River stage modeling with a Deep Neural Network using long-term rainfall time series as input data: Application to the Shimanto-River watershed

Wakatsuki Yuki[1], Nakane Hideaki[1], and Hashino Tempei[2]

[1]Kochi University of Technology
[2]School of Environmental Science and Engineering, Kochi University of Technology

November 16, 2022

Abstract

The increasing frequency of devastating floods from heavy rainfall associated with climate change has made river stage prediction more important. For steep, forest-covered mountainous watersheds, deep learning models may improve prediction of river stages from rainfall. Here we use the framework of multilayer perceptron (MLP) neural networks to develop such a river stage model. The MLP is constructed for the Shimanto river, which lies in southwestern Japan under a mild, rain-heavy climate. Our input for stage estimation, as well as prediction, is long-term rainfall time series. With a one-year time series of rainfall, the model estimates the stage with 50 cm RMSE for about 10 m of stage peaks as well as accurately simulate stage-time fluctuations. Furthermore, the forecast model can predict the stage without rainfall forecasts up to three hours ahead. To estimate the base flow stages as well as flood peaks with high precision we find the rainfall time series should be at least one year. This indicates that the use of a long rainfall time series enables one to model the contributions of ground water and evaporation. Given that the delay between the arrival time of rainfall at a rain-gauge to the outlet change is well simulated, the physical concepts of runoff appear to be soundly embedded in the MLP.

River stage modeling with a Deep Neural Network using long-term rainfall time series as input data: Application to the Shimanto-River watershed

Yuki Wakatsuki[1], Hideaki Nakane[2], and Tempei Hashino[3]

[1] Environmental and Mathematical Sciences Course, Kochi University of Technology

[2] Emeritus Professor, Kochi University of Technology

[3] School of Environmental Science and Engineering, Kochi University of Technology

Corresponding author: Tempei Hashino (hashino.tempei@kochi-tech.ac.jp)

Key Points:

- Flood runoff in a forest-covered watershed can be modeled with multilayer perceptron neural networks from long-term rainfall and stage data.
- Stages can be predicted with a few hours of lead time by using just the observed rainfall time series.
- Visualization of the learned network indicates a physically sound connection between rainfall history and flood stage.

Abstract

The increasing frequency of devastating floods from heavy rainfall associated with climate change has made river stage prediction more important. For steep, forest-covered mountainous watersheds, deep learning models may improve prediction of river stages from rainfall. Here we use the framework of multilayer perceptron (MLP) neural networks to develop such a river stage model. The MLP is constructed for the Shimanto river, which lies in southwestern Japan under a mild, rain-heavy climate. Our input for stage estimation, as well as prediction, is long-term rainfall time series. With a one-year time series of rainfall, the model estimates the stage with 50 cm RMSE for about 10 m of stage peaks as well as accurately simulate stage-time fluctuations. Furthermore, the forecast model can predict the stage without rainfall forecasts up to three hours ahead. To estimate the base flow stages as well as flood peaks with high precision we find the rainfall time series should be at least one year. This indicates that the use of a long rainfall time series enables one to model the contributions of ground water and evaporation. Given that the delay between the arrival time of rainfall at a rain-gauge to the outlet change is well simulated, the physical concepts of runoff appear to be soundly embedded in the MLP.

1 Introduction

Global warming is expected to increase the water vapor content in the atmosphere, thus fueling more intense tropical cyclones and producing heavier precipitation (IPCC, 2014). In Japan, the water vapor at about 1500 m asl is increasing, as is the number of days with precipitation exceeding 200 mm (Japan Meteorological Agency, 2020). One study argued that the precipitation from Tropical Cyclone Hagibis, which fell over the Kanto area in October 2019, was higher by 10.9% due to the increase of ocean temperature (Kawase et al. 2021). Furthermore, record breaking heavy rainfall events occur in Japan more frequently than before due to stagnant squall lines and other precipitation systems. For instance, quasi-stationary squall lines stayed over northern Kyushu from July 5th to 6th in 2017, setting new 6- and 12-hour precipitation records (Kato et al. 2018). Also, water vapor associated with a tropical cyclone circulation flowed into a stationary front from June 28th to July 8th 2018, releasing devastating heavy rainfall over western Japan and breaking 48- and 72-hour precipitation records (Uchida et al. 2021).

About 67% of the area on the Japanese main islands are covered by forests (MAFF), with most watersheds being composed of steep slopes from mountains to alluvial plains. In such a watershed, the concentration time, defined as the time when the rainfall most far in the drainage basin reach the river outlet, is estimated to be 400 minutes for the basin area of 1000 km^2 (240 minutes for 100 km^2) under an effective rain intensity of 30 mmhr^{-1} (Kadoya and Fukushima 1976). For the smallest mountain basins (area of 1–10 km^2), heavy rainfalls have concentration times less than 100 minutes (Kanda et al. 1990), which are too short to produce flood forecasts and thus basin residents must decide on their own whether or not to evacuate.

Although the flooding risk has increased under climate change, the willingness of residents to evacuate has not increased. For example, heavy rains in 2018 threatened flooding of the Monobe river in Kochi, Japan, triggering an evacuation advisory and order. Yet most of the residents did not evacuate, presumably because flooding had not occurred there for over

100 years (Kubo et al. 2019). One way that may help convince the residents to evacuate is to provide them with an accurate estimate of when a given heavy rainfall would produce a flood downstream. That is, we need a way to obtain accurate concentration times and stage forecasts. Probabilistic stage forecasts could be constructed from deterministic stage forecasts obtained by using short-term precipitation forecasts (such as Advanced Hydrologic Prediction Service products). Thus, it is necessary to improve numerical models for precipitation forecasting and stage forecasting, thereby gaining trust in the products among residents.

Recent availability of free full sets of machine-learning libraries and powerful GPUs has motivated the use of deep neural networks (DNN) in various fields including hydrology (e.g., Shen, 2018). For instance, multi-layer perceptron (MLP) models have produced accurate river flow forecasts (Rao and Giridhar, 2016; Chanu and Kumar, 2018; Oluwatobi et al., 2018). However, the previous studies considered daily to monthly forecasts, which are not applicable to the steep watersheds in Japan that have flood events over hourly time scales. MLP models using DNN for flood stage forecasts in Japanese rivers have reduced prediction errors (Hitokoto et al. 2016, 2017). Their approach has the limitation of requiring upstream stage observations as well as precipitation observations as the input data. Thus, it cannot be applied to the mountainous watersheds or small rivers where stage-gauge observation is not available in the upstream of the forecasting location.

When considering the development of DNN models for steep rivers in a moist and humid climate, the required input variables likely differ from those needed for rivers that are affected by early-spring melting of the snow that accumulate over the winter and that drains over a large catchment with small gradients. In general, the water stored in rivers and underground balances precipitation, evaporation, and runoff such that in any long-term analyses the storage term can be neglected. According to Nakano (1976), in the Japanese forest watersheds where annual precipitation is more than 1500 mm the annual evaporation is almost constant. Kochi prefecture, where the target watershed of this study is included, receives annual precipitation

of about 2500 mm, and thus it is speculated that the annual evaporation is constant and annual runoff and precipitation are in a linear relationship. This suggests that long-term precipitation and runoff time series can be digested into a DNN to construct a relationship that includes evaporation. Nakane and Wakatsuki (2018) proposed a modeling approach in which a river stage is estimated solely by inputting upstream long-term precipitation time series into DNN; this directly models characteristics of runoffs and infiltration processes in the watersheds from the time series. Their study shows that a DNN can learn the river stage (or discharge) at a certain time. Furthermore, Nakane et al. (2019) applied the same method to three rivers with different influences of dams, and showed that using a longer precipitation time series improved not only the estimated flood peak, but also the stage estimate during the dry season.

The purpose of this study is to discuss the estimation and forecast of river stages during flood events in the framework of the DNN model proposed by Nakane et al. (2019) in detail. When the estimation and forecast errors are not negligible, it is practically important to be able to explain the reasons behind, and the reliability of the modeling can be enhanced by showing that fundamental physical relationships are appropriately represented in the model. We, therefore, discuss if any physical concepts of runoffs are represented with the DNN model, and hereby we aim to establish the DNN modeling framework with long-term precipitation time series that enable us to automatically include runoff processes.

2 Study watershed and dataset

2.1 The Shimanto watershed

The drainage basin of this study is the Shimanto river, which is the longest river in Shikoku, the smallest of the four main Japanese islands. The official name of the river is the Watari river, but the name Shimanto is well-known to the public as “Japan’s last remaining limpid stream” (https://shikoku-tourism.com/en/see-and-do/10071). The Shimanto watershed lies in the island's southwest (Figure 1). Here, 95% of the land is forest, 4% is

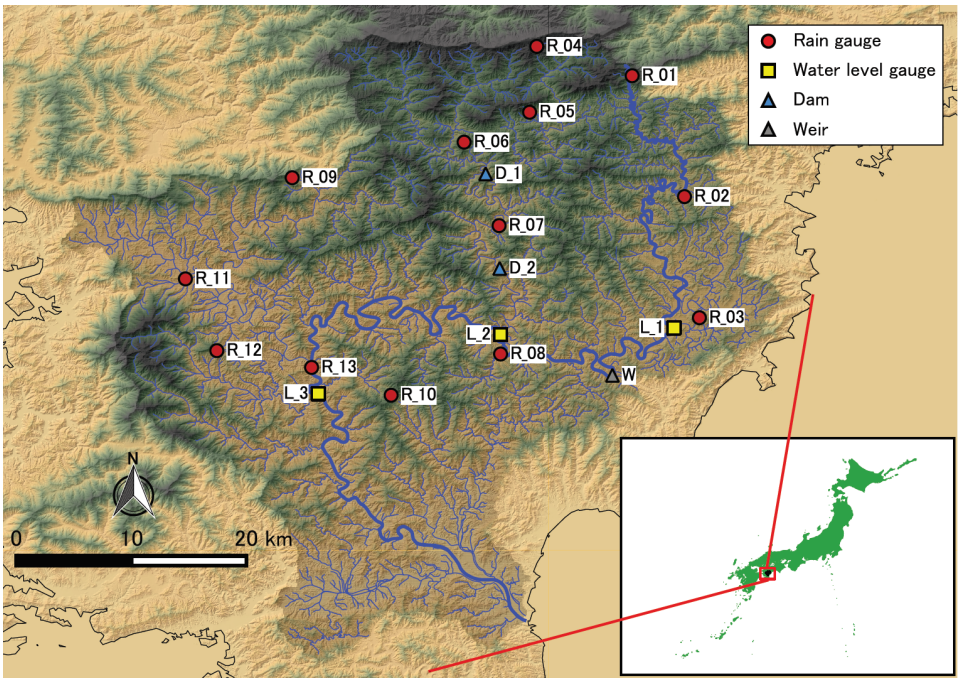

Figure 1. The Shimanto river watershed and location of rain gauges and stage observatories. Rain gauges are labeled R_01 to R_13. The stage observations are from L_3 (Tsunokawa).

farmland, and 1% is residential or urban. The drainage area is 2186 km^2. The annual precipitation reaches about 2900 mm in the headwaters, where it is designated a heavy rain area (MLIT). The source of the Shimanto has an elevation of 1336 m and the length of the main stream is 196 km, giving an average stream gradient of 0.68%. The average gradient of the upper basin ranges from 1/100 to 1/650, that of the middle basin from 1/380 to 1/1300, the lower basin from 1/1200 to 1/2200. It flows into the Pacific after 319 tributaries merge. Although one of the tributaries, the Yusuhara, has two reservoirs for power generation purpose, the main stream has a 8-m weir only. The weir cannot be operated for managing the river stage, making estimation of the stage during a flood event crucial for residents.

2.2 Dataset

The target estimation/forecast will be for the Tsunokawa stage observatory, which is 40 km upstream of the mouth (L_3 in Figure 1). The stage observations are from a hydrology and water quality database (Water Information System, Ministry of Land, Infrastructure,

Transport and Tourism: http://www1. river. go. jp/) and arranged to an hourly data set. The input precipitation observation data comes from 13 rain gauges upstream of Tsunokawa (Figure 1) that we obtained from the hydrology and water-quality database and from Japan Meteorological Agency (JMA, http://www. jma. go. jp/jma/indexe. html). This data covers 2002 to the present with resolution of either 10 min or 1 hr, but due to possible land use changes earlier in that period, we use just the hourly data for the 11 years of 2008–2018. Figure 2a shows the changes in the stage at Tsunokawa from 2008 to 2018. The stage is low during winter, then rises from summer to fall when the heavy rainfall events tend to occur, usually exceeding the level at which the flood prevention team is put on standby (yellow dashed line).

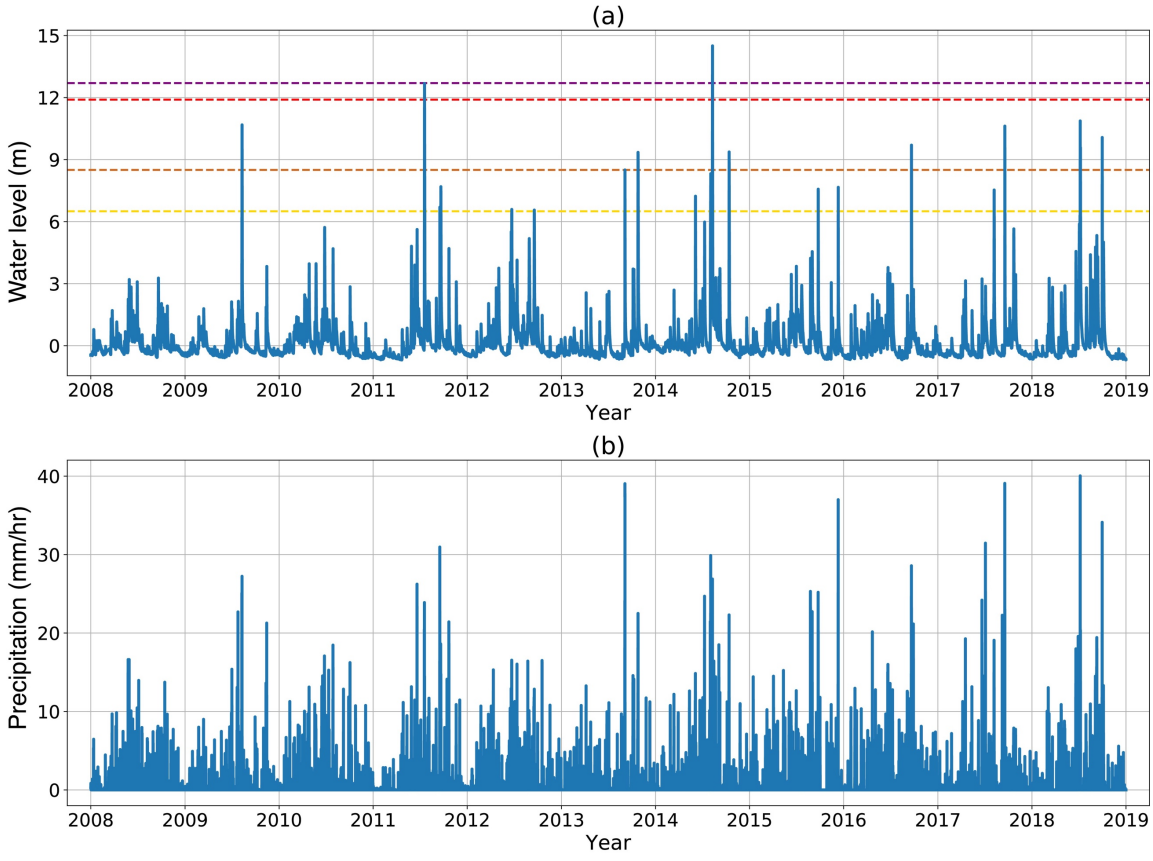

Figure 2. Observational data of (a) water levels at Tsunokawa Observatory, and (b) mean precipitation of the 13 rain gauges. Yellow dashed line indicates the water level at which the flood prevention team are put on standby (6.50 m). The orange dashed line is the water level for flood warning (8.50 m), the purple is the level for evacuation judgement (11.90 m), and the red is the flood danger level (12.70 m).

During this period, the red-dashed flood-danger level is exceeded twice. The average precipitation from the 13 rain gauges (Figure 2b) shows that an hourly rainfall of 30 mm is often exceeded after 2013 and short-term precipitation has intensified.

3 Method

3.1 Modeling concept

Rainwater falls in a watershed, infiltrates into the ground, moistens the soil, and then forms interflow. Some of the infiltrated water further percolates into the groundwater. Once the soil is saturated, the rain flows over the surface (overland flow) and directly flow into a channel. The interflow may emerge to the surface in a valley or saturated area, and then flow into the channel. When a pulse of rain falls to a regions with a short timescale for its overland runoff, then the river stage may form a pulse-like increase. However, depending on the depth of infiltration and geological characteristics, the interflow and the groundwater flow can possess various timescales that affect the river stage. In addition, the time series of spatially distributed rainfall will influence the river stage; that is, the stage is affected by soil moisture and groundwater from previous rainfall. Moreover, the river stage can respond nonlinearly to the upstream rainfall as suggested by the data in Figure 2. Therefore, the stage at a certain time t at an observatory is a complicated function of the spatial and temporal distribution of rainfall. Such a phenomenon can be inductively modeled with machine learning in which the precipitation time series going back from time t are set as the input data, and the downstream stage as the output data (labeled data). This downstream stage is not at the river's mouth to avoid influences from the tide.

3.2 Precipitation time series data as input data

As input, we use the individual hourly time series of the 13 rain gauges upstream, not the time series averaged over these gauges, so that the model may capture the spatial relationships of rain gauges to the Tsunokawa stage, which also provides hourly stage data.

Moreover, the specific slopes and geology of each branch where a rain gauge is located can presumably affect the stage.

Our preliminary investigation of the DNN modeling showed that the use of hourly rainfall time series directly as input produced a noisy time series of stage estimates. Also, after a pulse-like rainfall, the river stage would rise to a peak and then fall off with a long relaxation timescale. Rainfall events from further back in time can be represented as averages over a certain time window, assuming that such past rainwater can impact the current stage through changes in the base flow. Based on these considerations, we changed to a running average method in which the number of samples to be averaged increases with the time going back from the current time. The duration of the moving averaging y is determined for a period x hours previous to the time of evaluation to satisfy two conditions. 1) Precipitation reported within the past six hours has one-hour resolution. 2) Precipitation one year back (365 days times 24 hours) from the current time has one-month resolution (30 days times 24 hours).

$$y = \mathrm{floor}(6^{\frac{-\log 719}{\log 1460}} \cdot x^{\frac{\log 719}{\log 1460}} + 1) \quad (1)$$

As shown in Figure 3a, the duration of averaging (or time resolution) of rainfall that occurred 1-week back is 21 hours, and that of 3-month back is 204 hours. We assign “element” numbers to these averages (Figure 3b). This averaging essentially compresses the data; for example, a 1-year time series has 8760 samples, but the manipulation reduces the sample size to 69 (Figure 3b). Thus, the method reduces the computational burden for the DNN modeling.

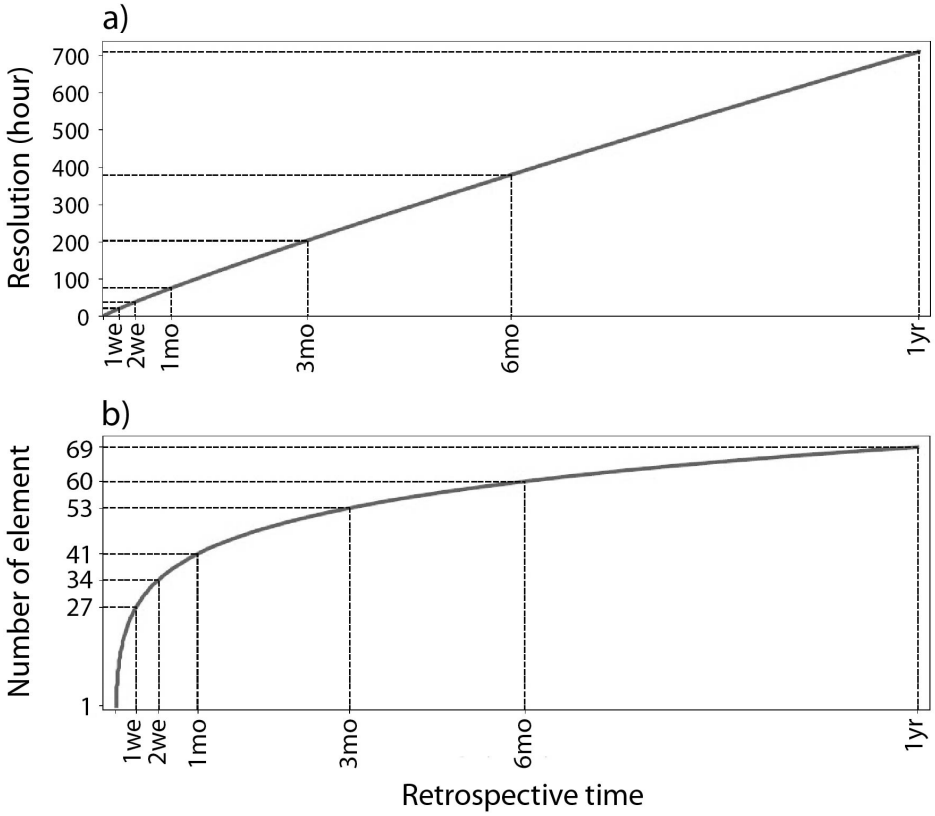

Figure 3. Time resolution a) and number of sample elements b) of input precipitation data. Abscissa is the time to which the precipitation time series goes back from the current time.

3.3 MLP model

We develop a DNN model with two or more hidden layers for the regression task, using a simple multilayer perceptron (MLP). A fully-connected MLP structure consists of input, hidden, and output layers as sketched in Figure 4. It is a feedforward network in which each node weights the inputs from the previous layer and the information moves from the input to the output layer. Each connecting line has a weight, and the linear sum of node values and the weights are incremented with the biases of the layer. An activation function acts on the resulting sum to propagate the information to the next layer. The weights and biases are updated to minimize the cost function (Rumelhart et al., 1986), a step that involves evaluating the gradients (partial derivatives). This procedure is repeated N times, where N is the number of batches times the learning epochs.

The number of nodes in the input layer of the MLP model is varied from 27 to 69 for

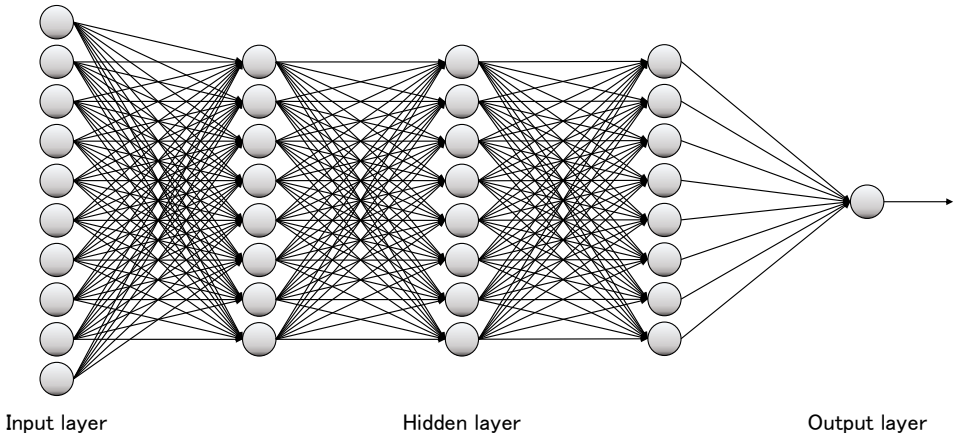

Figure 4. A fully connected multilayer perceptron (MLP) with three hidden layers. Here, the output layer is just the stage observatory water level. Input layer nodes are precipitation data.

each rain gauge, depending on the length of time gone back. The output layer has only one node. We apply the activation function PReLU (Maas et al., 2013) to each hidden layer before the output layer. As the river stage is regressed on the nodes of the last hidden layer, the mean squared error (MSE) is applied as the cost function without activation function on the output layer. To avoid overfitting, a penalty term of L1 norm is added to the cost function.

We divide the above data from 2008 January to 2018 December into training, validation, and test data sets. Most of the period is used for training. The validation set is then used to examine the model structures or hyper parameters (e.g., number of hidden layers), and the test set is used to evaluate the model performance. The exact data breakdown is 2008–2014: training; 2016: validation; and 2017–2018: test. As a result, the numbers of samples for training, validation, and test sets are 61368, 8784, and 17520, respectively.

To optimize the structure of the MLP model, the number of hidden layers, the number of nodes, and the parameter for L1 normalization are set as hyper-parameters, the combination of which are examined with the grid search method (Gorr et al., 1994). The search range is listed in Table 1. Concerning the number of nodes, the structures with layers that have more nodes than the previous layer are excluded from the search. (Such a search operation would increase the number of input dimensions, which seems inappropriate given that the input information has to be compressed from the first layer to the last layer.) The selection of hyper parameters are run for each length of the input time series to determine the best combination. In total, 480 combinations of hyper parameters are examined for each length of input time series. We set the batch size during the optimization to 100 and apply the Nesterov-accelerated Adam (Nadam) optimization method (Dozat, 2016). We found that for each combination of hyperparameters the errors started to converge after 25 epochs of learning, so to save on computational cost we set the number of epochs to 25. Then, the root mean squared errors (RMSEs) are calculated over the validation set by using MLP models having parameters that are obtained from 16 to 25 epochs. Among the above combinations of hyperparameters, we decided that the model with minimum average RMSE has the best structure for the length of input time series.

Table 1. Set of hyperparameters for the grid search method.

Hidden Layers	Nodes	L1-normalization
2	64	0
3	128	10^{-8}
4	256	10^{-7}
	512	10^{-6}
	1024	

3.4 Model evaluation criteria

To examine the model performance, we use the root mean squared error (RMSE), Nash-Sutcliffe efficiency (NSE) (Nash et al., 1970), percentage error of peak (PEP), and the time error of peak (TEP) calculated between observed stages and model estimates. RMSE is the measure of accuracy between the observation y_i and estimate $\hat{y_i}$. The NSE is equivalent to a coefficient of determination coefficient and indicates the fraction of the total variance explained by a regression. Thus, having NSE close to one indicates a smaller variance of residuals to the total. NSE can be considered as a relative measure of potential performance among cases with various magnitude of peak stages. The PEP is the fraction (in percentage) of the error between the estimated peak and observed peak stages during flood events. TEP equals the difference in time between the peak estimated in the model $\hat{t_p}$ and the peak observed t_p. As such, the TEP tells us to what degree the model has learned the time of flood arrival after the continuous precipitation that caused it. Specifically,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2} \quad , \tag{2}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y_i})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad , \tag{3}$$

$$\mathrm{PEP}(\%) = \frac{\hat{y}_{t_p} - y_{t_p}}{y_{t_p}} \cdot 100 \ \text{, and} \tag{4}$$

$$\mathrm{TEP} = \hat{t_p} - t_p \ . \tag{5}$$

4 Results and discussion

4.1 Stage estimation

First, we examine the highest four stages in the test set. These high-stage events are as

242 follows: case 1 on 2017 August 7th, case 2 on 2017 September 17th, case 3 on 2018 July 7th,
243 and case 4 on 2018 September 30th. The MLP model we discuss in this section was
244 constructed using a precipitation time series that goes back one year from the time of the high-
245 stage events, starting with hourly resolution. This model possesses small errors for the
246 training set and performs qualitatively better during low-stage periods, compared with the other
247 models with different input time lengths.

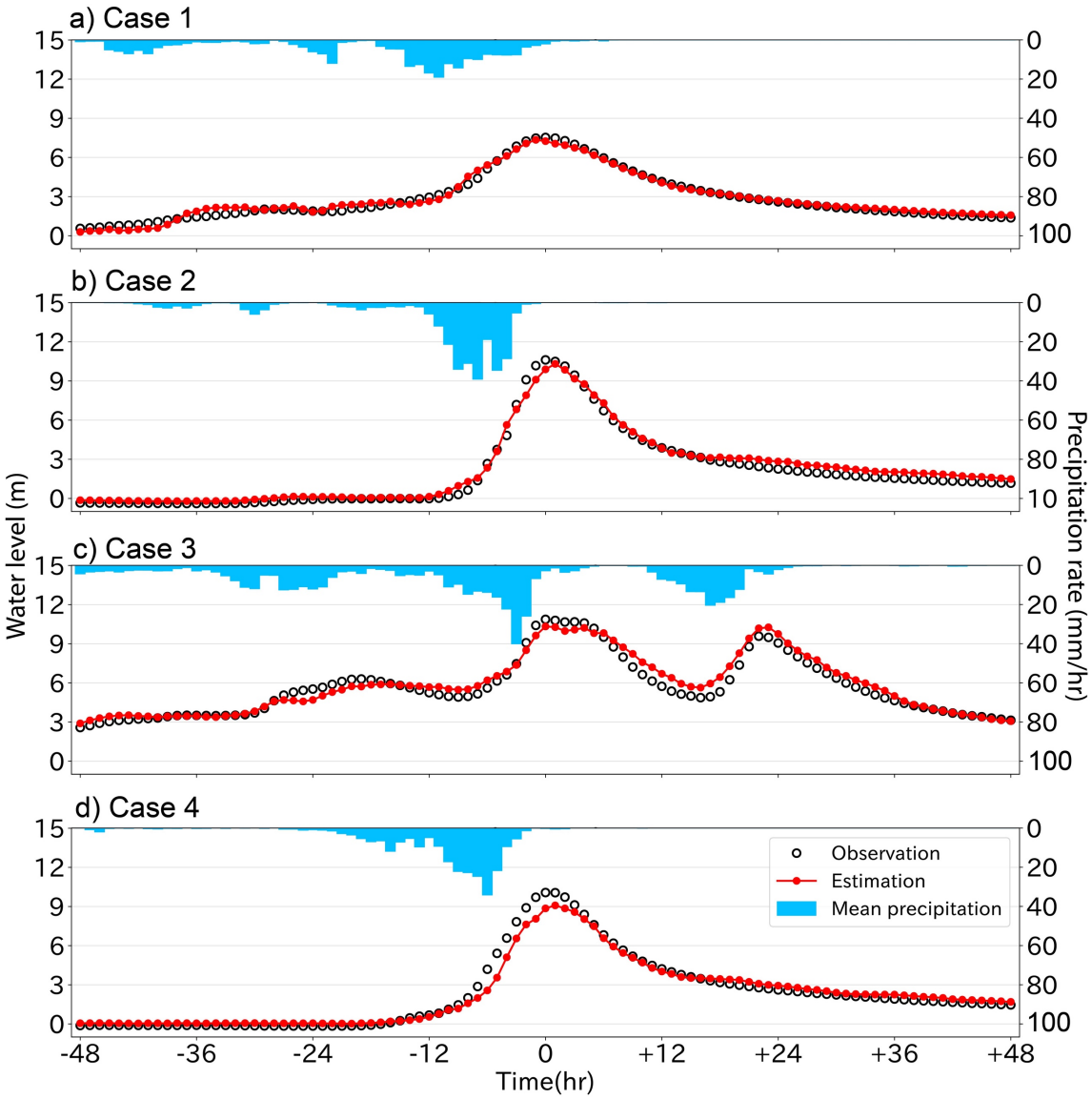

Figure 5. Stage estimation for the four flood events in the test dataset within 48 hr of the peak stage. The observed stage data are marked as open circles, the model estimate with red circles. Precipitation time series averaged over all gauges are in blue bars at top. Peak hours for cases 1–4 are 2017 August 7th 11 JST (=UTC+9), 2017 September 17th 22 JST, 2018 July 7th 11 JST, 2018 September 30th 21 JST.

Overall, the estimated stages match well with observed stages and without significant biases (Figure 5). The RMSE are about 30–50 cm (Table 2) for the stage variations, relatively small compared to the peaks of about 7–11 m. Also, the NSE exceed 0. 94, indicating a good temporal fit to the stage. Considering the individual cases, case 1 has relatively small rainfall and stage, which likely contribute to the small overall error. Cases 2 and 4, which had heavier

253 rain about 12 hours before their peaks, both show a delay of the rising limb, with the estimated
254 peak time behind about one hour, and both cases underestimate the peak stage. Later, as the
255 water levels ease, the estimated stages are slightly overestimated. Case 3 has three rainfall
256 peaks exceeding 10 mm hr^{-1}, with each stage peak well captured by the model. However, the
257 stages between the stage peaks are slightly overestimated.

258 In cases where a precipitation time series similar to the test data set is not available in the
259 training and validation sets, the estimation errors can be large. Even though case 2 shows a
260 similar change of stage over time to case 4, the latter had a larger underestimate and a much
261 larger PEP (Table 2). According to the time series in Figure 6, case 2 has little rainfall within
262 one month of the stage peak, whereas case 4 has several notable rainfall events within the
263 month preceding the flood. The difference in the precipitation pattern during that preceding
264 period probably led to a difference in soil moisture in the watershed, and hence to the retarded
265 accuracy of case 4. With regard to the soil moisture and precipitation occurrences, a condition
266 like case 2 was probably better learned from the training and validation sets, but a condition
267 like case 4, in which soil moisture increased with time, was probably not learned as well.
268 Physically, the model simulated a situation where the rainfall over-infiltrated into the soil,
269 leading to an underestimate of the overland flow and thus underestimation of the peak.

270 When there are not enough samples of heavy rainfall events in the training and validation
271 sets that are spatially similar to those in the test set, the estimation errors can be potentially

Table 2. Evaluation measures of the four flood events in the test dataset.

	CASE1	**CASE2**	**CASE3**	**CASE4**
RMSE (CM)	30.1	38.0	52.6	42.8
NSE	0.973	0.982	0.942	0.976
PEP(%)	3.44	6.42	4.70	11.3
TEP(HR)	-1	1	0	1

large. To compare the characteristics of precipitation in the training and test sets, we now investigate the fraction of total precipitation at each rain gauge for each flood event. For the statistics, we use the top 12 flood events from 2008–2014 in the training set, and the four events for the above test set. As defined below, the precipitation accumulation p_i^n for flood event n and for rain gauge i is calculated by the 24-hr window centered on the stage peak:

$$p_i^n = \sum_{t=t_p^n-12}^{t_p^n+12} r_{i,t}^n \ , \tag{7}$$

where t_p^n is the peak time of the stage. Summing over all 13 gauges, the total accumulation p^n is

$$p^n = \sum_{i=1}^{13} p_i^n. \tag{8}$$

Thus, the fraction for the gauge i is

$$P_i^n = \frac{p_i^n}{p^n} \ . \tag{9}$$

Figure 7 shows the fraction of the 24-hr accumulation centered on the time of stage peak P_i^n. A larger fraction may indicate a larger contribution to the flood (the area coverage of each gauge is not considered for simplicity). The rain gauges are arranged from the most upstream at left (R_01) to the most downstream at right (R_13). The generally decreasing trend shows

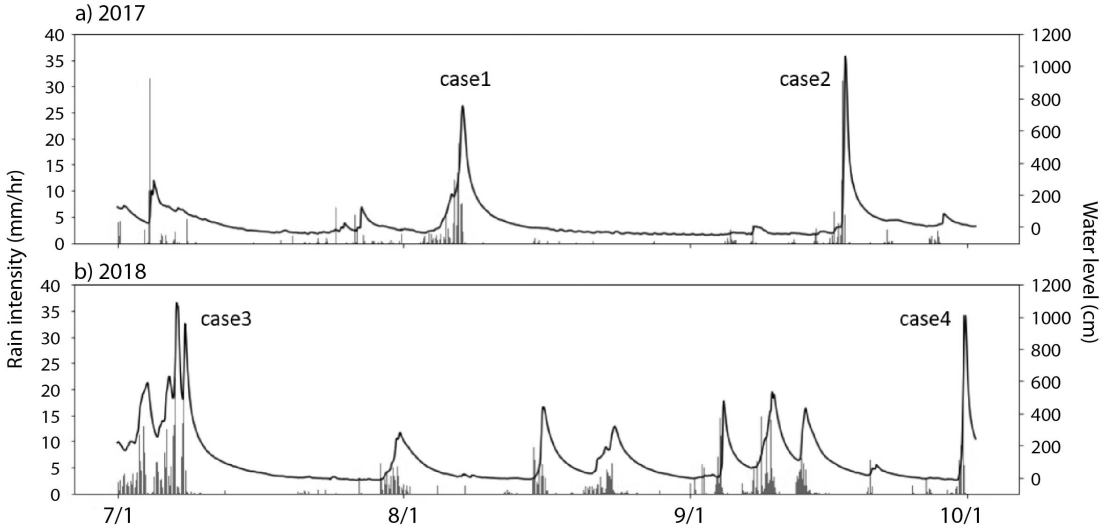

Figure 6. Time series of average precipitation and stage in the test dataset. a) For the cases in 2017. b) For the cases in 2018.

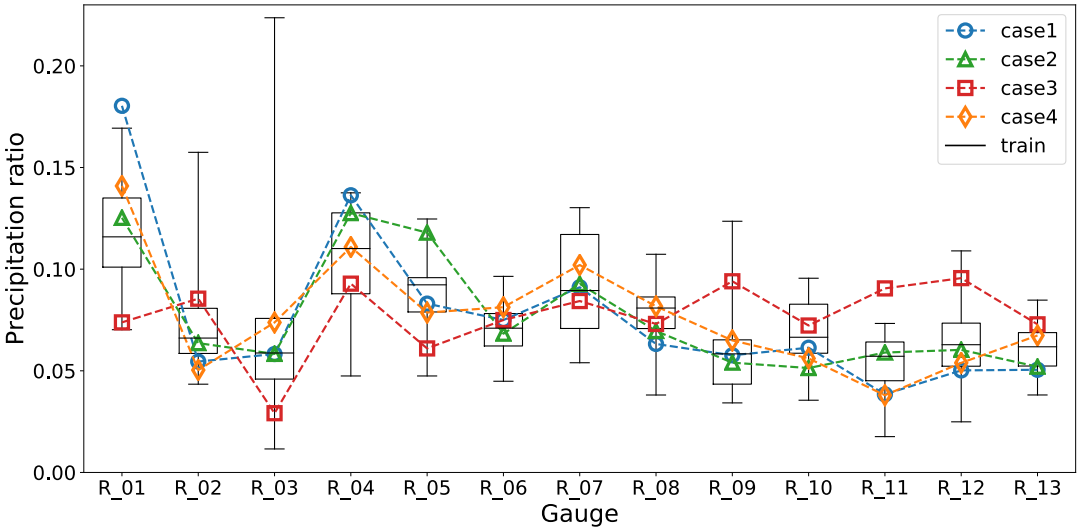

Figure 7. Ratios of 24-hour accumulated precipitation at each rain gauge to the total of all 13 gauges during the four flood cases in the test dataset. Ratios for 16 flood events in the training set are indicated with the whisker plots where 5, 25, 50, 75, and 95th percentiles are shown.

that the precipitation at the upstream gauges at higher altitudes contributed more than those further downstream. Cases 1, 2, and 4 show a pattern similar to the training set, and they are within the 25 to 75 percentiles except for R_01. However, case 3 indicates a quite different pattern compared with the training set, with the fraction at the upstream gauges relatively small and those downstream relatively large. Furthermore, several gauges have values outside the 25^{th} and 75^{th} percentiles. This difference in the spatial distribution of accumulation is most likely the reason why the RMSE of Case 3 is the worst.

4.2 Stage forecast

Here we examine an MLP model that learns the relationships between the precipitation time series up to the present and the river stage at δt later. In other words, this is the model that forecasts the stage at δt from the current time. The same precipitation time series and hyper parameters were used to construct this MLP model as used for that in the previous section.

The time series of the forecast stage with lead times δt of 1, 2, …, 6 hr are shown for case 2 in Figure 8. Forecasts with 1 and 2 hr lead times give an accurate stage time series similar to

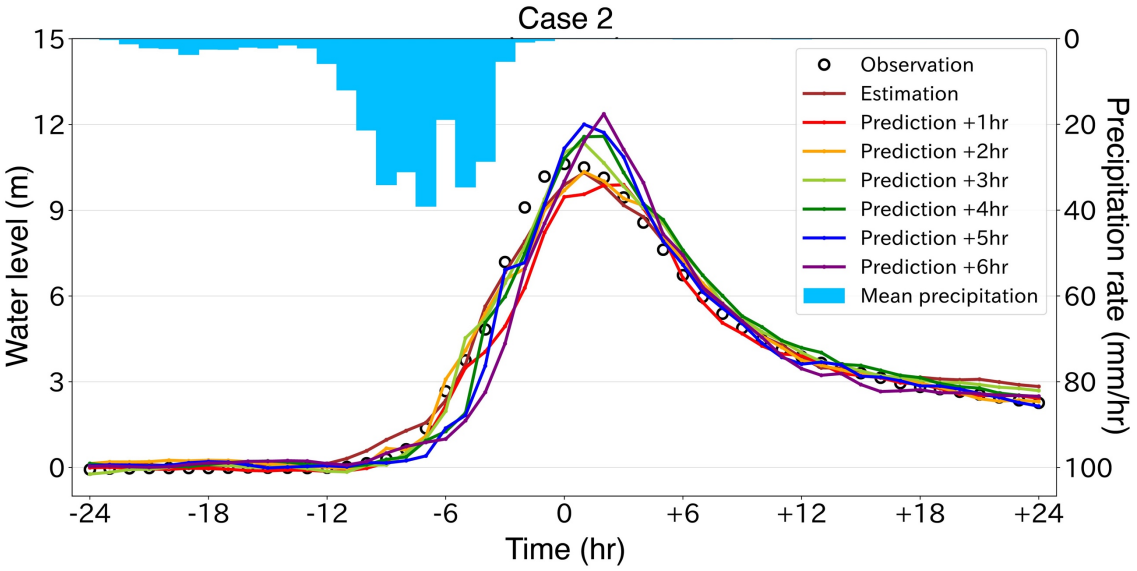

Figure 8. Case 2 observations and model runs. Water level is indicated by the left axis, precipitation by the right axis.

the estimation. However, the 3-hr forecast predicts a larger peak than the observation, as do those for the 4–6-hr forecasts. A closer look at the 4-hr forecast indicates that the stage is about a half that observed at −6 hr, and the predicted flood peak is late. The same issues also occur for the 5- and 6-hr forecasts. The errors are probably caused by a lack of precipitation information during δt. The accurate prediction from the 3-hr forecast suggests that the precipitation producing the stage peak arrives via surface runoff at Tsunokawa after about three hours.

Figure 9 shows the dependence of the RMSE on the lead time for all four cases. For cases 2–4 with the large magnitude of stage peak the RMSEs are similar among the forecasts when the lead time is three hours or less. However, case 3 shows a rapid increase of the RMSE after a 2-hr lead time. This case has a larger fraction of precipitation downstream (Figure 7). This suggests that when stage forecasts are made without the use of precipitation forecasts, especially cases with large fraction of precipitation occurring in the downstream, the errors can dramatically increase with the lead time.

Considering the MLP framework above, a MLP model for a δt-hour stage forecast may

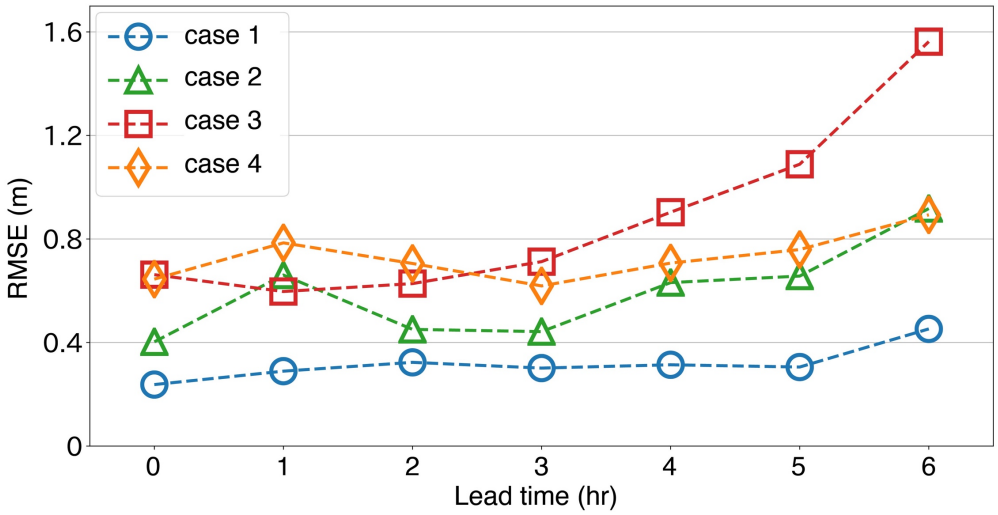

Figure 9. RMSEs for the four flood cases at 0- to 6-hr lead times. RMSEs were calculated over ±24 hours centered at the time of peak.

be developed with a combined time series of observed and forecast precipitation to learn the stage observed at δt hour. In this section, we discussed the case where the forecast precipitation data is not available up to the time when the stage forecast is made. Given that JMA produces precipitation forecasts up to 15 hr ahead, then by using their time series, a stage forecasting may be produced with a 15-hr lead time by using the current MLP framework for estimation. In such a case, any decrease in the accuracy of the precipitation forecast would likely decrease the accuracy of the river-stage forecast.

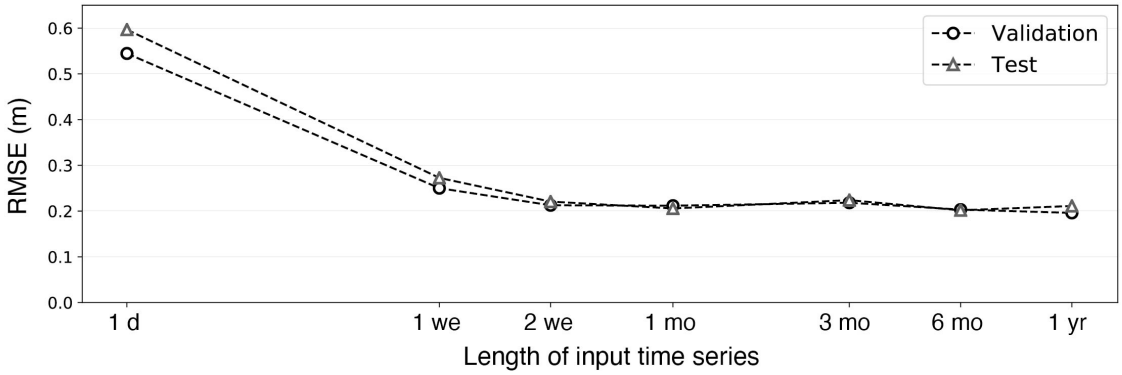

Figure 10. RMSEs for various lengths of the precipitation time series. A given length corresponds to an MLP model constructed with that length. Open circles are for the validation dataset, open triangles for the test dataset.

4.3 Effects of precipitation time series length on stage estimation

In above sections we discussed the stage estimation and forecast based on MLP models that use a one-year precipitation time series as the input dataset. Here, we examine how changing the precipitation time series length affects the stage estimation. This length was set to one day, one week, two weeks, one month, three months, six months, and one year. For a given length, an MLP model was developed and its hyperparameters were optimized.

Figure 10 shows the RMSEs between the estimated stage and observation for the seven MLP models for both the validation and the test datasets. For both datasets, the RMSE dramatically decreases from one day to one week, reaching about 0.2 m after one month. Note that Nakane et al. (2019) found that the RMSE continued to decrease monotonically to 0.156 m after 720 days. (They used a method involving similar MLP models, but shorter time-averaging.) Therefore, we cannot assert the information of precipitation that was observed more than one-month ago is not necessary.

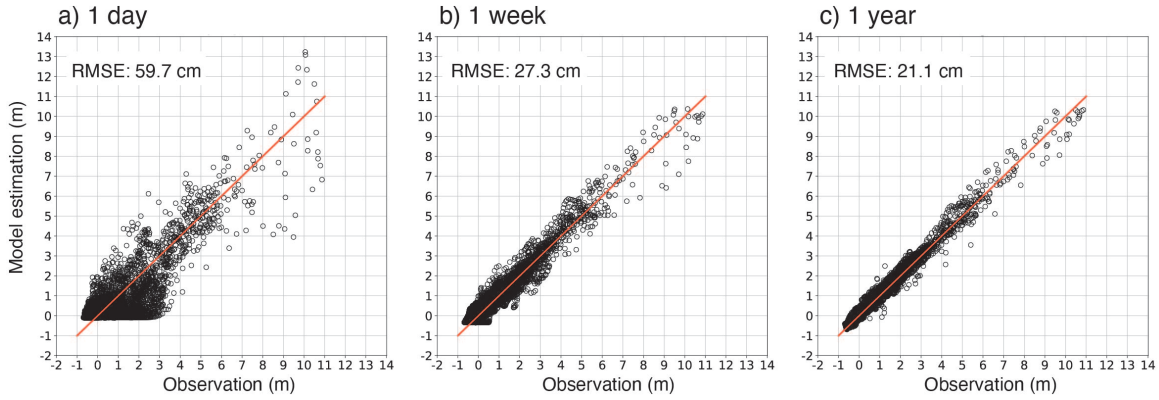

Figure 11. Scatter plots between stage observation and estimate for the test dataset for models using three lengths of the precipitation time series. a) With one-day time series. b) With one-week time series. c) With one-year time series. Red solid lines give the 1:1 relation. RMSEs are shown in the box.

The improvement in modelling fit for longer precipitation time series can also be seen in the scatterplots (Figure 11). For the one-year precipitation time series, the points are distributed close to the 1:1 line (red, in Figure 11c). The one-week case is similar, but with more spread, corresponding to the larger RMSE than the one-year case (Figure 11b). The case with one-day time series shows a large number of points are located below the red line over the range of observation of 3 m, indicating the model tends to underestimate the stage. The reason behind the large one-day RMSE compared to one week is examined next.

Consider the relation of stage estimate to precipitation pattern for the one-day, one-week and one-year MLP models for flood cases 2 and 3 in Figure 12. The importance of the length of input time series is evident in the one-day series for case 2 (Figure 12a) when the stage estimate suddenly drops just after 12 hr and then stays near 0 m. At 24 hr, the average precipitation during the past 24 hr is 0, which explains the estimate being about 0 m. Case 3 (Figure 12, right column) has intermittent heavy rainfall observed in the upstream. The one-day MLP model poorly estimates the second stage peak and essentially misses the third peak

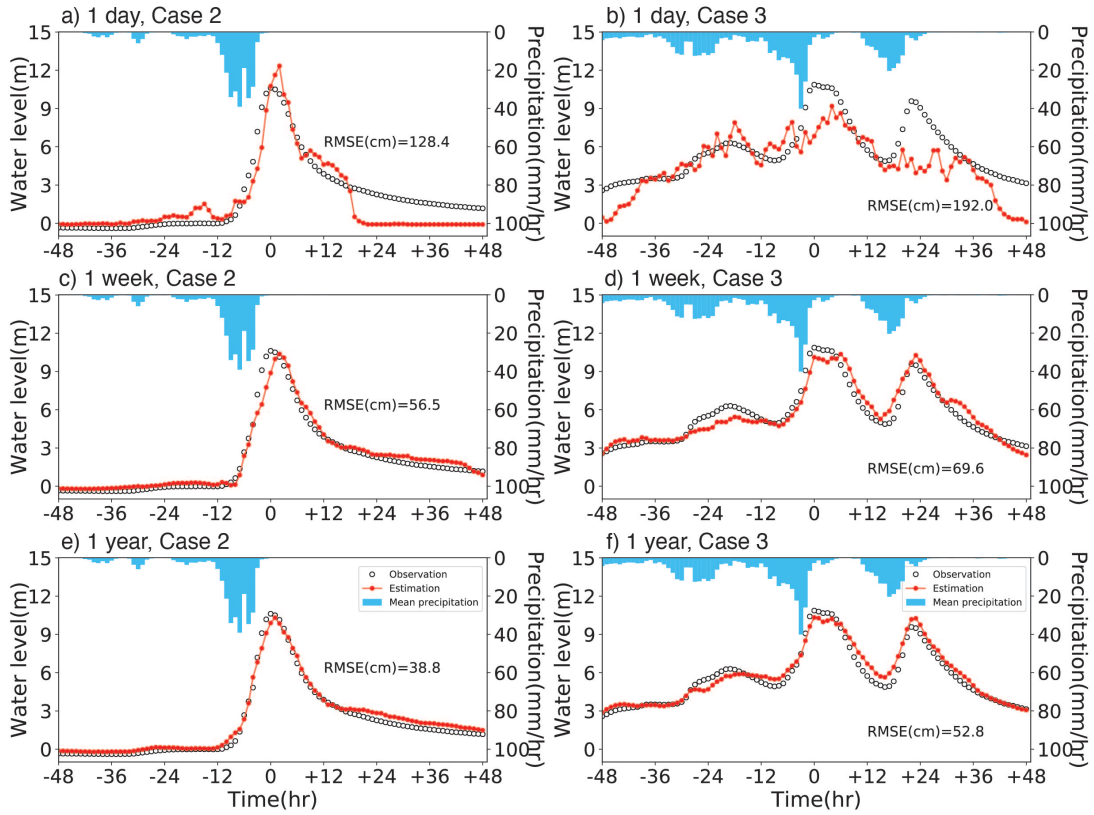

Figure 12. Stage estimates for three lengths of input precipitation time series, examined within 48 hours of two flood events. Left column: case 2. Right column: case 3. The top row is for the one-day time series, the middle for the one-week time series, and the bottom for the one-year time series.

(Figure 12b). This model's estimate is remarkably noisy presumably due to the absence of precipitation information before 24-hr back. When the length of input time series increases to one week, the three peaks in case 3 are reproduced nicely (Figure 12d), but the match between the estimates and observations even better with the one-year input time series (Figure 12f). From these results, we conclude that during flooding events the role of soil moisture and ground water is extremely important one day to one week before a rainfall event, and the runoff process with larger timescales associated with precipitation that occurred one week to one year in the past also cannot be ignored.

4.4 Stage estimation sensitivity to the rain gauges

How does a given rain gauge and rainfall event affect the flood stage? The typical way

to measure an output's sensitivity to an input is by perturbing the input and noting the change in output. However, in this case there are hundreds of input variables for the single output variable (the stage). Thus, we use layer-wise relevance propagation (LRP) (Bach et al., 2015), a visualization technique, to determine the relationship between stage and precipitation time series. The relationship enables us to examine the time of precipitation arrival to the stage, as well as the sensitivity of the stage, to a given precipitation event from a given rain gauge.

LRP measures the overall contribution of each input variable by propagating the contribution through the learned neural network for a specific instance of output. Rises in the river stage are in general formed not by the independent precipitation event at a certain moment in the past, but by a distribution of precipitation over a period of time. Therefore, for the MLP models here, it is more appropriate to measure the overall relevance of the past precipitation time series than to calculate the stage's sensitivity to the precipitation at a certain moment. Also, LRP can be run for estimates and forecasts for real flood events, providing an explanation of the model behavior for a given flood event. Therefore, it is an appropriate method to characterize the MLP models.

LRP quantifies the contribution of a given precipitation input to the model output, calling it the "relevance" of that input. A higher numerical value for the relevance indicates a higher contribution. Details on its calculation are in the appendix; here we focus on the results.

The relevance of the precipitation inputs for the stage estimate at a certain time was calculated for case 2 over 24 hours of the flood event. In Figure 13, the relevance of precipitation inputs for the stage at a certain time, shown as the black contours, can be examined by specifying the time on the abscissa and examining the contours from top down, which corresponds to the retrospective time. The top panel has results for the gauge furthest upstream, the bottom panel that for the gauge closest to Tsunokawa, L3 (see Figure 1). For the precipitation inputs, we analyze just the first 14 (of the 69 total).

In general, precipitation at a certain time shifts toward a larger element number as the

383 estimation time progresses, and the time resolution gets worse, combining more precipitation
384 events as the time goes back from the estimation time. For instance, in Figure 13a, the
385 precipitation (color fill) of E_01 at −9 hr is the same as that of E_02 at −8 hr, and at −3 hr the
386 precipitation is contained in E_07 element through the two-hour average. The sequence of
387 precipitation that starts at −10 hr lasts up to −4 hr, and the number of elements in the sequence
388 gradually decreases along the vertical axis toward +12 hr so as to remain as the input
389 information.

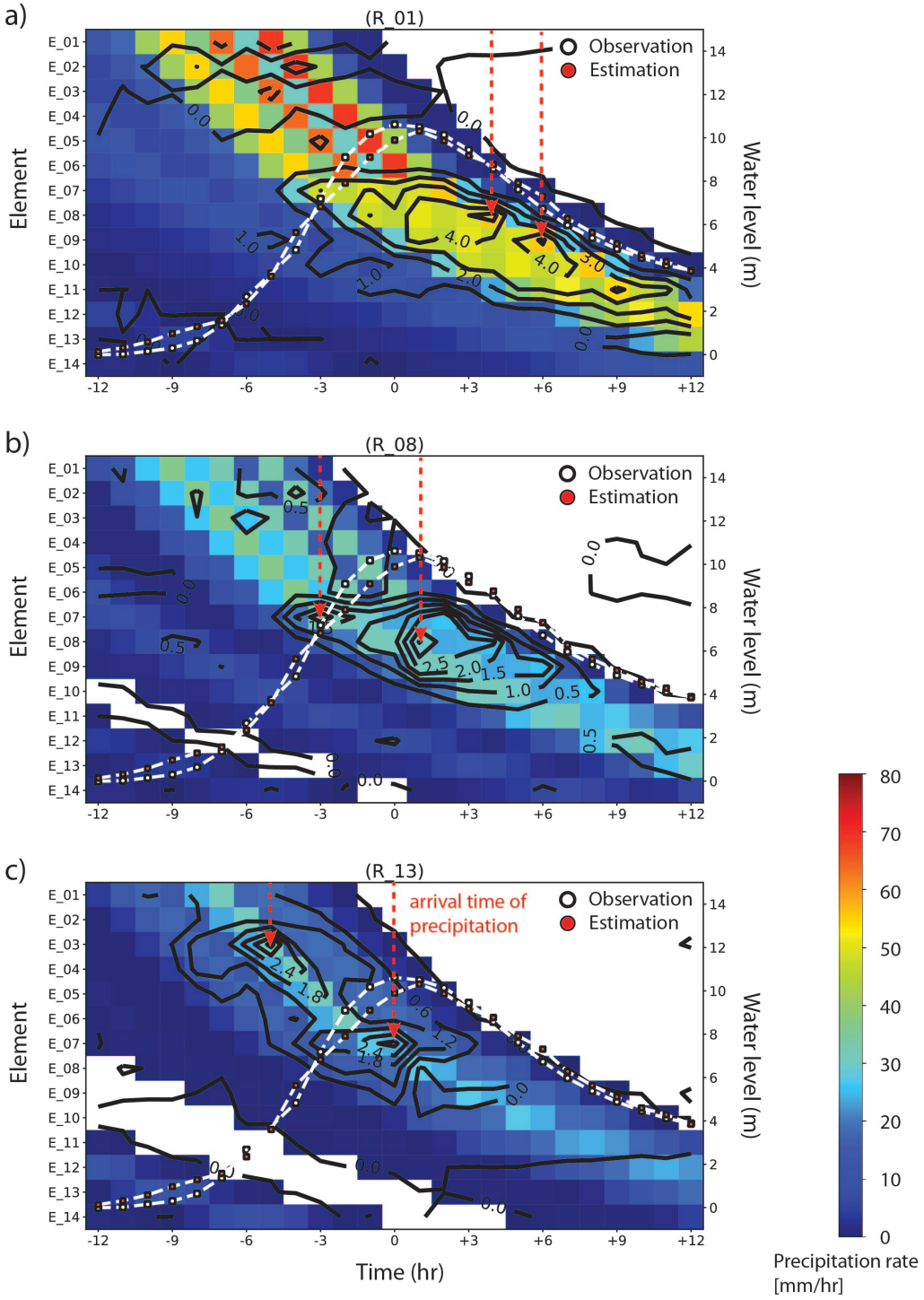

Figure 13. Relevance map of the stage estimation MLP model for case 2 and three rain gauges along the main stream (Figure 1). Abscissa is the time from the stage peak. Left ordinate is the precipitation element (Table 3), with its relevance marked as black contours. Right ordinate is the stage, with the white dashed curves showing observed (white-filled circles) and estimated (red-filled circles) stages. Color fill is input precipitation. a) Gauge R_01. b) Gauge R_08. c) Gauge R_13. The red arrows indicate estimation of arrival time of precipitation.

390

Consider first the relevance for R_01 (Figure 13a). The sequence of precipitation events that starts at −10 hr shows a rise in relevance (black contours) at −4 hr and at E_07. At −4 hr the most contributed precipitation to the stage estimate (about 5 m) is the one observed at −10 hr, the time when the stage starts to rise. For 0–4 hr the relevance shows a peak at E_08, indicating that the precipitation that fell 8 or 9 hr prior contributes to the stage up to four hr after the peak. Then, from 4 to 6 hr the maximum relevance values occur at E_08 and E_09, as marked with the red arrows. Based on this pattern, the time for precipitation that fell at R_01 to reach the Tsunokawa stage observatory is about 8–11 hr. Furthermore, there are large values of relevance associated with the sequence of precipitation events up to 7 hr, and precipitation peaks follow the relevance peaks closely. Moreover, as the relevance value is particularly large after the stage peak, this rain gauge has the geographical characteristic of being a significant contributor when the stage decreases rather than when it increases.

Now consider the relevance for gauge R_08 (Figure 13b). The relevance starts to increase at −4 hr and at E_06, reaching a maximum value at 1 hr and E_08. This pattern indicates an arrival time of precipitation of about 8–9 hours. In contrast with R_01 further upstream, this gauge makes its maximum contribution at the time of the stage peak. Thus, for the same arrival time of precipitation, the gauge further downstream has a different timing of its largest contribution. In addition, it has a narrower spread of relevance along the estimation

Table 3. Time elements prior to current time *t* and their time resolutions.

ELEMENT	TIME (HR)	RESOLUTION (HR)
E_01	t	1
E_02	t-1	1
E_03	t-2	1
E_04	t-3	1
E_05	t-4	1
E_06	t-5	1
E_07	t-6~ t-7	2
E_08	t-8~ t-9	2
E_09	t-10~ t-11	2
E_10	t-12~ t-13	2
E_11	t-14~ t-16	3
E_12	t-17~ t-19	3
E_13	t-20~ t-22	3
E_14	t-23~ t-26	4

time, suggesting that the precipitation that contributes to the flood peak occurs over a shorter time period.

Finally, consider R_13, the gauge furthest downstream and closest to the stage. Its relevance starts to increase at −6 hr and E_02, reaching a maximum at either −5 hr and E_03 or at 0 hr and E_07. This pattern indicates an arrival time of precipitation of about 2–7 hours. The relevance is larger before the peak, so for this gauge the precipitation contributes most from the rising limb to the peak.

Let d_{R_n} be the distance between the Tsunokawa station and R_n. Figure 1 indicates

that $d_{R\_13} < d_{R\_08} < d_{R\_01}$. The relevance analyzed above suggests the arrival time of precipitation t_{R_n} follows t_{R_13} (2–7) < t_{R_08} (8–9) ≤ t_{R_01} (8–11). The t_{R_13} and t_{R_08} are clearly different, and at most about 6–7 hours traveling time in between. This suggests that the MLP model represents the geographical distance in the arrival time of precipitation. On the other hand, the nearly equal timescales between gauges 1 and 8 is unclear.. Gauge R_01 experienced about twice the precipitation than that at R_08 during case 2 and lies in a steeper, higher region than R_08 (Figure 1). The steeper slope could reduce the arrival time of R_01, but other factors such as vegetation, soil types, and local topography may also influence the relevance. Future research is required to clarify the physical implication of the relevance in detail.

5 Conclusion

In this study, we developed multilayer perceptron (MLP) models to estimate and forecast a river stage in the Shimanto river watershed. This watershed lies in a mild, rain-heavy mountain region covered with forests. The models were developed solely based on the observed long-term precipitation and stage time series, and then tested. The main findings are

- Models that estimate the stage at the latest time of the input precipitation time series captured the time fluctuation of stages with an RMSE of 50 cm for flood peaks of about 10 m.
- Stage forecasts were made 1 to 6 hours after the latest precipitation observation with the MLP framework. The performance was highly accurate with up to a 3 hour lead time. This suggest that the current precipitation information in the watershed contributes significantly to the stage 3 hours later.
- Input of precipitation that occurred one day to one week prior to a flood plays influences the river stage estimate during flood events, which is likely related to the infiltration to soils and interflow processes. Precipitation further back, up to one year, has non-negligible impacts on the base flow, which implies that the MLP

models learned the ground water flow over long timescales.

- Use of LRP (layer-wise relevance propagation) enabled us to estimate the arrival time of precipitation based on the increase of the contribution (called relevance). The arrival time correlated to the distance between rain gauge and stage observatory, indicating that the MLP models likely capture the geographical characteristics of the watershed. However, more detailed analysis is required to relate the arrival time to physical parameters such as gradients and vegetation.

The inductive (or empirical) modeling proposed in this study does not include relevant physical processes explicitly. Even so, the models exhibited reasonable physical behavior. The visualization technique such as LRP helps users interpret the model characteristics, and it can enhance the reliability of the MLP models as a practical hazard prevention tool. However, due to the inductive nature of the method, accurate predictions require that the model developers carefully select training sets that are expected to be similar to any anticipated flood events.

The advantage of the MLP modeling is in allowing one to construct models for forest-covered watersheds with complicated topography only from long-term time series of precipitation and stage. In contrast, physical-based modeling typically requires information on topography, gradient, vegetation, and soil type that can be cumbersome to acquire. Also, the MLP modeling does not require involved parameter tuning, which can save on manpower and cost. Furthermore, if accurate precipitation forecast datasets are available as the input, a stage forecast MLP model can be constructed. Future research is warranted to examine the ability of the MLP model to remove biases in the precipitation forecast and to effectively use the precipitation forecasts.

6 Appendix

6.1 Development environment

The computational hardware is a personal desktop computer with Intel Core i7-7700 CPU and GeForece GTX1080Ti GPU. The OS is Linux OS (Ubuntu 17). The software framework for Deep Learning is Keras that uses Google Tensorflow as the backend. The code is written with Python (Version 3. 6). The MLP model of this study uses fully connected networks, so sequential programing is possible. We used Keras functional API for ease of future development.

6.2 LRP

The basic LRP concept is shown in Figure A1. The subscript gives the node number, the superscript gives the layer number. The left diagram shows the feedforward propagation of information from input to output layers. The product of inputs and weights z_{xx} and the addition of the products and bias terms z_x are defined by Eqs. (50) and (51) of Bach et al. (2015). The right diagram shows the backward propagation of information from output to input layers with relevance $R_x^{(x)}$ shown for the feedforward network. The relevance is calculated using Eqs. 10–14 below, which are based on Eqs. (57), (58), and (62) of Bach et al. (2015). Then, z_{xx} and z_x obtained in the feedforward network are used to calculate backpropagating information for the connected nodes in the upstream layer. The ratio of z_{xx} in the upstream layer to the input before activation z_x is calculated, and then multiplied by the relevance of the node in the layer. The relevance of a node is the sum of the abovementioned ratio over all the connections to the node. Specifically,

$$R_l^{(5)} = f(x), \quad l = 1 \ , \tag{10}$$

$$R_k^{(4)} = \sum_{l=1}^{1} \frac{z_{kl}}{z_l} R_l^{(5)}, \ k \in \{1, \dots, 128\} \ , \tag{11}$$

$$R_j^{(3)} = \sum_{k=1}^{128} \frac{z_{jk}}{z_k + \mathrm{sign}(\varepsilon, z_k)} R_k^{(4)},\ j \in \{1, \dots, 128\}\ , \tag{12}$$

$$R_i^{(2)} = \sum_{j=1}^{128} \frac{z_{ij}}{z_j + \mathrm{sign}(\varepsilon, z_j)} R_j^{(3)},\ i \in \{1, \dots, 512\}\ , \text{ and} \tag{13}$$

$$R_d^{(1)} = \sum_{i=1}^{512} \frac{z_{di}}{z_i + \mathrm{sign}(\varepsilon, z_i)} R_i^{(2)},\ d \in \{1, \dots, 897\}\ , \tag{14}$$

where $\mathrm{sign}(\varepsilon, z_i)$ gives the sign of z_i to the absolute value of ε. This constant ε is introduced to prevent the relevance from vanishing or exploding. For appropriate correction, ε was set to $\varepsilon = 10^{-6}$, which is about 1/100 of absolute value of z_x. These calculations are repeated layer-by-layer toward the upstream to obtain the relevance of the input layer $R_d^{(1)}$.

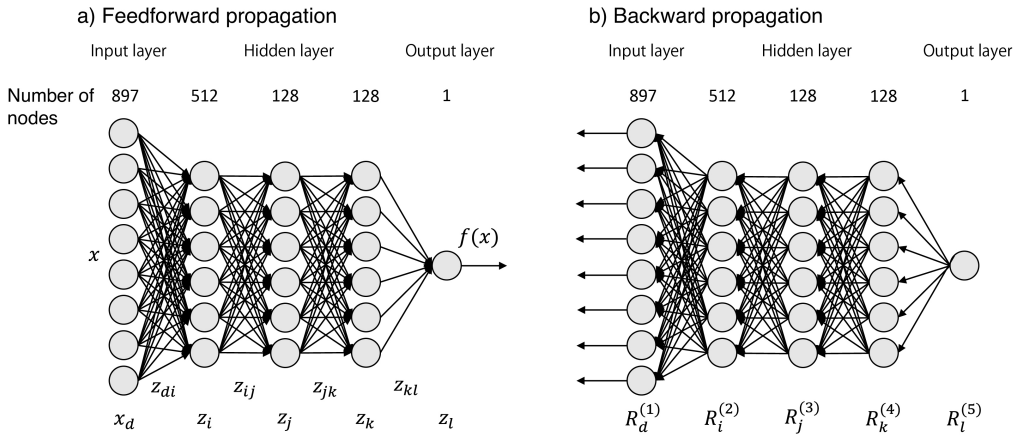

Figure A1. Feedforward propagation (left) and back propagation (right) for an MLP model with a one-year precipitation time series.

Acknowledgement

We are grateful for the public access and maintenance of precipitation and stage data provided by Ministry of Land, Infrastructure, Transport and Tourism, and Japan Meteorological Agency. We would also like to thank River Division, Civil Engineering department of Kochi Prefecture

for providing detailed datasets. This study was partly supported by the Social Implementation Program on Climate Change Adaptation Technology (SI-CAT) of the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT).

References

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, KR., Samek, W. (2015), On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLoS ONE, 10, (7), e0130140.

Chanu, S. N., & Kumar, P. (2018), Modelling of Daily Rainfall-Runoff Using Multi-Layer Perceptron Based Artificial Neural Network and Multi-Linear Regression Techniques in A Himalayan Watershed. Indian Journal of Hill Farming. 31, 1, 166-176.

Dozat, Timothy (2016), INCORPORATING NESTEROV MOMENTUM INTO ADAM. ICLR 2016 workshop paper, 107, reviewer 10.

Gorr, W. L, Nagin, D, & Szczypula, J. (1994), Comparative study of artificial neural network and statistical models for predicting student grade point averages. International Journal of Forecasting, 10, 17-34.

Hitokoto, M., Sakuraba, M., & Sei, Y. (2017), DEVELOPMENT OF THE REAL-TIME RIVER STAGE PREDICTION METHOD USING DEEP LEARNING. Journal of JSCE. 5, 422-429.

Intergovernmental Panel on Climate Change (IPCC), Climate Change 2013, The Physical Science Basis, https://archive.ipcc.ch/report/ar5/wg1/.

Japan Meteorological Agency (JMA) (2020), https://www.jma.go. jp/jma/kishou/books/hakusho/2020/index1. html

Kato, R., Shimose, K., & Shimizu, S. (2018), Predictability of precipitation caused by linear precipitation systems during the July 2017 Northern Kyushu heavy rain event using a cloud-resolving numerical weather prediction model. J. Disaster Research, Vol. 13,

No. 5.

Kawase, H., Yamaguchi, M., Imada, Y., Hayashi, S., Murata, A., Nakaegawa, T., Miyasaka, T, & Takayabu, I. (2021), Enhancemnt of extremely heavy precipitation induced by Typhoon Hagibis (2019) due to historical warming. SOLA, Vol 17A, 7-13, doi:10. 2151/sola. 17A-002.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013), Rectifier Nonlinearities Improve Neural Network Acoustic Models. Stanford University, CA 94305 USA.

Ministry of Agriculture, Forestry and Fisheries, https://www.rinya.maff.go.jp/j/keikaku/genkyou/h29/attach/pdf/1-3.pdf

Ministry of Land, Infrastructure, Transport and Tourism (MLIT), river maintenance plan, overview of Watari-gawa basin.
http://www.skr.mlit.go.jp/nakamura/seibikeikaku/about/outline. html

Nakane, H., Wakatsuki, Y. (2018), Startup of Deep Learning Application to Environmental Research. Kochi University of Technology research bulletin. 15, 1, 111-120. http://hdl. handle. net/10173/1949

Nakane, H., Wakatsuki, Y., Yamamoto, K., Takeda, T., Hashino, T. (2019), Application of Deep Learning to River Disaster Prevention and Environmental Conservation - on the Shimanto River and Kagami River Water Levels, and the Ohdo Dam Inflow of the Niyodo River -. Kochi University of Technology research bulletin. 16, 1, 227-244. http://hdl. handle. net/10173/00002125

Oluwatobi, A., Gbenga, O., Joy, A., & Oluwole, A. (2018), MODELING AND SIMULATION OF RIVER DISCHARGE USING ARTIFICIAL NEURAL NETWORKS. Journal of Science. 20, 2, 362-370.

Rao, G. S., & Giridhar, M. V. S. S. (2016), Daily Runoff Forecasting using Artificial Neural Network. International Journal of Scientific & Engineering Research. 7, 6, 478-484.

Rumelhart, D. E., Hinton, G., & Williams, R. J. (1986), Learning representations by back-

551 propagating errors. Nature, 323(6088), 533–536. https://doi.org/10.1038/323533a0
552 Shen, C. (2018), A Transdisciplinary Review of Deep Learning Research and Its Relevance for
553 Water Resources Scientists. Water Resources Research. 561, 918-929.
554 The Japan Times, (2019), At least 35 killed and 17 missing after Typhoon Hagibis tears through
555 country, flooding rivers and submerging cities. https://www. japantimes. co.
556 jp/news/2019/10/13/national/typhoon-hagibis-japan-tokyo/#. XdjKFuj7SUn
557 Uchida, T., Akamatsu, Y., Suzuki, Y., Moriguchi, S., Oikawa, Y., Shirahata, H., & Izumi, N.
558 (2021), Special issue on the heavy rain event of July 2018 in western Japan. Journal of
559 JSCE. vol. 9, 1-7.