data for the explanatory variables.

Observed items	Data description	Number of explanatory variables
Amount of inflow (m^3)	Time value of *t*	7
Average rainfall of the basin (mm)	6-hour cumulative value for the period from *t*-6 to *t*	7
Point rainfall (mm)	6-hour cumulative value for the period from *t*-6 to *t*	7
Radar rainfall (mm)	6-hour cumulative value for the period from *t*-6 to *t*	7
Assumed predicted rainfall (mm)	6-hour cumulative value for the period from *t* to *t*+6	7
Soil water index (mm)	Time value for the period from *t*-6 to *t*	7

t indicates current time.

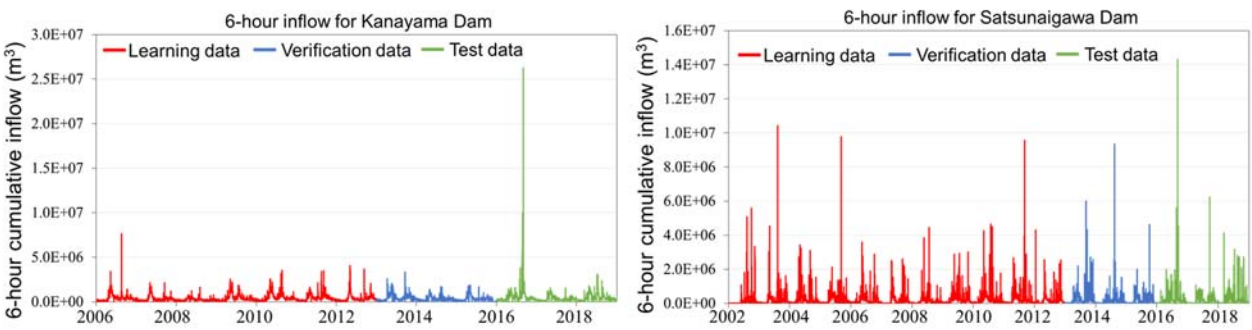

Figure 2. Hydrograph for 6-hour cumulative inflow.

2.3 Methods used for the prediction

This study uses the EN algorithm (Hui, Z et al. (2005)) for dam inflow prediction. The EN algorithm is a common method of sparse modeling. It is a multi-regression analysis method that uses regularization. Unlike conventional multiple-regression equations, the equation we use is distinguished by its ability to give zero as the weight (coefficient) of unnecessary explanatory variables. $J(w)$, which is the cost function of the EN algorithm, is expressed by the following equation, which uses the weight w of the regression equation.

$$J(w) = \sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2 + (1-\alpha)\lambda_1 \sum_{j=1}^{m} w_j^2 + \alpha\lambda_2 \sum_{j=1}^{m}|w_j| \tag{1}$$

where, $y^{(i)}$ is the observed value, and $\hat{y}^{(i)}$ is the predicted value obtained from the regression equation. The first term is the square error. The second term, which is a Ridge regression called L2 Penalty, is the restricting condition that controls the number of variables to be selected. The third term, which is a Lasso regression called L1 Penalty, is the restricting condition that is intended to reduce the number of explanatory variables which have weights other than 0. The EN algorithm is a method for obtaining a good regression equation from a small (sparse) number of observed values (sparse values) by optimizing the two penalty weights, α, as hyper parameters. Unlike in DNN and similar methods, the number of parameters in the model is small, and the completed model is characterized by high interpretability because it is a linear combination of explanatory variables. In this study, we use α, which is a value optimized by using the validation data. The EL algorithm is able to take only one objective variable, so we generated dam inflow prediction models for each LT.

For the comparison, the DNN algorithm used in a previous study (Tamura et al. (2018)) was used. The DNN algorithm is a method that is typical of artificial neural networks and that is characterized by having all the nodes of a layer closely connected to each node of the next layer. In this paper, we based our optimization on the parameters of Tamura et al. (2018). The batch sizes were 250, 500, and 1000. The middle layers were 2, 3, and 4. The node numbers were 125, 250, and 500. The dropout rates were 0, 0.3, and 0.5. Each hyper parameter was optimized within this range. In addition, the DNN algorithm is able to predict multiple objective variables, so we trained a model that simultaneously outputs 6-hour LT and 12-hour LT. These hyper parameters have values that are optimized in greater detail than in our previous work (Sakamoto et al. (2019)). As mentioned in the introduction, the DNN algorithm has been used in a number of studies. In contrast, the EN algorithm is not used in dam or river inflow prediction even though it enables a practical linear combination formula with small parameters to be obtained.

3. RESULTS

For the accuracy indexes of prediction inflow, relative error of peak inflow (J_{pe}) and Nash-Sutcliffe (NS) coefficient are used, as shown below.

$$J_{pe} = \frac{V_{Q_{inop}} - V_{Q_{incp}}}{V_{Q_{incp}}} \tag{2}$$

$$NS = 1 - \frac{\sum(V_{oi} - V_{ci})^2}{\sum(V_{oi} - V_{oave})^2} \tag{3}$$

where, $V_{Q_{inop}}$ is the peak 6-hour accumulated inflow observed during the period (m^3), $V_{Q_{incp}}$ is the peak 6-hour accumulated inflow predicted for the period (m^3), V_{oi} is the 6-hour accumulated observed inflow (m^3), V_{ci} is the 6-hour accumulated predicted inflow (m^3), and V_{oave} is the average 6-hour accumulated observed inflow (m^3). In this calculation, negative values for J_{pe} mean that the prediction is on the safe side. The closer the NS coefficient value is to 1.0, the more accurate is the model. The accuracy of the model when the NS coefficient is 0.7 or higher is considered to be high.

3.1 Prediction results and accuracy of each method

Figure 3 shows results of dam inflow prediction using both algorithms for the two dams. The results for the evaluation index values are shown in the table in this figure; the blue-shaded areas in the table indicate that the values in these areas are those with good accuracy. We chose the following J_{pe} best value ±0.1, and NS coefficient ≥ 0.8. The prediction results using the DNN model tend to be highly accurate. The left column of Table 3 summarizes the values of each evaluation index. As a result, the DNN prediction results are slightly better than those of our previous work (Sakamoto et al. (2019)). These results probably owe to the best solution being included in the search range of hyper parameters. Determining the strictly optimal values for hyper parameters requires considerable time and calculation, which is a problem to be solved in predictions using the DNN algorithm.

The prediction result using EN model shows the same or better prediction accuracy than the DNN model. The results for the prediction model that uses EN indicate that prediction with good accuracy is possible through the use of a linear model which has a simple structure similar to that for the water level correlation method, based on the idea that the rainfall in the basin flows into the dam.

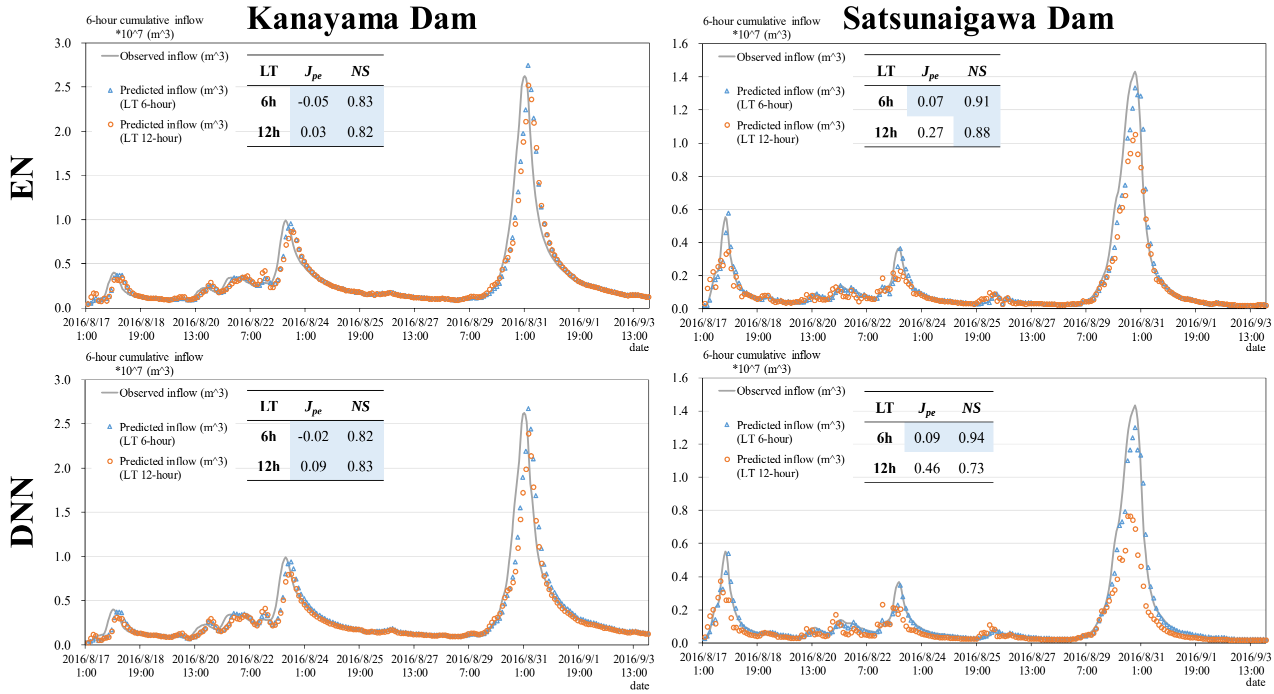

Figure 3. Dam inflow prediction results for each dam and algorithm; the small table in the figure shows the evaluation index values.

Table 3. Prediction accuracy list: The blue-shaded areas in the table indicate that the values in these areas are those with good accuracy: J_{pe} of within ±0.1 and *NS* coefficients ≥ 0.8.

			Including predicted rainfall		Removed predicted rainfall		Hourly prediction	
		LT	J_{pe}	***NS* coefficients**	J_{pe}	***NS* coefficients**	J_{pe}	***NS* coefficients**
Kanayama Dam	EN	**6-hour**	-0.05	0.83	-0.08	0.82	0.00	0.59
		12-hour	0.03	0.82	-0.002	0.81	0.15	0.71
	DNN	**6-hour**	-0.02	0.82	-0.21	0.77	-0.02	0.60
		12-hour	0.09	0.83	-0.10	0.83	0.05	0.58
Satsunaigawa Dam	EN	**6-hour**	0.07	0.91	0.08	0.89	0.00	0.98
		12-hour	0.27	0.88	0.26	0.82	0.26	0.77
	DNN	**6-hour**	0.09	0.94	0.53	0.61	-0.39	0.83
		12-hour	0.46	0.73	0.21	0.83	-0.11	0.79

3.2 Results for predictions that exclude assumed predicted rainfall

In the previous section, the actual rainfall was regarded as the predicted rainfall in addition to the explanatory variables. In actual dam inflow prediction, predicted rainfall will be used as an explanatory variable. The predicted rainfalls used are not sufficiently accurate. The inaccuracy of the predicted rainfalls may adversely affect the prediction results. Hence, we thought it would be ideal to avoid using inaccurate information in the prediction. In this section, the predicted rainfall was excluded from the explanatory variables shown in Table 2, and a new dam inflow prediction model was generated.

The upper part of Figure 4 shows the results of a dam inflow prediction model that uses both machine learning algorithms. The middle column of Table 3 summarizes the values of each evaluation index in this section. Prediction results using the DNN model at Kanayama Dam caused the NS coefficients to decrease from 0.82 to 0.77 from the removal of predicted rainfall. In contrast, the prediction results using the EN model at Kanayama Dam had the same *NS* coefficient and J_{pe} values as the results before the removal of predicted rainfall. The prediction results for 12-hour LT had the same accuracy as those for 6-hour LT. The EN algorithm can compensate for a shortage of variables, and it was able to make predictions with the explanatory variables remaining after predicted rainfall was excluded. These prediction results are roughly equivalent to those of the previous section. Accordingly, it was shown that that EN algorithm is able to obtain highly accurate prediction results without using predicted rainfall.

3.3 Prediction results for every hour

According to a previous study (Sakamoto et al. (2019)), the predicted rainfall accuracy is low in certain cases. Therefore, we used the accumulated value of 6 hours. However, the actual dam discharge is predicted for every hour, and the predicted reservoir level is determined. Therefore, we conducted prediction using the rainfall for each hour.

The lower side of Figure 4 shows the prediction results given by an hourly dam inflow prediction model that uses EN. The right column of Table 3 summarizes the values of each evaluation index in this section. The predict results for 6-hour LT at Satsunaigawa Dam were *NS* coefficients of 0.89 to 0.98 and J_{pe} of 0.08 to 0.00. However, the peak time at Kanayama Dam was delayed by LT, and the *NS* coefficient was 0.82 to 0.59, reducing the overall accuracy. Furthermore, the prediction results using DNN were generally worse than those for the EN model, with a negative J_{pe}.

As a result of the whole this paper, 6-hour accumulated values had better prediction accuracy. It is thought to have been difficult to predict the rainfall, which fluctuates with time. However, in the case of actual predictions, it is useful for judgments of operation that the hourly value come from the cumulative 6-hour values. Therefore, we want to improve the accuracy by changing the explanatory variables and training data in the future.

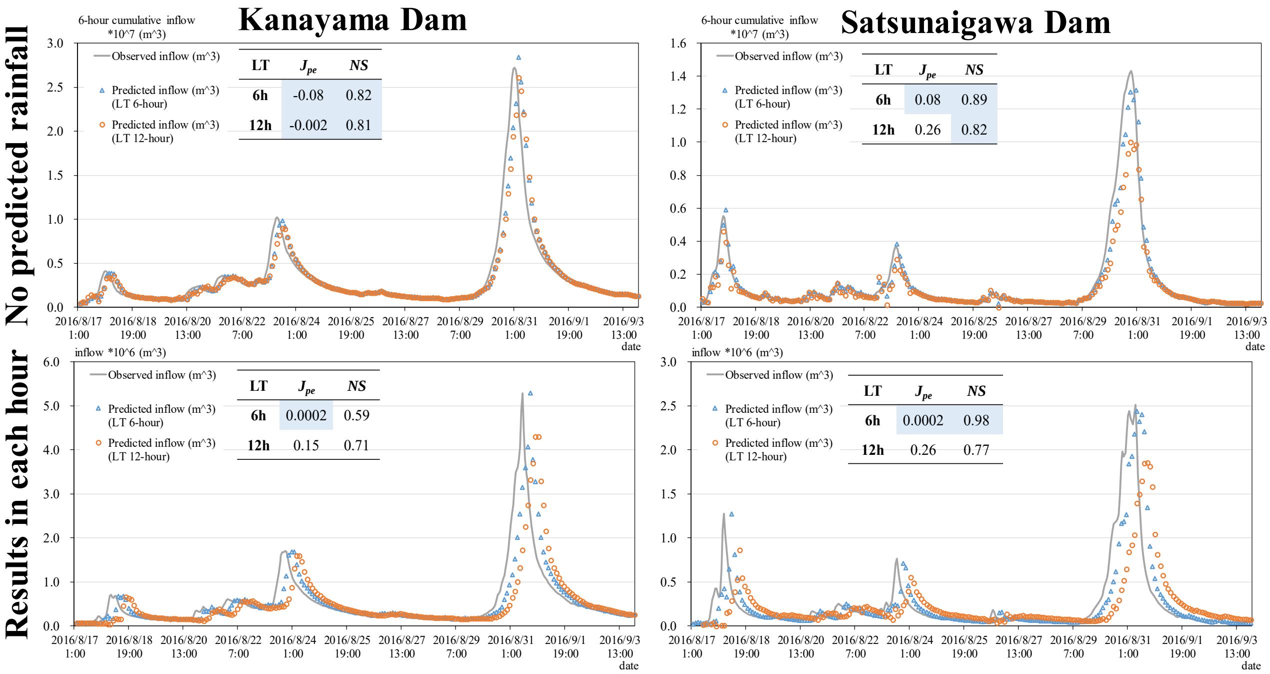

Figure 4. Dam inflow prediction results using the EN algorithm for each dam; the upper figures are 6-hour accumulated prediction. Predicted rainfall was removed from the explanatory variables. The lower figures are hourly dam inflow predictions.

4. CONCLUSIONS

The results of this research are as follows.

1) By excluding predicted rainfall from the prediction of 6-hour accumulated inflow, the prediction model using the EN algorithm improved the accuracy of *NS* coefficients from 0.10 to 0.20 and improved J_{pe} from 0.20 to 0.50, relative to the prediction model using the DNN algorithm.

2) According to the hourly dam inflow prediction, the 6-hour LT for the Satsunaigawa Dam was 0.98 for *NS* coefficients and 0.00 for J_{pe}, which is about the same as the 6-hour accumulated value. On the other hand, the *NS* coefficients for Kanayama Dam were 0.82 to 0.59. As a result, it was demonstrated that dam inflow can be properly predicted by Elastic Net.

In future works, we will continue to examine whether it is possible to prolong LT of prediction and to perform similar dam inflow predictions at other dams.

ACKNOWLEDGMENTS

We would like to express our gratitude to the Sapporo Development and Construction Department and the Obihiro Development and Construction Department of the Hokkaido Regional Development Bureau, MLIT for providing various data on the dams.

REFERENCES

Grant-in-Aid for Scientific Research on Innovative Areas, Term of Project FY 2013-2018, Initiative for High-Dimensional Data-Driven Science through Deepening of Sparse Modeling, http://sparse-modeling.jp/. Last accessed December 20, 2019.

Hui, Z. and Trevor, H. (2005). Regularization and variables selection via the elastic net, *Journal of the Royal Statistical Society*, 67, 301-320. DOI: 10.1111/j.1467-9868.2005.00527.x.

Sakamoto R., Kobayashi Y., and Nakatsugawa M. (2019). Comparison of Machine Learning Methods for the Prediction of Dam Water Level during an Abnormal Flood. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), 75, 2, I_85-I_90. (in Japanese)

Sparse Modeling. (2013). Initiative for High-Dimensional Data-Driven Science through Deepening of Sparse Modeling. http://sparse-modeling.jp/. Last accessed December 20, 2019.

Sunohara T, Utumi H, Inoue K, Mama S, Yoshida T, and Takemura H. (2006). Water Volume Estimation of Dams and the Flood Forecasting System on the Sagami River using Artificial Neural Networks. Advances in River Engineering, 12, 229-234. (in Japanese)

Tamura K., Kanou S., Miura S., Yamawaki M., and Kaneko H. (2018). Application of Deep Learning to Long-term Prediction of Dam Inflow toward Efficient Flood Control Operations. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), 74, 2, I_1327-I_1332, DOI https://doi.org/10.2208/jscejhe.74.5_I_1327 (in Japanese)

The Event Horizon Telescope Collaboration. (2019). First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole, The Astrophysical Journal Letters, 875, L6 (44pp). DOI:10.3847/2041-8213/ab1141.