



OPEN

Applicability of a nationwide flood forecasting system for Typhoon Hagibis 2019

Wenchao Ma^{1✉}, Yuta Ishitsuka², Akira Takeshima¹, Kenshi Hibino¹, Dai Yamazaki¹, Kosuke Yamamoto³, Misako Kachi³, Riko Oki³, Taikan Oki^{4,5} & Kei Yoshimura^{1✉}

Floods can be devastating in densely populated regions along rivers, so attaining a longer forecast lead time with high accuracy is essential for protecting people and property. Although many techniques are used to forecast floods, sufficient validation of the use of a forecast system for operational alert purposes is lacking. In this study, we validated the flooding locations and times of dike breaking that had occurred during Typhoon Hagibis, which caused severe flooding in Japan in 2019. To achieve the goal of the study, we combined a hydrodynamic model with statistical analysis under forcing by a 39-h prediction of the Japan Meteorological Agency's Meso-scale model Grid Point Value (MSM-GPV) and obtained dike-break times for all flooded locations for validation. The results showed that this method was accurate in predicting floods at 130 locations, approximately 91.6% of the total of 142 flooded locations, with a lead time of approximately 32.75 h. In terms of precision, these successfully predicted locations accounted for 24.0% of the total of 542 locations under a flood warning, and on average, the predicted flood time was approximately 8.53 h earlier than a given dike-break time. More warnings were issued for major rivers with severe flooding, indicating that the system is sensitive to extreme flood events and can issue warnings for rivers subject to high risk of flooding.

As one of the most frequently occurring natural disasters, floods threaten millions of people and significantly damage socioeconomic development. Under the current warmer climate, flood risks have increased in most of the world^{1–3}. Japan has lost billions of homes and businesses and hundreds of lives to frequent typhoons. According to the Japan Meteorological Agency (JMA), approximately 799 typhoons approached Japan, with 206 landing in the country from 1951 to 2019. In 2019, Typhoon Hagibis swept central, eastern, and northern Japan from October 11 to 13. This typhoon increased the damage caused by Typhoon Faxai, which destroyed most of the residential regions in Chiba Prefecture (eastern part of Japan). The typhoon's trajectory covered 15 prefectures, and heavy rain warnings were issued in these regions. It resulted in 86 deaths, three missing persons, nearly 500 people injured, and approximately 400 billion dollars of damage. According to reports from the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), 142 locations sustained structural damage, such as dike failure (<https://www.mlit.go.jp/common/001313204.pdf>). This enormous disaster was explained by a sufficient supplement of precipitable water under very humid conditions⁴, which caused a strong convergence of runoff in comparison with other extreme flood events, such as the 2018 event⁵.

In Japan, most rivers and river reaches are closely associated with densely populated regions. These regions become vulnerable when flooding occurs due to heavy rainfall associated with events such as typhoons and storms. Floods are the inevitable rapid and dangerous results of typhoon events because most urban areas lie on a floodplain. However, no effective river flood forecasting method is available to provide a sufficiently long warning time with high accuracy. The MLIT provides accurate predictions with a few hours of warning before flooding occurs based on the observed upstream water level, but this is too late for people to respond effectively, and the situation is even worse if flooding occurs late in the night. Therefore, precise forecasting with longer lead times is extremely important for densely populated, low-elevation coastal regions. Furthermore, as Japan is mountainous and includes many small basins, a validation study is needed to determine whether numerical flood forecasting can be effective in such a challenging region. However, no numerical flood forecasting system has been tested in Japan.

¹IIS, The University of Tokyo, Tokyo, Japan. ²University of Massachusetts, Amherst, USA. ³Japan Aerospace Exploration Agency (JAXA), Tsukuba, Japan. ⁴United Nations University, Tokyo, Japan. ⁵Graduate School of Engineering, The University of Tokyo, Tokyo, Japan. ⁶These authors contributed equally: Wenchao Ma and Yuta Ishitsuka. ✉email: wma@iis.u-tokyo.ac.jp; kei@iis.u-tokyo.ac.jp

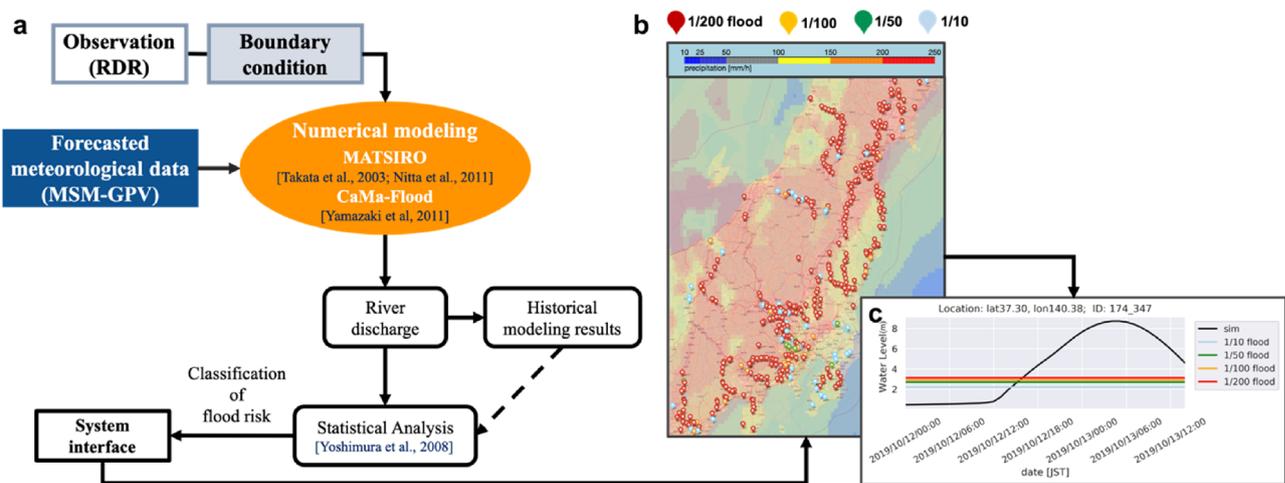


Figure 1. Schematic of the flood forecasting system. (a) Flowchart of the flood forecasting system. (b) Snapshot of the system interface at 00:00 JST on Oct. 12, 2019. Pins of different colors represent once per 200-year floods (red) and once per 100-, 50-, and 10-year floods (orange, green, and light blue, respectively). The real-time forecasting interface can be accessed at <http://apps.diasjp.net/tjpn/training/latest/jpn.html>. (c) The forecasted 1/200-year hydrograph starting from 00:00 JST on Oct. 12, 2019 at 140.38° N, 37.30° E. This figure was generated through Python2.7 (<https://www.python.org/>), and Microsoft PowerPoint Version 16.47 provided by University of Tokyo.

Several flood forecasting systems, with different lengths of lead time, have been developed to cover global or regional scales^{6,7}. For example, at the global scale, there are two flood forecasting systems: the Global Flood Forecasting and Information System (GLOFFIS) run by Deltares⁸ and the Global Flood Awareness System (GloFAS)⁹ developed jointly by the European Commission and the European Centre for Medium-range Weather Forecasts. In addition to GLOFFIS and GloFAS, the Global Flood Monitoring System (GFMS) aims to produce real-time global maps of flood events^{10,11}. There are also several regional-scale forecasting systems, such as the European Flood Awareness System (EFAS) of the European Commission¹², the Hydrologic Ensemble Forecasting Service (HEFS) covering the continental USA by the U.S. National Weather Service¹³, Hydrological Predictions for the Environment (E-HYPE) operated by the Swedish Meteorological and Hydrological Institute¹⁴, and the Flood Forecasting and Warning Service (FFWS) run by the Bureau of Meteorology of Australia. Among these flood forecasting systems, GloFAS can achieve prediction lengths in excess of 25 days for some large basins and up to 20 days for some small basins¹⁵. In comparison with global systems, HEFS, covering the continental USA, strives to forecast with a longer prediction length of up to 1 week⁶, EFAS issues a national warning with lead times of up to 2 days, and the Bureau of Meteorology of Australia issues warnings with a minimum prediction length of 6 h. The wide range of forecasted prediction lengths given by these systems depends on the target of the systems. If the target is decision support for evacuation, high accuracy is required, so the prediction length can be short. However, if the target is early warning to improve preparedness among citizens, longer prediction length is prioritized over accuracy.

More specifically, in the case of flood forecasting in Japan, as stated above, alarms issued by the MLIT have reasonably high accuracy but short prediction length because their goal is to obtain precise and specific flood control locations to evacuate citizens. It is important that citizens are prepared for disaster, so a system with longer prediction length but relatively low prediction accuracy might be desirable, but Japan has no other official flood forecasts due to legal restrictions.

Furthermore, the prediction lengths of the aforementioned systems are clearly related to the predictability of the target, i.e., the lead time, so the definition of lead time is dependent on the catchment structure and the forecasting and warning system facilities¹⁶. Therefore, there is no specific way to validate predictability, which hinders improvements to flood forecasting systems through assessments of the accuracy of forecasting results¹⁷. As an indicator of flooding, the time of a dike break can directly distinguish between a flooded and non-flooded area and thus provide valid information about the time of flooding.

Here, we newly present a flood forecasting system with longer prediction length developed by Ishitsuka^{18,19} and Yoshimura et al.²⁰. Flood forecasting results of Typhoon Hagibis in 2019 were validated using the forecast flooding time and all dike-break times for flooded locations. Specifically, we used 39-h predictions of the JMA Meso-Scale Model grid point values (MSM-GPVs) as forcing data to run a land surface model, the Minimal Advanced Treatments of Surface Interaction and RunOff (MATSIRO) model, to obtain runoff values (Fig. 1a). MATSIRO is a physical-based land surface model, and the simulation covers a horizontal resolution of approximately 5 km (0.05 degrees) from 24° to 46° N latitude and 123° to 148° E longitude in Japan. Then, we employed a catchment-based macroscale floodplain model (CaMa-Flood) to estimate river water depth and flood area for all rivers and streams with an approximately 5-km (0.05 degrees) horizontal resolution. CaMa-Flood calculates the river discharge of a 1-dimensional river channel. The river parameters were calculated from Multi-Error-Removed Improved-Terrain (MERIT) DEM and hydrography (MERIT Hydro)^{21,22} datasets. Subsequently, the statistical distribution of river water depth given by CaMa-Flood was analyzed for comparison with return period

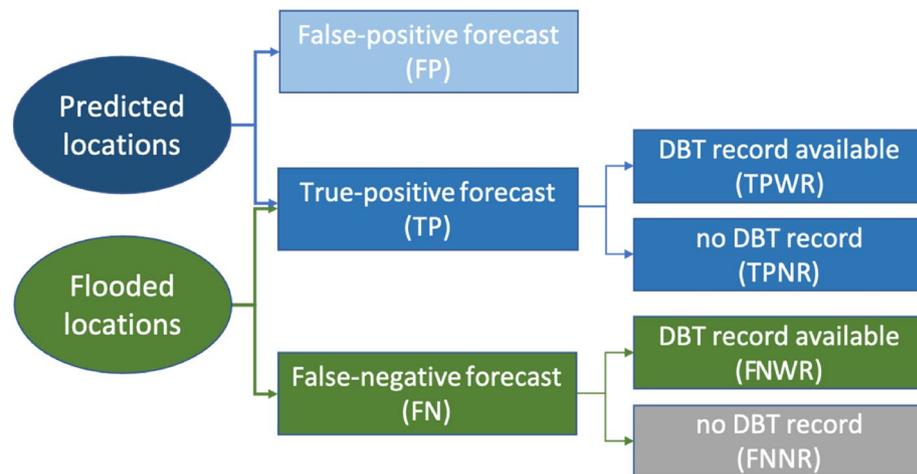


Figure 2. Classification of flood locations related to Typhoon Hagibis. This figure was generated through Microsoft PowerPoint Version 16.47 provided by University of Tokyo.

values used for generating flood alarms. Here, the Gumbel distribution^{23–27} was applied because of its better fitting for extreme value analysis, such as values of extreme flood events^{26,28}.

Results

Simulated flood locations and flood time. Figure 1 shows the procedures of this forecasting system and a screenshot of the interface featuring Typhoon Hagibis. The red pins are flood alarms that occurred at 00:00 JST on October 12 (Fig. 1b). These alarms are updated every 3 h. For each alarm, a hydrograph is archived to show the exact flooding alarms for 1/10-, 1/50-, 1/100-, and 1/200-year return periods (Fig. 1c). To obtain the forecasting results, we first applied numerical modeling to a 39-h forecasting dataset comprising MSM-GPVs to force the land surface model MATSIRO²⁹ and the hydrological model CaMa-Flood³⁰. Then, the estimated river water depth was analyzed via comparison with the return period. In this study, we chose locations with a return period of 200 years as forecast locations because the occurrence of flood levels in major Japanese rivers is typically set to once during a 100–200-year event (Fig. 1b,c). This method was first tested by Yoshimura et al.²⁰, who assessed six river predictions in 2003 and 2004 using the previous version of the MSM-GPV dataset, in which 18-h predictions were made every 6 h.

Dike-break time. To evaluate the forecasting performance, we obtained the dike-break times (DBTs) for all flood locations. We used dike breaks to represent all locations where flooding might occur with various inundation patterns related to the time of levee or river dike breakage. To obtain DBTs, we collected official reports, issued by the MLIT (<https://www.ktr.mlit.go.jp/kisha/index00000134.html>), of JMA disaster prevention information in XML format (<http://agora.ex.nii.ac.jp/cps/weather/river/>), as well as information from Twitter and personal websites. According to the public broadcaster Nippon Hoso Kyokai (NHK) and the MLIT, there were floods at 142 locations. Among these flood locations, only 80 records of DBTs could be found. This finding indicates that proper records were lacking for many of the floods or inundations that were identified.

The classification scheme for predicted and flooded locations of Typhoon Hagibis are shown in Fig. 2. The predicted locations are locations with a more than 1/200-year-flood alarm issued by the flood forecasting system, and the flooded locations are areas where floods actually occurred. With reference to the DBT records, true positives (TPs), correctly forecasted locations, were further classified as true positives with DBT records (TPWRs) and true positives with no DBT records (TPNRs). Among the unsuccessfully predicted flood locations, only those with DBT records (FNWRs) were verifiable. Those locations determined to be incorrectly predicted with no DBT record (FNNRs) could not be verified by location, but they were traceable via flood reports.

Figure 3 compares the forecasted 1/200-year flooding times and DBTs. We considered the following three outcomes: TPWR, FNNR, and FNWR. To compare the time differences between each predicted 1/200-year flooding time and DBT, we plotted these as color-gradient circles (Fig. 3). The circle size indicates the lead time for a given location, and the redness of the circles indicates the difference between the predicted 1/200-year flood time and the DBT. In the figure, locations with longer lead times are generally concentrated in the upper and middle reaches of rivers, where most flooding originate. We performed additional temporal analysis for the TPWR sites and found that the average lead time for the predicted 1/200-year flood time was approximately 32.75 h. Moreover, the predicted 1/200-year flood time was on average approximately 8.53 h earlier than the DBT, indicating that the predicted flood time was earlier than the real flood time. Because the goal of this system is to generate longer lead times, we argue that it is reasonable to accept this advancement, which is helpful because it allows more time for further evaluation and decision-making, such as for evacuation and disaster preparation.

Moreover, 12 locations (false negatives) were not predicted, but the occurrence of floods was recorded. Among these, five locations had DBT records (FNWR, blue crosses in Fig. 3), and seven sites had no DBT records (FNNR,

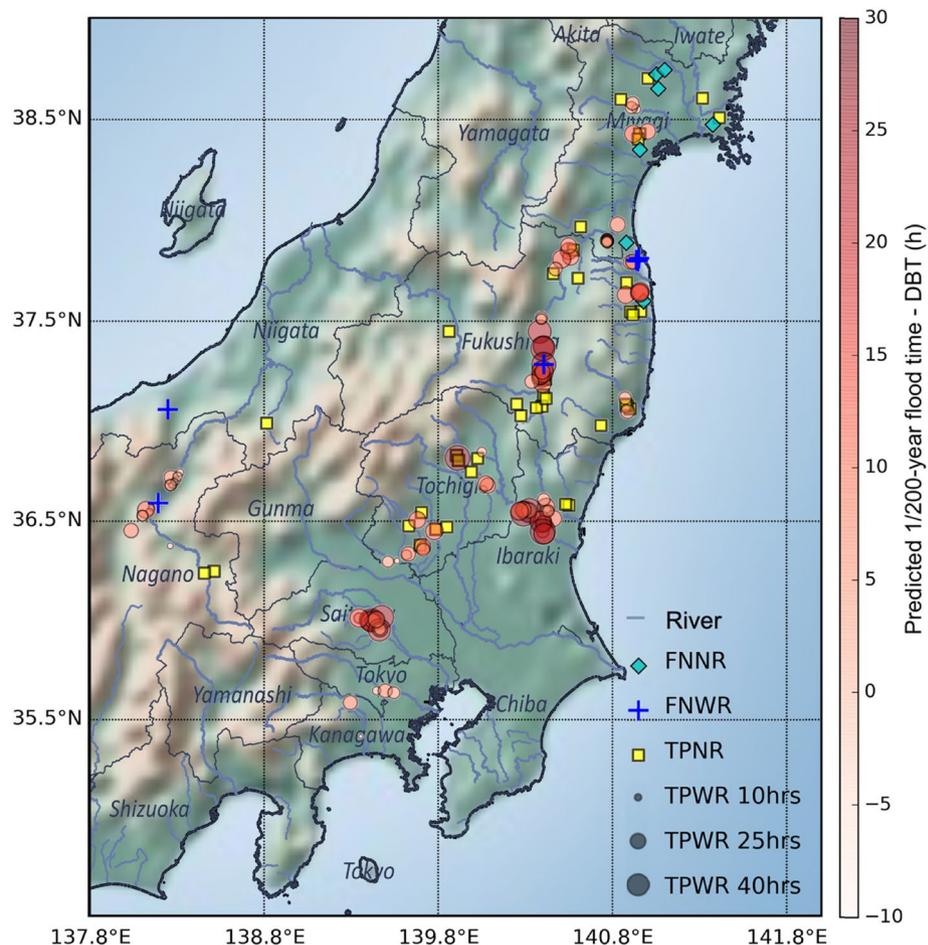


Figure 3. Forecasted locations and lead time distribution. FNNR indicates false-negative sites with no DBT records (blue diamonds, 8 spots). FNWR indicates false-negative locations with DBT records (blue crosses, 4 spots). TPNR (50 spots) and TPWR (80 spots) indicate locations that were successfully predicted (true positive) with and without DBT records, respectively. The color of the circle indicates how much the predicted 1/200-year flood time preceded the DBT at a given location. The size of the circle indicates the lead time in comparison with the predicted 1/200-year flood time. This figure was generated through Python2.7 (<https://www.python.org/>).

green diamonds in Fig. 3), which were mainly located in Miyagi Prefecture and near downstream portions. It is reasonable to assume that the five FNWR sites were technical prediction failures, which means that 2.30% of the 142 flooded locations were not successfully predicted by this system. In addition to biases or uncertainties that might have been present in the forecasted meteorological data, the spatial resolution of this system, which was designed as 0.05 degrees and is thus relatively coarse for regional forecasts, was a potential source of these mispredictions. However, taking these false-negative locations (FNWRs and FNNRs) into account, this system successfully predicted 130 flood locations, approximately 91.55% of the 142 flood locations, with a gain of approximately 32.75 h of lead time.

Merits of applying the forecasting system. This system has several merits. It can predict floods over an extensive number of locations, because some locations are not observed for various reasons, such as a lack of instruments, gaps in temporal observations, or dangers associated with collecting information. Moreover, because of flood damage, multiple problems may lead to a loss of observations. Therefore, one of the merits of a flood forecasting system is the provisioning of forecasting results without spatial or temporal limitations and the avoidance of physical risk. In Fig. 4, we show 80 locations (TPWR) from all predicted results for Typhoon Hagibis. Fifty TPNR locations did not have a DBT record (TPNR, Fig. 4). For example, in the Abukuma River, one of the hardest-hit areas³¹, 12 TPNR sites had no DBT records. As a result of this flood forecasting system, a warning time of more than 31.0 h could be achieved at these 12 TPNR locations, despite the fact that the physical detection method would not fully cover these areas. In the case of the Uda River (Fig. 4), flood alarms were issued at 3:00 a.m. on October 11, 2019. At that moment, the system predicted that 1/200-year flooding would occur at approximately 33.0 h later (at approximately 12:00 p.m. on October 12), which was validated as being approximately 9.1 h earlier than the DBT. Therefore, the actual lead time for the Uda River was approximately

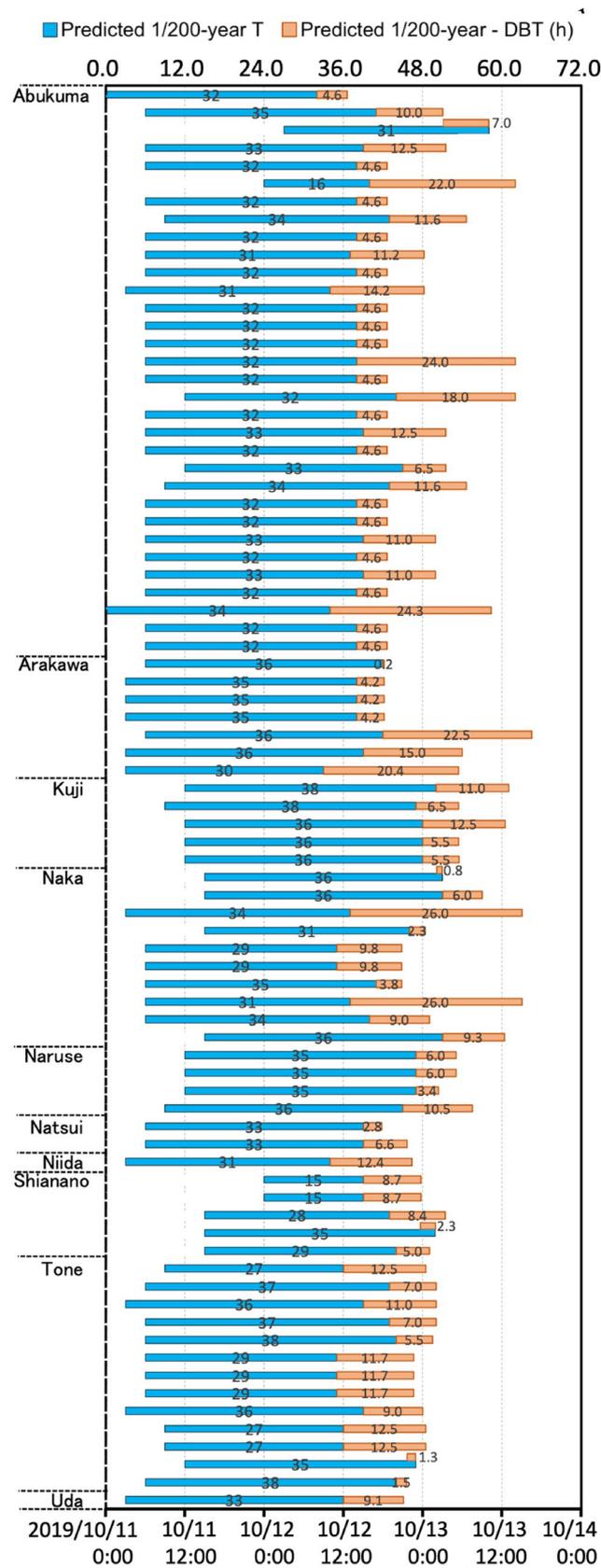


Figure 4. Comparison of predicted 1/200-year flood times and DBTs. The vertical axis shows the location of each flooded river. Each blue bar begins at the time when a 1/200-year flood was first predicted by the system. The end of each blue bar is the predicted flood time. The length of each blue bar is the lead time, which is approximately 32.75 h on average. The orange bars show differences between the DBT and the predicted 1/200-year flood time. The average of these differences indicates that the predicted 1/200-year flood time was approximately 8.53 h earlier than the DBT.

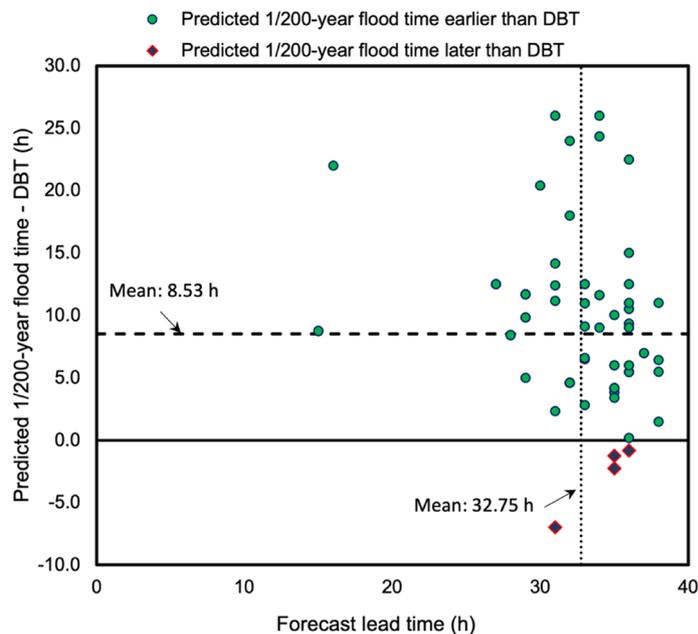


Figure 5. Differences between predicted 1/200-year flood times and DBTs, and the corresponding lead times. The vertical dashed line indicates the average lead time, which is approximately 32.75 h. The horizontal dashed line indicates the average difference between the 1/200-year flood time and the DBT, which is approximately 8.53 h.

42.1 h. However, warnings for 1/200-year floods forecasted at eight TPWR locations were issued later than the DBT by 2.3 h on average (Fig. 4). With the forecasted 33.0-h lead time for a 1/200-year flood, there was approximately 30.7 h of actual lead time for these eight TPWR locations. Figure 5 shows a comparison between 1/200-year flood times and DBTs, along with the lead times. Points representing a 1/200-year flood time later than the DBT are shown as diamonds with a red outline. On average, the flood times provided approximately 8.53 h of advanced warning relative to the DBT, with a predicted 32.75-h lead time. Overall, for 80 TPWR locations, this system provided at least 32.75 h of lead time. For some locations, such as flooded locations along the Abukuma River, Ara River, and Naka River, the actual lead time was more than 50 h. Furthermore, although we could not calculate the lead time for 50 TPWR locations due to the lack of observed DBTs, it is important to state that these floods were forecasted successfully by our system.

In this flood forecasting system, estimations are conducted every 3 h. Therefore, we inspected the forecast lead time for each estimation between October 11 and 13, 2019 (Fig. 6). Differences between the predicted 1/200-year flood times and DBTs for each estimation period were plotted for all 80 TPWR locations. The earliest time when an alarm was issued was 00:00 on October 11, 2019. Differences between the predicted 1/200-year flood time and the DBT varied from -7.2 h to approximately 32.0 h, with mean values varying from 3.7 to 11.6 h. As time passed, the mean values decreased, and the range of difference between the predicted 1/200-year flood time and the DBT decreased. This change indicates that the forecast accuracy is higher when the forecast time is closer to the time of typhoon landing.

Discussion

This study performed temporal and spatial assessments of flood forecasting, which has been challenging to validate. Observing disasters remains a great challenge because the unpredictable and devastating effects may lead to missing in situ observations. Thus, flood forecast modeling is of particular importance. It is also capable of quantitatively estimating the water volume in each layer over the land surface, which is an advantage in comparison with satellite observations. Although the number of possible observations was limited, this system successfully predicted 91.55% (130/142) of the flood locations during Typhoon Hagibis, including 50 sites that had no recorded DBTs, with a lead time of approximately 32.75 h. We argue that this forecast lead time for flood locations is much longer than the lead time of traditional forecasts issued in Japan. The high accuracy demonstrated in this study will be critical for disaster preparation and evacuation. The forecast results also demonstrated that the combination of MSM-GPV forecasted forcing data, the MATSIRO land surface model, the CaMa-Flood hydrological model, and statistical analysis is an effective solution for predicting floods in Japan. It is also reasonable to expect that this method can be applied to flood forecasting in other regions with available forcing data. Furthermore, using ensemble forecasting may help improve reliability and identify uncertainties in forecasting.

In addition to the successful alarms, we also analyzed false alarms, including false-positive (FP) locations. From the assessment of Typhoon Hagibis, we found that all 542 red pins (Fig. 1b) issued as indications of 1/200-year flooding achieved a hit ratio of 91.55% (130/142) and a precision of 24.0% (130/542). In addition, there were four FNWR locations (2.30%) where the system failed to forecast floods (Fig. 2). We plotted a relative operating

The small FP rate indicates that FPs made up only a small proportion of true-negative (TN) locations at all time steps considered by this forecasting system. The TN locations were grids with precipitation but without alarms or observed flooding. Therefore, the results imply that the system did not provide false alarms for most non-flooded grids. The FP rate for 1/200-year floods was better than that for other values, indicating that the 1/200-year threshold was suitable for the case of Typhoon Hagibis, which produced a large amount of precipitation.

To check the false alarms, we evaluated them in two aspects. The first aspect was to check false alarms of *exact location but inexact time*, which is the most direct way to assess forecasting accuracy. To do so, we assessed forecast alarms issued for exact flooding locations (Figs. 4, 5). However, it is undeniable that insufficient data can cause assessment deviations. Furthermore, flood risk tends to increase naturally along lengthy parts of a channel, so a number of alarms covering a large area are forecast by our system. Nevertheless, when a dike break occurs at one location, it significantly decreases the chance of flooding at other areas. In such a case, with the *exact location but inexact time* aspect, most of the alarms are counted as false alarms and there is a single good alarm. Therefore, we also considered a second aspect of false alarms of *inexact location but exact time*, which has great significance for flood forecasting over a large area. Forecast alarms indicate high flood risk from rivers or catchments. To assess false alarms related to *inexact location but exact time*, we overlaid forecast alarms onto the relevant catchments, which is helpful for determining the spatial distribution of information. As shown in Fig. 7b, the plotted main streams of flooded rivers (red) and non-flooded rivers (blue) during Typhoon Hagibis were densely surrounded by forecast alarms. Most of these forecast alarms indicate risk areas for flooding near these 21 major rivers. Four major rivers actually had no flooding (blue), which was considered an overestimation in terms of a warning forecast. However, flooding was still possible around these four major rivers, because of a lack of observations or actual reports. In addition to being along major rivers, some alarms appeared near the outlets of streams or downstream from catchments (Fig. 7b). This study applied a 0.05-degree resolution, which is relatively coarse for the small channels that may coexist within one modeling grid. By comparing our spatial distribution with the flood locations shown in Fig. 3, it is obvious that some locations, such as those of eastern Fukushima, southern Miyagi, southern Chiba, and southwestern Shizuoka, shown as black dots in Fig. 7b, were actually not flooded. These alarms are reasonable selections for considering FP alarms of *exact location but inexact time* and *inexact location but exact time*. These FP locations may be attributable to one of the following three causes: the meteorological forcing data, hydrological model, or statistical analysis. In terms of meteorological forcing, higher precision³² and spatial resolution for each forecasted data point are required.

One possible way to improve flood forecasting accuracy is to adopt an ensemble prediction approach¹⁶, which should be analyzed in future work. Moreover, there is still opportunity for improvement in the resolution of the hydrological models used in forecasting systems. In this study, the results provided by CaMa-Flood have resolution of 0.05 degrees, which is the same as the resolution of EFAS and finer than that of other forecasting systems⁶. However, this resolution is relatively coarse for the many small channels that may coexist within one modeling grid. A finer resolution would produce a more reliable representation of hydrological states; however, it is subject to the resolution of the meteorological forcing data. In addition to the uncertainty due to forcing data and model resolution, some of the false alarms in this forecast might have been generated by underestimated 200-year return periods. Studies on the return period have demonstrated that analyzing homogeneous return periods may result in bias³³. Although we used a 200-year return period as a threshold, the actual designed flood level exhibited particular variability. A survey and arrangement of designed flood level data are required in a future study.

Finally, validation is a great obstacle in flood forecasting because a shortage of observation data is a common difficulty. In particular, disaster monitoring systems are greatly challenged when disasters occur. Observation shortages may lead to underestimation of the accuracy of modeling and cause deviations in the validity of forecasted results. How to extend in situ observations remains a problem that can only be answered by taking cost into account. There is still potential for using satellite observations to enhance the quantitative information analyzed in hydrological studies.

In this paper, we present a flood forecasting system that is more useful in forecasting extreme flood events, such as the events of Typhoon Hagibis, compared to conventional forecasting based on gauged water levels. Despite system deficiencies, including limited modeling spatial resolution and forecasting precision, such long lead-time flood forecasting is urgently needed for early warning in Japan. At present, the JMA is issuing flood forecast alarms no earlier than 3 h ahead of time, which may result in difficulties in evacuation at night or for people who find it inconvenient to evacuate in such a short amount of time. A flood forecasting system with more than 30 h of lead time is helpful in many ways. Particularly, given the increasing tendency of extreme precipitation events worldwide^{34–38}, an accurate flood forecasting technique is urgently needed in Japan and the rest of the world.

Methods

System description. This flood forecasting system was developed by Ishitsuka^{18,19}. Its performance is shown in Supplementary Fig. 1. The modeling framework includes a land surface model, the MATSIRO²⁹, and a global river routing model, CaMa-Flood³⁰. The river water depth from CaMa-Flood was compared based on its statistical distribution across various return period values. The forecasting system makes hydrographic predictions for all rivers in Japan, which are integrated into a model mesh with 0.05-degree resolution.

MATSIRO is a physically based land surface model analyzing an environment consisting of a single-layer canopy, three layers of snow (at maximum), and six layers of soil. It simulates vertical movement of water and energy at the global scale. In Japan, it covers the area from 24° to 46° N latitude and 123° to 148° E longitude with a horizontal resolution of approximately 5 km (0.05 degrees)²⁹. The input atmospheric forcing data include precipitation, temperature, surface pressure, wind speed, and radiation²⁰. The output runoff from MATSIRO is used to run CaMa-Flood, the river routing model. CaMa-Flood was originally developed as a global hydrodynamic

model that solves the local inertial Eq.³⁹. It calculates the river discharge of a one-dimensional river channel with a rectangular riverbed and trapezoid floodplain storage. The river network, routing direction, and river parameters were calculated from the Multi-Error-Removed Improved-Terrain (MERIT) DEM and hydrography (MERIT Hydro)^{21,22} datasets with approximately 5-km (0.05-degree) horizontal resolution in Japan³⁰. CaMa-Flood calculates river water depth (D_r) from the total water stored (S_r) at each grid point, as shown in Eq. 1:

$$D_r = \frac{S_r}{WL} \quad (1)$$

where W is the channel width, and L is the channel length. Each grid point has a river channel reservoir and a floodplain reservoir, which make up the unit catchment of the river channel.

Forcing data preparation. In this study, MSM-GPV data^{40,41} provided by the JMA were used as meteorological forcing data. The MSM-GPV dataset includes 39-h forecast data around Japan, with a horizontal grid of 5 km and 50 vertical layers, which are released every 3 h (00, 03, ..., 21 UTC). MSM-GPV data have been widely applied in meteorological and hydrological research in Japan on precipitation^{42–44} and typhoons⁴⁵, wind^{42,46}, energy^{44–50}, and others^{51–54}. In this study, we applied humidity, cloud cover, precipitation, surface air pressure, downward shortwave radiation, downward longwave radiation, wind speed/direction, and air temperature from the MSM-GPV dataset as meteorological forcing data. To minimize bias caused by precipitation, initial boundary data were estimated using radar data (Fig. 1) provided by the JMA at the same resolution. This method has been tested by Yoshimura et al.²⁰.

Data collection for alarm locations and DBTs. To assess the validity of the alarms forecasted by the model, we simply compared the alarm times and locations with the corresponding DBTs and locations of flood areas, which were obtained from the MLIT (<http://xml.kishou.go.jp/xmlpull.html>), JMA disaster prevention information in XML format (<http://agora.ex.nii.ac.jp/cps/weather/river/>), and NHK (<https://www.nhk.or.jp/>). The DBTs from MLIT and JMA were our primary data sources, as they provided the most reliable and rigorous information, including spatiotemporal details. The secondary source was the media (NHK), which had quickly broadcasted news of severe inundation, including the general location and timing. To prepare a systematic list of flood information, flood locations were mostly obtained from data from MLIT but supplemented by data from JMA and NHK.

Statistical analysis. Significant systematic biases inevitably exist between naïve simulations and reality²⁶ because of errors in the forcing data and inherent uncertainty in the models⁵⁵. Instead of improving only the accuracy of forecasting, an accessible and practical method to mitigate the problem is to combine modeling results and statistical analysis because the results of hydrological modeling inevitably contain uncertainties^{56–58}. Simulated results from a land surface model can reproduce hydrological processes to some extent. However, the output can be used more effectively when combined with statistical analysis. Many statistical distributions, such as the generalized extreme value distribution^{59,60}, Gumbel distribution^{23–27}, log-normal distribution^{61,62}, and log Pearson type-III distribution⁶³, have been tested in flood studies. Based on the characteristics of historical flood distributions, the Gumbel distribution²⁸ is widely accepted as representing extreme flood events well because of its better fit in extreme value analysis²⁶. Application of the Gumbel distribution was consistent with the estimation carried out by Yoshimura et al.²⁶. First, the following equation was applied to estimate the probability distribution of the annual maximum discharge for each grid:

$$F_{(D)} = \exp[-\lambda(1 - G_{(D)})] = \exp\left(-\exp\left(-\frac{D - \mu}{\beta}\right)\right) \quad (2)$$

where F is the cumulative distribution function (CDF) of annual maximum values, D is the discharge, G is the CDF of the values that exceed a specific threshold value, λ is a constant representing annual occurrence frequency, and β and μ are the scale and location parameters of the Gumbel distribution, respectively.

Second, the scale and location parameters of the Gumbel distribution were estimated as follows:

$$\hat{\beta} = \frac{1}{M} \sum_{i=1}^M (D_i - D_M), \hat{\mu} = D_M + \hat{\beta} \ln \lambda, \lambda = M/N \quad (3)$$

where D_i indicates the i th maximum and M and N are the numbers of samples and years, respectively, which give λ , a constant representing annual occurrence frequency. DPI (Π) for all the daily maximum discharges was calculated as follows:

$$\Pi = (1 - F_{(D)})^{-1} = \left(1 - \exp\left(-\exp\left(-\frac{D - \mu}{\beta}\right)\right)\right)^{-1} \quad (4)$$

The unit for DPI is years, meaning that the probability of exceeding discharge D in a year is $1/\Pi$ and the expected occurrence is once in Π years if the discharge occurs at an annual maximum value.

In this study, 10 years of flood events were collected, and the Gumbel distribution was analyzed for each grid. The river water depth estimated by CaMa-Flood was compared with the 1/200-year flood water depth. For river water depths exceeding the 1/200-year threshold, alarms appeared in the forecasting system interface (Fig. 1b).

ROC curve. The ROC curve is an effective way of assessing forecast ability in terms of hit rate and false warning rate. The ROC curve plots the proportion of occurrences that have been forecasted successfully (TP rate, y-axis) versus the proportion of false alarms (FP rate, x-axis) with reference to different thresholds^{64–67}. In this study, the hit ratio (TP rate) indicates the probability of successfully forecasting grids (TP) among both TP and FN grids (Eq. 5). The false alarm rate refers to FP detection among both FP and TN grids (Eq. 6). The precision is indicated by the positive predictive value (PPV), defined in Eq. (7).

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

$$PPV = \frac{TP}{TP + FP} \quad (7)$$

Both the TP and FN rates were estimated from observation data. The values of the total grids including all negative and positive locations were selected from the grids with precipitation. According to the ROC plot, if the curve approaches the top-left corner, the forecasting system has a greater ability to forecast floods. Conversely, if the ROC curve lies close to the diagonal, the forecasting ability is considered weak⁶⁸.

Data availability

The datasets generated during and/or analyzed during the current study are available in the “Zenodo” repository, (<https://doi.org/10.5281/zenodo.4604483>).

Received: 8 January 2021; Accepted: 26 April 2021

Published online: 13 May 2021

References

- Hirabayashi, Y. *et al.* Global flood risk under climate change. *Nat. Clim. Chang.* **3**, 816–821 (2013).
- Chang, L. *et al.* flood forecasts up to two days in advance. *Nat. Commun.* <https://doi.org/10.1038/s41467-020-15734-7> (2020).
- Paprotny, D., Sebastian, A., Morales-Nápoles, O. & Jonkman, S. N. Trends in flood losses in Europe over the past 150 years. *Nat. Commun.* **9**, (2018).
- Takemi, T. & Unuma, T. Environmental factors for the development of heavy rainfall in the eastern part of Japan during Typhoon Hagibis (2019). *Sci. Online Lett. Atmos.* **16**, 30–36 (2020).
- Sayama, T., Yamada, M., Sugawara, Y. & Yamazaki, D. 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. (2020) <https://doi.org/10.21203/rs.3.rs-40714/v1>.
- Emerton, R. E. *et al.* Continental and global scale flood forecasting systems. *Wiley Interdiscip. Rev. Water* **3**, 391–418 (2016).
- Adams, Thomas E.; Pagano, T. C. *Flood forecasting. A Global Perspective. Academic Press is an imprint of Elsevier* vol. 16 (2016).
- van der Knijff, J. M., Younis, J. & de Roo, A. P. J. LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *Int. J. Geogr. Inf. Sci.* **24**, 189–212 (2010).
- Alfieri, L. *et al.* A global network for operational flood risk reduction. *Environ. Sci. Policy* **84**, 149–158 (2018).
- Wu, H. *et al.* Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.* **50**, 2693–2717 (2014).
- Yilmaz, K. K., Adler, R. F., Tian, Y., Hong, Y. & Pierce, H. F. Evaluation of a satellite-based global flood monitoring system. *Int. J. Remote Sens.* **31**, 3763–3782 (2010).
- Bartholmes, J. C., Thielen, J., Ramos, M. H. & Gentilini, S. The european flood alert system EFAS “Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol. Earth Syst. Sci.* **13**, 141–153 (2009).
- Demargne, J. *et al.* The science of NOAA’s operational hydrologic ensemble forecast service. *Bull. Am. Meteorol. Soc.* **95**, 79–98 (2014).
- Donnelly, C., Andersson, J. C. M. & Arheimer, B. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrol. Sci. J.* **61**, 255–273 (2016).
- Alfieri, L. *et al.* GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* **17**, 1161–1175 (2013).
- World Meteorological Organization. *Manual On Flood Forecasting and Warning P-ClW_102107*. (2011).
- Biondi, D., Freni, G., Iacobellis, V., Mascaro, G. & Montanari, A. Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Phys. Chem. Earth* **42–44**, 70–76 (2012).
- Entire area along major northeastern Japan river flooded as typhoon path matched flow.* (2018, 10 19). Retrieved from The Mainichi: <https://mainichi.jp/english/articles/20191019/p2a/00m/0na/003000c>.
- Ishitsuka, Y. *Building an ensemble flood prediction system in Japan using numerical weather prediction datasets* (University of Tokyo, 2016).
- Yoshimura, K. *et al.* Development and verification of a predicting system of river discharge of Japan using JMA-MSM-GPV. *Proc. Hydraul. Eng.* **51**, 403–408 (2007).
- Yamazaki, D. *et al.* A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* **44**, 5844–5853 (2017).
- Yamazaki, D. *et al.* MERIT hydro: a high-resolution global hydrography map based on latest topography dataset. *Water Resour. Res.* **55**, 5053–5073 (2019).
- Haktanir, T. Comparison of various flood frequency distributions using annual flood peaks data of rivers in Anatolia. *J. Hydrol.* **136**, 1–31 (1992).
- Onen, F. & Bagatur, T. Prediction of flood frequency factor for gumbel distribution using regression and GEP model. *Arab. J. Sci. Eng.* **42**, 3895–3906 (2017).
- Rasmussen, P. F. & Gautam, N. Alternative PWM-estimators of the gumbel distribution. *J. Hydrol.* **280**, 265–271 (2003).
- Yoshimura, K., Sakimura, T., Oki, T., Kanae, S. & Seto, S. Toward flood risk prediction: a statistical approach using a 29-year river discharge simulation over Japan. *Hydrol. Res. Lett.* **2**, 22–26 (2008).
- Katz, R. W., Parlange, M. B. & Naveau, P. Statistics of extremes in hydrology. *Adv. Water Resour.* **25**, 1287–1304 (2002).

28. Gumbel, E. The Return Period of Flood Flows Author (s): E. J. Gumbel Source: The Annals of Mathematical Statistics, Vol. 12, No. 2 (Jun., 1941), pp. 163–190 Published by: Institute of Mathematical Statistics Stable. <http://www.jstor.org/stable/223.223>. *Statistics (Ber)*. **12**, 163–190 (1941).
29. Takata, K., Emori, S. & Watanabe, T. Development of the minimal advanced treatments of surface interaction and runoff. *Glob. Planet. Change* **38**, 209–222 (2003).
30. Yamazaki, D., Kanae, S., Kim, H. & Oki, T. A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.* **47**, 1–21 (2011).
31. Ishitsuka, Y. *Toward a seamless application of global flood forecasting: a development and validation of global and regional prediction systems* (University of Tokyo, 2018).
32. Bhomia, S., Jaiswal, N. & Kishtawal, C. M. Accuracy assessment of rainfall prediction by global models during the landfall of tropical cyclones in the North Indian Ocean. *Meteorol. Appl.* **24**, 503–511 (2017).
33. Metin, A. D. *et al.* The role of spatial dependence for large-scale flood risk estimation. *Nat. Hazards Earth Syst. Sci. Discuss.* <https://doi.org/10.5194/nhess-2019-393> (2019).
34. Mills, E. Insurance in a climate of change. *Science* **309**, 1040–1044 (2005).
35. Arnell, N. W. & Lloyd-Hughes, B. The global-scale impacts of climate change on water resources and flooding under new climate and socio-economic scenarios. *Clim. Change* **122**, 127–140 (2014).
36. Winsemius, H. C. *et al.* Global drivers of future river flood risk. *Nat. Clim. Chang.* **6**, 381–385 (2016).
37. Endo, H., Kitoh, A., Mizuta, R. & Ishii, M. Future changes in precipitation extremes in East Asia and their uncertainty based on large ensemble simulations with a high-resolution AGCM. *Sci. Online Lett. Atmos.* **13**, 7–12 (2017).
38. Tanoue, M., Hirabayashi, Y. & Ikeuchi, H. Global-scale river flood vulnerability in the last 50 years. *Sci. Rep.* **6**, 1–9 (2016).
39. Bates, P. D., Horritt, M. S. & Fewtrell, T. J. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *J. Hydrol.* **387**, 33–45 (2010).
40. Technical, W. M. O., Report, P., Data-processing, G. & Prediction, N. W. Outline of the operational numerical weather prediction at japan meteorological agency. (2019).
41. 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. The Operational JMA Nonhydrostatic Mesoscale Model. *Mon. Weather Rev.* 1266–1298 (2006).
42. Yoshikane, T., Yoshimura, K., Chang, E. C., Saya, A. & Oki, T. Long-distance transport of radioactive plume by nocturnal local winds. *Sci. Rep.* **6**, 1–7 (2016).
43. Akatsuka, S., Susaki, J. & Takagi, M. Estimation of precipitable water using numerical prediction data. *Eng J* **257**, 268. <https://doi.org/10.4186/ej.2018.22.3.257> (2018).
44. Shimadera, H., Kondo, A., Shrestha, K. L., Kitaoka, K. & Inoue, Y. Numerical Evaluation of the Impact of Urbanization on Summertime Precipitation in Osaka, Japan. *Adv. Meteorol.* **2015**, (2015).
45. Tada, H., Uchiyama, Y. & Masunaga, E. Deep-Sea Research Part I Impacts of two super typhoons on the Kuroshio and marginal seas on the Pacific coast of Japan. *Deep. Res. Part I*(132), 80–93 (2018).
46. Kitajima, T. & Member, S. Study on output prediction system of wind power generation using complex-valued neural network with multipoint GPV data. *IEEEJ. Trans. Electr. Electron. Eng.* **33**, 39. <https://doi.org/10.1002/tee.21788> (2013).
47. Yamanaka, Y. *et al.* Nearshore dynamics of storm surges and waves induced by the 2018 Typhoons Jebi and Trami based on the analysis of video footage recorded on the Coasts of Wakayama, Japan. *J. Mar. Sci. Eng.* **7**, (2019).
48. Suzuki, T., Goto, Y., Terazono, T., Wakao, S. & Oozeki, T. Forecasting of solar irradiance with just-in-time modeling. *Electr. Eng. Jpn.* **182**, 912–919 (2013).
49. Goto, Y., Suzuki, T., Shimoo, T., Hayashi, T. & Wakao, S. Operation design of PV system with storage battery by using next-day residential load forecast. *Conf. Rec. IEEE Photovolt. Spec. Conf.* <https://doi.org/10.1109/PVSC.2011.6186427> (2011).
50. Ohtake, H. *et al.* Accuracy of the solar irradiance forecasts of the Japan Meteorological Agency mesoscale model for the Kanto region Japan. *Sol. Energy* **98**, 138–152 (2013).
51. Ishida, H. *et al.* Scheme for detection of low clouds from geostationary weather satellite imagery. *Atmos. Res.* **143**, 250–264 (2014).
52. Tsurushima, D., Sakaida, K. & Honma, N. Spatial distribution of cold-season lightning frequency in the coastal areas of the Sea of Japan. *Prog. Earth Planetary Sci.* <https://doi.org/10.1186/s40645-017-0122-0> (2017).
53. Shimadera, H. *et al.* Contribution of transboundary air pollution to ionic concentrations in fog in the Kinki Region of Japan. *Atmos. Environ.* **43**, 5894–5907 (2009).
54. Katata, G., Ota, M., Terada, H., Chino, M. & Nagai, H. Atmospheric discharge and dispersion of radionuclides during the Fukushima Dai-ichi Nuclear Power Plant accident. Part I: Source term estimation and local-scale atmospheric dispersion in early phase of the accident. *J. Environ. Radioact.* **109**, 103–113 (2012).
55. Oki, T., Nishimura, T. & Dirmeyer, P. Assessment of Annual Runoff from Land Surface Models Using Total Runoff Integrating Pathways (TRIP). *J. Meteorol. Soc. Japan. Ser. II*(77), 235–255 (1999).
56. Butts, M. B., Payne, J. T., Kristensen, M. & Madsen, H. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *J. Hydrol.* **298**, 242–266 (2004).
57. Haddeland, I. *et al.* Multimodel estimate of the global terrestrial water balance: setup and first results. *J. Hydrometeorol.* **12**, 869–884 (2011).
58. Lohmann, D. *et al.* Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *J. Geophys. Res. D Atmosph.* vol. 109 (2004).
59. Wang, Q. J. Using higher probability weighted moments for flood frequency analysis. *J. Hydrol.* **194**, 95–106 (1997).
60. Martins, S. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.* **36**, 737–744 (2000).
61. Singh, V. P. Three-Parameter Lognormal Distribution BT - Entropy-Based Parameter Estimation in Hydrology. in (ed. Singh, V. P.) 82–107 (Springer Netherlands, 1998). https://doi.org/10.1007/978-94-017-1431-0_7.
62. Chaibandit, K. & Konyai, S. Using Statistics in Hydrology for Analyzing the Discharge of Yom River. *APCBEE Proc.* **1**, 356–362 (2012).
63. Griffis, V. W. & Stedinger, J. R. Log-pearson type 3 distribution and its application in flood frequency analysis. I: Distribution characteristics. *J. Hydrol. Eng.* **12**, 482–491 (2007).
64. Atger, F. Estimation of the reliability of ensemble-based probabilistic forecasts. *Q. J. R. Meteorol. Soc.* **130**, 627–646 (2004).
65. Golding, B. W. Quantitative precipitation forecasting in the UK. *J. Hydrol.* **239**, 286–305 (2000).
66. DeLeo, J. M. Receiver operating characteristic laboratory (ROCLAB): Software for developing decision strategies that account for uncertainty. *Proc. - 2nd Int. Symp. Uncertain. Model. Anal. ISUMA 1993* 318–325 (1993) <https://doi.org/10.1109/ISUMA.1993.366750>.
67. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38–49 (1997).
68. Jha, S. K., Shrestha, D. L., Stadnyk, T. A. & Coulibaly, P. Evaluation of ensemble precipitation forecasts generated through post-processing in a Canadian catchment. *Hydrol. Earth Syst. Sci.* **22**, 1957–1969 (2018).

Acknowledgements

This study was supported by the water environment and resource research project at the Earth Observation Research Center, Japan Aerospace Exploration Agency (JAXA EORC); Cross-ministerial Strategic Innovation Promotion Program (SIP); the Integrated Research Program for Advancing Climate Models (TOUGOU program) from the MEXT, Japan; the development of an application towards water-related problems in the Data Integration & Analysis System (DIAS) supported by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan. The JMA data used in this study was provided by way of “Meteorological Research Consortium”, a framework for research cooperation of JMA and MSJ. We thank Dr. Tomoko Nitta, Dr. Yuki-hiko Onuma, Dr. Takao Yoshikane, Ms. Xiaoxing Wang and Ms. Risa Hanazaki for their help to prepare this manuscript.

Author contributions

K.Y. supervised the experiments, K.Y. and W.M. conceived of the presented idea. Y.I. and K.Y. developed and performed the computations. W.M. prepared the data collection and analysing, A.T. assist data preparation, K.H. provided a high resolution of flood area data, D.Y. assist CaMa-Flood modelling, K.Y., M.K., and R.O. assist the system preparation, W.M., took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research analysis, and manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-89522-8>.

Correspondence and requests for materials should be addressed to W.M. or K.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021