Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

The DOI for this manuscript is
DOI:10.2151/jmsj.2025-033
J-STAGE Advance published date: July 16, 2025

Advanced impact-based forecast for a severe flood event using a 1000-member ensemble weather prediction at convective-scale

Tsutao Oizumi[1], Takuya Kawabata[1], Le Duc[2,1], Kenichiro Kobayashi[3], Kazuo Saito[4,1,5], and Takuma Ota[1]

[1]Meteorological Research Institute, Tsukuba, Ibaraki, Japan
[2]Graduate School of Engineering, the University of Tokyo, Bunkyo, Tokyo, Japan
[3]Graduate School of Science and Engineering, Saitama University, Saitama, Saitama, Japan
[4]Atmosphere and Ocean Research Institute, the University of Tokyo, Kashiwa, Chiba, Japan
[5]Japan Meteorological Business Support Center, Chiyoda, Tokyo, Japan

23 June, 2025

---

1) Corresponding author: Tsutao Oizumi, Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan
E-mail: oizumi@mri-jma.go.jp

Abstract

Impact-based forecasts provide information on exposure, vulnerability, and risk, which are essential for quantifying risks and facilitating timely evacuation. For small- and medium-sized rivers, accurate deterministic flood forecasting more than three hours in advance is quite difficult owing to forecast errors in the location and intensity of precipitation systems. Operational weather forecasting services employ ensemble prediction systems for early flood warning, although the number of ensemble members is limited to a few dozen. From the perspective of probability prediction with minimized sampling errors, increasing the number of ensemble members for flood forecasting is an important research topic. This study investigated the impacts of the number of ensemble members on flood prediction. The precipitation dataset was 100 and 1000-member local ensemble transform Kalman filter (LETKF100 and LETKF1000, respectively) and 21-member operational weather forecast (MEPS). The flood forecasting model was the Japan Meteorological Agency's (JMA) operational flood forecasting system, "Runoff Index Model (RIM)", which incorporates a tank model and the Manning equation. The case study was the extreme flooding caused by record-breaking rainfall in the Kuma River Basin (1880 km^2), Japan, that occurred in July 2020. Using LETKF1000, the RIM successfully forecasted a high flood risk in the flood damaged area, with a probability of 60% half a day ahead; six hours earlier than the JMA's operational flood forecast. Furthermore, we

investigated the number of members required for ensemble flood forecasting. The prediction accuracy for the occurrence of risks during the flood forecasting period was similar between LETKF1000 and LETKF100. Whereas, LETKF1000 has higher prediction accuracy than LETKF100 regarding the timing of flood peaks. When selecting 500 members from the initial 1000 members, results nearly identical to those from the 1000 members were obtained. These results demonstrate that the LETKF1000 has the potential to provide valuable information for facilitating early evacuation.

Keywords ensemble forecast; flood risk; ensemble flood forecast; impact-based forecast

Introduction

The accuracy of weather forecasts provided by weather bureaus have improved in recent years owing to significant advances in numerical weather prediction (NWP) systems and supercomputing. However, losses and damage caused by floods have also been increasing (UNESCO, UN-Water 2020, JMA 2021). One of the reasons is that many weather bureaus issue flood warnings based solely on rainfall amounts. They occasionally fail to issue appropriate flood warnings because they do not consider hydrological processes such as water level forecasts. To address this issue, the WMO (2015, 2019) has proposed a three-step strategy: "weather forecasts and warnings," "impact-based forecasts and warnings (IBFW)," and "impact forecasts and warnings." The first step involves only weather forecasts, whereas the second step incorporates meteorological and hydrological processes and forecasts flood risk. The first and second steps are the responsibilities of national governments, whereas the third step is the responsibility of local governments, citizens, and businesses, where the actions are based on IBFW, such as closing schools, providing evacuation announcements, opening shelters, and stopping public transportation.

As shown by WMO, in flood forecasting, the collaboration between meteorological and hydrological models is vital. NWP models predict the amount of precipitation and rainfall location. Using these data, hydrological models predict whether rainfall will cause flooding. One difficulty in flood forecasting is that many river basins are much smaller than that of

NWP model grid. Even small errors in predicting the locations and amounts of rainfall in NWP models can lead to significant errors in flood forecasting. In particular, small and medium-sized rivers tend to be strongly affected by these errors. Since it is difficult to accurately predict heavy rainfall with deterministic forecasting, many advanced weather bureaus have implemented ensemble weather forecasting.

Recent advances in high-performance supercomputers in Japan have led to breakthroughs in Meso-/Micro-scale meteorological studies using numerical weather prediction models (Saito et al, 2023). Significant improvements in data assimilation techniques have improved the forecasting skills of NWP (e.g., Kawabata et al. 2011, Kunii et al. 2016, Ikuta et al. 2020, Duc et al. 2021). Some studies conducted extensive ensemble forecasts in research on the development of supercomputers (Kunii 2014; Duc and Saito 2017; Kawabata and Ueno 2020; Neckar et al. 2020; Duc et al. 2021; Terasaki and Miyoshi 2022). These studies showed that using a large number of ensemble members improved the precipitation forecast skills by using different models and model resolutions at the mesoscale. Kunii (2014) evaluated the influence of sampling noise on the background error covariance estimated from a 1000-member ensemble forecast and found that a larger ensemble size reduced the sampling error. Necker et al. (2020) performed the first convective-scale 1,000-member ensemble simulation over Central Europe and carried out sensitivity experiments using 40, 200, and 1000 members. They found that approximately 200 ensemble members are required to estimate the potential

impact of precipitation forecasts and that a 1000-member ensemble can detect small-scale traceable features in both space and time.

Two studies performed large ensemble forecasts for heavy rainfall in Kyushu, Japan, on 4 July 2020, which was the target rainfall event in this study. Duc et al. (2021) conducted 100- and 1000-member local ensemble transform Kalman filter (LETKF100 and LETKF1000, respectively) experiments using the 2-km resolution by using the Japan Meteorological Agency (JMA) mesoscale model Nonhydrostatic Model (NHM; Saito et al., 2006) with 50 members of the JMA global ensemble forecasts as the lateral boundary conditions. Quantitative verification showed that the 1000-member forecasts outperformed the operational forecasts of the JMA. Terasaki and Miyoshi (2022) showed that 1024-member ensemble forecasts captured the probability of a heavy rainfall event approximately five days before using the NICAM-LETKF system (Terasaki and Miyoshi 2017; Terasaki et al. 2019). These studies have shown that an ensemble size of 1000 members is sufficient to neglect sampling errors.

Ensemble flood forecasting has also been developed with progress in ensemble weather forecasting (e.g., Cloke and Pappenberger 2009; Wu et al. 2020). Many studies have conducted ensemble flood forecasting for severe disaster events using various ensemble datasets. Those researches demonstrated that the ensemble forecasts provide valuable information for decision-making of emergency responses (e.g., Kobayashi et al. 2016; Hanasaki et al. 2019; Sayama et al. 2020; Ma et al. 2021; Ito, T. et al. 2021; and

119 Nakayama et al. 2025).

120 In a pioneering work on probabilistic flood forecasting using a large ensemble weather
121 forecast, Kobayashi et al. (2020) performed a large ensemble flood simulation with a
122 1600-member precipitation dataset. They demonstrated the necessity of flood forecasting
123 by using large-ensemble weather forecasts for emergency flood operations. They also
124 demonstrated the difficulty of selecting the best ensemble members. Kobayashi et al.
125 (2023) conducted an ensemble rainfall-runoff and inundation simulation using ensemble
126 precipitation data (Duc et al., 2021), which is the same dataset used in this study. They
127 predicted the inflow volumes of the Ichifusa Dam and the Kawabegawa Dam using MEPS,
128 100-member, 1000-member of rainfall data and the storage function model. The 1000-
129 member experiments showed the best results. Furthermore, they investigated the
130 accuracy of selective 100-member ensembles out of 1000 members. The results showed
131 that 100-member ensembles showed a similar tendency with 1000-member ensembles.

132 In Japan, the plains are often densely populated, making them more likely to be affected
133 by floods. The distribution of river basin areas in Japan is as follows: basin area (larger
134 than 500 km^2) is approximately 5%, basin area (500–200 km^2) is approximately 10%,
135 basin area (200–50 km^2) is approximately 25%, and the basin area (smaller than 50 km^2)
136 is approximately 60% (Japan Institute of Country-ology and Engineering, 1999).
137 Generally, rivers with catchment areas of smaller than 200 km^2 are insufficient for
138 monitoring floods and flood damage. Therefore, it is necessary to monitor and predict the

flood risk in these rivers using a numerical model. The JMA has been developing the Runoff Index Model (hereafter RIM, Tanaka et al., 2008; Ohta, 2017; Ohta and Makihara, 2018, Ohta et al. 2023), which has been in operational use since 2008. This model predicts the risk of flooding for almost all small and medium-sized rivers (approximately 21,000 rivers) in Japan by comparing a predicted “Index” value (square root of discharge) with an original criterion value, which indicates the risk of flooding statistically computed from flooding events since 1991. The main difference between the RIM and other hydrological models is that RIM does not directly predict water levels or river discharge. RIM predicts the level of flood risk and it uses them as a basis for issuing warnings and alerts. As mentioned earlier, IBFW aims to predict the potential impacts (risks) based on weather forecasts. From this perspective, RIM, which predicts flood risks, can be regarded as a model that performs IBFW.

However, the accuracy of the RIM is limited by its inability to predict risk with sufficient lead time if weather forecasts are unreliable. Currently, the JMA operates this system using precipitation nowcasting with radar observations and NWP outputs for 6 hours. However, more than the 6 hours lead time is required to ensure safe evacuation. One way to extend the lead time is to use a large ensemble dataset and quantify the risk. For example, the 1000-member forecasts of Duc et al. (2021) showed that the accumulated rainfall over the Kuma River was well predicted with a lead time of half a day.

This study proposed a probabilistic flood risk forecasting system using RIM and a large-

159 ensemble weather forecast at a convective scale. The case study is a flood event that
160 occurred in Kyushu in July 2020, and which was declared as a "Designated Disaster of
161 Extreme Severity" flood by the Japanese government. A quasistationary convective band
162 (QSCB) caused this flood. Kato (2020) defined QSCBs as “band-shaped heavy rainfall
163 areas with a length of 50–300 km and a width of 20–50 km, produced by successively
164 formed and developed convective cells, lining up to organize multicell clusters, and
165 passing or stagnating at almost the same place for a few hours.” Meteorological disasters
166 associated with QSCB frequently occur in Japan (Tsuguti and Kato 2014; Hirockawa et
167 al. 2020a). Several studies have been conducted to understand the mechanisms and
168 predictability of QSCBs (e.g., Ogura 1991; Kunii 2014; Oizumi et al. 2018; Oizumi et al.
169 2020; Hirockawa et al. 2020b; Ito, J. et al. 2021). The forecasting skills of the QSCB have
170 improved in recent years. However, more observations (especially of water vapor over
171 the sea) and improvements in the NWP model (understanding the QSCB process from
172 occurrence to disappearance) are required for a more accurate forecast.
173 This study investigates the impacts of the number of ensemble members on flood risk
174 predictions. We perform a probabilistic flood risk forecast with three different ensemble
175 forecast datasets (Duc et al. 2021) on the actual flood event in Kyushu in July 2020. We
176 investigate the probabilistic flood risk for the small- and medium-sized rivers and the
177 Kuma River with three ensemble weather prediction systems. For statistical robustness,
178 we quantify the flood risk by expressing the probability of occurrence in 10% increments.

Note that Kobayashi et al. (2023) used the same dataset and investigated uncertainties in hydrological models using ensemble weather forecasts. In contrast, the present study focused on the impact of the large ensemble weather forecast on the quantification of flood risk.

The remainder of this paper is organized as follows: Section 2 describes the heavy rainfall events in the Kuma River. The results of deterministic and ensemble predictions by Duc et al. (2021) are also shown. Section 3 describes the RIM and the experimental settings. Section 4 presents the results, and Section 5 presents the discussion. Finally, the conclusions are presented in Section 6.

2. Heavy Rainfall in July 2020 at the Kuma River

On 4 July 2020, a flood occurred on the Kuma River, which is in the central-western part of Kyushu. Figure 1a shows the location of the Kuma River Basin, and the light blue line represents the target rivers (2,536 rivers) in the RIM. Figure 1b shows the Kuma River Basin, where the blue line represents the Kuma River (with a basin area of 1880 km^2, and length of around 115 km). The Kuma River is one of the steepest rivers in Japan and has experienced several severe floods in the past. Most of its tributaries have basin areas smaller than 100 km^2. The flooded area in the middle stream covered four municipalities (Kuma Village, Hitoyoshi City, Nishiki Town, and Sagara Village). Hitoyoshi City and Kuma Village, in particular, suffered severe damage. Figure 1c shows the detected

199 flooding areas in pink mesh areas by the Geospatial Information Authority of Japan (GSI,
200 2020). Rectangle A indicates the locations of the 14 casualties caused by flooding.
201 According to the Akamatsu (2021), overflow occurred in rectangle A at 0555 JST (UTC
202 +9) on 4 July. Rectangle B indicates the confluence of the Kawabe River and the Kuma
203 River. The Kawabe River basin, with a basin area of 533 km^2, and a river length of
204 approximately 115 km, is the second largest sub-basin within the Kuma River Basin.
205 Rectangle C indicates the confluence of the Ogawa River (basin size 29 km^2, river length
206 approximately 12 km) and the Kuma River. The overflow of the Ogawa River was caused
207 by backwaters because the water level of the Kuma River is higher than that of the Ogawa
208 River (Nihei 2021, Onaka et al. 2021).

209 On 3 July, a low-pressure system developed over a seasonal rainfront in the East China
210 Sea and advanced to the northern Kyushu region before dawning on 4 July. As the low
211 pressure moved eastward, the seasonal rain front moved northward to the northern
212 Kyushu region on the night of 3 July and warm and humid air flowed towards the low
213 pressure and front, making atmospheric conditions very unstable in Kyushu. Some
214 QSCBs that caused the Kuma River flood have been observed (Hirockawa et al. 2020;
215 Araki et al. 2021) in the central part of Kyushu.

216 Figure 2 shows the three-hour accumulated precipitation valid at 0300 and 0600 JST
217 on 4 July. The analyzed rainfall by the JMA Radar-AMeDAS system (Nagata 2011,
218 hereafter RA) shows QSCB in the central part of Kyushu and an intense rainfall area (over

219 100 mm $3h^{-1}$) in the Kuma River basin (Figs. 2a and b). The deterministic forecasts shown
220 in Figs. 2c–2f were adopted from Duc et al. (2021). These forecasts were initialized at
221 1800 JST (09 UTC) on 3 July 2020 using the ensemble data assimilation system NHM-
222 LETKF with 100 and 1000 ensemble members (LETKF100 and LETKF1000). The initial
223 conditions for the deterministic forecasts of LETKF1000 and LETKF100 were created by
224 averaging the initial conditions for each member. The vertical localization of both
225 LETKF1000 and 100 was turned off during these experiments. In LETKF1000, intense
226 rain areas (over 100 mm $3h^{-1}$) covered the basin at both 0300 and 0600 JST (Figs. 2c
227 and 2d). In LETKF100, the intense rain area covered the river basin at 0300 JST (Fig.
228 2e), and the rain intensity increased at 0600 JST (Fig. 2f). Figures 2g and 2h show the
229 results of the operational Mesoscale Model of JMA (MSM: JMA, 2019) with the same
230 initial time. It should be noted that the deterministic forecasts of MEPS and MSM are
231 identical. The intense rain area was predicted to be south of the river basin at 0300 JST
232 (Fig. 2g), and its intensity was underestimated. Subsequently, more intense rain occurred
233 around the river basin at 0600 JST (Fig. 2h). The model resolutions of LETKF1000 and
234 LETKF100 are on a convective scale of 2-km while MSM is 5-km, and this difference may
235 have affected the forecast of heavy rainfall.

236 Figure 3 shows the probabilities of rainfall exceeding 50 ($mm3h^{-1}$) at 0300 and 0600
237 JST on 4 July by ensemble forecasts. In LETKF1000, the probabilities in the northern part
238 of the Kuma River Basin exceeded 40% at 0300 JST (Fig. 3a). However, the center of

239 the precipitation area is shifted northward compared to Fig. 2a. At 0600 JST (Fig. 3b), the
240 center of the precipitation area shifts towards the central region of the Kuma River Basin,
241 and the probability in the northern part appears higher than in other areas, resembling the
242 pattern shows in Fig. 2b. LETKF100, the probabilities were approximately 20% across
243 the river basin at 0300 JST (Fig. 3c); at 0600 JST, they had increased to over 40% across
244 the river basin (Fig. 3d). Figures 3e and 3f show the results of the JMA's operational 20-
245 member mesoscale ensemble prediction system (MEPS; Ono et al., 2020). Since the
246 MEPS is operated every 6-hours only at 00, 06, 12, and 18 UTC, the MEPS with an initial
247 time of 2100 JST on 3 July 2020, was used for this study. At 0300 JST (Fig.3e),
248 probabilities exceeded 20% covered the Kuma River Basin, but only low probabilities
249 were predicted for the basin at 0600 JST (Fig. 3f). The model resolutions of LETKF1000
250 and LETKF100 were 2-km, whereas that of MEPS was 5-km, and the difference in
251 resolution and initial time may have influenced the forecast skill for heavy rainfall.
252
253 **3. JMA Runoff Index Model and experimental setting**
254 *3.1 JMA Runoff Index Model*
255 The JMA has operated RIM since 2008. RIM, officially named by the JMA, is designed
256 to predict runoff, but also river discharge. RIM predicts an increasing flood risk for
257 medium- and small-sized rivers managed by municipalities and the JMA. Flood advisories
258 and warnings are issued operationally under the consideration of RIM. Before the

operation of RIM, the advisory and warning were only issued based on the rainfall amount, whereas after implementing RIM, the transition from rainfall-based flood forecasting to flood forecasting considering rainfall-runoff processes improved forecasting accuracy. The JMA has started providing the “Real-time Flood Risk Map” since 2017 (Ohta 2017, Ota et al. 2023).

RIM consists of two components: rainfall that flows into the river (runoff component) and rainfall that flows downstream (river routing component). The runoff component was a tank model consisting of three tanks for non-urban areas and a tank consisting of five tanks for urban areas. The tank models were located every 1 km in Japan. The parameters of the three-tank model were set based on five geological types (Ishihara and Kobatake 1979). The five-tank model had one hole at the bottom, and the value of the tank runoff rate changed according to the slope.

The river routing was simulated using a kinematic wave model based on Manning’s roughness formula (Manning, 1891). The square root of the river flow rate at each grid point is named “Index”. Instead of explicitly considering the factors that influence river discharge, such as dams, levees, the confluence of rivers, tides, etc., the RIM considers these factors at the “Criteria level”, which links the model’s computational output with actual disasters. The values of the criteria levels were determined based on actual flooding events that have occurred in heavy rainfall cases since 1991. The value is adjusted (increased or decreased) by the effect of infrastructure facilities and several flood

phenomena. Criterion level 1 (L1) captures minor damages (e.g., inundation in roads or agricultural fields). Criterion level 2 (L2) indicates the occurrence of severe flood damage, and criterion level 3 (L3) targets severe flood damage with high accuracy. If there have been no floods since 1991, L2 and L3 were set to a simulation result using 30- and 50-year return period rainfall, respectively. The criterion levels were determined through consultations between the JMA and local governments and renewed every year to consider the changing factors described above. The details of the RIM are described in the Appendix.

The performance of the flood risk was verified in three rivers of different sizes: the Kuma River (1880 km^2), the Kawabe River (553 km^2), and the Ogawa River (29 km^2). Different criterion values were used for large rivers such as the Kuma River and other medium- and small-sized rivers such as the Kawabe and Ogawa Rivers. In this study, rivers with basin area larger than 1000 km^2, between 200 and 1000 km^2, and smaller than 200 km^2 were defined as large-, medium-, and small-sized river, respectively. In Section 4, the Index values of the Kuma River are compared with 30-year (30Y), 50-year (50Y) and historical maximum (HM) values. The 30Y and 50Y values were obtained from the Gumbel distributions of the Index values using RA from 1991 to the present. In Section 5, the Index values of the Kawabe and Ogawa rivers are compared with those of the 30Y, 50Y, HM, L1, L2, and L3. HM is the observed value; thus, it is higher than 50Y and 30Y, depending on the location. According to the Basic Law of River Management, the Ministry

of Land, Infrastructure, Transport, and Tourism (MLIT) manages the Kuma River but does not set criteria levels. Meanwhile, the municipalities and the JMA manage the Kawabe and Ogawa Rivers and set the criteria levels.

3.2 Experimental settings

The spin-up run of RIM was from 0000 JST on 1 June to 2200 JST on 3 July using RA. The forecast run with LETKF1000 and LETKF100 were from 1900 JST on 3 July to 1500 JST on 4 July. The forecast run with MEPS was from 2200 JST on 3 July to 1500 JST on 4 July. Duc et al. (2021) began forecasting at 1800 JST, but in the RIM experiment, the first hour was treated as a spin-up period, starting at 1900 JST. Similarly, the MEPS experiment adopted this approach and commenced at 2200 JST. The RIM computes approximately 21,000 rivers, and the elapsed time for a 21-hour forecast is approximately 4 minutes. The forecasts used precipitation data from LETKF1000, LETKF100, and MEPS. The resolutions of LETKF100 and LETKF1000 were 2 km, and that of the MEPS was 5 km. The precipitation data were interpolated to the domain of the RIM with 1-km grid spacing.

4. Results

4.1 Flood risk in the Kuma River basin

Figure. 4 shows the hourly precipitation at 0500, 0700, and 0900 JST on 4 July and the

corresponding probabilities exceeding HM by the RIM using RA. At 0500 JST, the river basin was covered with QSCB, and the intense rain areas (greater than 50 mm h^{-1}) covered the middle part of the river basin (Fig. 4a). The areas of intense rain were in the eastern and western parts of the basin at 0700 JST Fig. 4b). QSCB shifted south at 0900 JST (Fig. 4c).

In Fig. 3d, many rivers exceeded their HM (black) in the central part of the basin at 0500 JST. The locations of these rivers correspond to the location of the intense rainfall areas in Fig. 4a. Around Hitoyoshi City, where severe damage occurred (rectangle in Fig. 4d), and the Kuma River did not exceed its HM at 0500 JST. At 0700 JST, the Index value exceeded HM for the entire Kuma River and its tributaries in the eastern part of the basin (Fig. 4e). At 0900 JST (Fig. 4f), most of the rivers in the southern part of the Kuma River basin exceeded their HM value.

Figures 5a–5c show the probabilistic map of the intense rainfall area (greater than 20 mm h^{-1}) by LETKF1000 at 0500, 0700, and 0900 JST. The probability for most parts of the Kuma River Basin was exceeded 40% from 0500 JST to 0700 JST (Figures 5a–b). The high-probability areas moved south at 0900 JST (Fig. 5c). The temporal change in the spatial pattern of the probability area in the simulation is similar to that of the hourly precipitation observed in RA data (Figs. 4a-c).

In RIM simulation, the probability of exceeding HM was 20% in several rivers in the northern and central parts of the Kuma River basin at 0500 JST (Fig. 5d). At 0700 JST,

several rivers in the western and southern parts of the basin also exceeded a probability of 20% (Fig. 5e). At 0900 JST, these probabilities decreased in the northern part of the river basin but increased in the south. Overall, the tendency of the high-probability river distribution is similar to that of RA, except for the rivers in the northeastern part (Figs. 4d–4f). The probability of the Kuma River being in Hitoyoshi City was 20% to over 40% during this period (rectangle in Figs. 5d–5f).

Figure. 6 shows the detected inundation area in Hitoyoshi City and the probability of exceeding the HM (same data as the rectangles in Figs. 5d–5f). The GSI (2020) detected and confirmed inundation areas using aerial and social networking service photographs taken at 1500 JST on 4 July. Most of the riverside sites were inundated. Some studies have reported that backwaters in small rivers on the north side of the Kuma River caused this flood (Akamatsu 2021; Ohmoto; 2021, Ohnaka 2021; Nihei 2021).

RIM with LETKKF1000 showed that the probability of exceeding HM in the Kuma River was exceeded 20 % at 0500JST and over 40% after 0700 JST. Circle A in Fig. 6 shows the Watari water-level gauge station (MLIT Water Information System). At 0000 JST on 4 July, the observed water level was 3.16 m. At 0500 JST, the observe water level reached 10.53 m, and it exceeded the historical maximum water level of 9.98 m, recorded on 22 July 2006. At 0700 JST, the water level reached 12.55 m, after which data were unavailable. This result indicated that RIM with LETKF1000 could predict the exceedance of the historical record at 0500 JST with a probability of 20%.

359

360 *4.2 Evaluation for Watari area*

361 Figure 7 shows the Index values obtained by the RIM using RA and LETKF1000 for the
362 criterion levels (30Y, 50Y, and HM) in rectangle A (see Figure 1c). RA exceeds the HM
363 threshold (orange dashed line) at 0500 JST. This timing approximately corresponds to
364 exceeding the previously observed water level at the Watari water level gauge station
365 (Circle A in Fig. 6). RA also exceeds 30Y (red dashed line), and 50Y (purple dashed line)
366 thresholds at 0500 JST. The peak of RA is at 0900 JST and then fall below the HM
367 threshold by 1100 JST.

368 The deterministic forecast of LETKF1000 exceeded HM at 0600 JST, whereas the
369 ensemble mean exceeded the HMs at 0800 JST. This difference was due to the
370 accumulated rainfall over the basin; the rainfall amount of the deterministic forecast was
371 approximately 50 mm greater than the ensemble mean after 0400 JST (see Figure 6d in
372 Duc et al. 2021).

373 Figure 7, the bar graphs show the exceedance probabilities for HM (orange), 30Y (red),
374 and 50Y (purple). Those maximum probabilities were 60%, 50%, and 40%, respectively.
375 The RIM with LETKF1000 well forecasted this event, and the Index value of the
376 deterministic forecast exceeded that of HM at almost the same time as RA. The RIM with
377 LETKF1000 can predict an actual flood event 11 hours in advance.

378 As shown in Fig. 5, a-c, the temporal change in the spatial pattern of the probability area

379 in LETKF1000 was similar to that of the hourly precipitation observed in RA data (Figs.
380 4a-c). It likely contributes to accurate flood forecasting. Furthermore, a good prediction in
381 the RIM is attributed to the fact that the Kuma River has a large catchment area that
382 reduces the impact of positional and temporal errors in rainfall forecasts on flood forecasts.
383 These results were identical to the JMA's report (JMA, 2018). RIM tends to be more
384 precise as the distance from the headwaters increases (i.e., as the basin area increases).
385 It is related to the calculation characteristics of RIM. As the basin area expands, the
386 absorption of positional errors in predicted rainfall becomes more pronounced. As a result,
387 the prediction accuracy of RIM tends to be higher for rivers with larger basin areas.
388
389 **5. Discussion**
390 *5.1 Predictability of flood warnings*
391 In the JMA flood forecasting operations, RIM predicts flood risk only for small- and
392 medium-sized rivers. In this chapter, the prediction accuracy of RIM was examined using
393 RA and LETKF1000 for small- and medium-sized rivers in the Kuma River Basin.
394 Figure 8 shows the probabilities exceeding L2, which indicates a severe flood risk. For
395 all of small- and medium-sized rivers, the JMA sets L2 values based on the 30Y values.
396 These L2 values were slightly modified based on actual flood events in the past, and the
397 values were further adjusted to account for infrastructure facilities and phenomena that
398 implicitly caused the past few floods. In the results with RA (Figs. 8a–8c), the black line

399 shows that it exceeded L2. The Index gradually exceeded L2 from north to south along
400 the Kuma River (blue line) from 0500 to 0900 JST on 4 July. These results correspond to
401 the time when the JMA issued a heavy-rain emergency warning for this area from 0450
402 to 1155 JST on 4 July. The Kuma River is indicated by the blue line in these figures
403 because L2 was not set for the Kuma River. In the results with LETKF1000 (Figs. 8d–8f),
404 the probability of exceeding L2 also increased from north to south between 0500 and
405 0900 JST on 4 July. Many rivers exhibited probabilities ranging from 20% to over 40%.
406 The results indicated that the RIM with LETKF1000 demonstrated a similar tendency to
407 that of the RIM with RA.

408 Figures 9a–9f show enlarged views of the rectangles in Figs 8a–f. Figs. 9a–9c, the
409 Index of the Kawabe River (rectangle B, medium-sized river) exceeded L2 between 0500
410 and 0900 JST on 4 July. The small-sized rivers (rectangles C to F) also exceeded L2 at
411 0500 JST. The Ogawa River (rectangle C) and Mae River (rectangle E) continued to
412 exceed L2 at 0700 JST, and the Index was below L2 at 0900 JST. Some studies have
413 reported that backwaters occur at the outlets of these small rivers. Nihei (2021)
414 investigated the backwater occurrence mechanism at the outlets of small rivers. The
415 mechanism is as follows: first, the water level rose in the narrow section of the Kuma
416 River downstream of the Watari area (rectangle A), which in turn caused the water level
417 in Hitoyoshi City to rise. As a result, the tributaries were unable to drain into the Kuma
418 River and subsequently overflow, leading to flooding. Akamatsu (2021) conducted a

419 detailed simulation with a resolution of 3 m of flooding at the outlet of the Ogawa River
420 (rectangle C) and investigated the effect of backwater on flooding. They reported that
421 overflow started at 0545 JST on 4 July and the backwater caused flooding of the Ogawa
422 River. The RIM with RA at the outlet of the Ogawa River exceeded L3 from 0400 to 0700
423 JST (not shown).

424 Figures 9d–9f show the results for the LETKF1000. The probability of exceeding L2 in
425 the Kawabe Rive (rectangle B) was over 40%. In the Ogawa River (rectangle C), the
426 probability of exceeding L2 was 10–20%. A more detailed analysis is presented in the
427 following section.

428

429 *5.2 Comparison of LETKF1000, LETKF100, and MEPS*

430 *a. Medium-size rivers*

431 In Japan, approximately 10% of rivers are medium-sized (basin area 200–500 km^2),
432 often flowing through populated areas and causing significant flood damage. Therefore,
433 it is necessary to verify the performance of RIM in medium-sized rivers. Figure 10 shows
434 the indices of the Kawabe River (533 km^2) outlet (rectangle B in Fig. 1c), where the
435 criterion values for Figs. 10a–10c are HM, 30Y, and 50Y, and Figs. 10d–10f are L1, L2,
436 and L3, respectively. HM and Y50 had similar values. Therefore, only Y30 and HM are
437 discussed.

438 Figures 10a–10c, the peak period when the Index value of RA exceeded Y30 and HM

was between 0600 and 09:0 on 4 July. Focusing on the peak period, the deterministic forecasts of LETKF1000 and LETKF100 performed well, exceeding those of Y30 and HM, and this peak was higher than that of RA. The probability of exceeding Y30 and HY for LETKF1000 was 10% higher than that for LETKF100 (bar graph). This difference was also observed in the ensemble mean, with LETKF1000 being higher than LETKF100. In MEPS, only deterministic forecasts exceeded Y30 and HM. The LETKF1000 exhibited the best performance under severe flooding. This difference may be due to the different performances of the ensemble forecasts shown in Fig. 3. It should also be noted that the initial times of LETKF1000 and LETKF100 were 1800 JST and the MEPS was 2100 JST. The model resolutions of LETKF1000 and LETKF100 were 2-km, whereas MEPS had a resolution of 5 km. These differences may also have contributed to the lower performance of the MEPS.

Figures 10d–10f, the values of HM and L3 are similar, and L2 is slightly smaller than Y30. Therefore, the trend was similar to Figs. 10a–10c: LETKF1000 performed well for L2 and L3, and the probabilities of L2 and L3 were 10% higher than those of LETKF100. L1 threshold is set to ensure that even minor flood damage is not overlooked and to detect potential flooding. The performance in relation to L1 is discussed following.

Figures 10d–10f, the RA exceeded L1 between 0400 and 1200 JST on 4 July. In the LETKF1000 result, the probability of exceeding L1 ranged from 40% to 90% between 0400 and 1200 JST. In particular, the peak of 90% corresponded to the peak period of

RA from 0600 to 0800 JST. In LETKF100, the exceedance probabilities of L1 were 30%–80% during 0400–1200 JST. However, the peak of 80% from 0900 to 1000 JST occurred later than the peak period of RA from 0600 to 0800 JST. Thus, it can be concluded that the performance of LETKF1000 was better than that of LETKF100 for L1. The deterministic forecast of MEPS showed good performance due to the accuracy of the precipitation forecast (see Figure 6a in Duc et al. 2020); but it could only predict exceedance L1 with a maximum probability of 30%. Owning to its poor ensemble performance (Figs. 3e and 3f). These results show that LETKF1000 outperformed LETKF100 in terms of severe flood risk, even in medium-sized rivers.

b. Small rivers

Approximately 80% of the rivers in Japan are small-sized river (basin area smaller than 200 km^2). These rivers are inadequately monitored and RIM prediction is vital for early evacuation. Therefore, it is important to verify the performance of RIM in small-sized rivers.

Figure 11 presents the same as Fig. 10, but for the Ogawa River (29 km^2) outlet (Rectangle C in Figure 1c). The values of Y30 and HM were almost identical; therefore, only Y50 and HM are discussed. Figures, 11a–11c, RA exceeded Y50, HM was from 0400 to 0900 JST with a peak at 0500 JST. The maximum Y50 exceedance probabilities for LETKF1000, LETKF100, and MEPS were all 10%. Similarly, the maximum probability of HM for LETKF1000, LETKF100, and MEPS was 10%. One reason it was difficult to predict the HM exceedances is that the basin size of the Ogawa River is only 29 km^2. The

basin was covered by only eight NWP model grids in LETKF1000 and LETKF100, and just two model grids in MEPS. Accurately predicting conditions using such a limited number of NWP model grids is virtually impossible. This result suggests that if the number of the NWP model grids is insufficient for the size of the river basin, the accuracy of the maximum predicted value may be limited.

Akamatsu (2021) reported that the overflow of the Ogawa River continued from 0545 to 1455 JST on 4 July. The time required to exceed the Y50 of RA was similar to the report. However, the time of passing below the Y50 of the RA was at 0900 JST, it was much earlier than that in the report. This difference was due to the fact that the RIM did not account for the back water phenomena. LETKF1000 exhibited a probability value of 10% for exceeding Y30 during the peak period; LETKF100 showed a delayed probability response, also around 10%, while MEPS did not show any notable probability during the peak period. This tendency was similar to that observed in the case of the Kawabe River.

Figures11d–11f, L3 and L2 are similar to Y50 and Y30; note that the interval of the Y-axis is different from Figs. 11a–11c. Here, only L1 is discussed. The RA exceeded the L1 levels from 0300 to 0900 JST. During the period exceeding L1, in LETKF1000, the exceedance probability of L1 was 10–50%; in LETKF100, the exceedance probability of L1 was 10–40%; and in MEPS, the exceedance probability of L1 was 10–20%. When focusing on the peak period exceedance of L3 (0400–0800 JST) in the RA, in LETKF1000, high probability values (0700–0800 JST) appeared during the peak period. The peak of

LETKF100 (0900 JST) was delayed until the peak period of RA. In the MEPS, the timing of the peak period (0200 JST) was very different from that in the RA.

Since the basin size of the Ogawa River is much smaller than the area that the NWP model can resolve, it is difficult to accurately predict floods in terms of L2 and L3 using LETKF1000 unless the rainfall forecast is highly accurate. Both LETKF1000 and LETKF100 produced good predictions for L1. In particular, LETKF1000 exhibited relatively high probability values during peak periods. It is important to treat maximum values with caution, especially in the case of the small-sized river basins. To improve the forecasting performance of the RIM for small-sized rivers, a model that incorporates backwater phenomena needs to be developed.

c. Forecast Accuracy at other tributaries

In this subsection, we compare the prediction accuracy for 36 tributaries between the Ichifusa Dam and the Kuma River estuary. First, we investigated the probability of the HM value being exceeded at least once during the forecast period (1900 JST on 3July to 1500 JST on 4 July). Figure 12 shows the probability of HM value exceedance for LETKF1000, LETKF100, and MEPS at the Kuma River estuary and at the grid points where the Kuma River and 36 tributaries connect. The probability values are rounded to the nearest 10%. The river names are arranged on the horizontal axis in order of their basin areas. In the simulation with RA, the indices at the Nakatani, Tagashira, Dozan, and Fukamizu Rivers

519 did not exceed the HM values, and the names of such rivers are indicated in italic bold.
520 In the simulations using LETKF1000 and LETKF100, the probability of exceeding the HM
521 value during the forecast period was predicted to be 10% or higher for all rivers.
522 Meanwhile, using MEPS, the predicted probability was less than 5% for some rivers, such
523 as the Kuragi, Yoshino, Aburatani, and Kawabe rivers.
524 The probability values for LETKF1000 were higher than those for LETKF100 in 11 rivers,
525 while LETKF100 exceeded LETKF1000 in 12 rivers, and the probability values were the
526 same in 14 rivers. Thus, the prediction accuracy for the occurrence of risks during the
527 flood prediction period was similar between LETKF1000 and LETKF100. These results
528 indicate that floods exceeding HM can be predicted by both LETKF1000 and LETKF100
529 in this case.
530 In the previous section, it was considered that the accurate prediction in the Kuma River
531 (Fig.7) was attributed to its extensive basin area, which mitigated the impact of positional
532 and temporal errors in rainfall forecast on flood forecasts. However, in Fig. 12, no clear
533 relationship was found between basin area and forecast accuracy for the small-sized
534 rivers. This result indicates that precipitation accuracy is more important than the size of
535 the basin area in simple hydrological model such as RIM, especially in small-sized rivers.
536
537 *d. Timing of flood peaks*
538 From the perspective of disaster risk management, the timing of flood peaks is critical

information. In this section, the period during which the HM value exceeded the threshold is defined as “Ex_period” using RA-based simulations. Figure 13 shows the probability that the forecasted indices exceed HM at least once during the Ex_period. LETKF1000 captured all peaks except for the Kawauchi River. LETKF100 failed to predict the exceedance of the HM in the Kajiki River, the Magouri River and the Yamada River. MEPS failed to predict HM exceedances in 20 rivers. In 16 rivers, the probability values of LETKF1000 were higher than those of LETKF100, while the probability values of LETKF100 were higher than those of LETKF1000 in 5 rivers, and in the remaining 16 rivers the same probability values were predicted. These results indicate that LETKF1000 has higher prediction accuracy than LETKF100 regarding the timing of flood peaks. Both models showed a tendency towards slightly higher prediction accuracy in rivers with a catchment area of 20 km² or more.

5.3 Number of members required for ensemble flood prediction

To evaluate the impact of the number of ensemble members on flood prediction, we compared cases where the same number of members were sampled from LETKF1000 and LETKF100. Table 1 shows the HM exceedance probability for the Ogawa River during the Ex_period when 100, 250, and 500 members were sampled from LETKF1000 and LETKF100. Sampling was performed sequentially from 1 to 100 and from 101 to 200. Random sampling was also performed, but no difference was observed. Probability

559 values are not rounded to facilitate understanding of the results. When all 1,000
560 members of LETKF1000 were used, the probabilities were 13%, 13%, and 15% at 0500,
561 0600, and 0700 JST on June 4, respectively. When 100 members were sampled, the
562 probability values ranged from 9% to 20%, for 250 members from 10% to 16%, and were
563 nearly the same as the value for 1000 members for 500 members. Similar tendency was
564 found in other 35 rivers. The range of deviation for each 100 members was ±10%, ±5%
565 for each 250 members, and ±2% for each 500 members. When comparing the results
566 for 100 members of LETKF1000 and LETKF100, LETKF1000 showed better results as
567 shown in Table 1. In the 16 rivers shown in Figure 13, LETKF1000 showed superior
568 performance to LETKF100 during the Ex_period. This result is likely because
569 LETKF1000 more accurately represents atmospheric uncertainties, leading to improved
570 precipitation forecasts. The performance of LETKF1000 is considered to have been
571 passed down to the sampled 100 members. This trend is similar to Kobayashi et al.
572 (2023). Even when selecting 500 members from the initial 1000 members, results nearly
573 identical to those from the 1000 members are obtained. These results also suggest
574 that it is not always necessary to calculate all 1000 members; selecting 500 members
575 from the initial 1000 members yields results nearly identical to those for 1000 members
576 in flood calculations using a hydrological model. Terasaki and Miyoshi (2022) conducted
577 a similar verification using NICAM in an ensemble experiment with 1024 members. Their
578 experiment showed similar trends to both the 1024-member and 512-member sampling

results, supporting the validity of this study.

Table 1. HM exceedance probability for the Ogawa River during the Ex_period when 100, 250, and 500 members were sampled from LETKF1000 and LETKF100.

LETKF1000							
Averaged member	Probability (%) for every 1000 and 100 member			Averaged member	Probability (%) for every 50 member		
	7/4 5:00	7/4 6:00	7/4 7:00		7/4 5:00	7/4 6:00	7/4 7:00
001-1000	**13%**	**13%**	**15%**	001-50	6%	10%	18%
001-100	9%	10%	18%	051-100	12%	10%	18%
101-200	12%	14%	10%	Probability (%) for every 250 member			
201-300	13%	13%	13%	001-250	10%	12%	13%
301-400	14%	19%	13%	251-500	14%	15%	15%
401-500	14%	11%	16%	501-750	12%	15%	16%
501-600	10%	13%	16%	751-1000	12%	12%	13%
601-700	15%	19%	17%	Probability (%) for every 500 member			
701-800	9%	12%	14%				
801-900	9%	9%	12%	001-500	12%	13%	14%
901-1000	20%	10%	20%	501-1000	12%	13%	15%

LETKF100			
Averaged member	Probability (%) for every 100 and 50 member		
	7/4 5:00	7/4 6:00	7/4 7:00
001-100	**4%**	**6%**	**14%**
001-050	4%	6%	16%
051-100	4%	6%	12%

6. Conclusion and future work

On 4 July 2020, the Kuma River in Kyushu, Japan, was flooded, causing severe damage. Many meteorological researchers have investigated this event, and some studies have demonstrated the importance of large-ensemble forecasting. Several hydrological studies have investigated these events. Kobayashi et al. (2023) investigated the uncertainties in flood models using LETKF100, LETKF1000, and MEPS. However, no studies have investigated the impact of large-ensemble weather forecasts at a convective scale on flood prediction and flood risk quantification. In this study, the impact of ensemble members on flood prediction was investigated. The operational flood forecasting system,

the “Runoff Index Model (RIM)”, developed by the Japan Meteorological Agency and ensemble datasets of varying size (LETKF1000, LETKF100, and MEPS) by Duc et al. (2020) were used.

First, the results of the RIM using RA were compared with some previous field reports to confirm the model's reproducibility of the flood event. Forecasts using the LETKF1000 predicted a trend of increasing flood risk from north to south in the Kuma River Basin. Flood prediction accuracy was investigated at flooding points in the Kuma River. LETKF1000 predicted a 60% probability of a flood exceeding the HM of 11 hours before the flood event. The performance of LETKF1000 was also evaluated for medium- and small-sized rivers in the Kuma River Basin. The LETKF1000 demonstrated good predictive performance for medium-sized rivers as well.

The impact of different numbers of ensemble members on flood prediction was investigated. LETKF100 indicated a high probability of flooding in medium-sized rivers. In small-sized rivers, LETKF100 showed high probabilities of exceeding of L1; however, compared with LETKF1000, the probability of exceeding L2 was 10–20% lower, and the timing of exceedance L2 was delayed. These differences are due to the accuracy of the weather forecasts. Furthermore, these results confirmed that the prediction skills of LETKF1000 are higher than those LETKF100. The MEPS could predict flooding only with respect to the L1 levels for both small- and medium-sized rivers.

The prediction accuracy for the occurrence of risks during the flood prediction period

was similar between LETKF1000 and LETKF100. However, regarding the timing of flood peaks, LETKF1000 has higher prediction accuracy than LETKF100. Finally, the impact of the number of ensemble members on flood prediction was evaluated. The results showed that selecting 500 members from the initial 1000 members, results nearly identical to those from the 1000 members are obtained.

This is the first study to investigate the impact of the large number of ensemble members on flood prediction. It was confirmed that a large number of ensemble members can improve flood forecasting and extend lead time. The RIM with the LETKF1000 also shows that the probability of flood risk increases at three levels in small- and medium-sized rivers. The IBFW for small- and medium-sized rivers is an important task in mitigating flood damage. The results of this study provide a way to solve this task.

The large ensemble weather-forecasting methods require substantial resources. As a result, only one flood event could be validated in this study. Future work will focus on the following aspects: 1) To confirm the robustness of the findings, they need to be examined with other data and different case studies. 2) As shown in Section 5.2b, the RIM must be improved. 3) Finding the optimal number of members and testing another case are topics for future work. 4) Development of improved flood forecasting skills with fewer ensemble members, determination of the optimal number of members, and testing of another case.

Finally, we discuss the potential for implementing probabilistic flood forecasting in society. If a meteorological agency has powerful supercomputers, we believe that it is

technically possible to implement large-scale ensemble weather and flood forecasting in society, despite the many challenges involved. Meanwhile, in the author's interviews with some crisis managers, all respondents wanted deterministic rather than probabilistic information, especially in the case of an imminent crisis. As this study has shown, it is possible to obtain reliable probability information. However, the next major issue is how to use the flood risk probability information appropriately for decision making.

Data Availability Statement

The output data from this study are archived and available upon request from the corresponding author. The observation data and data assimilation system are made available under contract with the Japan Meteorological Agency because these data are collected and developed for operational purposes.

Supplement

Details of JMA Runoff Index Model

1. Over view of the JMA Runoff Index Model

The Runoff Index Model (RIM) is the operational flood forecasting model for medium- and small-sized rivers (21,394 rivers) by the JMA (Tanaka et al. 2008; Ohta 2017; Ohta and Makihara 2018; Ota et al. 2023). The analysis interval for RA was 10 min. The forecast interval using the high-resolution precipitation nowcasts was 10 minutes for one hour ahead.

654 The other forecast interval using short-range precipitation nowcasts (Tsujimura and Nagata,
655 2018) was 30 minutes, six hours ahead. The RIM consists of an “Index” and a “Criterion
656 value.” The Index is defined as the square root of the river discharge for at each grid point.
657 Discharge is calculated using runoff and river routing models, and model resolution is set at
658 1 km. The geographic data used in the model include river channels, geology, slope, and
659 land use from the National Land Information System. The Criterion value is determined
660 based on the Index values from heavy rainfall events since 1991. RIM predicts an increase
661 in flood risk for small- and medium-sized rivers. Its fundamental concept of the RIM is that
662 both rainfall and hydrological factors, such as water level and discharge, are crucial for flood
663 forecasting. This is aligned with the principle of IBFW. In other words, RIM implements the
664 concept of IBFW.
665
666 *2. Runoff process*
667 A “three-tank model,” consisting of three tanks connected vertically in series, is used in the
668 runoff process for non-urban areas (Figs. A1a-A1c). Moreover, a "five-tank model,"
669 consisting of five tanks connected vertically in series, is used for urban areas (Fig. A1d). In
670 the three-tank model, the upper, middle, and lower tanks represent the surface, midstream,
671 and groundwater runoffs, respectively. The accuracy of the tank model depends on its
672 parameter values; however, as tank models are positioned every 1 km across Japan, it is
673 impractical to set the appropriate parameters for each tank. Instead of tuning the parameters

674 for each tank model, representative parameters sets were applied based on geological
675 classification, as suggested in previous studies (Ishihara and Kobatake 1979). During heavy
676 rainfall, the tank model did not reproduce the hydrograph well. Specifically, the Index was
677 excessive compared with the actual flow rate (the peak time was early). To solve these
678 problems, the model incorporates a method for varying the infiltration coefficient according
679 to the water level in the tank. Specifically, when the water level in the first tank is S1 and the
680 infiltration coefficient is F1, the value of F1 is used as-is for S1≤L1. However, when
681 L1<S1≤L2, the value is changed to F1×2, and when L2<S1, the value is changed to F1×3.
682 The second tank was treated similarly (JMA 2017).

683 In urban areas, the ground surface is generally covered with artificial structures such as
684 concrete, resulting in less infiltration than natural soil. Therefore, surface runoff was the
685 primary source of runoff. Urban areas were defined as building sites, roads, and railways.
686 The five-tank model had a hole on the side, and the value of R5 (tank runoff rate) was set
687 according to the maximum slope I in each grid. For slopes of 1% or more and for those less
688 than 1%, R5 is expressed as follows:

689 $$Fi = R5 = 0.6910 \cdot \ln(I) + 8.1234, I < 1\%.$$
690 $$Fi = R5 = 0.9879 \cdot \ln(I) + 8.1234, I \geq 1\%.$$

691 In actual runoff processes in urban areas, even in fully urbanized regions contain green
692 spaces; therefore, infiltration into the ground must be considered. The RIM defines the
693 "urbanization rate" as the percentage of artificial structures on the ground surface within
694 each grid. Both urban and non-urban tank models were used to calculate runoff for all grids,

and the final runoff volume was obtained by averaging results from each model according to the urbanization rate. For example, in grids with no artificial structures, such as those in mountainous areas only the three-tank model was applied. In fully urbanized areas, 70% of the runoff was calculated using the urban tank model and 30 % of the runoff was calculated using the three-tank model.

3. River routing process

The water flow process in the river for each grid is as follows: Manning's formula was used to calculate the flow velocity.

$$v = \frac{1}{n} R^{2/3} I^{1/2}.$$

where v is the average cross-sectional velocity (m/s). Manning's roughness coefficient n is set to 0.04 in most rivers. A lower value of n=0.02 is applied to specific rivers that pass-through areas with an urbanization rate of 80% and river lengths of less than 30 km. R is the hydraulic radius (m), and I is the channel slope. The flow rate Q (m^3/s) is the flow velocity multiplied by the cross-sectional area of the flowing water A (m^2).

$$Q = vA.$$

R and A are related by the cross-sectional shape of the river, but the RIM assumes a river channel that widens in proportion to the depth, as shown in the formula below.

$$A = mR^2.$$

Here, coefficient m is the embankment coefficient, which is the ratio of the width of the embankment to its depth. The cross-sectional shape was assumed to be such that the river

width increased linearly with depth. The coefficient m varies according to river length: 10 for lengths less than 10 km and 40 for those exceeding 100 km. For lengths between 10 km and 100 km, the values are proportionally distributed. Consequently, the flow velocity in the river channel can be expressed as follows:

$$v = \left(\frac{Q}{m}\right)^{1/4} \left(\frac{1}{n} I^{1/2}\right)^{3/4}.$$

In other words, if the slope I, embankment coefficient m, and Manning's roughness coefficient n are set in advance, the velocity v in the river channel can be determined from the flow Q. The continuity equation is as follows:

$$\frac{d \int A dx}{dt} = \int (r - i) dS + Q_{in} - Q_{out}$$
$$= q(r) + Q_{in} - Q_{out}.$$

Here, r is rainfall intensity, i is infiltration amount, q(r) is runoff volume from the tank models, Qin is inflow from upstream, Qout is outflow to downstream, ∫dS is area integral concerning the basin.

The points to be noted in the river routing calculations are as follows. (1) when rivers join within a grid, they are assumed to enter the main river trunk. (2) rainfall occurring on a river within a grid is treated as flowing downstream within the grid. (3) when multiple rivers are present in the same grid, the runoff is evenly allocated among them in proportion to the number of rivers. (4) When the water level in the main river was higher than that of the tributary river at the confluence point, the model assumed continuous inflow from the tributary. (5) the river channel is discretized into a single grid composing of six regions along

736 the flow direction for calculation purpose; and (6) the model does not incorporate the
737 influence of the effects of artificial water quantity controls, such as dams, lake storage effects,
738 tidal levels, and snowmelt. These factors were considered the Criterion values.
739
740 *4. Criterion value*
741 The Criterion values were calculated based on past catastrophic flood events at site. The
742 specific procedure is as follows:
743 (1) Disaster data collection: Flood event data, including date of event and extent of damage,
744 are gathered for municipalities since 1991. Corresponding meteorological data are also
745 collected. The past Index values were obtained from these data for each river.
746 (2) Definition of relevant disaster events and statistical analysis: In consultation with
747 municipalities (prefectures), representative disaster events for warnings and advisories are
748 predefined. The relationships between flood damage and the Index from RIM is statistically
749 analysed, and the Criterion values are determined using "cost-loss model" (e.g. Tatehira
750 1999). According to this model, if no countermeasures are taken during a disaster, the loss
751 corresponds to the magnitude of the damage. However, if countermeasures are taken, the
752 loss considered to zero. These determinations are based on a combination of the
753 announcement cost and the extent of damage.
754 (3) Classification and interpretation of criterion levels: Criterion values are categorized into
755 three levels: level 1 for advisories, level 2 for warnings, and level 3 for situations that

significantly exceed the warning thresholds. Criterion Level 1 is designed with a high detection rate to ensure that even minor disasters, such as road or agricultural field inundation are not overlooked. Criterion 2 sets a coverage rate to avoid missing severe disasters and uses the number of flooded houses as the primary indicator. For rivers with no recorded flood disasters, a 30-year return-period value is applied. Criterion 3 sets the coverage rate that accurately captures severe disasters.

Acknowledgments

This study was supported by JSPS KAKENHI (Grant Number: 22K20455, 25K08248). This research used computational resources of the supercomputer Fugaku and other computers of the High Performance Computing Infrastructure (HPCI) system provided by the RIKEN R-CCS (ID: hp200128, hp210166, and hp220167) supported by MEXT (JPMXP1020200305) as “Programme for Promoting Research on the Supercomputer Fugaku” (Large Ensemble Atmospheric and Environmental Prediction for Disaster Prevention and Mitigation). Valuable comments from the editor and two anonymous reviewers are appreciated. We thank the Japan Meteorological Agency for providing the Runoff Index Model (RIM). An MRI supercomputer system was used for RIM calculations. We would like to thank Editage (www.editage.jp) for English language editing.

References

775 Akamatsu, T. 2021: Chapter 5 Flood inundation simulation on the Kuma River, The disaster
776 investigation report of the Kyushu heavy rainfall in July 2020, *Committee on*
777 *Hydroscience and Hydraulic Engineering*, 73-81 (in Japanese). [Available at;
778 https://committees.jsce.or.jp/hydraulic05/system/files/令和 2 年 7 月九州豪雨災害調査団
779 報告書 HP.pdf.]

780 Araki, K., T. Kato, Y. Hirockawa, and W. Mashiko, 2020: Characteristics of atmospheric
781 environments of quasi-stationary convective bands in Kyushu, Japan during the July
782 2020 heavy rainfall event. *SOLA*, **17**, 8−15.

783 Cloke, H. L. and Pappenberger, F., 2009: Ensemble flood forecasting: A review, *J. Hydrol.*,
784 **375**, 613–626.

785 Duc L., and K. Saito, 2017: On Cost Functions in the Hybrid Variational–Ensemble Method,
786 *Mon. Wea. Rev.* **145**, 2071-2082.

787 Duc L., T. Kawabata, K. Saito, and T. Oizumi, 2021: Forecasts of the July 2020 Kyushu
788 heavy rain using a 1000-member ensemble Kalman filter, *SOLA*, **17**, 41-47.

789 Geospatial Information Authority of Japan, 2020: Estimated inundation map. [Available at;
790 https://www1.gsi.go.jp/geowww/saigai/202007/shinsui/02_shinsui_kuma_tougou.pdf.]

791 Hanasaki, R., Y.Ishitsuka, D. Yamazaki, and K. Yoshimura, 2019: Implementing reservoir
792 operation in a probabilistic flood forecast system and its application to 2015 Kinu river
793 flood. *J. of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, **75** (2),
794 L_151-L156 (in Japanese).

795 Hirockawa, Y., T. Kato, K. Araki, and W. Mashiko, 2020a: Characteristics of an extreme
796 rainfall event in Kyushu district, southwestern Japan in early July 2020. *SOLA*, **16**,
797 265−270.

798 Hirockawa, Y., T. Kato, H. Tsuguti, and N. Seino, 2020b: Identification and classification of
799 heavy rainfall areas and their characteristic features in Japan. *J. Meteor. Soc. Japan,* **98**,
800 835−857.

801 Ikuta, Y., H. Seko, and Y. Shoji, 2020: Assimilation of shipborne precipitable water vapour
802 by Global Navigation Satellite Systems for extreme precipitation events, *Quart. J. Roy*.
803 Meteor. Soc., **148**, 57-75.

Ito J., H. Tsuguchi, S. Hayashi, and H. Niino., 2021: Idealized high-resolution simulations of a back-building convective system that causes torrential rain. *J. Atmos. Sci.*, **78**. 117-132.

Ito.T., K. OTA, and Y. Nihei.,2021: Flood risk probability prediction using a rainfall index equivalent warning level 4 for each river system. *J. of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, **77** (2), L_295-L_300.

Japan Institute of Country-ology and Engineering, 1999: Guidance for Small and Medium River Plans. *JICE* (in Japanese). [Available at: https://www.jice.or.jp/cms/kokudo/pdf/tech/material/pla-boo-1-01.pdf.]

Japan Meteorological Agency, 2017: "Chapter 2: A New Index for Heavy Rain and Flood Warning Operations. *Training note of weather forecast techniques*, Japan Meteorological Agency, 33-34. (in Japanese). [Available at: https://www.jma.go.jp/jma/kishou/books/yohkens/22/chapter2.pdf.]

Japan Meteorological Agency, 2018: Chapter 2: Verification of accuracy of "Indices and Criteria" used in heavy rain and flood warnings. *Training note of weather forecast techniques*, Japan Meteorological Agency, 45 pp. (in Japanese). [Available at: https://www.jma.go.jp/jma/kishou/books/yohkens/24/chapter2.pdf.]

Japan Meteorological Agency 2019: Outline of the operational numerical weather prediction at the Japan Meteorological Agency. Appendix to WMO technical progress report on the global data-processing and forecasting system and numerical weather prediction. Japan Meteorological Agency, 89-113.

Kato, T. 2020: Quasi-stationary band-shaped precipitation systems, named “Senjo-Kousuitai”, causing localized heavy rainfall in Japan. *J. Met Soc. Japan*, **98**, 485−509.

Kawabata, T., T. Kuroda, H. Seko, and K. Saito, 2011: A cloud-resolving 4D-Var assimilation experiment for a local heavy rainfall event in the Tokyo metropolitan area, *Mon. Wea. Rev.*, **139**, 1911-1931.

Kawabata, T. and G. Ueno, 2020: Non-gaussian probability densities of convection initiation and development investigated using a particle filter with a storm-scale numerical weather prediction model. *Mon. Wea. Rev.*, **148**, 3-20.

Kobatake, S., and Y. Ishihara, 1983: Synthetic runoff model for flood forecasting Kobayashi. Proceedings of the Japan Society of Civil Engineers, **337**, 35 (in Japanese). https://doi.org/10.2208/jscej1969.1983.337_129.

836 Kobayashi, K., S. Otsuka, Apip, and K. Saito, 2016: Ensemble flood simulation for a small
837 dam catchment in Japan using 10 and 2 km resolution nonhydrostatic model rainfalls,
838 *Nat. Hazards Earth Syst. Sci*., **16**, 1821–1839.

839 Kobayashi, K., L. Duc, APIP, T. Oizumi, and K. Saito, 2020: Ensemble flood simulation for
840 a small dam catchment in Japan using nonhydrostatic model rainfalls. Part 2: Flood
841 forecasting using 1600-member 4D-EnVar-predicted rainfalls. *Nat. Hazards Earth Syst.*
842 *Sci.*, **20**, 755-770.

843 Kobayashi, K., L. Duc, T. Kawabata, A. Tamura, T. Oizumi, K. Saito, D. Nohara, and T.
844 Sumi, 2023: Ensemble rainfall–runoff and inundation simulations using 100 and 1000
845 member rainfalls by 4D LETKF on the Kumagawa River flooding 2020. *Prog. Earth*
846 *Planet Sci.,* **10**. https://doi.org/10.1186/s40645-023-00537-3

847 Kunii, N., 2014: The 1000-member ensemble Kalman filtering with the JMA nonhydrostatic
848 mesoscale model on the K Computer. *J. Met Soc. Japan*, 92, 623-633.

849 Kunii, M., M. Otsuka1, K. Shimoji, and H. Seko, 2016: Ensemble data assimilation and
850 forecast experiments for the September 2015 heavy rainfall event in Kanto and Tohoku
851 regions with atmospheric motion vectors from Himawari-8, *SOLA*, **12**, 209-214.

852 Ma W, Y. Ishitsuka, A. Takeshima, K. Hibino, D. Yamazaki, K. Yamamoto, M. Kachi, R. Oki,
853 T. Oki, and K. Yoshimura., 2019: Applicability of a nationwide flood forecasting system
854 for Typhoon Hagibis 2019. Sci Rep. **11**, https://doi.org/10.1038/s41598-021-89522-8.

855 Manning, R. (1891). "On the flow of water in open channels and pipes". Transactions of the
856 Institution of civil engineers of Ireland. **20**: 161–207.

857 Nagata, K., 2011: Quantitative precipitation estimation and quantitative precipitation
858 forecasting by the Japan meteorological agency. *RSMC Tokyo Typhoon Center*
859 *Technical Review*. **13**, 37-50.

860 Nakayama, S., R. Seto, S. Kanae, T. Oizumi, T. Ota, T. Kawabata, O. Sasaki, and L. DUC,
861 2024: Landslide hazard prediction using a 200-member ensemble for the July 2020
862 heavy rainfall in Kyushu. *J. of Japan Society of Civil Engineers,* **81 (16)** (in Japanese).

863 Necker, T., S. Geiss, M. Weissmann, J. Ruiz, T. Miyoshi, and G.Y. Lien, 2020: A convective-
864 scale 1000-member ensemble simulation and potential applications. *Quart. J. Roy.*
865 *Meteor. Soc*., **146**, 1423-1442, https://doi.org/10.1002/qj.3744.

866 Nihei, Y. 2021: Chapter 8 Relation between damage of building and human damage on the
867 Kuma River basin, The disaster investigation report of the Kyushu heavy rainfall in July

2020, *Committee on Hydroscience and Hydraulic Engineering*, 100-113 (in Japanese). [Available at: https://committees.jsce.or.jp/hydraulic05/system/files/令和 2 年 7 月九州豪雨災害調査団報告書 HP.pdf.]

Ogura, Y., 1991: Analyses and mechanisms of intense precipitation. *Tenki*, **38**, 276–287 (in Japanese). [Available at: https://www.metsoc.jp/tenki/pdf/1991/1991_05_0276.pdf.]

Ohmoto, T. 2021: Chapter 1 A summary of damage on the Kuma River basin, The disaster investigation report of the Kyushu heavy rainfall in July 2020, *Committee on Hydroscience and Hydraulic Engineering*, 35-46 (in Japanese). [Available at: https://committees.jsce.or.jp/hydraulic05/system/files/令和 2 年 7 月九州豪雨災害調査団報告書 HP.pdf.]

Ohta, T. 2017: Chapter 2 New index for heavy rainfall and flood warning, Technical document of quantitative prediction, **24**, 30-41 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/yohkens/22/chapter2.pdf.]

Ohta, T., and Y., Makihara, 2018: Technical developments for mitigating disasters from Inundation and flood and dissemination of their results. *Tenki*, **66**, 723-742 (in Japanese). [Available at https://doi.org/10.24761/tenki.66.11_723.]

Ota, T, S. Hashiguchi, J. Hotta, A. Sobajima, A. Yano, G. Nagaoka, S. Tateno, Y. Murakami, T. Okamoto, S. Chijimatsu, Y. Tsuboi, and J. Take, 2023: JMA's hazard potential index and real-time risk map related to heavy rain. *Weather service bulletin.* **90,** 1-96 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/sokkou/90/vol90_8.pdf.]

Oizumi T., K. Saito, J. Ito, T. Kuroda, and L. Duc, 2018: Ultra-high-resolution numerical weather prediction with a large domain using the K computer: A case study of the Izu Oshima heavy rainfall event on october 15-16, 2013, *J. Meteor. Soc. Jpn*., **96**, 25-54.

Oizumi, T., K. Saito, L. Duc, and J. Ito, 2020: Ultra-high-resolution numerical weather prediction with a large domain using the K computer: Part 2 Case of the Hiroshima heavy rainfall event on August 2014 and size dependency of simulated convection cores on model resolutions. *J. Meteor. Soc. Jpn*., **98**, 1163-1182.

Onaka N., Y. Akamatsu, S. Yano, and Y. Nihei, 2021: Elucidation of the actual state on the flood in the upstream part of the Kuma River during the heavy rain in July 2020. *J.*

of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), **77**, L_203-L214 (in Japanese).

Ono, K., M. Kunii, and Y. Honda, 2020: The regional model-based mesoscale ensemble prediction system, MEPS, at the Japan Meteorological Agency. *Quart. J. Roy*. Meteor. Soc., **147**, 465-484.

Saito, K., T. Fujita, Y. Yamada, J. Ishida, Y. Kumagai, K. Aranami, S. Ohmori, R. Nagasawa, S. Kumagai, C. Muroi, T. Kato, H. Eito and Y. Yamazaki, 2006: The operational JMA nonhydrostatic mesoscale model. *Mon. Wea. Rev.*, **134**, 1266-1298.

Saito, K., T. Kawabata, H. Seko, T. Miyoshi, L. Duc, T. Oizumi, M. Kunii, G. Chen, K. Ito, J. Ito, S. Yokota, W. Mashiko, K. Kobayashi, S. Fukui, E. Tochimoto, A. Amemiya, Y. Maejima, T. Honda, H. Niino, and M. Satoh, 2023: Forecast and numerical simulation studies on meso/micro-scale high-impact weathers using high-performance computing in Japan. Numerical Weather Prediction: East Asian Perspectives. 461-481.

Sayama, T., M. Yamada., Y. Sugawara, and D. Yamazaki, 2020: 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. *Prog. Earth Planet Sci.* **7,** https://doi.org/10.1186/s40645-020-00391-7

Tatehira, R., 1999 : Decision making on weather forecast. *Tokyodo Shuppan Co. Ltd*, p.142. (in Japanese)

Tanaka, N., T Ohta, and Y. Makihara, 2008: Flood warning/advisory improvement based on JMA Runoff Index. *Weather service bulletin*. **75**, 35-50 (in Japanese). [Available at: https://www.jma.go.jp/jma/kishou/books/sokkou/75/vol75p035.pdf.]

Terasaki, K., and T. Miyoshi, 2017: Assimilating advanced microwave sounding unit-A satellite radiances with the NICAMLETKF. *J. Meteor. Soc. Japan*, **95**, 433−446.

Terasaki, K., S. Kotsuki, and T. Miyoshi, 2019: Multi-year analysis using the NICAM-LETKF data assimilation system. *SOLA*, **15**, 41−46.

Terasaki, K., and T. Miyoshi, 2022: A 1024-Member NICAM-LETKF Experiment for the July 2020 heavy rainfall event. *SOLA*, **18A**, 8-14.

Tsuguti, H., and T. Kato, 2014: Objective extraction of heavy rainfall events and statistical analysis on their characteristic features. *Tenki*, **61**, 455-469 (in Japanese).

United Nations Educational, Scientific and Cultural Organization, 2020; United Nations World water development report 2020: Water and climate change. *UNESCO*. [Available at: https://en.unesco.org/themes/water-security/wwap/wwdr/2020.]

World Meteorological Organization, 2015; WMO Guidelines on multi-hazard impact-based forecast and warning services. *WMO- No. 1150*, 21 pp.

World Meteorological Organization, 2021; WMO Guidelines on multi-hazard impact-based forecast and warning services (WMO-No. 1150), Part II: Putting Multi-Hazard IBFWS into Practice. *WMO- No. 1150*, 47 pp.

Wu, W., Emerton, R., Duan, Q., Wood, A. W., Wetterhall, F., and Robertson, D. E., 2020: Ensemble flood forecasting: Current status and future opportunities. *WIREs Water,* **7.** [Available at: https://doi.org/10.1002/wat2.1432.]

List of Figures

952 Fig. 3. Probability distribution map of 50 mm $3h^{-1}$ at 0300 and 0600 JST on July 4. LETKF1000
953 are (a) and (b), LETKF100 are (c) and (d), and MEPS are (e) and (f). Note that the
954 initial time of MEPS is not 1800 JST but 2100 JST. Reproduced from Duc et al. (2021).
955 Fig. 4. (a)–(c) Hourly precipitation distributions by RA (modified Fig. 3 in Duc et.al (2021)).
956 Blacklines in (d)–(f) illustrate rivers which exceeded HM in the RIM using RA at 0500,
957 0700, and 0900 JST on July 4.
958 Fig. 5. (a)–(c) Hourly probability distributions of rain greater than 20 mm from LETKF1000
959 at 0500, 0700, and 0900 JST on July 4 (modified Fig. 4 in Duc et al. (2021)). (d)–(f)
960 Exceedance probability (%) to HM for each river by the RIM using LETKF1000.
961 Fig. 6. Enlarged view of the black rectangles in Figs. 5d-f. Circle A is a water level gauge
962 station at Watari.
963 Fig. 7. Calculated indices predicted by the RIM using RA, and LETKF1000, for the criterion
964 levels (30Y, 50Y, and HM) in the rectangle A of Fig. 1c. Red line is RA, the black line
965 is deterministic from LFTKF1000, and the gray dash line is the ensemble mean.
966 Thresholds for the HM, 30Y and 50Y probability rainfall are orange, red, and purple
967 dash lines, respectively. The lines are corresponding left axis. Orange, red, and
968 purple bars indicate the probability (right axis) of exceeding the HM, 30Y and 50Y
969 thresholds, respectively.
970 Fig. 8. Rivers in black in (a)–(c) are exceeded the Criterion level 2 by RA on 0500, 0700,
971 and 0900 on July 4, respectively. (d)–(f) Same as Fig. 5 (d)–(f) but for L2. The Kuma

River (blue line) is excluded from the calculation due to the absence of L2 settings.

Fig. 9. Enlarged view of Figs. 8a-f. Rectangle (A) is the same in Fig. 1c. "The rectangular markers indicate the outlets of the Kawabe River (B), Ogawa River (C), Magouri River (D), Mae River (E), and Yamada River (F). Among these, (B) represents a medium-sized river, while the others are classified as small-sized rivers. Pink area: detected inundation area.

Fig. 10. (a) to (c) Same as Fig.7 but for the Kawabe River. Lines of (d) to (f) are the same as those of (a)-(c) but the bars and thresholds are for the Criterion level. (a) and (d), (b) and (e), (c) and (f) are for LETKF1000, LETKF100, and MEPS, respectively. The index by the RA (red line) is common for all.

Fig. 11. Same as Fig. 10 but for the Ogawa River.

Fig.12. Probability of HM values being exceeded at least once during the forecast period, 1900 JST on 3 July to 1500 JST on 4 July. Probability values are rounded to the nearest 10%. The names of rivers that did not exceed the HM values in the simulation with RA are indicated in italic bold.

Fig.13. Same as in Fig. 12 but the probability that forecasted indices exceed HM value at least once during the Ex_period.

Fig. A1. (a)–(c) Parameters setting of the three-tank model at different rain intensities; (d) Five-tank model for an urban area.

992 List of Tables

993 Table 1. HM exceedance probability for the Ogawa River during the Ex_period when 100,

994 250, and 500 members were sampled from LETKF1000 and LETKF100.

995

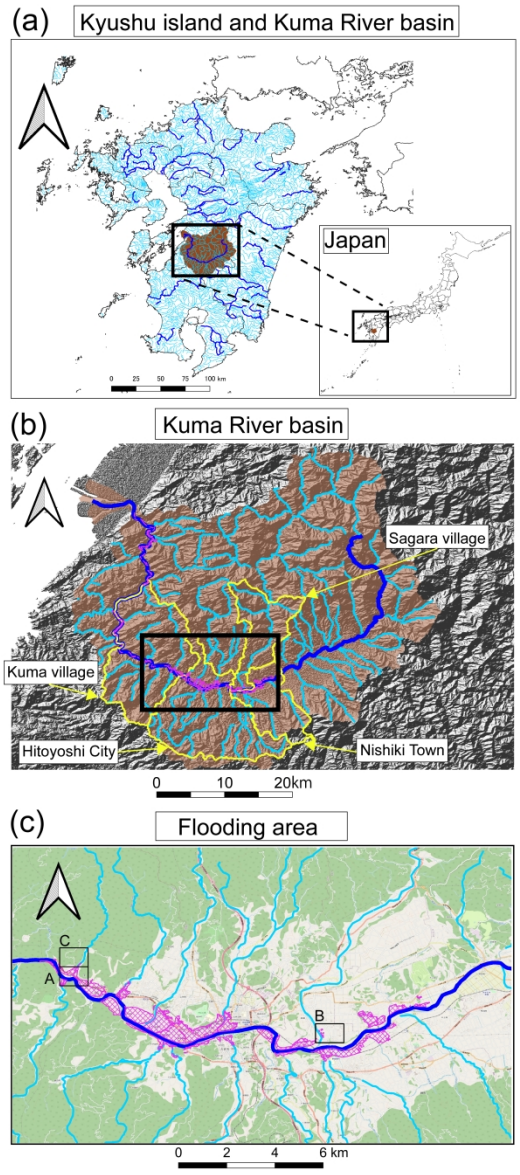

Fig. 1. (a) River networks in Kyushu on RIM and location of the Kuma River Basin. Blue lines: designated forecast river; light blue lines: small and middle-sized rivers. (b) Enlarged view around the Kuma River. Black rectangle indicates the most damaged area. (c) Enlarged view of the black rectangle in (b). Pink mesh area: detected inundation area. Blue line: Kuma River; light blue line: tributaries. Rectangle A is the Watari area, the most human-lost area at the Kuma River. Rectangle B is at the confluence of the Kuma River and Kawabe River. Rectangle C is at the confluence of the Kuma River and Ogawa River.

788x1686mm (118 x 118 DPI)

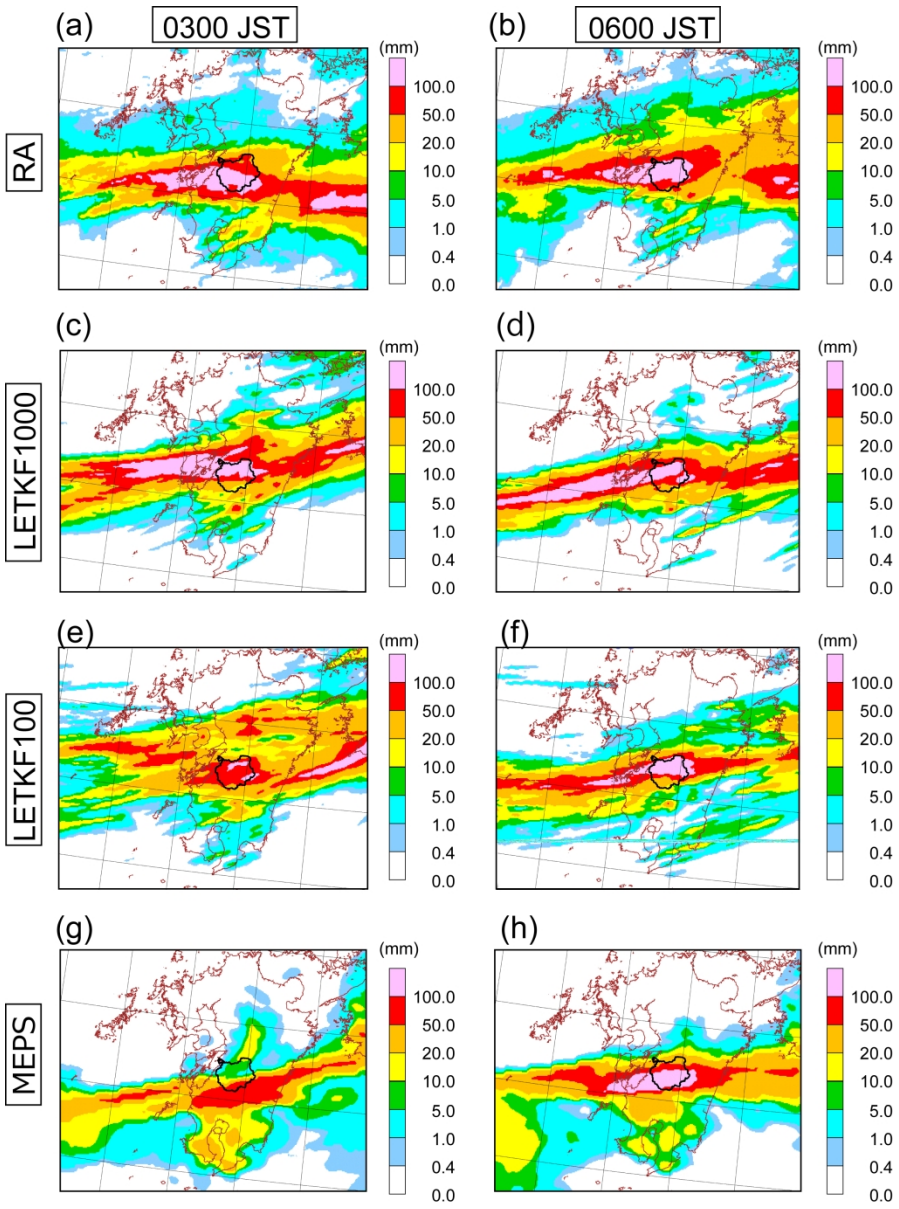

Fig. 2. Three-hour observed precipitation at (a) 0300 and (b) 0600 JST on July 4. (c, d) The deterministic forecasts by LETKF1000, (e, f) Same as in (c,d) except by LETKF100. (g, h) Same as in (c, d) except by MEPS. Reproduced from Duc et al. (2021).

562x758mm (118 x 118 DPI)

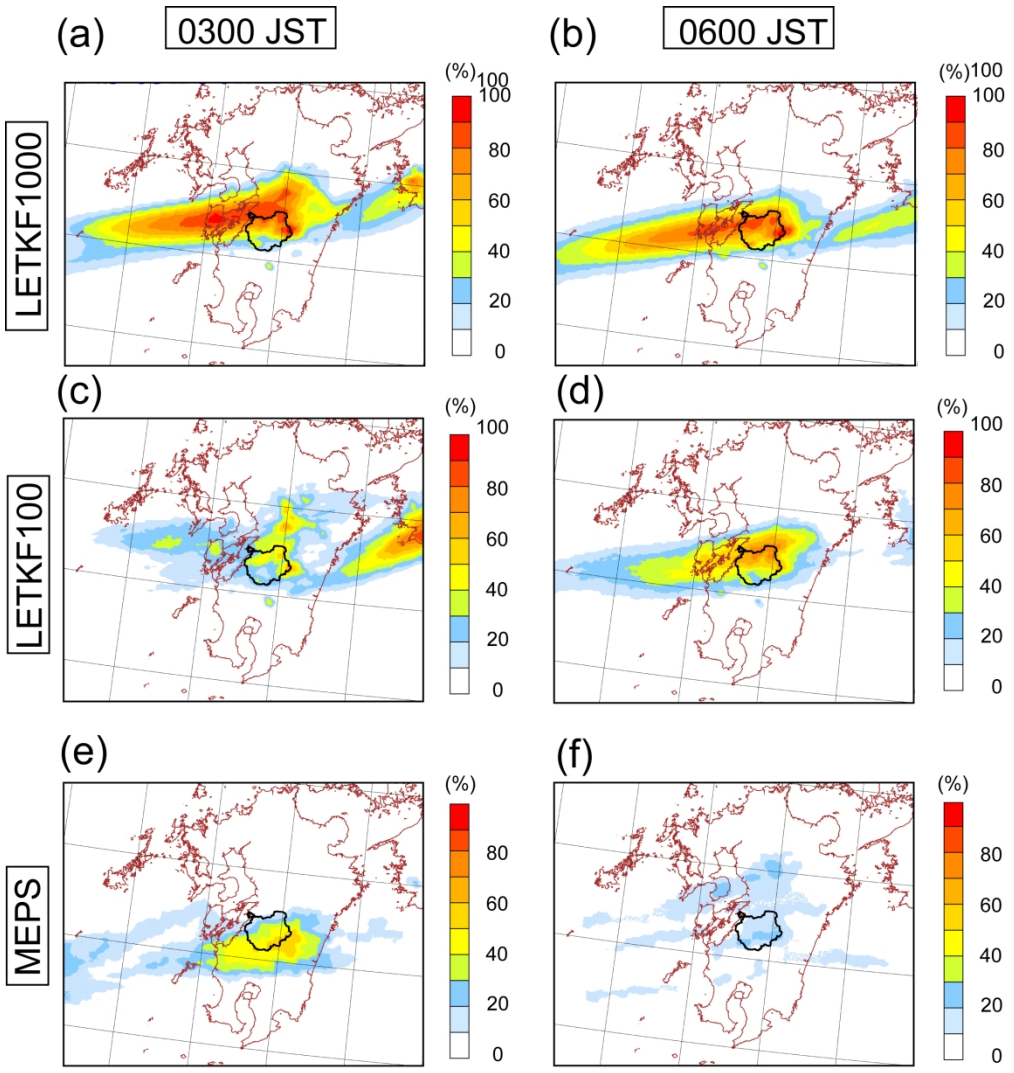

Fig. 3. Probability distribution map of 50 mm 3h-1 at 0300 and 0600 JST on July 4. LETKF1000 are (a) and (b), LETKF100 are (c) and (d), and MEPS are (e) and (f). Note that the initial time of MEPS is not 1800 JST but 2100 JST. Reproduced from Duc et al. (2021).

538x572mm (118 x 118 DPI)

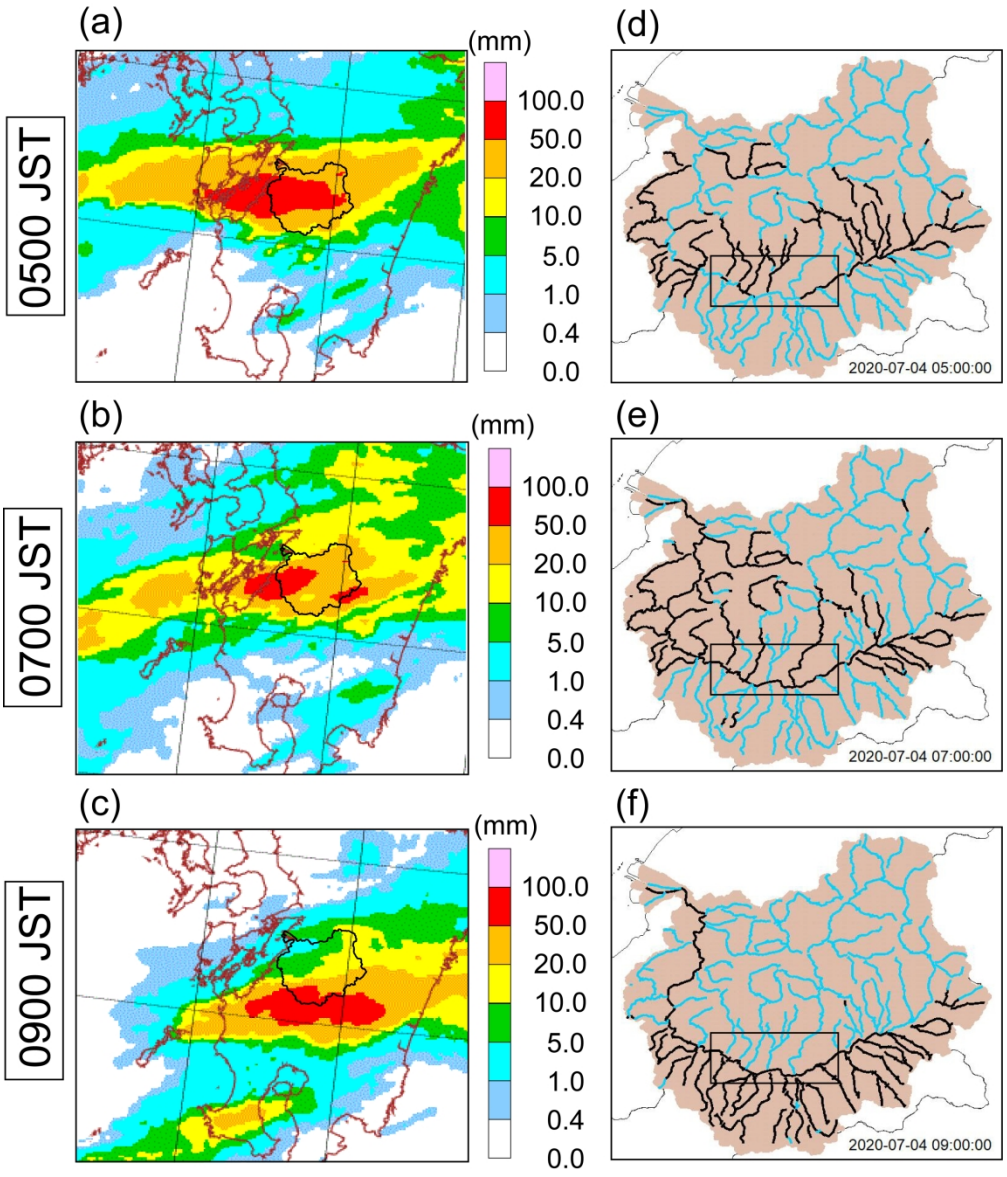

Fig. 4. (a)–(c) Hourly precipitation distributions by RA (modified Fig. 3 in Duc et.al (2021)). Blacklines in (d)–(f) illustrate rivers which exceeded HM in the RIM using RA at 0500, 0700, and 0900 JST on July 4.

589x692mm (118 x 118 DPI)

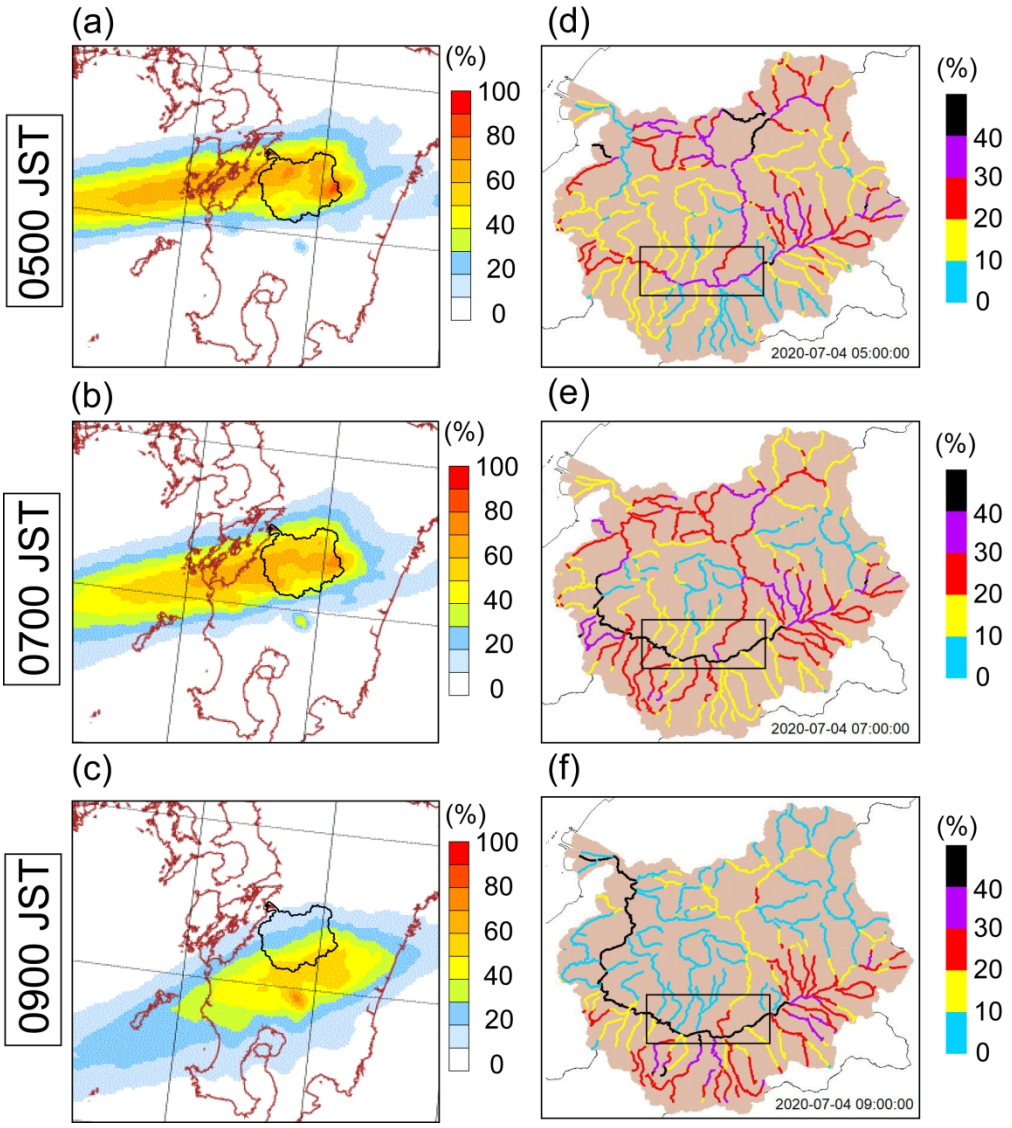

Fig. 5. (a)–(c) Hourly probability distributions of rain greater than 20 mm from LETKF1000 at 0500, 0700, and 0900 JST on July 4 (modified Fig. 4 in Duc et al. (2021)). (d)–(f) Exceedance probability (%) to HM for each river by the RIM using LETKF1000.

609x683mm (118 x 118 DPI)

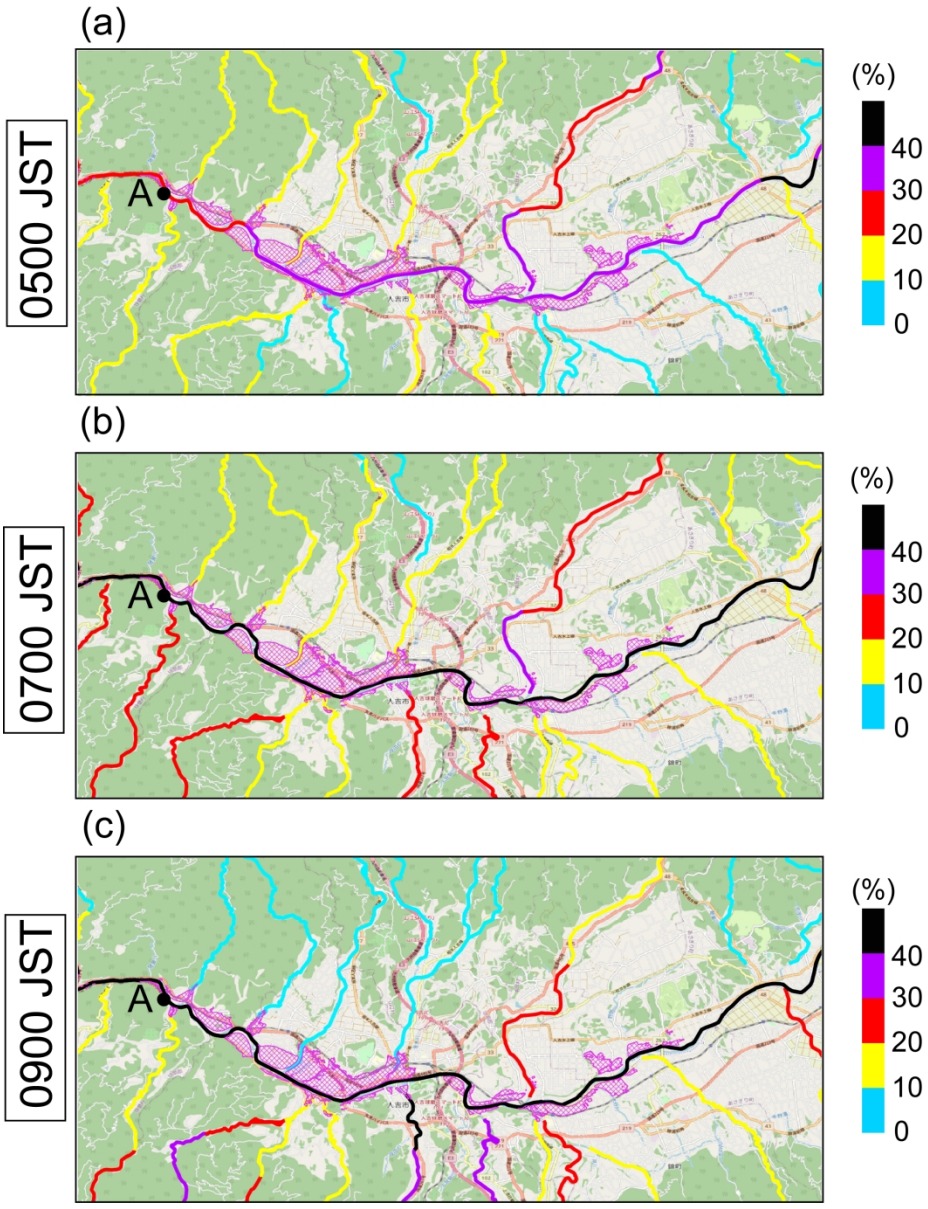

Fig. 6. Enlarged view of the black rectangles in Figs. 5d-f. Circle A is a water level gauge station at Watari.

522x680mm (118 x 118 DPI)

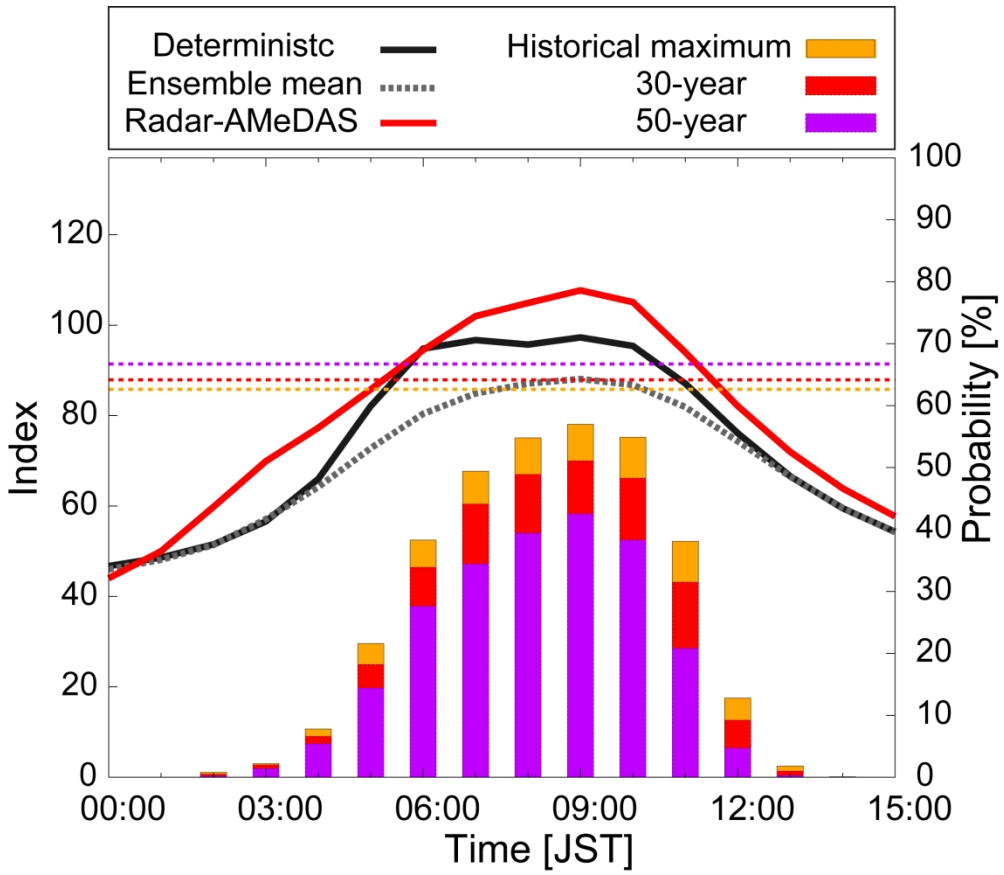

Fig. 7. Calculated indices predicted by the RIM using RA, and LETKF1000, for the criterion levels (30Y, 50Y, and HM) in the rectangle A of Fig. 1c. Red line is RA, the black line is deterministic from LFTKF1000, and the gray dash line is the ensemble mean. Thresholds for the HM, 30Y and 50Y probability rainfall are orange, red, and purple dash lines, respectively. The lines are corresponding left axis. Orange, red, and purple bars indicate the probability (right axis) of exceeding the HM, 30Y and 50Y thresholds, respectively.

547x475mm (118 x 118 DPI)

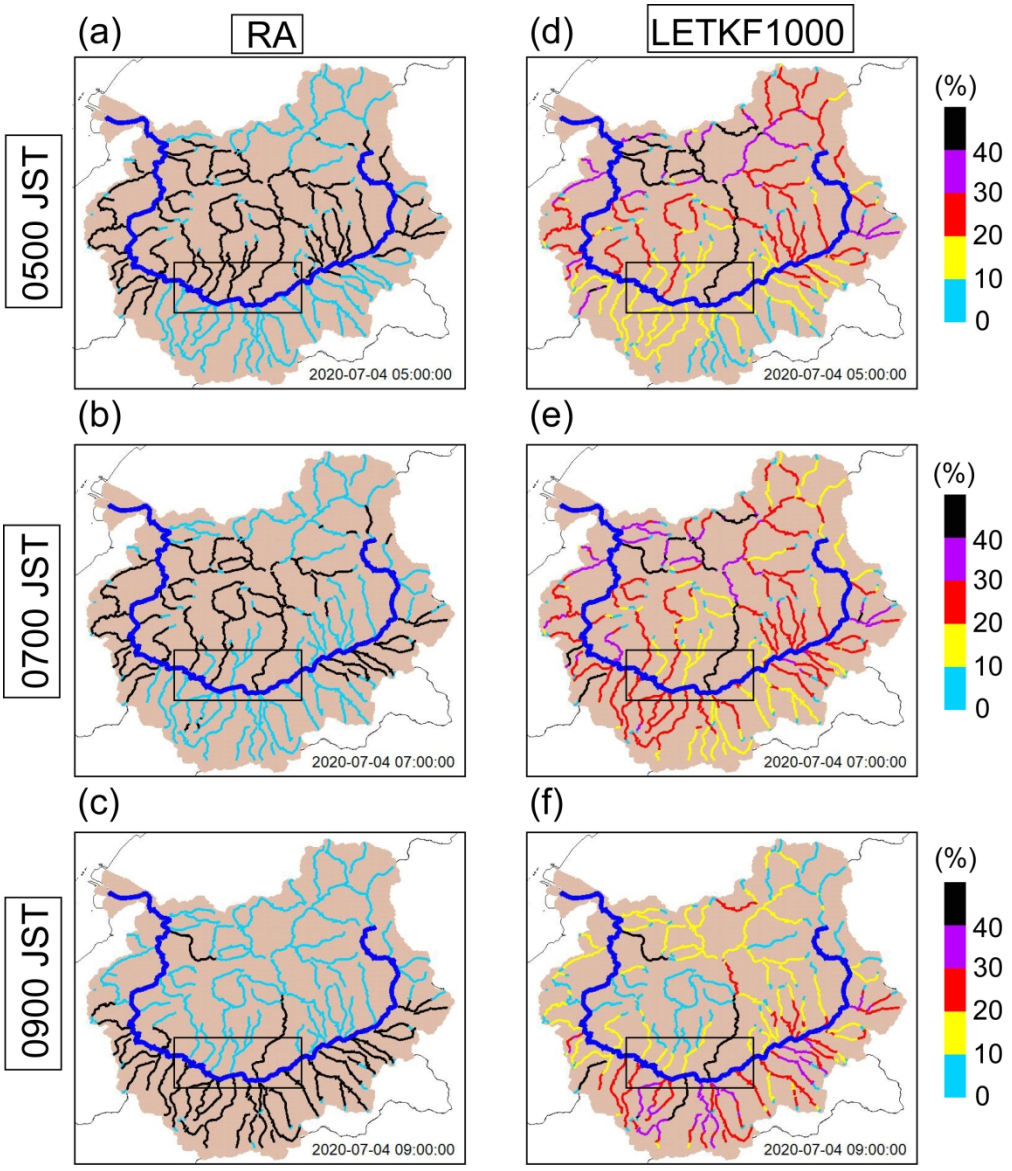

Fig. 8. Rivers in black in (a)–(c) are exceeded the Criterion level 2 by RA on 0500, 0700, and 0900 on July 4, respectively. (d)–(f) Same as Fig. 5 (d)–(f) but for L2. The Kuma River (blue line) is excluded from the calculation due to the absence of L2 settings.

590x685mm (118 x 118 DPI)

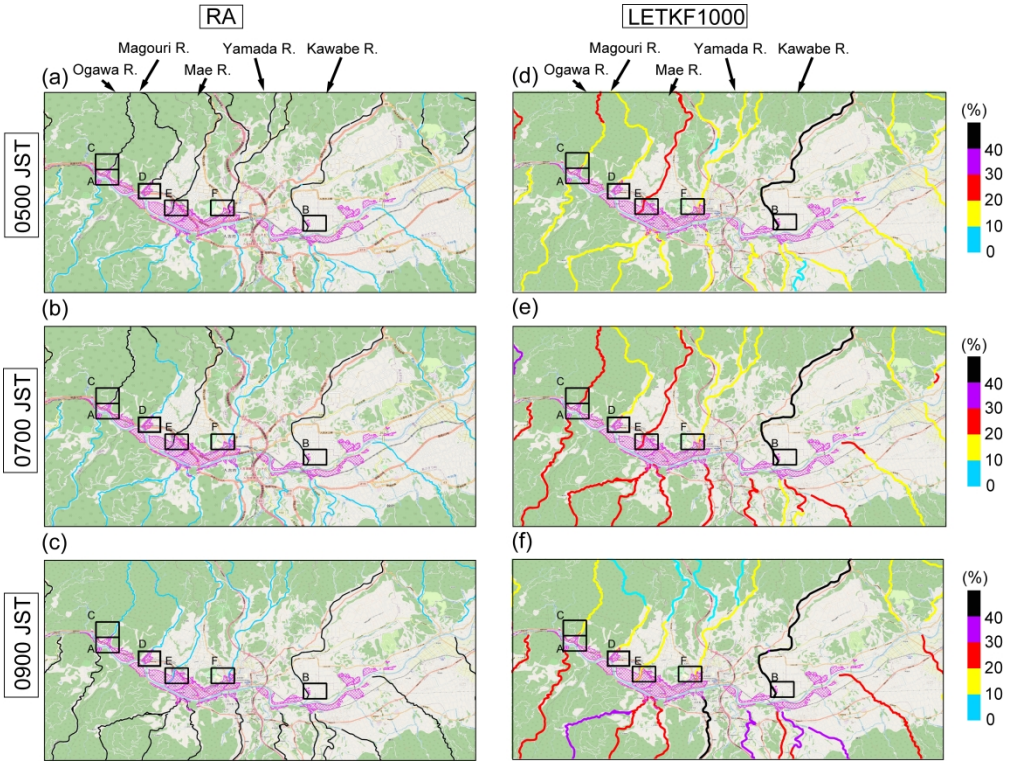

Fig. 9. Enlarged view of Figs. 8a-f. Rectangle (A) is the same in Fig. 1c. "The rectangular markers indicate the outlets of the Kawabe River (B), Ogawa River (C), Magouri River (D), Mae River (E), and Yamada River (F). Among these, (B) represents a medium-sized river, while the others are classified as small-sized rivers. Pink area: detected inundation area.

977x740mm (118 x 118 DPI)

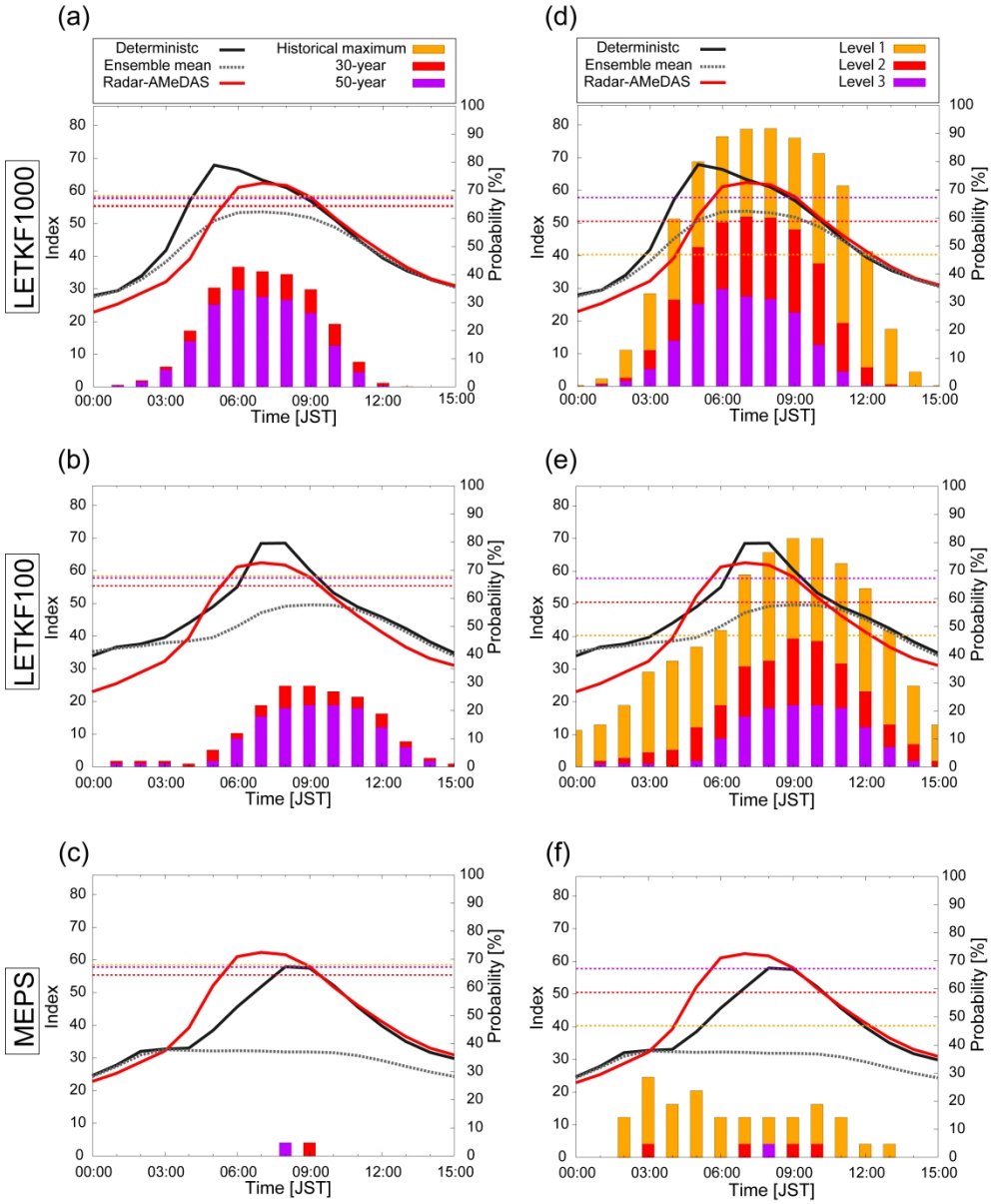

Fig. 10. (a) to (c) Same as Fig.7 but for the Kawabe River. Lines of (d) to (f) are the same as those of (a)-(c) but the bars and thresholds are for the Criterion level. (a) and (d), (b) and (e), (c) and (f) are for LETKF1000, LETKF100, and MEPS (MSM for deterministic run), respectively. The Index by the RA (red line) is common for all.

1187x1445mm (118 x 118 DPI)

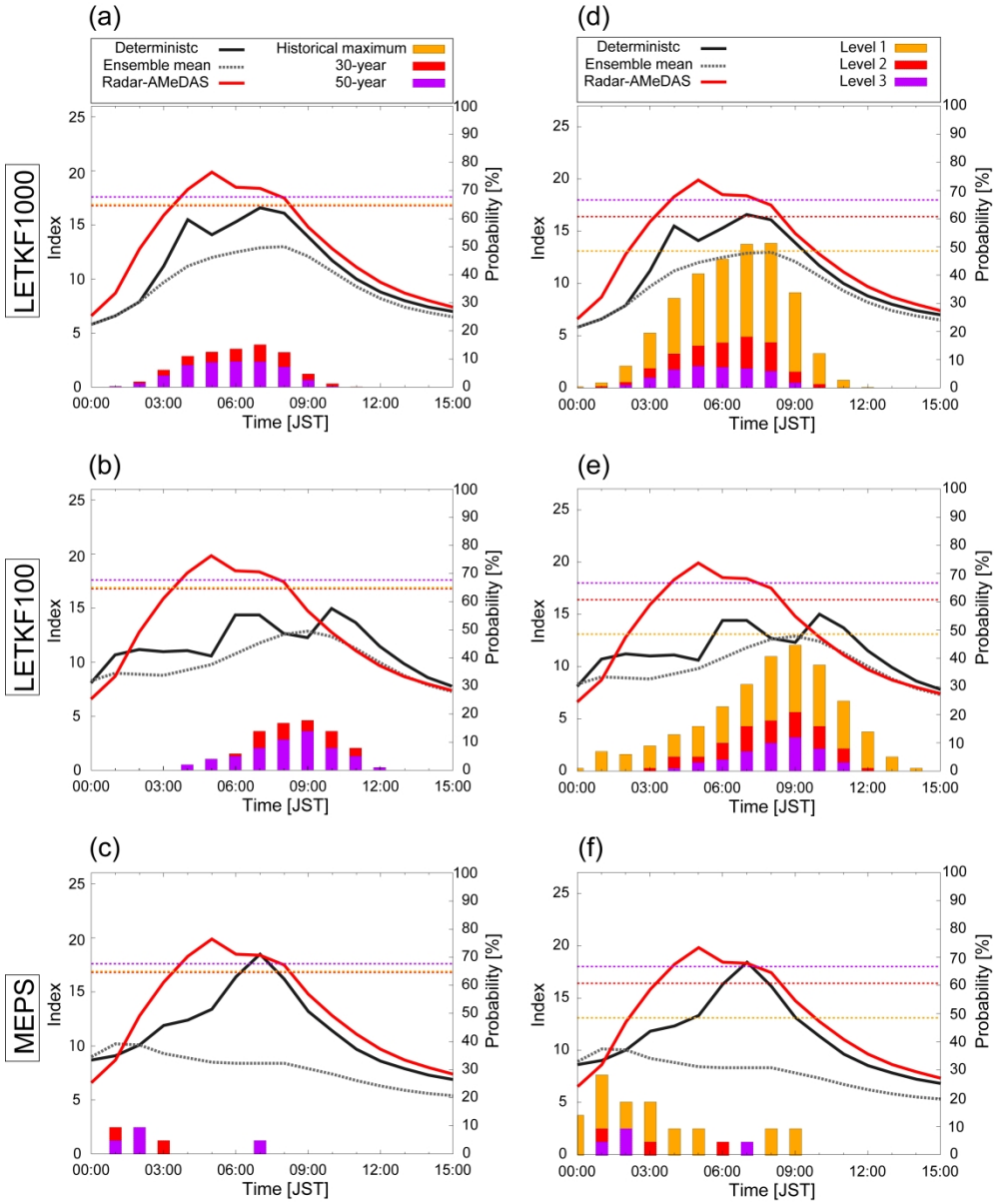

Fig. 11. Same as Fig. 10 but for the Ogawa River.

1187x1439mm (118 x 118 DPI)

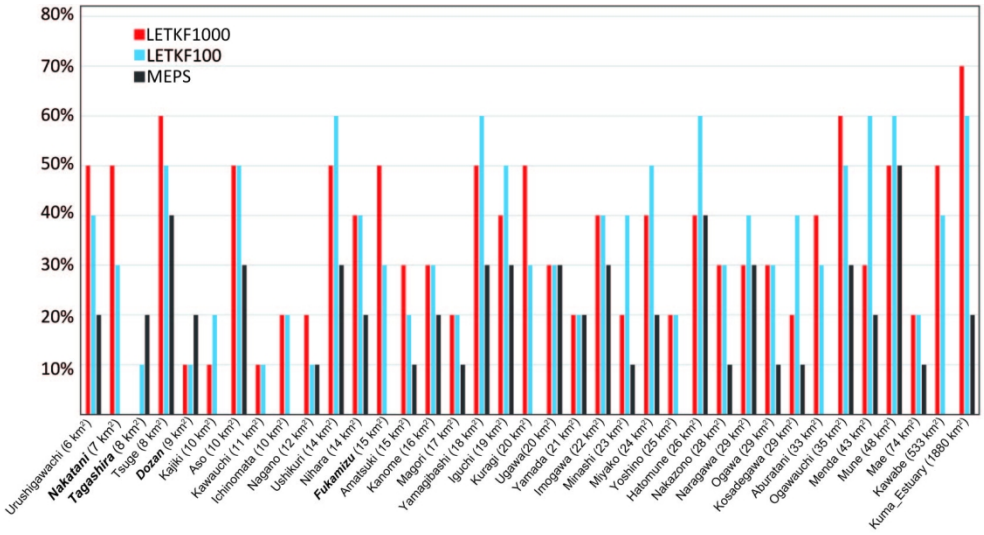

Fig. 12. Probability of HM values being exceeded at least once during the forecast period, 1900 JST on 3 July to 1500 JST on 4 July. Probability values are rounded to the nearest 10%. The names of rivers that did not exceed the HM values in the simulation with RA are indicated in italic bold.

287x169mm (150 x 150 DPI)

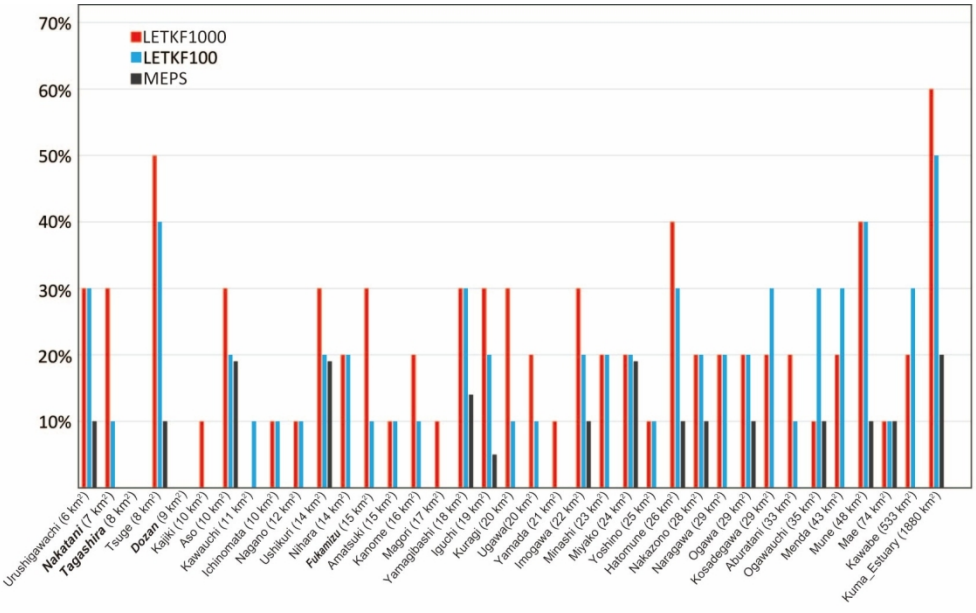

Fig.13. Same as in Fig. 12 but the probability that forecasted indices exceed HM value at least once during the Ex_period.

289x178mm (150 x 150 DPI)

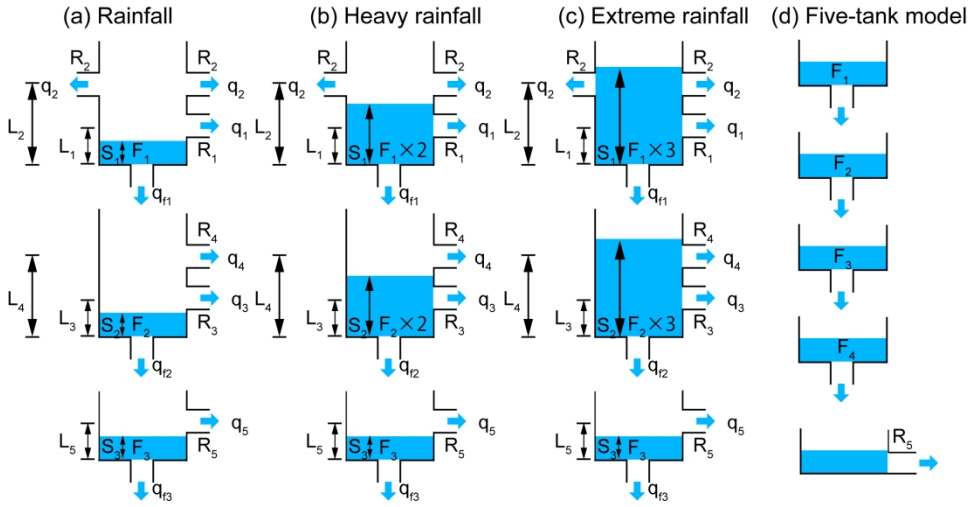

Fig. A1. (a)–(c) Parameters setting of the three-tank model at different rain intensities; (d) Five-tankmodel for an urban area.

551x277mm (118 x 118 DPI)

LETKF1000							
Averaged member	Probability (%) for every 1000 and 100 member			Averaged member	Probability (%) for every 50 member		
	7/4 5:00	7/4 6:00	7/4 7:00		7/4 5:00	7/4 6:00	7/4 7:00
001-1000	**13%**	**13%**	**15%**	001-50	6%	10%	18%
001-100	9%	10%	18%	051-100	12%	10%	18%
101-200	12%	14%	10%		Probability (%) for every 250 member		
201-300	13%	13%	13%	001-250	10%	12%	13%
301-400	14%	19%	13%	251-500	14%	15%	15%
401-500	14%	11%	16%	501-750	12%	15%	16%
501-600	10%	13%	16%	751-1000	12%	12%	13%
601-700	15%	19%	17%		Probability (%) for every 500 member		
701-800	9%	12%	14%				
801-900	9%	9%	12%	001-500	12%	13%	14%
901-1000	20%	10%	20%	501-1000	12%	13%	15%

LETKF100			
Averaged member	Probability (%) for every 100 and 50 member		
	7/4 5:00	7/4 6:00	7/4 7:00
001-100	**4%**	**6%**	**14%**
001-050	4%	6%	16%
051-100	4%	6%	12%

Table 1. HM exceedance probability for the Ogawa River during the Ex_period when 100, 250, and 500 members were sampled from LETKF1000 and LETKF100.

283x115mm (150 x 150 DPI)