Ensemble Flash Flood Predictions Using a High-Resolution Nationwide Distributed Rainfall-Runoff Model: Case Study of the Heavy Rain Event of July 2018 and Typhoon Hagibis in 2019

Takahiro Sayama[1]

Corresponding author

Email: sayama.takahiro.3u@kyoto-u.ac.jp

Masafumi Yamada[1]

Email: yamada.masafumi.7n@kyoto-u.ac.jp

Yoshito Sugawara[2]

Email: 4410sy@gmail.com

Dai Yamazaki[3]

Email: yamadai@rainbow.iis.u-tokyo.ac.jp

(Institutional addresses)

[1] Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

[2] Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8530, Japan

[3] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Abstract

The heavy rain event of July 2018 and Typhoon Hagibis in October 2019 caused severe flash flood disasters in numerous parts of western and eastern Japan. Flash floods need to be predicted over a wide range with long forecasting lead time for effective evacuation. The predictability of flash floods caused by the two extreme events are investigated by using a high-resolution (~150 m) nationwide distributed rainfall-runoff model forced by ensemble precipitation forecasts with 39-h lead time. Results of the deterministic simulation at nowcasting mode with radar and gauge composite rainfall could reasonably simulate the storm runoff hydrographs at many dam reservoirs over western Japan for the case of heavy rainfall in 2018 (F18) with the default parameter setting. For the case of Typhoon Hagibis in

2019 (T19), a similar performance was obtained by incorporating unsaturated flow effect in the model applied to Kanto region. The performance of the ensemble forecast was evaluated based on the bias scores and the relative operating characteristic curves, which suggested the higher predictability in peak runoff for T19. For the F18, the uncertainty arises due to the difficulty in accurately forecasting the storm positions by the frontal zone; as a result, the actual distribution of the peak runoff could not be well forecasted. Overall, this study showed that the predictability of flash floods was different between the two extreme events. The ensemble spreads contain quantitative information of predictive uncertainty, which can be utilized for the decision making of emergency responses against flash floods.

Keywords

Ensemble forecasting, Flash floods, Rainfall-Runoff-Inundation model, Typhoon Hagibis, Uncertainty, Flood forecasting, Quantitative precipitation forecasting, Saturated subsurface flow, Distributed hydrological model, Meso-scale Ensemble Prediction System (MEPS)

Introduction

In the current decade, worldwide occurrences of floods and extreme rainfall become four times that in the 1980s (EASAC 2018; UNESCO, UN-Water 2020). In Japan, flood disasters

occur every year, causing devastating damages (Udmale et al. 2020). In the last two years in particular, the *Heavy Rain Event of July 2018* (hereafter F18) caused by an intensified Baiu frontal zone killed or missed 245 people in the western part of Japan (Cabinet Office 2018), and *Typhoon Hagibis* in October 2019 (hereafter T19) killed or missed 114 people in the eastern part of Japan (Cabinet Office 2019). The two extreme events caused flood damages in 315 rivers including 37 levee breaching sections by F18 (MLIT 2018) and 325 rivers including 142 levee breaching sections by T19 (MLIT 2019), mostly along small-to-medium sized rivers. In countries like Japan, which have a steep topography and upland catchments receiving heavy torrential rain, flash floods occur frequently. Flash floods are characterized by a rapid increase in flood peaks after rainfall onset (Collier 2007; Georgakakos 1986), whose intervals are shorter than 12 to 24 hours (Georgakakos 1986; Raynaud et al. 2015). During such flash floods, safe evacuation is difficult to be performed because of the short flood concentration time and road submersions (Terti et al. 2019; Vincendon et al. 2016). Thus, site-specific flash flood mapping and forecasting are required for safe evacuations (Collier 2007; Gourley et al. 2017; Hapuarachchi et al. 2011).

Early warnings for flash flood cannot focus only on well gauged main rivers. Many small-to-medium sized rivers are typically poorly gauged and get damaged more frequently. Therefore, selecting a single river basin for flash flood prediction is not practical, and instead

a large-scale distributed approach forecasting all river sections including ungauged rivers is required (Grimaldi et al. 2013; Javelle et al. 2014; Reed et al. 2007). With the recent advancements in quantitative precipitation estimate (QPE) and quantitative precipitation forecast (QPF) models, large-scale flood forecasting systems have been developed (Emerton et al. 2016). Such systems have been operated at the continental to nationwide scales; for example, the European Flood Awareness System (EFAS) (Bartholmes et al. 2009; Thielen et al. 2009) in Europe, the Community Hydrologic Prediction System (CHPS) in USA (Demargne et al. 2014), the Hydrological Forecasting System (HyFS) in Australia (Hapuarachchi et al. 2017), the Grid-to-grid Model (G2G) in England and Wales (Anderson et al. 2019; Price et al. 2012) and Scotland (Cranston et al. 2012), and AIGA in France (Javelle et al. 2016). For site-specific flood predictions at a fine scale, for example, EC-JRC provides a rainfall-driven flash flood indicator within the EFAS framework called the European Precipitation Index based on Climatology (EPIC) (Alfieri and Thielen 2015) and a runoff-driven indicator called the European Runoff Index based on Climatology (ERID) (Raynaud et al. 2015). The NWS has adopted the Flash Flood Guidance (FFG), which is based on the estimated amount of rainfall causing floods at a specific site by running a hydrologic model in the backward mode (Clark et al. 2014; Georgakakos 2006).

Recent advances in high-performance computing and geographic information systems

have motivated the application of large-scale distributed hydrologic models to predict flash floods at any river sections including ungauged sites. Gourley et al. (2017) reported the latest research project called the Flooded Locations and Simulated Hydrographs (FLASH) across the Conterminous United States. One of their targets was to apply a distributed hydrologic model with 1-km resolution to estimate flood peak discharge forced by the latest QPE. Such model applications are important from the viewpoint of hydrological science because physically sound parsimonious distributed modes are necessary for the purpose as well as finding the dominant runoff processes (Antonetti et al. 2019), regionalizing parameters (Vergara et al. 2016) and assessing the impact of dataset resolution (Lovat et al. 2019).

As noted by many previous studies, the largest uncertainty is associated with the precipitation forecast (Hapuarachchi et al. 2011). In particular, even with the state-of-art NWPs, accurately predicting severe storms with sufficient prediction lead time is challenging. Instead of deterministic forecasting, probabilistic forecasting with EPS has advanced in the last decade (Cloke and Pappenberger 2009; Wu et al. 2020). When the possibility of a severe storm becomes high because a typhoon is approaching or because a frontal line is stagnant, if we can predict the occurrence probability of flash floods leading to local damage, we could prepare for the extreme weather as a society (Terti et al. 2019). For ensemble flood predictions, global scale EPSs such as European Centre for Medium-Range Weather Forecasts

– Ensemble of forecast (ECMWF-ENS) have been widely used (Alfieri et al. 2014). Meanwhile, for ensemble flash flood predictions, some case studies have demonstrate the importance of meso-scale ensemble forecasting (Alfieri et al. 2012; Hsiao et al. 2013; Roux et al. 2020; Ushiyama et al. 2014) and their combinations by the Meteorological Ensemble Forecast Processor (MEFP) approach adopted by Hydrologic Ensemble Forecast Service (HEFS) (Brown et al. 2014a; Brown et al. 2014b).

In Japan, the Japan Meteorological Agency (JMA) recently started operation of the Meso-scale Ensemble Prediction System (MEPS) with a 39-h lead time with 21 ensemble members based on a meso-scale numerical weather prediction model with a 5-km resolution. The test operation began just before July 2018, and official operation started since June 2019. Since heavy storm is spatiotemporally concentrated in mountainous areas and the flood concentration time is much shorter in Japan, we believe it is important to use fine spatial resolution forecasting. For hydrological modeling, we used the Rainfall-Runoff-Inundation (RRI) model (Sayama et al. 2012; Sayama et al. 2015a; Sayama et al. 2015b) applied recently to all over Japan by dividing the whole of Japan into 14 regions with a 5 sec (about 150 m) spatial resolution. We suppose that the impact of complex rainfall-runoff phenomena become comparatively less important during such extreme flood events; then the simple model structure and the default RRI model parameter setting, which considers only saturated

subsurface flow and overland flow, may be able to reproduce the flash floods over a wide range. Because the two extreme flood events described above affected large areas over Japan, and because the latest ensemble forecasting product is now available, it is important to investigate the performance of the newly developed flash flood forecasting model by taking the two extreme events as a case study.

The objective of this study is to evaluate the predictability of flash floods caused by the two extreme events. In particular, the specific questions addressed in this study are described below.

1) Can the default RRI model, representing lateral saturated subsurface flow and surface flow, forced by radar and gauge composite rainfall product reproduce the observed storm runoff hydrographs at many upstream dam reservoirs over wide ranges?

2) Do the spatial distributions of peak runoff (i.e., peak discharge normalized by the upstream contributing area) correspond to the actual damage by the flash floods?

3) How well can we forecast the spatial distributions of peak runoff with a 39-h lead time based on ensemble precipitation forecasting?

Methods

Nationwide Application of the RRI Model over Japan

In this study, the RRI model was applied to the whole of Japan with a spatial resolution of approximately 150 m (5 sec). The RRI model is a two-dimensional model, which can simulate both rainfall-runoff and flood inundation simultaneously (see the details in the supplement). For model application, we used the Japan Flow Direction Map (J-FlwDir) developed by Yamazaki et al. (2018). The dataset is based on a digital elevation model and water map provided by the Geospatial Information Authority of Japan and prepared with a 30 m (1 sec) spatial resolution. The dataset contains elevation, flow direction and upstream contributing area, similar to HydroSHEDs available worldwide based on satellite-based topographic data. To upscale the flow direction of J-FlwDir, we used an upscaling algorithm developed by Masutani *et al.* (2006), which maintains the location of the main stream regardless of the upscaling. The RRI assumes that the river starts when the upstream contributing area becomes greater than 50 grid-cells (i.e., 1.1 km^2). Hence, all river channels including small tributaries are explicitly modelled by the RRI. The cross sections of the rivers are assumed to be rectangular, whose widths and depths were estimated using the following formulae.

$$W = C_W A^{S_W} \tag{1}$$

$$D = C_D A^{S_D} \tag{2}$$

where A is the upstream contributing area [km^2] and the values of C_W, S_W, C_D and S_D were estimated from our previous model application in Japan: C_W = 4.73, S_W = 0.4, C_D = 1.57, and S_D = 0.3 (Sayama et al. 2017). The depth parameters were intentionally set to be larger than many actual cases, so that the model had enough river cross section capacity. With this setting, the RRI model does not calculate the overtopping inundation effect and focuses primarily on the rainfall-runoff processes (i.e., river discharge). In terms of the model parameter settings, we attempted to prepare a parsimonious model to simulate extreme events. As described in the supplement, by taking d_m = 0 in Equation (S1), the number of parameters becomes only four. The default uncalibrated parameters are shown in Table 1. The model parameters were set to avoid strong non-linearity between storage and discharge relationship to ensure the model represented quick runoff responses to storm events. With the settings (i.e., d_m = 0), the model may overestimate the discharge if catchments store large amount of rainfall in soil layers.

Flood Events in July 2018 and October 2019

This study focuses on two extreme flood events that occurred in July 2018 (F18) and October 2019 (T19). F18 was caused by long-lasting Baiu frontal rain covering most of the western part of Japan. T19 was due to the Typhoon Hagibis which mostly damaged the eastern part of Japan. Detailed information regarding the meteorological conditions and

analysis can be obtained from recent studies (Enomoto 2019; Kotsuki et al. 2019; Takemi and Unuma 2020; Tsuguti et al. 2018). The periods of the two simulations were 0:00 July 5, 2018 to 0:00 July 9, 2018 (JST) for F18 and 9:00 October 11, 2019 to 9:00 October 14, 2019 (JST) for T19. The input rainfall data are JMA's radar and gauged composite products, whose spatial and temporal resolutions are 1 km and 30 min, respectively.

In the ensemble flood forecasting experiments, we used the MEPS dataset provided by JMA. The forecasting lead time of this product was 39 h with a spatial resolution of approximately 5 km. The JMA non-hydrostatic model (NHM) was used for forecasting with 21 members by perturbating initial conditions. Although the forecasting is updated every 6 h, we decided to focus only on a single initial time for each event to cover the whole event. The period of the MEPS forecasting rainfall is 21:00 July 5, 2018 to 12:00 July 7, 2018 (JST) for F18, and 9:00 October 12, 2019 to 0:00 October 13, 2019 (JST) for T19.

After the model was run with the simulation and forecasting modes, model performance was evaluated by comparing the observed and simulated discharge at dam reservoirs. Note that in this nationwide RRI model application, so far any dam reservoir operation model has not been incorporated. Furthermore, to visualize the spatial distribution of flood discharge, we compute peak discharge at all river grid-cells and normalized it by the upstream contributing area (peak runoff). The evaluations were conducted focusing on all the river grid-cells over

the target areas. The simulation of F18 was conducted in Kansai, Chugoku, Shikoku and Kyushu regions. The total area of these regions is 125,864 km^2, which is composed of 6,364,492 grid-cells including 674,291 river grid-cells. The simulation of T19 was conducted in Kanto, Tohoku and part of Hokuriku regions. The total area of these regions is 121,129 km^2, which is composed of 6,435,707 grid-cells including 751,102 river grid-cells.

Verification Metrics of Simulated Hydrographs

The performance of the hydrologic model was evaluated by the following three measures: the Kling-Gupta efficiency (*KGE*) (Gupta et al. 2009), the Nash-Sutcliffe efficiency (*NSE*) and the relative peak error (*PE*) defined as (3), (5) and (8).

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\alpha-1)^2} \tag{3}$$

$$r = \frac{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})(Q_s^t - \overline{Q_s})}{\sqrt{\left(\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2\right)\left(\sum_{t=1}^{T}(Q_s^t - \overline{Q_s})^2\right)}} \tag{4}$$

$$\beta = \frac{\overline{Q_s}}{\overline{Q_o}} \tag{5}$$

$$\alpha = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Q_s^t - \overline{Q_s})^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2}} \tag{6}$$

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_s^t - Q_o^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2} \tag{7}$$

$$PE = \frac{Q_{p,s}}{Q_{p,o}} \tag{8}$$

where Q_o^t is the observed discharge at time *t*; Q_s^t is the simulated discharge at time *t*; $\overline{Q_o}$ is

the mean observed discharge in an event; $\overline{Q_s}$ is the mean simulated discharge in an event; β is a measure of bias; α is a measure of the variability error; r is the correlation coefficient between Q_o^t and Q_s^t; $Q_{p,s}$ is the simulated peak discharge; and $Q_{p,o}$ is the observed peak discharge. *NSE* has been widely used for hydrologic modeling, while *KGE* is being increasingly used as an alternative. *KGE* was developed to decompose the normalized mean squared error represented by the *NSE* and addressed a shortcoming of the *NSE*, which is maximized when $\alpha = r$ (Gupta et al. 2009).

Verification Metrics of Ensemble Forecasting

To evaluate ensemble forecasting, this study uses the following the bias score (*BI*) defined as (9) and the relative operating characteristic (ROC) curve.

$$BI = \frac{\sum_{k=1}^{N} Q_{p,f}(k)}{\sum_{k=1}^{N} Q_{p,s}(k)} \tag{9}$$

where $Q_{p,f}(k)$ is the ensemble mean of the forecasted peak discharge at each river grid cell k; and N is the number of river grid cells. The *BI* score is evaluated depending on the ranges of $Q_{p,s}(k)$, which is the simulated peak discharge by the nowcasting mode used as a reference.

The ROC curve is obtained by plotting the false alarm rate $F(p_t)$ on the x-axis and the hit rate $H(p_t)$ on the y-axis. For ensemble forecasting, a continuous variable, in this case, the peak

runoff at each river grid-cell, must be converted to binary data based on a certain threshold (e.g., peak runoff exceeding 20 mm/h (1) or not (0)) (Toth et al. 2003). Note that both the false alarm rate and hit rate depend on the decision threshold probability p_t within the range of ensemble members. $F(p_t)$ and $H(p_t)$ decrease from 1 to 0 as the decision threshold p_t increases from 0 to 1. The ROC curve is close to the perfect forecast $H(p_t) = 1$ and $F(p_t) = 0$ for better performance.

Results

Simulation Results with Radar Rainfall

We run the RRI model using the observed radar rainfall data during the F18 event with the default parameter setting (hereafter F18-UC). Figure 1 compares the observed and simulated hydrographs at ten dam reservoirs. The results suggest that the uncalibrated RRI model simulates the flood inflows generally well for the F18 event regardless of the geographic locations. Figure 2, Table 2 and Table 3 summarize the quantitative evaluations of the simulations. The average and standard deviations of KGE and NSE for F18-UC are 0.78±0.11 and 0.75±0.10, respectively. The boxplot of Figure 2 further suggests that among the three components of KGE, the correlation coefficients are approximately one (r = 0.94±0.04), suggesting that the shapes of the hydrographs are well represented. In contrast, the standard

deviations of α and β are relatively high (α = 1.02±0.17, β = 1.00±0.16). Since the averages of α and β are close to one, there is no consistent overestimation or underestimation. In terms of simulated peak discharge, the result of *PE* = 0.87±0.15 indicates 87 % of underestimations of the peak discharge by the simulation.

The same experiment was conducted for T19 in Kanto, Tohoku and Hokuriku regions. The uncalibrated model (T19-UC) could not represent the observed hydrographs in this case. Figure S1 shows that at the eight catchments, the simulation generally overestimated the observed hydrographs. The values of α, β and *PE* were higher than one (α = 1.45±0.29, β = 1.36±0.22, *PE* = 1.31±0.29) indicating hydrograph variance, total volumes and peak discharges were all overestimated by 30 to 40 percent. The shapes of the hydrographs represent the observed patterns, which are quantified as r = 0.96±0.02. As a result, both KGE and NSE are much lower (0.42±0.36 and 0.50±0.46) for T19-UC compared with those for F18-UC even for the same model settings. For the case of T19 in the Kanto region, we changed the model parameter setting by introducing unsaturated flow component (see Table 1). By introducing the unsaturated flow component, the model performance improved (Figure 3) with KGE = 0.76±0.16 and NSE = 0.86±0.10, and the performance was almost equivalent to that for the F18-UC case, as shown in Figure 2.

Ensemble Forecasting Results at Dam Reservoirs

Hydrographs in Figure 1 and Figure 3 show forecasted dam inflows by the RRI model with MEPS rainfall. The two sets of graphs indicate higher convergence among the 21 ensemble members for T19 compared to F18. For the F18 case, different members showed different hydrograph patterns with larger spreads in the forecasted peak runoff. On the other hand, for the T19 case, the flood peaks were well estimated about half day before the peak arrivals.

To analyze the characteristics of the forecasted peak runoff, Figure 4 and Figure 5 show plots of the descending orders of the forecasted peak runoff by 21 ensemble members and the simulated peak runoff estimated with the observed radar rainfall. The figures can show how the peak runoff is spread (a steeper line shows higher variation compared to a flatter line which represents a smaller variation among the members). Furthermore, if the simulated peak runoff points are within the range and are positioned close to the center, the median of ensemble forecast is suggested to cover what actually happened. The results support the above described point between F18 and T19; the former case has higher ensemble variations than the latter case.

Evaluations Over the Regions Including Tributaries

Figure 6 and Figure 7 compare the forecasted and simulated peak runoff over the entire

simulation domains. Figure 7 shows that the simulated peak runoff for T19 can generally capture the areas of flash flood tributaries such as the tributaries in Tochigi prefecture with white cross marks representing levee breaching points along the tributaries. Furthermore, it shows the higher peak runoff in west Kanto including the upstream of the Arakawa river. Figure 7 shows the forecasted result based on MEPS rainfall at 9:00 on October 12. The spatial variability such as high peak runoff in Tochigi prefecture and the upstream of the Arakawa river is well represented.

The performance of the F18 shown in Figure 6 is not as clear as that of T19. First, the simulated peak runoff and actual flood damages did not agree very well. According to our simulation, the rivers in eastern Kochi and norther Kyushu islands show comparatively higher peak runoff. However, flash flood damages were concentrated mostly in the Chugoku region and western Shikoku island during F18.

To quantify the patterns, we used the *BI* and the ROC curve. The evaluations were conducted focusing on all river grid-cells over the target areas. In case of the *BI* score shown in Figure 8, the x-axis shows the peak runoff based on the simulation mode and the y-axis depicts the corresponding biases (forecasted / simulated peak discharges). The shades show the ranges of standard deviations of the mean biases for each region. For F18, the computed biases were 0.5-0.7 at the high peak runoff ranges (10-25 mm/h), suggesting underestimation

by the ensemble forecasting. Smaller biases (i.e., *BI* closer to 1.0) were confirmed for T19. Especially for the case of Kanto, for example, the biases were nearly one for almost the entire range, which indicates the high predictability of peak runoff. The ROC curves in Figure 9 show also higher accuracy in T19. Moreover, the ROC curves from the three regions for T19 almost overlap each other. All the above evaluations are normally performed not for a single event but should be performed with many flood events or for long-terms. The present evaluation does not indicate the overall model performance, but the figures are used to quantify the results of the ensemble forecasting of the case studies.

Discussions

Can the Default RRI Model Reproduce the Observed Storm Events?

For a regional flash flood prediction system, since it is impossible to calibrate hydrologic model parameters at individual river basins, limited sets of parameters should be applied over a wide range and to produce reliable results (Collier 2007). The simulation results for F18 showed that the default parameter representing lateral subsurface flow and surface flow perform fairly well to reproduce the observed hydrographs in many dam reservoirs. In case of T19, the same model with the same default parameter overestimated the flood peaks and the performance of T19-UC was unacceptable. Introducing the unsaturated soil layer into the

model was necessary to reproduce the patterns observed hydrographs of the Kanto region.

Here, we discuss the possible reason of the difference is related to the geologic settings in different regions. Large parts of the Chugoku and Shikoku regions affected by F18 belong to Granite rocks, Mesozoic or Paleozoic formations. The volcanic landscapes formed in Paleogene and Cretaceous are mostly distributed in central and western Japan. In contrast, mountainous area in northeastern Japan, including Kanto and Tohoku regions affected by T19, is dominated mostly by the volcanic rocks formed in the Quaternary and Neogene. The geologically younger catchments typically exhibit stable groundwaterm while the geologically older catchments show more flashy runoff patterns (Yoshida and Troch 2016), (Shimizu 1980; Mushiake et al. 1981). These previous studies follow data driven approach focusing more on flow duration curves and baseflow. This study indicated the effect of geology on storm runoff, previously evaluated at small catchments in comparative studies (e.g., Onda et al. 2001; Onda et al. 2006) or modeling studies focusing on a selected river basin (Sayama et al. 2017)

Do the Spatial Distributions of Peak Runoff Correspond to Flash Flood Damages?

The simulated peak runoff distribution in Figure 7 can help to visualize severely flood areas. The results of T19 showed that Tochigi prefecture and the upper Arakawa river basin reached

about 30 mm/h in the peak runoff. Furthermore, it shows that the peak runoff exceeded 40 mm/h in Marumori-township in Miyagi Prefecture in Tohoku region, where severe damage was reported due to flash floods with many slope failures and debris flow.

To quantify the relationship between the peak runoff and actual flood damages, we collected location information of levee breaching points in small-to-medium sized rivers mainly in Tochigi prefecture by T19. Figure 10 plots the relationship between upstream contributing areas and the corresponding peak runoff simulated by the nowcasting mode at the levee breaching points. From this plot, we can roughly estimate that the levee breaches occur when the peak runoff exceeds approximately 30 mm/h for the tributaries with upstream areas contributing less than 300 km^2. The figure also indicates that the threshold becomes smaller for larger catchments. This pattern is common; the peak runoff tends to be smaller due to the normalization by the upstream contributing area (Amponsah et al. 2018). Meanwhile, the bank-full discharge (i.e., the threshold) also varies depending on climatic and other geographic conditions as well as the status of river management works such as constructions of flood defiance structures. In fact, for the case of T18, the spatial pattern of high peak runoff did not represent the actual distribution of the flash flood damage over western Japan (Figure 6). To advance the system, evaluating the peak runoff relative to the actual local bank-full capacity or converting the peak runoff to stream water levels are necessary. The other

approach is to evaluate the frequency of the peak runoff at each location compared with the climatology as presented by previous studies (Alfieri et al. 2012; Yoshimura et al. 2008).

How Well Can We Forecast the Peak Runoff Distributions?

Operational flood forecasting in Japan typically focuses on stream water levels at important cross sections, where the water levels are monitored. The lead time of flood forecasting is normally 3 to 6 h. Recently, the JMA released a new type of flood forecasting information based on hydrologic simulation with a spatial resolution of 1 km covering all over Japan. The predicted streamflow is converted to a flood risk index and its lead time is set to be 6 h. Although our approach is similar to the JMA's new product, the demonstrated flash flood predictions for the two extreme cases with total of 21 ensemble members with 39-h lead time could provide different type of information that could be useful for the better preparedness. The estimated uncertainty bands by the two events were contrasting: smaller spreads for T19 and larger spreads for F18. The high predictability of the T19 event in particular is discussed from meteorological viewpoint. Takemi and Unuma (2020) reported that the moist absolutely unstable status under very humid conditions and a sufficient precipitable water were responsible for the heavy rainfall. For the F18 case, Kotsuki et al. (2019) indicated high predictability of intense rainfall at the synoptic-scale with long lead-time (3 days) using their

data assimilation system. However, forecasting accurate locations of the rain band causing heavy rainfall still has a high uncertainty (Matsunobu and Matsueda 2019). For this particular case, in addition to focus on the ensemble mean, we should pay attention to the worse cases (e.g., the runoff map of the fifth order). Further investigations are required on the prediction accuracy and how to utilize such ensemble flood forecasting information.

Conclusions

This study examined the predictability of flash floods on a nationwide scale in Japan using the new operational meso scale ensemble precipitation forecast and a high-resolution distributed rainfall-runoff model. Two extreme events were selected as a case study, since both of them caused levee breaching and overtopping in various regions because of different rainfall mechanisms; i.e., frontal rain in 2018 and the typhoon in 2019. Based on the numerical experiment at the nowcasting and forecasting modes, we obtained the following conclusions.

The nationwide model could reasonably reproduce extreme floods at many dam reservoirs over the wide range in the nowcasting mode without individual parameter tuning. However, in certain areas, such as the Kanto region, remarkable basin storage effects were observed even under extreme events. The model and parameters had to be tuned to reflect these effects in these areas. In terms of the forecasting with lead time, the predictability was

different between the two event cases. In case of the Typhoon Hagibis in October 2019, the prediction accuracy was high, and the spatial distribution of peak runoff estimated by the ensemble mean corresponded well with the results of the nowcasting mode. For the frontal heavy rain in July 2018, the detailed locations of high peak runoff could not be forecasted by the ensemble mean. Instead, the spread of 21 ensemble members showed the flash flood potential areas and their possible magnitudes. The large spread quantifies the higher uncertainty in the predictions.

The model presented in this study has not been operated on a real-time basis. The real-time system and its continuous verification should be performed in future studies. If such a system can be developed, stochastic, high-resolution, long-lead time flash flood predictions can be realized. Such a system can be useful for pre-imaging the situations of flash flood disasters, especially before a typhoon strikes or when a frontal rain stagnates, to realize early and safe evacuation and other preparations.

Abbreviations

BI: BIas score; BoM: Bureau of Meteorology; CHPS: Community Hydrologic Prediction System; CORNUS: the Conterminous United States; ECMWF-ENS: European Centre for

Medium-Range Weather Forecasts – Ensemble of forecast; EC-JRC: Joint Research Center of the European Commission; EFAS: European Flood Awareness System; EPIC: European Precipitation Index based on Climatology; EPS: Ensemble Prediction System; ERID: European Runoff Index based on Climatology; FFC: Flood Forecasting Centre; FFG: Flash Flood Guidance; FLASH: Flooded Locations and Simulated Hydrographs; G2G: Grid-to-grid Model; HEFS: Hydrologic Ensemble Forecast Service; HydroSHEDs: Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales; J-FlwDir: Japan Flow Direction Map; KGE: Kling-Gupta Efficiency; MEFP: Meteorological Ensemble Forecast Processor; MEPS: Meso-scale Ensemble Prediction System; NHM: Non-hydrostatic Model; NSE: Nash-Sutcliffe efficiency; NWP: Numerical Weather Prediction; NWS: National Weather Service; QPE: Quantitative Precipitation Estimates; QPF: Quantitative Precipitation Forecast; PE: relative Peak Error; ROC: Relative Operating Characteristic; RRI: Rainfall-Runoff-Inundation

Declarations

Availability of data and material

The simulation datasets created during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interest.

Funding

This work was supported by Cabinet Office, Government of Japan, Cross-ministerial Strategic Innovation Promotion Program (SIP), “Enhancement of Societal Resiliency against Natural Disasters” (PI: Prof. Hori) and JSPS KAKENHI Grant Number 19H02248 (PI: Sayama) and 19K24678 (PI: Prof. Nihei).

Authors' contributions

TS proposed the topic, conceived and designed the study. MY and YS collaborated the data analysis and field work. DY carried out the hydro-geographic data used in this study. All authors read and approved the final manuscript.

Acknowledgements

The authors express deep gratitude to FRICS of Japan and Mitsui Consultants Co. Ltd. for their technical assistants, discussions and data sharing. The MEPS forecasting data was kindly

provided by Japan Meteorological Agency. We thank also Prof. Tachikawa at Kyoto University for his leading the SIP project and research advice.

References

Alfieri L, Pappenberger F, Wetterhall F, Haiden T, Richardson D, Salamon P (2014) Evaluation of ensemble streamflow predictions in Europe. Journal of Hydrology 517:913-922.

Alfieri L, Thielen J (2015) A European precipitation index for extreme rain-storm and flash flood early warning. Meteorological Applications 22:3-13.

Alfieri L, Thielen J, Pappenberger F (2012) Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. Journal of Hydrology 424-425:143-153.

Amponsah W, Ayral PA, Boudevillain B, Bouvier C, Braud I, Brunet P, Delrieu G, Didon-Lescot JF, Gaume E, Lebouc L, Marchi L, Marra F, Morin E, Nord G, Payrastre O, Zoccatelli D, Borga M (2018) Integrated high-resolution dataset of high-intensity European and Mediterranean flash floods. Earth Syst Sci Data 10:1783-1794.

Anderson SR, Csima G, Moore RJ, Mittermaier M, Cole SJ (2019) Towards operational joint river flow and precipitation ensemble verification: considerations and strategies given limited ensemble records. Journal of Hydrology 577.

Antonetti M, Horat C, Sideris IV, Zappa M (2019) Ensemble flood forecasting considering dominant runoff processes - Part 1: Set-up and application to nested basins (Emme,

Switzerland). Natural Hazards and Earth System Sciences 19:19-40.

Bartholmes JC, Thielen J, Ramos MH, Gentilini S (2009) The european flood alert system EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrol Earth Syst Sci 13:141-153.

Brown JD, He M, Regonda S, Wu L, Lee H, Seo D-J (2014a) Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 2. Streamflow verification. Journal of Hydrology 519:2847-2868.

Brown JD, Wu L, He M, Regonda S, Lee H, Seo D-J (2014b) Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 1. Experimental design and forcing verification. Journal of Hydrology 519:2869-2889.

Cabinet Office, Government of Japan (2018) On disaster damage by the Heavy Rainfall in July 2018 (in Japanese), http://www.bousai.go.jp/updates/h30typhoon7/pdf/310109_1700_h30typhoon7_01.pdf. Accessed 18 April 2020.

Cabinet Office, Government of Japan (2019) On disaster damage by the Typhoon No. 19 (Hagibis) in 2019 (in Japanese), http://www.bousai.go.jp/updates/r1typhoon19/pdf/r1typhoon19_45.pdf. Accessed 18 April 2020.

Cash JR, Karp AH (1990) A Variable Order Runge-Kutta Method for Initial-Value Problems with Rapidly Varying Right-Hand Sides. Acm Transactions on Mathematical Software

16:201-222.

Clark RA, Gourley JJ, Flamig ZL, Hong Y, Clark E (2014) CONUS-Wide Evaluation of National Weather Service Flash Flood Guidance Products. Weather and Forecasting 29:377-392.

Cloke HL, Pappenberger F (2009) Ensemble flood forecasting: A review. Journal of Hydrology 375:613-626.

Collier CG (2007) Flash flood forecasting: What are the limits of predictability? Quarterly Journal of the Royal Meteorological Society 133:3-23.

Cranston M, Maxey R, Tavendale A, Buchanan P, Motion A, Cole S, Robson A, Moore RJ, Minett A (2012) Countrywide flood forecasting in Scotland: challenges for hydrometeorological model uncertainty and prediction. In: Moore RJ, Cole SJ, Illingworth AJ (eds) Weather Radar and Hydrology, Proc. Exeter Symp., April 2011. International Association of Hydrological Sciences.

Demargne J, Wu L, Regonda SK, Brown JD, Lee H, He M, Seo D-J, Hartman R, Herr HD, Fresch M, Schaake J, Zhu Y (2014) The Science of NOAA's Operational Hydrologic Ensemble Forecast Service. Bulletin of the American Meteorological Society 95:79-98.

EASAC (European Academies' Science Advisory Council) (2018) Extreme Weather Events in Europe: Preparing for Climate Change Adaptation: An Update on EASAC's 2013 Study. https://easac.eu/publications/details/extreme-weather-events-in-europe/. Accessed 18 April 2020.

Emerton RE, Stephens EM, Pappenberger F, Pagano TC, Weerts AH, Wood AW, Salamon P, Brown JD, Hjerdt N, Donnelly C, Baugh CA, Cloke HL (2016) Continental and global scale flood forecasting systems. Wiley Interdisciplinary Reviews: Water 3:391-418.

Enomoto T (2019) Influence of the Track Forecast of Typhoon Prapiroon on the Heavy Rainfall in Western Japan in July 2018. Sola 15A:66-71.

Georgakakos KP (1986) On the Design of National, Real-Time Warning Systems with Capability for Site-Specific, Flash-Flood Forecasts. Bulletin of the American Meteorological Society 67:1233-1239.

Georgakakos KP (2006) Analytical results for operational flash flood guidance. Journal of Hydrology 317:81-103.

Gourley JJ, Flamig ZL, Vergara H, Kirstetter P-E, Clark RA, Argyle E, Arthur A, Martinaitis S, Terti G, Erlingis JM, Hong Y, Howard KW (2017) The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bulletin of the American Meteorological Society 98:361-372.

Grimaldi S, Petroselli A, Arcangeletti E, Nardi F (2013) Flood mapping in ungauged basins using fully continuous hydrologic–hydraulic modeling. Journal of Hydrology 487:39-47.

Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377:80-91.

Hapuarachchi HAP, Kabir A, Zhang XS, Kent D, Bari MA, Tuteja NK, Hasan MM, Enever D,

Shin D, Plastow K, Ahmad Z (2017) Performance evaluation of the national 7-day water forecast service. Proceedings of 22nd International Congress on Modelling and Simulation:1815-1821.

Hapuarachchi HAP, Wang QJ, Pagano TC (2011) A review of advances in flash flood forecasting. Hydrological Processes 25:2771-2784.

Hsiao L-F, Yang M-J, Lee C-S, Kuo H-C, Shih D-S, Tsai C-C, Wang C-J, Chang L-Y, Chen DY-C, Feng L, Hong J-S, Fong C-T, Chen D-S, Yeh T-C, Huang C-Y, Guo W-D, Lin G-F (2013) Ensemble forecasting of typhoon rainfall and floods over a mountainous watershed in Taiwan. Journal of Hydrology 506:55-68.

Javelle P, Demargne J, Defrance D, Pansu J, Arnaud P (2014) Evaluating flash-flood warnings at ungauged locations using post-event surveys: a case study with the AIGA warning system. Hydrological Sciences Journal 59:1390-1402.

Javelle P, Organde D, Demargne J, Saint-Martin C, de Saint-Aubin C, Garandeau L, Janet B (2016) Setting up a French national flash flood warning system for ungauged catchments based on the AIGA method. E3S Web Conf 7:18010.

Kotsuki S, Terasaki K, Kanemaru K, Satoh M, Kubota T, Miyoshi T (2019) Predictability of Record-Breaking Rainfall in Japan in July 2018: Ensemble Forecast Experiments with the Near-Real-Time Global Atmospheric Data Assimilation System NEXRA. Sola 15A:1-7.

Lovat A, Vincendon B, Ducrocq V (2019) Assessing the impact of resolution and soil datasets on flash-flood modelling. Hydrology and Earth System Sciences 23:1801-1818.

Masutani, K., Akai, K., Magome, J., 2006, A new scaling algorithm of gridded river networks, Journal of Japan Soc. Hydrol. & Water Resour., 19(2), 139-150 (in Japanese with English abstract).

Matsunobu T, Matsueda M (2019) Assessing the Predictability of Heavy Rainfall Events in Japan in Early July 2018 on Medium-Range Timescales. Sola 15A:19-24.

MLIT (Ministry of Land Infrastructure, Transport and Tourism) (2018) On disaster damage by the Heavy Rainfall in July 2018 (in Japanese), http://www.mlit.go.jp/common/001268344.pdf, Accessed 18 April 2020.

MLIT (Ministry of Land Infrastructure, Transport and Tourism) (2019) the Typhoon No. 19 (Hagibis) in 2019 (in Japanese), http://www.mlit.go.jp/common/001340439.pdf, Accessed 18 April 2020.

Mushiake, K., Takahashi, Y., and Ando, Y.: Effects of basin geology on river flow regime in mountainous areas of Japan, Proc. Jpn. Soc. Civ. Eng., 1981, 51–62, 1981, (In Japanese with English abstract)

Onda Y, Komatsu Y, Tsujimura M, Fujihara J-i (2001) The role of subsurface runoff through bedrock on storm flow generation. Hydrological Processes 15:1693-1706.

Onda Y, Tsujimura M, Fujihara J-i, Ito J (2006) Runoff generation mechanisms in high-relief mountainous watersheds with different underlying geology. Journal of Hydrology 331:659-673.

Price D, Hudson K, Boyce G, Schellekens J, Moore RJ, Clark P, Harrison T, Connolly E, Pilling C (2012) Operational use of a grid-based model for flood forecasting.

Proceedings of the Institution of Civil Engineers - Water Management 165:65-77.

Raynaud D, Thielen J, Salamon P, Burek P, Anquetin S, Alfieri L (2015) A dynamic runoff co-efficient to improve flash flood early warning in Europe: evaluation on the 2013 central European floods in Germany. Meteorological Applications 22:410-418.

Reed S, Schaake J, Zhang Z (2007) A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. Journal of Hydrology 337:402-420.

Roux H, Amengual A, Romero R, Bladé E, Sanz-Ramos M (2020) Evaluation of two hydrometeorological ensemble strategies for flash-flood forecasting over a catchment of the eastern Pyrenees. Natural Hazards and Earth System Sciences 20:425-450.

Sayama T, McDonnell JJ (2009) A new time-space accounting scheme to predict stream water residence time and hydrograph source components at the watershed scale. Water Resources Research 45:W07401.

Sayama T, Ozawa G, Kawakami T, Nabesaka S, Fukami K (2012) Rainfall–runoff–inundation analysis of the 2010 Pakistan flood in the Kabul River basin. Hydrological Sciences Journal 57:298-312.

Sayama T, Tatebe Y, Iwami Y, Tanaka S (2015a) Hydrologic sensitivity of flood runoff and inundation: 2011 Thailand floods in the Chao Phraya River basin. Natural Hazards and Earth System Science 15:1617-1630.

Sayama T, Tatebe Y, Tanaka S (2015b) An emergency response-type rainfall-runoff-inundation simulation for 2011 Thailand floods. Journal of Flood Risk

Management 10:65-78.

Sayama, T., Tanaka, S. and Takara, K.: Rainfall-Runoff Analysis of the Kinu River Flood with a Distributed Model and Time-Space Accounting Scheme, J. Japan Soc. Hydrol. And Water Resour., Vol. 30, No. 3, 2017, doi: https://doi.org/10.3178/jjshwr.30.161, (In Japanese with English abstract)

Shimizu, T,: Relation between Scanty Runoff from Mountainous Watershed and Geology, Slope and Vegetation, Bulletin of the Government Forest Experiment Station, No. 301, pp. 109-128, 1980.

Tachikawa Y, Nagatani G, Takara K (2004) Development of storage-discharge relationship equation incorporating saturated-unsaturated flow mechanism (in Japanese). Annual Hydraulic Engineering, Japan Society for Civil Engineers 48:7-12.

Takemi T, Unuma T (2020) Environmental Factors for the Development of Heavy Rainfall in the Eastern Part of Japan during Typhoon Hagibis (2019). Sola.

Terti G, Ruin I, Gourley JJ, Kirstetter P, Flamig Z, Blanchet J, Arthur A, Anquetin S (2019) Toward Probabilistic Prediction of Flash Flood Human Impacts. Risk Anal 39:140-161.

Thielen J, Bartholmes J, Ramos MH, de Roo A (2009) The European Flood Alert System – Part 1: Concept and development. Hydrol Earth Syst Sci 13:125-140.

Toth Z, Talagrand O, Candille G, Zhu Y (2003) Probability and ensemble forecast. In: Jolliffe I., Stephenson D (ed) Forecast verification, a practitioner’s guide in atmospheric science, Wiley, England.

Tsuguti H, Seino N, Kawase H, Imada Y, Nakaegawa T, Takayabu I (2018) Meteorological overview and mesoscale characteristics of the Heavy Rain Event of July 2018 in Japan. Landslides 16:363-371.

Udmale P, Tachikawa Y, Kobayashi K, Sayama T (2019) Flood Hazard Mapping in Japan, In: Kobayashi K, Sutapa I, Tabios G, Tachikawa Y, Thulstrup H (ed) Catalogue of Hydrologic Analysis for Asia and the Pacific, vol. 1, UNESCO-IHP, pp. 16-36.

UNESCO, UN-Water (2020) United Nations World Water Development Report 2020: Water and Climate Change, Paris, UNESCO, https://en.unesco.org/themes/water-security/wwap/wwdr/2020. Accessed 18 April 2020.

Ushiyama T, Sayama T, Tatebe Y, Fujioka S, Fukami K (2014) Numerical Simulation of 2010 Pakistan Flood in the Kabul River Basin by Using Lagged Ensemble Rainfall Forecasting. Journal of Hydrometeorology 15:193-211.

Vergara H, Kirstetter P-E, Gourley JJ, Flamig ZL, Hong Y, Arthur A, Kolar R (2016) Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States. Journal of Hydrology 541:421-433.

Vincendon B, Édouard S, Dewaele H, Ducrocq V, Lespinas F, Delrieu G, Anquetin S (2016) Modeling flash floods in southern France for road management purposes. Journal of Hydrology 541:190-205.

Wu W, Emerton R, Duan Q, Wood AW, Wetterhall F, Robertson DE (2020) Ensemble flood forecasting: Current status and future opportunities. Wiley Interdisciplinary Reviews:

Water.

Yamazaki D., S. Togashi, A. Takeshima, T. Sayama (2018) HIGH-RESOLUTION FLOW DIRECTION MAP OF JAPAN, Journal of Japan Society of Civil Engineers (B1), vol.74(5), I_163-I_168, (In Japanese with English abstract)

Yoshida T, Troch PA (2016) Coevolution of volcanic catchments in Japan. Hydrology and Earth System Sciences 20:1133-1150.

Yoshimura K, Sakimura T, Oki T, Kanae S, Seto S (2008) Toward flood risk prediction: a statistical approach using a 29-year river discharge simulation over Japan. Hydrological Research Letters 2:22-26.

Figure legends

Figure 1. Observed and simulated hydrographs at dam reservoirs for F18. The “Radar” shows the simulated hydrographs, and “MEPS” shows the forecasted results initialized at 21:00 on July 5, 2018, by 21 ensemble members. Names in the parentheses show the regions to which each dam belongs.

Figure 2. The box plot of verification metrics of simulated hydrographs evaluated at dam reservoirs. The verification metrics include the correlation coefficient (r), the measure of the variability error (γ) and bias (β), the Kling-Gupta efficiency (*KGE*), the Nash-Sutcliffe efficiency (*NSE*) and the relative peak error (*PE*). “UC” denotes the uncalibrated cases for

F18 and T19, and “C” stands for the calibrated case only in the Kanto region for T19.

Figure 3. Observed and simulated hydrographs at dam reservoirs for T19 (calibrated case). The “Radar” shows the simulated hydrographs, and “MEPS” shows the forecasted results initialized at 9:00 on October 11, 2019 by 21 ensemble members. These dams are positioned in the Kanto region.

Figure 4. Peak runoff by ensemble forecasting (lines) and the simulation (dots) at all the dam reservoirs for F18. The x-axis shows the 21 ensemble members in descending order of forecasted peak runoff shown on the y-axis.

Figure 5. The same plot as shown in Figure 4 but for T19.

Figure 6. The spatial distributions of the simulated and forecasted (ensemble mean) peak runoff for F18. (a) The simulated result is shown only along the main rivers. (b) The simulated result for all rivers. (c) The forecasted result for all rivers.

Figure 7. The same graph as Figure 6 but for T19. The figures also show the major levee

breaching points (red marks) and levee breaching points including small tributaries in Tochigi prefecture (white marks).

Figure 8. The mean and standard deviation of bias in the forecasted peak runoff compared to the simulated one for (a) F18 and (b) T19. The bias ratios are computed depending on the peak runoff shown in x-axis. The dotted lines without the shading show the relative frequency of the simulated peak runoff.

Figure 9. ROC curves for (a) F18 and (b) T19. Ensemble forecasting is evaluated at all river grid-cells in each region performance. A line closer to the top left corner shows better performance of ensemble forecasting.

Figure 10. Simulated peak runoff at levee breaching points in small to medium sized rivers in Tochigi prefecture (and Saitama prefecture for the Oppe river). The red points are rivers that originated from mountains, while the blue points are rivers that originated from plains, whose bank-full discharges are typically smaller.

Figure S1. Observed and simulated/hindcasted hydrographs at dam reservoirs for T19

(uncalibrated case). The “Radar” shows the simulated hydrographs, and “MEPS” denotes the hindcasted results at 9:00 on October 11, 2019.