TITLE:

Simultaneous flood risk analysis and its future change among all the 109 class-A river basins in Japan using a large ensemble climate simulation database d4PDF

AUTHOR(S):

Tanaka, Tomohiro; Kobayashi, Keita; Tachikawa, Yasuto

CITATION:

Tanaka, Tomohiro ...[et al]. Simultaneous flood risk analysis and its future change among all the 109 class-A river basins in Japan using a large ensemble climate simulation database d4PDF. Environmental Research Letters 2021, 16(7): 074059.

ISSUE DATE:
2021-07

URL:
http://hdl.handle.net/2433/276619

RIGHT:
© 2021 The Author(s). Published by IOP Publishing Ltd; Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 license. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

京都大学学術情報リポジトリ
KURENAI
Kyoto University Research Information Repository

ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Simultaneous flood risk analysis and its future change among all the 109 class-A river basins in Japan using a large ensemble climate simulation database d4PDF

To cite this article: Tomohiro Tanaka *et al* 2021 *Environ. Res. Lett.* **16** 074059

View the article online for updates and enhancements.

You may also like

- Impact of Stellar Superflares on Planetary Habitability
Yosuke A. Yamashiki, Hiroyuki Maehara, Vladimir Airapetian et al.

- Economic growth dominates rising potential flood risk in the Yangtze River and benefits of raising dikes from 1991 to 2015
Wei Qi, Lian Feng, Hong Yang et al.

- The effect of surge on riverine flood hazard and impact in deltas globally
Dirk Eilander, Anaïs Couasnon, Hiroaki Ikeuchi et al.

This content was downloaded from IP address 130.54.110.24 on 04/10/2022 at 02:21

京都大学
KYOTO UNIVERSITY

A Self-archived copy in
Kyoto University Research Information Repository

京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

IOP Publishing Environ. Res. Lett. **16** (2021) 074059 https://doi.org/10.1088/1748-9326/abfb2b

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
30 September 2020

REVISED
18 February 2021

ACCEPTED FOR PUBLICATION
23 April 2021

PUBLISHED
16 July 2021

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

LETTER

Simultaneous flood risk analysis and its future change among all the 109 class-A river basins in Japan using a large ensemble climate simulation database d4PDF

Tomohiro Tanaka[1,*], **Keita Kobayashi**[2] and **Yasuto Tachikawa**[2]

[1] Graduate School of Global Environmental Studies, Kyoto University, C1-4, KyotoDaigaku-Katsura, Nishikyoku, Kyoto, Kyoto 615-8540, Japan
[2] Graduate School of Engineering, Kyoto University, C1-1, KyotoDaigaku-Katsura, Nishikyoku, Kyoto, Kyoto 615-8540, Japan
[*] Author to whom any correspondence should be addressed.

E-mail: tanaka.tomohiro.7c@kyoto-u.ac.jp

Keywords: simultaneous flood occurrence, ensemble climate simulation, d4PDF, association rule analysis, graph theory

Supplementary material for this article is available online

Abstract

This study investigated simultaneous flood risk among all the 109 class-A river basins over Japan using the big data of (over 1000 years) annual maximum hourly flow simulated from a large ensemble climate simulation database for policy decision making for future climate change, and proposed a novel approach in its geospatial analysis by applying two informatics techniques: the association rule analysis and graph theory. Frequency analysis of the number of rivers with the annual maximum flow over the flow capacity in the same year (defined as simultaneous flooding here) indicated that simultaneous flood risk will increase in the future climate under 4-degree rise scenarios in Japan, whose increment is larger than the variation of sea surface temperature projections. As the result, the return period of simultaneous flooding in eight river basins (the same number as in a severe storm in western Japan, 2018, causing the second worst economic damage since 1962) is estimated at 400 years in the historical experiment, 25 years in the 4-degree rise experiment. The association rule and graph theory analyses for the big data of annual maximum flows in the future climate scenarios indicated that simultaneous flood occurrence is dominated by spatial distance at a national scale as well as by the spatial relation between mountainous ridges and typhoon courses at a regional scale. Large ensemble climate simulation data combined with the informatics technology is a powerful approach to simultaneous flood risk analysis.

1. Introduction

Flood is one of the most hazardous natural disasters around the world and its economic impact keeps increasing in the last decades (Swiss Re Institute 2018). In particular, Japan has intensively experienced severe floods almost every year over this decade: the worst and second worst flood damage were caused by Typhoon Hagibis in 2019 (1.86 trillion Japanese yen) and a heavy rainfall event over western Japan in 2018 (1.35 trillion Japanese yen), respectively (MLIT 2019a, 2019b). These extreme flood damages were caused by river overflow and dike breach in multiple river basins (six and eight river basins among 109 class-A rivers, respectively). Such simultaneous flood occurrence results in devastating disaster through the supply chain and thus massive insurance payment as well as disaster recovery expenditure. In fact, the accumulation risk (Bull-Kamanga *et al* 2003) is of great attention in the insurance industry. Climate change impact on flood disaster risk has been intensively investigated in the literature at national to regional scales (e.g. Hirabayashi *et al* 2013, Arnell and Gosling 2016); nevertheless, the impact on the coincidence of flood in multiple river basins has rarely been discussed in the literature. Rather, more attention is paid to compound risk of river flooding with other types of disasters such as storm surge (Wahl *et al* 2015, Saleh *et al* 2017) and/or wildfires/heatwaves (Zscheischler *et al* 2018, Moftakhari and AghaKouchak 2019).

Spatial dependence of flood risk has been evaluated in the context of hydrological frequency analysis

© 2021 The Author(s). Published by IOP Publishing Ltd

(Keef *et al* 2009, Chen *et al* 2012, Neel *et al* 2013, Geertsema *et al* 2018). Keef *et al* (2009) analyzed spatial dependence of flood peaks and precipitation totals over the whole UK. Neal *et al* (2013) estimated the joint probability of flood peaks among the River Eden and its two tributaries. They successfully estimated joint probabilities of multi-site flood peaks and the resulting inundation depths at various return periods with the help of copula functions. Geertsema *et al* (2018) presented another approach, dynamic time warping as an efficient tool in time series analysis that detects analogy in two different flood waves, and clarified that rapid drainage in a lowland tributary reduces the likelihood of coincident discharge peak with the main stream. These statistical techniques are essential to discuss simultaneous flooding because the observation period is far shorter than the return periods of such events. However, it has been a challenge to apply these statistical approaches to inter-catchment and/or national-scale simultaneous flood analysis due to high dimension of their joint probability distribution.

Recently, some studies expanded flood spatial dependence to continental scales by accumulating all the available observation data over the countries. Quinn *et al* (2019) generated many synthetic flood events by constructing spatial flood dependence structures with conditional multivariate extreme value analysis. They were then converted into a flood risk curve through depth-damage functions. Diederen *et al* (2019) constructed a statistical model of spatially coherent discharge peaks from block maxima data and produced pan-European flood events. Using these state-of-the-art statistical models for increased number of gauges is a major way to do a large-scale flood risk assessment (Brunner *et al* 2020, Winter *et al* 2020). There will be, however, still challenges for extending this analysis to assess climate change impact on large-scale flooding, which would be addressed in the near future.

As an alternative approach, this study explores the use of large ensemble climate simulations with global climate models. In the last few years, to detect the climate change signals in extreme weather events such as heavy rainfall and/or storm surge, a large ensemble climate simulation database has been developed by the meteorological research community such as Half A degree additional warming, Prognosis and Projected Impacts (Mitchell *et al* 2017), database for(4) policy decision making for future climate change (d4PDF) (Mizuta *et al* 2017), and the ClimEx project (Leduc *et al* 2019). They realized a tremendous number of climate simulations and as a result, produced thousands of climate simulation data for the present and future climate conditions. They are considered long enough to estimate low-frequency phenomena including river overflow at the return period of several hundred years. Recent studies have widely applied these data to impact assessments on river overflow, storm surge, and/or wind disasters (Doll *et al* 2018, Faye *et al* 2018, Lavender *et al* 2018, Yang *et al* 2018, Mori *et al* 2019). These large ensemble climate simulation datasets also have potential of giving coincident flood frequencies over a wide range in a more straightforward manner.

This research aims to estimate future changes of simultaneous flood occurrence probability and its geospatial characteristics among all the 109 class-A river basins in Japan (see figure 1) using the annual maxima of hourly river discharge simulated with a rainfall-runoff model from d4PDF. The climate change signal of simultaneous flood risk is detected by comparing the probability distribution of the number of flooded river basins per year between the past and 4-degree rise experiments of d4PDF.

Furthermore, we also analyze spatial dependence in flood occurrence among all the basins. This can be challenging because d4PDF provides 3000-year and 5400-year annual maximum flow data in the 109 river basins, which makes a tremendous number of combinations to examine. We overcome this big data issue by employing association rule analysis (ARA) and a graph theory that efficiently identify a group of river basins which are likely to be flooded in the same year. Finally, two examples of regional community extraction are demonstrated in the Kyushu and Hokkaido regions to examine the geospatial characteristics of simultaneous flood occurrence at a regional scale. Note that the community extraction approach is presented only for the 4-degree rise experiment because enough simultaneous flood events did not occur in the past experiment.

2. Model and data

2.1. Rainfall-runoff model 1K-DHM

We employ a rainfall-runoff model 1K-DHM (Tanaka and Tachikawa 2015), a distributed rainfall runoff model based on a kinematic wave flow approximation, constructed at all the 109 class-A rivers basins in Japan in Kobayashi *et al* (2020). 1K-DHM calculates the slope runoff and river routing at each cell at around 1 km (30 s) resolution over a two-dimensional domain. Slope runoff is simulated with the continuity and momentum equations of the kinematic wave theory, considering saturated and unsaturated subsurface flows and surface flow with the following momentum equation and discharge–storage relation (Tanaka and Tachikawa 2015)

$$\frac{\partial h}{\partial t}+\frac{\partial q}{\partial x}=r_{\mathrm{e}} \tag{1}$$

$$q=\begin{cases} v_{\mathrm{c}}d_{\mathrm{c}}\left(\frac{h}{d_{\mathrm{c}}}\right)^{\beta}\,(0\leqslant h\leqslant d_{\mathrm{c}}) \\ v_{\mathrm{c}}d_{\mathrm{c}}+v_{\mathrm{a}}\left(h-d_{\mathrm{c}}\right)\left(d_{\mathrm{c}}\leqslant h\leqslant d_{\mathrm{a}}\right) \\ \alpha(h-d_{\mathrm{c}})^{\mathrm{m}}+v_{\mathrm{a}}\left(h-d_{\mathrm{c}}\right)+v_{\mathrm{c}}d_{\mathrm{c}}\left(d_{\mathrm{a}}\leqslant h\right) \end{cases} \tag{2}$$

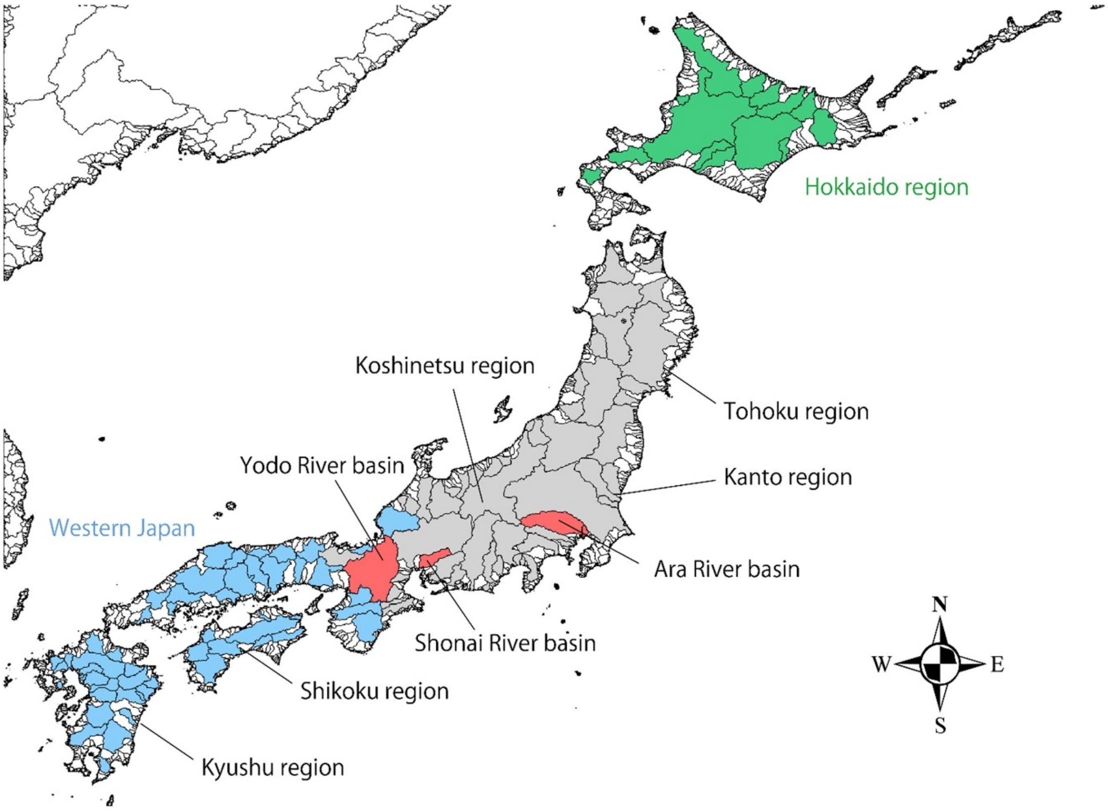


Figure 1. River basin map in Japan. The filled areas show the class-A river basins. The red river basins show the Ara, Shonai, and Yodo River basins having a mega city (Tokyo, Nagoya, and Osaka, respectively). The blue areas show the western Japan river basins, the most southern island of which is the Kyushu region; the green areas show the Hokkaido River basins. The exceedance probability of number of simultaneously inundated basins is investigated all Japan and the western Japan; community extraction is demonstrated all Japan, in the Kyushu and Hokkaido regions.

where, q is the slope runoff discharge per unit width; h is the water depth; d_a and d_c are the water depth corresponding to the maximum water content of saturated and unsaturated soil layers, respectively; v_a and v_c are the water velocity of saturated and unsaturated soil layers, respectively; $\alpha = n/\sqrt{\sin\theta}$; n is the roughness coefficient; and θ is the slope gradient. River flow is simulated with a kinematic wave model with the following momentum equation and discharge–storage relation

$$q_{\mathrm{L}} = \frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} \tag{3}$$

$$Q = \alpha_{\mathrm{c}} A^{\mathrm{m}} \tag{4}$$

where, A is cross-sectional area of a river, Q is discharge. q_{L} is lateral inflow per unit width and given by calculated slope runoff discharge per unit. α_{c} is parameter and defined $\alpha_{\mathrm{c}} = (1/B_{\mathrm{c}})^{2/3}\sqrt{i_{\mathrm{c}}}/n_{\mathrm{c}}$ where i_{c} river gradient, n_{c} is roughness coefficient of river, B_{c} is river width. Figure 1 shows the concept of 1k-DHM. A kinematic wave flow approximation with thin subsurface flow components well represents flash flood characteristics over steep mountainous terrain in a temperate climate, which are widely seen in Japan river catchments (Shakti 2017).

Dams whose catchment area is larger than 5% basin area are modeled at each basin. Dam operation is assumed to be constant release operation, by which outflow of dam is calculated with the following equation

$$Q(t) = \begin{cases} I(t)\ (I(t) < Q_{\mathrm{ctl}}) \\ Q_{\mathrm{ctl}}\ (I(t) \geqslant Q_{\mathrm{ctl}}, (S(t) < S_{\mathrm{F}}) \\ I(t)\ (I(t) \geqslant Q_{\mathrm{ctl}}, S(t) \geqslant S_{\mathrm{F}}) \end{cases} \tag{5}$$

where, $I(t)$, $S(t)$, $Q(t)$ is inflow, storage, outflow of a dam at time t. Initial storage $Q(0)$ is assumed to be capacity for water utilization in flooding period, S_{F}.

The model parameters are identified in each basin to maximize the Nash–Sutcliffe coefficient using the SCE-UA algorithm (Duan *et al* 1994) as an optimization method. This study examines simultaneous occurrence probability of extreme flood peaks at the standard gauging station of each river system (located near by the major city within its catchment); thus, the parameters are calibrated to the hydrograph of the station during the largest flood event between 1981 and 2017 (when enough observation data was available in most river systems). If the observed rainfall/discharge data has problems (most data are missing, runoff ratio exceeds 1.0, …etc), the second largest one was targeted. Calibrated parameters are summarized in supplementary table S1 (available online at stacks.iop.org/ERL/16/074059/mmedia). As the result, for the calibrated event, the model showed the Nash–Sutcliffe efficiency (NSE) coefficient ranging from 0.730 to 0.997 (0.923 on average) and peak discharge error within ±20% among all the basins. The calibrated model was validated for the second largest event (the third largest one if the second largest event was used for calibration) in each river basin. As the result, the NSE values ranged from 0.523 to 0.988 (0.816 on average), and peak error ratios were within ±30% among all the basins. Hydrographs of the calibration and validation events at one river system from each region, i.e. Hokkaido, Tohoku, Kanto, Koshinetsu, Shikoku and Kyushu (see figure 1), were shown in supplementary figures S2 and S3, respectively.

2.2. Large ensemble climate simulation data d4PDF

d4PDF is a huge ensemble climate simulation database produced using MRI-AGCM 3.2 s at 60 km resolution, which was downscaled to 20 km with the non-hydrostatic regional climate model (NHRCM) (Sasaki 2008) over Japan. This study used this 20 km regional experiment. The downscaled d4PDF consists of 50 ensembles of climate simulations under the observed boundary conditions with observable small perturbations in sea surface temperature (SST) and sea ice from 1951 to 2010 (historical experiment, hereinafter HPB), and six ensembles of different future SST projections corresponding to around 4-degree rise in global mean temperature under 15 boundary condition ensembles (900 ensembles in total), in which the NHRCM was run for 60 years (4-degree rise experiment, hereinafter HFB) (Mizuta *et al* 2017). Therefore, HPB and HFB have 3000- and 5400-year data, respectively.

This dynamically downscaled dataset resolved typhoons, which is the major driver of flooding in Japan, and well reproduced their genesis frequency with the spatial distribution (Yoshida *et al* 2017) and the central pressure to some degree (Nakajo *et al* 2020) while it still has bias from observation with similar to other climate models. The model bias in d4PDF was corrected by Kobayashi *et al* (2020) in terms of rainfall intensity (direct input to the rainfall-runoff model 1K-DHM) so that annual maximum basin rainfall reproduces observation in all the class-A river basins (the target rivers in this study) as follows.

2.3. Bias correction and rainfall-runoff simulation

Kobayashi *et al* (2020) corrected the bias of rainfall intensity during a storm event causing the annual maximum basin rainfall in each year (hereinafter, the annual maximum storm) in all the class-A rivers using the radar AMeDAS rainfall (RAR) as reference. Bias correction ratio was calculated as the difference ratio of D-hour rainfall between HPB of d4PDF (D is the design rainfall duration of each river ranging between 6 and 72 h) and RAR at the same quantile (Piani *et al* 2010) and applied to rainfall intensity of the original 20 km grid rainfall during D-hour. The bias-corrected rainfall was converted into river discharge using the aforementioned rainfall-runoff model (Kobayashi *et al* 2020). This study utilizes the simulated discharge data for the following simultaneous flood risk analysis.

3. Simultaneous flood risk analysis

3.1. Definition of occurrences of flooding in each river basins and simultaneous flooding

Performing 3000- and 5400-year continuous simulation for both the past (HPB) and 4-degree rise (HFB) experiments consumes huge time and computer resources for its completion. To avoid this, we focus on annual maximum storms in each river basin, i.e. simulate 3000 and 5400 flood events for all the target river basins in total. Then, we define, in a river basin of interest, flooding as a phenomenon that the flood peak discharge of an annual maximum storm exceeds the one corresponding to the design return period (100, 150 and 200 years depending on each river basin, hereinafter, flow capacity). This procedure implies that we assume the flood occurrence at most once per year per one river basin. Simultaneous flooding is, accordingly, defined as a phenomenon that annual maximum river discharge exceeds the flow capacity in multiple river basins in the same year.

Regulating simultaneous flooding in the same year (between January and December) may miss simultaneous floods crossing the year. Although it is well known that Japanese river basins under the Asian Monsoon climate receive extreme precipitation from typhoons and frontal rain in the middle of the year, the frequency of annual maximum storms (years) for each month in d4PDF was examined. Among all the target river basins, around 60%–68% to 98% of events (years) occurred between June and October in HPB and HFB, respectively (see supplementary figure S4). It became 88%–100% when it comes to floods larger than the 100-year event (minimum level of the design return period) (see supplementary figure S5), justifying the treatment of every year as individual isolated time units both in the present and future climates.

3.2. Annual flood occurrence probability at a single river basin

The estimated annual probability of flood occurrence in HFB at all the 109 class-A river basins is shown in figure 2. The probability in HFB is much higher than the design level of ranging from 0.005 to 0.01 (corresponding to 100- to 200-year return periods) in the target river basins. Increase of the occurrence probability is higher in the northern (Hokkaido and Tohoku) regions where design flood is smaller corresponding to smaller flood discharge in historical records and

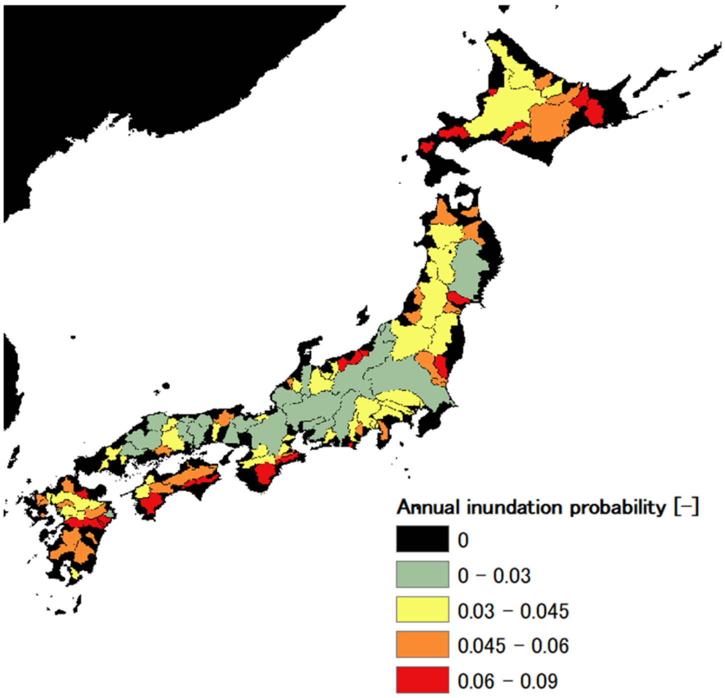


Figure 2. Annual probability of flood occurrence in HFB at all the 109 class-A river basins.

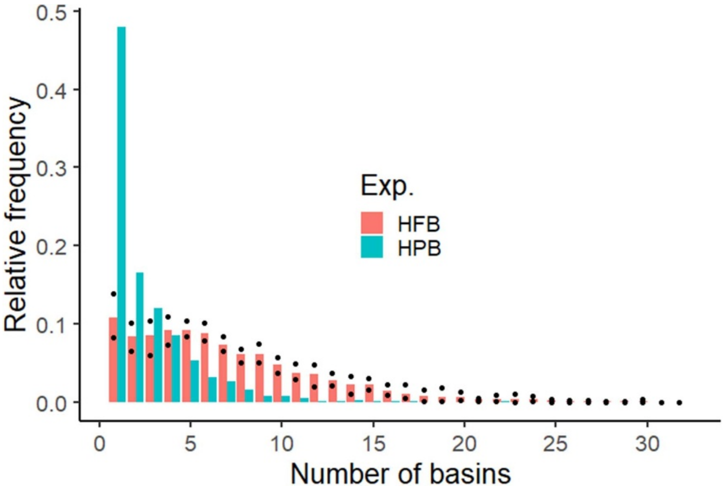


Figure 3. The histogram of the number of inundated river basins in the same year. The blue and red bins show the historical (HPB) and the mean of the 4-degree rise (HFB) experiment, respectively. The black dots show the SST ensemble range (the minimum and maximum values) in HFB.

the southern (Kyushu and Shikoku) regions where severe typhoons approach frequently (see figure 1 for locations). The one-sided binominal distribution test showed that the change is statistically significant at 1% significance level.

3.3. Probability of simultaneous flood occurrence

The histograms of the number of inundated river basins in the same year in HPB and HFB are shown in figure 3. HFB shows much higher frequencies of simultaneous flood occurrence among more than three river basins. Although there are some variations in frequencies among the SST ensembles of HFB (between the two black dots), its length is far shorter than the future increase, indicating that future changes of simultaneous flood occurrence probability is far larger than the uncertainty of future SST projections. The tail of the distributions become long; consequently, there are some extreme events with around 30 basins inundated in the same year. In the top five years with the largest number of inundated basins for both the experiment, heavy rainfall is caused by a typhoon.

The exceedance probability of the number of flooded river basins among the 41 basins in western Japan is shown in figure 4. The black line in the

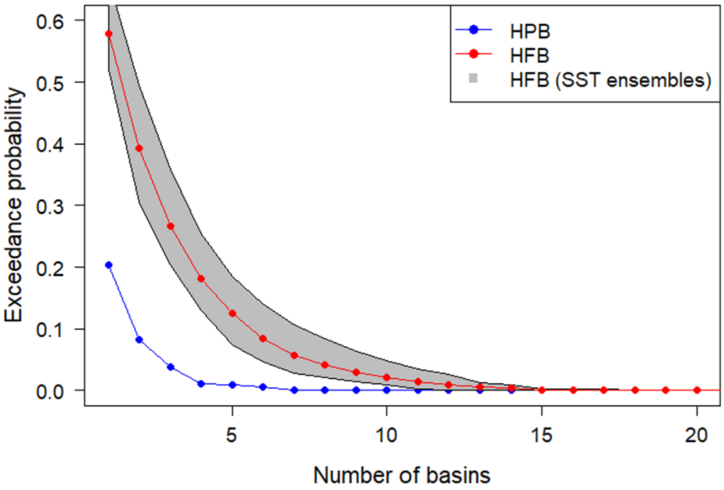


Figure 4. Annual exceedance probability (AEP) of the number of flooded river basins in HPB and HFB among the 41 basins in western Japan. The gray bar shows the range of SST ensembles in HFB.

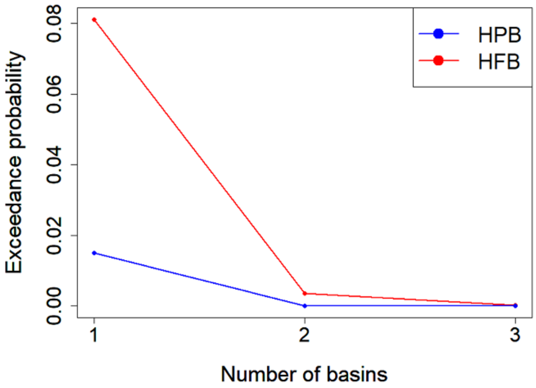


Figure 5. The exceedance probability of the number of flooded river basins among the Ara, Yodo, and Shonai River basins. As the probability is much smaller than all Japan and the western Japan cases, it was estimated using all the SST ensemble members (5400 years in total).

figure indicates the number in the rainfall event in 2018 (eight rivers were inundated), whose exceedance probability in HPB is around 0.002 corresponding to about 500-year return period, whereas one in HFB is around 0.035 corresponding to 28-year return period. For the Tohoku, Kanto and Koshinetsu Regions where six river basins were inundated by Typhoon Hagibis in 2019, the similar analysis (without the SST ensemble range) was performed by Tanaka *et al* (2020) showing that the exceedance probability in HPB is around 0.0027 (370-year return period) while that in HFB is around 0.05 (20-year return period). Both results show that a large-scale flooding at hundreds of return period will occur around once in a few decades in the 4-degree rise climate.

The same plot for the Ara, Yodo, and Shonai River basins having the mega cities: Tokyo, Osaka, and Nagoya, respectively, is plotted in figure 5. Note that this probability was estimated using all the SST ensemble members (5400-year data) because simultaneous flood occurrence among the three river basins rarely occurs; thus, uncertainty range is not shown in this case. In the historical experiment HPB, simultaneous flood did not occur among them. HFB shows that two river basins are inundated with probability of 0.0011 (around 910-year return period) and all the basins are inundated in only one year, in which there are two major flood events: one caused flood in the Yodo and Shonai River basins; the other did in the Ara River basin. This is simply because the Yodo and Shonai Rivers are closer to each other than the Ara River. In Japan, simultaneous flooding in multiple mega cities is found to rarely occur in the present climate while it might happen more often than once in 1000 years under the 4-degree rise climate. Note, however, that this estimation is under high uncertainty due to the absence of SST ensemble range.

Return periods of simultaneous flood events have been rarely discussed due to the lack of observation data or climate simulation ensembles. The presented analysis demonstrated the usefulness of large ensemble climate simulation data for the frequency analysis of simultaneous flood occurrence, i.e. the accumulation risk of floods, which is different information from large-scale flood risk map in a literature (Alfieri *et al* 2014, Dottori *et al* 2016).

4. Geospatial analysis of simultaneous flood occurrence

4.1. ARA

As seen in a literature, simultaneous flood risk has been analyzed by multivariate analysis of flood peak data among multiple gauges (Keef *et al* 2009, Diederen *et al* 2019). As Quinn *et al* (2019) pointed out, statistical flood generators need to consider that dependence structure is different between regular and extreme floods. To focus on flood exceeding the design level, this study deals with the flood occurrence data, i.e. whether annual maximum peaks exceed the flow corresponding to the design return period or not. As such data is not a continuous but discrete variable (flooding occurs or not), typical correlation analysis cannot be applied (Changhai and Shenping 2019). Furthermore, it is unrealistic to examine dependence between all the combinations from 109 river basins (only a pair of two basins makes 5886 patterns) in each of which 5400 flood occurrence data are included in the 4-degree rise experiment. Dependence structure in such discrete, big data has been analyzed in the field of market analysis using the ARA. This study first applies the ARA to geospatial analysis of simultaneous flood occurrence.

The ARA is a method to identify interesting relations between two sets of variables from big data and first proposed and applied in the field of market basket analysis to find items bought together frequently by Agrawl *et al* (1993). In this analysis, problems are formulated as follows: let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of items and D be a set of transactions where each transaction T is a set of items ($T \in I$). Then, the combination of two transactions X and Y such that $\emptyset \neq X, Y \subset I$ and $X \cap^{Y} \neq \emptyset$ is defined as an association rule $X \Rightarrow Y$, where X is the antecedent; Y is the consequent. An association rule is featured as a higher value of indices: 'support', 'confidence', and 'lift value'. Support $\sup(X \Rightarrow Y)$ is defined as $\sup(X \Rightarrow Y) = P\left(X \cap^{Y}\right)$, which means the frequency of transaction(s) $T = \{X, Y\}$ in all the transactions $|D|$. The confidence $\mathrm{conf}(X \Rightarrow Y)$ is defined as: $\mathrm{conf}(X \Rightarrow Y) = P(X \cap Y)/P(X) = P(Y|X)$, indicating the conditional probability of Y on X.

In this study, a set of items I consists of all the river basins $\{\mathrm{RB}_1, \mathrm{RB}_2, \ldots, \mathrm{RB}_{109}\}$ and $\{0\}$ (no occurrence of flooding); a transaction T is a set of river basins flooded ($T = \{0\}$ when no river is flooded) in a certain year. In this analysis, the 4-degree rise experiment data (5400-year simulations) was used as it showed larger number of flood occurrences; thus, 5400 transactions were obtained. This study deals with flood occurrences between two river basins in each transaction. For example, between river basins RB_2 and RB_5 (denote that $X = \{\mathrm{RB}_2\}$ and $Y = \{\mathrm{RB}_5\}$, let RB_2 and RB_5 be flooded in 108 and 270 events (year) in 5400 years and both are flooded in 54 years. Then, $\sup(X \Rightarrow Y) = P(X \cap Y) = 54/5400 = 1/100$ and $\mathrm{conf}(X \Rightarrow Y) = P(X \cap Y)/P(X) = 54/108 = 1/2$. Conditional probability is used to avoid from extracting rules with week relations but high support due to the high frequency of transaction X. On the other hand, only the confidence may extract trivial rules such that the transaction X rarely occurs. In the above example, $\sup(X \Rightarrow Y)$ is 1/100, meaning that both RB_2 and RB_5 are flooded once in 100 years, and $\mathrm{conf}(X \Rightarrow Y) = 1/2$, indicating that RB_5 is flooded once in two times when RB_2 is flooded. If both $\sup(X \Rightarrow Y)$ and $\mathrm{conf}(X \Rightarrow Y)$ are high, flooding in RB_2 and RB_5 is strongly related and not rare. ARA aims to extract interesting but non-trivial rules by selecting those with both $\sup(X \Rightarrow Y)$ and $\mathrm{conf}(X \Rightarrow Y)$ over the threshold (see the values in this study in section 4.3). This selection is efficiently implemented with the Apriori algorithm (Agrawal *et al* 1994) in R package 'arules'. Finally, an additional criterion called 'lift value'

$$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{conf}(X \Rightarrow Y)}{\sup(Y)} = \frac{P(Y|X)}{P(Y)} = \frac{P(Y \cap X)}{P(X)\,P(Y)} \tag{6}$$

is used to cut out rules extracted due to frequent occurrence of Y. After all, interesting and meaningful rules are identified as the high support and confidence above the thresholds and the high lift value.

4.2. Community extraction by graph theory

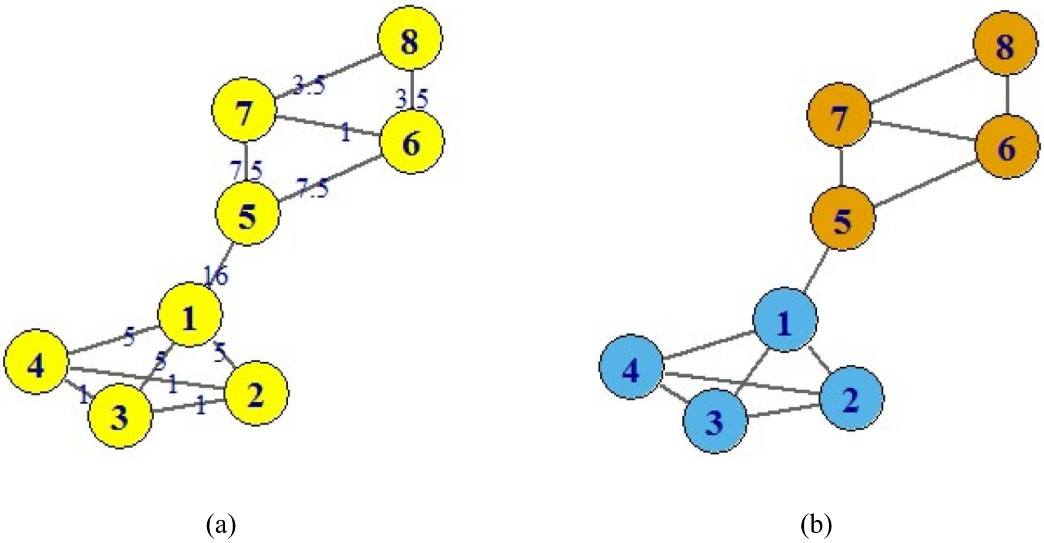
A rule (transaction) extracted as support and confidence above threshold is expressed as a link and the connected two river basins are expressed as nodes, making undirected graph (see figure 6(a) as an example). A community of simultaneous flood occurrence, composed of elements strongly connected with each other (i.e. high support, confidence, and the lift value), is then extracted from the constructed graph based on the intermediation centrality C_{v} defined as:

$$C_{\mathrm{v}}(e) = \frac{1}{2} \sum_{s,t} \frac{\sigma_{st}(e)}{\sigma_{st}} \tag{7}$$

where e is a link targeted; σ_{st} is the total number of the shortest paths between nodes s and t; $\sigma_{St}(e)$ is the paths with the link e. High intermediation centrality indicates that a link between nodes s and t works as a hub connecting two individual communities. Such communities are identified by cutting a link with the largest C_{v} (see figure 6(b) as an example). This study applies the GN algorithm (Girvan and Newman 2002) which repeats the above process until the network is split into strongly connected communities using the modularity as criterion. Modularity is a measure of connectivity in communities and high when links connecting nodes within a community are dense but those between communities are sparse, i.e. nodes (river basins in this study) are divided into strongly related groups (see Newman and Girvan 2004). In figure 7(b), the modularity is the highest when dividing into two communities $\{1, 2, 3, 4\}$ and $\{5, 6, 7, 8\}$. All these algorithms are implemented in R package 'igraph'.

4.3. Geospatial characteristics of simultaneous flood occurrence

We applied the ARA to simultaneous flood occurrence among all the 109 river basins and extracted interesting rules by the Apriori algorithm. The threshold values depend on the problem. This study


8
7
6
5
1
4
2
3
3.5
3.5
1
7.5
7.5
16
5
5
5
1
1
1
8
7
6
5
1
4
2
3

(a) (b)

Figure 6. An example of community extraction: (a) the numbers indicate the intermediation centrality defined as equation (7); (b) two communities are extracted by cutting the whole network at a link with the largest intermediation centrality of 16.

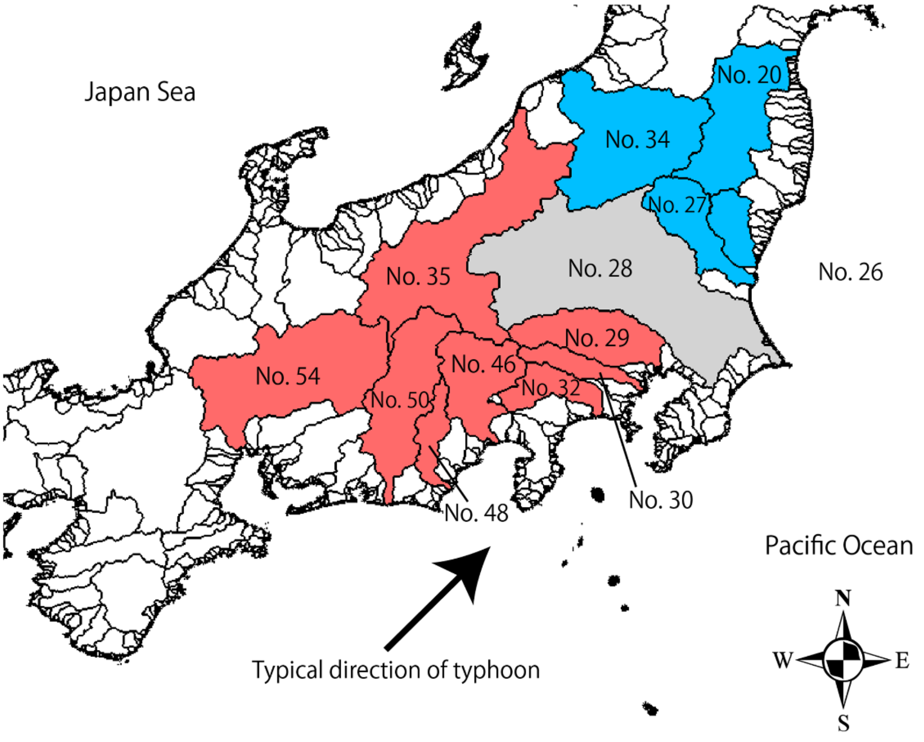

Japan Sea
No. 20
No. 34
No. 27
No. 28
No. 26
No. 35
No. 29
No. 54
No. 46
No. 32
No. 50
No. 48
No. 30
Pacific Ocean
N
W
E
S
Typical direction of typhoon

Figure 7. The locations of river basins around the Tone River basin (No. 28 shown in gray). The southeastern river basins of the Tone River basin (red) have higher lift value whereas the value is lower for the northeastern basins (blue).

set the minimum threshold of the support and confidence as 0.001 and 0.1, respectively and by selecting the top 25 ones so that geospatial characteristics can be seen. To see the characteristics of association rules, the top ten rules fixing the precedent to the Tone River basin having the largest catchment area in Japan are shown in table 1. The location of each basin is displayed in figure 7. Obviously, the combinations of nearer river basins have higher lift value. Interestingly, the consequent of the most rules is located in the southeastern of the Tone River basin (as shown in red in figure 7) whereas its northern basins No. 34, No. 20, No. 27 and No. 26 are ranked at 8th, 10th and out of the top tenth, respectively (as shown in blue in figure 7). As typhoons approach from the south Pacific Ocean in Japan, when the Tone River is flooded, its southeastern basins are more likely to be flooded together than the northern basins. The ARA

Table 1. Top 10 association rules from the Tone River basin. See figure 7 for the locations.

Association rule	Lift value	Support	Confidence
28 ⇒ 35	14.0	0.0094	0.364
28 ⇒ 46	10.1	0.0100	0.386
28 ⇒ 29	8.42	0.0080	0.307
28 ⇒ 32	7.97	0.0070	0.271
28 ⇒ 50	7.71	0.0033	0.129
28 ⇒ 54	7.01	0.0030	0.114
28 ⇒ 30	6.69	0.0070	0.271
28 ⇒ 34	6.29	0.0072	0.279
28 ⇒ 48	5.95	0.0069	0.264
28 ⇒ 20	5.85	0.0059	0.229

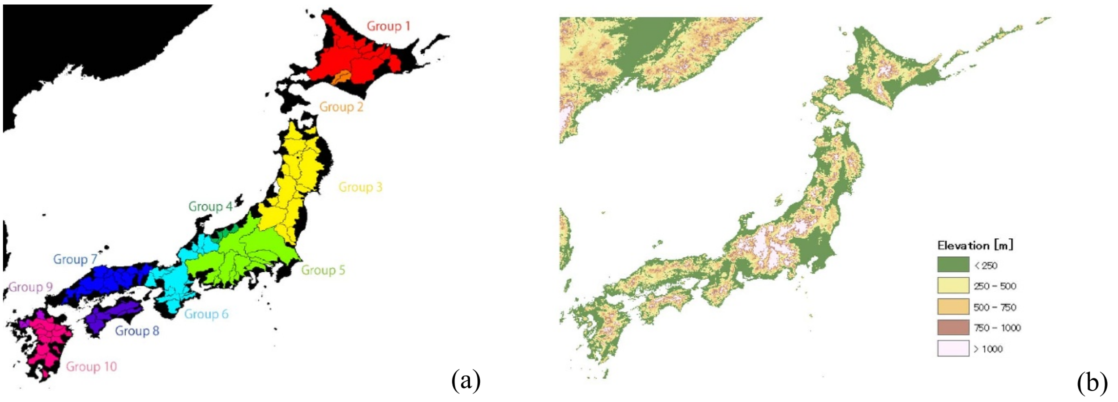


Figure 8. (a) Extracted communities of simultaneous flood occurrence and (b) elevation in Japan.

clarified that such geographical feature still applies in extreme storm events that cause large-scale precipitation. This finding will be informative for large-scale flood risk management and/or insurance planning.

Based on the extracted rules, the 109 river basins are classified into ten groups as shown in figure 8(a). River basins with the same color belong to the same community. With similar to the ARA, spatially closer basins are classified into the same group; however, its boundary does not always the same as mountain ridges (figure 8(b)). This indicates that at a national scale, spatial distance is more dominant than topographic features in extreme storms. On the other hand, the Group 6 (light-blue region in figure 7) expands to the north of the Group 5, whose boundary exactly correspond to the ridges of the Japan Alps probably because of their notably high altitudes. More regional investigation is demonstrated in Hokkaido and Kyushu regions (see figure 1) by applying the community extraction within the region as shown in figures 9 and 10, respectively, which clearly shows that the communities are consistent with mountain ridges. In summary, the grouping of simultaneous flood occurrence at a national scale is affected mainly by spatial distance and partly by mountain ridges with high altitudes while at a regional scale, the latter dominates it.

5. Summary, concluding remarks and future work

This study estimated the present and future simultaneous flood occurrence probability and its geospatial characteristics using big data of simulated flood peak discharge of annual maximum rainfall events in a large ensemble climate simulation database d4PDF and informatics techniques: ARA and graph theory. Even in Japanese river basins suffering from severe typhoons every year, simultaneous flood occurrence is a low-frequency phenomenon at least in the last decades. Frequencies of such events are first quantified in this study owing to a large number of ensembles in d4PDF and a rainfall-runoff model calibrated in all the class-A river basins in Japan.

The probability of simultaneous flood occurrence in the 4-degree rise experiment HFB is significantly larger than that in the historical experiment HPB in all the three cases examined: all Japan, western Japan that experienced large-scale inundation in 2018, and mega cities. The significant increase is largely attributed to the increased magnitude and frequencies of extreme discharge in each river basin. The future change of their correlation might have some impact, which should be analyzed in future research. Statistical techniques for multivariate

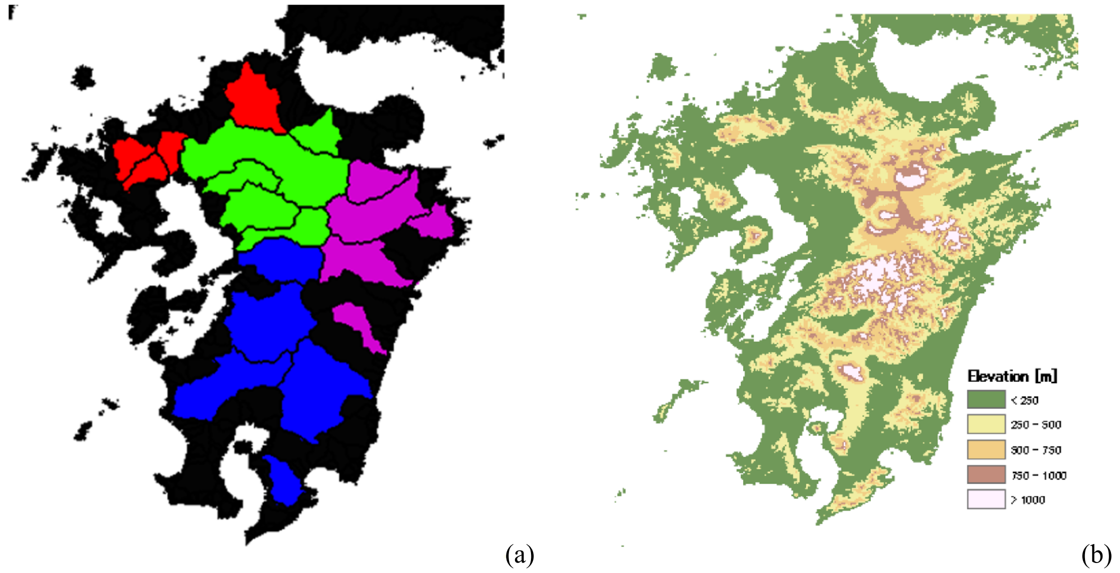


Figure 9. (a) Extracted communities of simultaneous flood occurrence and (b) elevation in the Kyushu region (see figure 1 for the location).

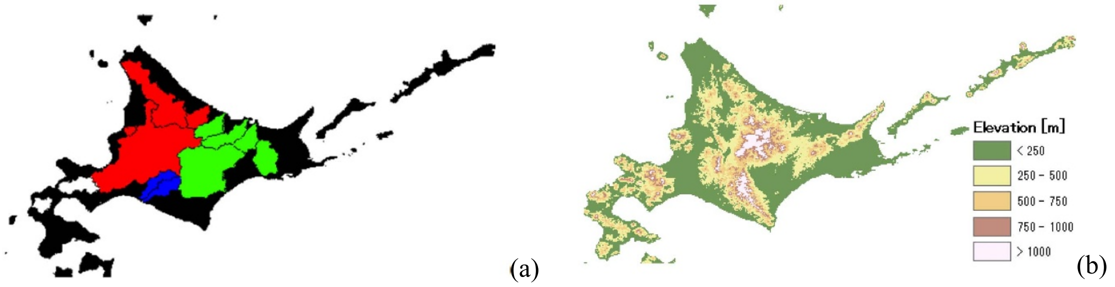


Figure 10. (a) Extracted communities of simultaneous flood occurrence and (b) elevation in the Hokkaido region (see figure 1 for the location).

probability distribution such as the copula approach might be helpful (Bevacqua *et al* 2017). The uncertainty analysis of the probability estimation is kept as the future work because it focused on first proposing a big data-based approach to multi-basin flood risk analysis. The Bayesian approach and/or the MCMC approach could support to quantify the uncertainty range of simultaneous flood occurrence probability (Zhou *et al* 2017).

The geospatial characteristics of simultaneous flood risk were first analyzed using the ARA and graph theory, which provided the degree of correlation with surrounding areas while being unable to detect various combinations of river basins at high simultaneous flood risk. This study, using the graph theory, identified the groups of river basins to which insurance design and/or urban planning should pay much attention.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This study was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research(A) JP17H01294, Grant-in-Aid for Early-Career Scientists JP18K13837 and the Integrated Research Program for Advancing Climate Models (TOUGOU) Grant Number JPMXD0717935498 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. A rainfall-runoff model 1K-DHM was developed in collaborative research with Tokio Marine dR Co., Ltd. This study used d4PDF produced with the Earth Simulator jointly by science programs (SOUSEI, TOUGOU, SI-CAT, DIAS) of the MEXT, Japan. This dataset was collected and provided under the Data Integration and Analysis System (DIAS, Project No. JPMXD0716808999), which has been developed and operated by the MEXT.

References

Agrawal R, Imieliński T and Swami A 1993 Mining association rules between sets of items in large databases *Proc. 1993 ACM SIGMOD Int. Conf. on Management of Data* (June) pp 207–16

Agrawal R and Srikant R 1994 Fast algorithms for mining association rules *Proc. 20th Int. Conf. Very Large Data Bases, VLDB* (September) vol 1215 pp 487–99

Alfieri L, Salamon P, Bianchi A, Neal J, Bates P and Feyen L 2014 Advances in pan-European flood hazard mapping *Hydrol. Process.* **28** 4067–77

Arnell N W and Gosling S N 2016 The impacts of climate change on river flood risk at the global scale *Clim. Change* **134** 387–401

Bevacqua E, Maraun D, Hobæk Haff I, Widmann M and Vrac M 2017 Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy) *Hydrol. Earth Syst. Sci.* **21** 2701–23

Brunner M I, Papalexiou S, Clark M P and Gilleland E 2020 How probable is widespread flooding in the United States? *Water Resour. Res.* **56** 10

Bull-Kamanga L *et al* 2003 From everyday hazards to disasters: the accumulation of risk in urban areas *Environ. Urban* **15** 193–204

Changhai H and Shenping H 2019 Factors correlation mining on maritime accidents database using association rule learning algorithm *Cluster Comput.* **22** 4551–9

Chen L, Singh V P, Shenglian G, Hao Z and Li T 2012 Flood coincidence risk analysis using multivariate copula functions *J. Hydrol. Eng.* **17** 742–55

Diederen D, Liu Y, Gouldby B, Diermanse F and Vorogushyn S 2019 Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment *Nat. Hazards Earth Syst. Sci.* **19** 1041–53

Döll P, Trautmann T, Gerten D, Schmied H M, Ostberg S, Saaed F and Schleussner C F 2018 Risks for the global freshwater system at 1.5 °C and 2 °C global warming *Environ. Res. Lett.* **13** 044038

Dottori F, Salamon P, Bianchi A, Alfieri L, Hirpa F A and Feyen L 2016 Development and evaluation of a framework for global flood hazard mapping *Adv. Water Resour.* **94** 87–102

Duan Q, Sorooshian S and Gupta V K 1994 Optimal use of the SCE-UA global optimization method for calibrating watershed models *J. Hydrol.* **158** 265–84

Faye B *et al* 2018 Impacts of 1.5 versus 2.0 °C on cereal yields in the West African Sudan Savanna *Environ. Res. Lett.* **13** 034014

Geertsema T J, Teuling A J, Uijlenhoet R, Torfs P J and Hoitink A J 2018 Anatomy of simultaneous flood peaks at a lowland confluence *Hydrol. Earth Syst. Sci.* **22** 5599–613

Girvan M and Newman M E J 2002 Community structure in social and biological networks *Proc. Natl Acad. Sci.* **99** 7821–6

Hirabayashi Y, Mahendran R, Koirala S, Konoshima L, Yamazaki D, Watanabe S, Kim H and Kanae S 2013 Global flood risk under climate change *Nat. Clim. Change* **3** 816–21

Keef C, Svensson C and Tawn J A 2009 Spatial dependence in extreme river flows and precipitation for Great Britain *J. Hydrol.* **378** 240–52

Kobayashi K, Tanaka T, Shinohara M and Tachikawa Y 2020 Analyzing future changes of extreme river discharge in Japan using d4PDF *J. Japan Soc. Civil Eng. B1* **76** 140–52

Lavender S L, Walsh K J, Caron L P, King M, Monkiewicz S, Guishard M, Zhang Q and Hunt B 2018 Estimation of the maximum annual number of North Atlantic tropical cyclones using climate models *Sci. Adv.* **4** eaat6509

Leduc M *et al* 2019 The ClimEx project: a 50-member ensemble of climate change projections at 12-km resolution over Europe and northeastern North America with the Canadian Regional Climate Model (CRCM5) *J. Appl. Meteorol. Climatol.* **58** 663–93

Mitchell D *et al* 2017 Half a degree additional warming, prognosis and projected impacts (HAPPI): background and experimental design *Geosci. Model Dev.* **10** 571–83

Mizuta R *et al* 2017 Over 5000 years of ensemble future climate simulations by 60 km global and 20 km regional atmospheric models *Bull. Am. Meteorol. Soc.* **98** 1383–98

MLIT 2019a Economic damage by Typhoon Hagibis, 2019 (50th report) (available at: www.mlit.go.jp/common/001319108.pdf) (Accessed 29 September 2020)

MLIT 2019b Economic damage by heavy rainfall in July, 2018 (available at: www.mlit.go.jp/common/001301033.pdf) (Accessed 29 September 2020)

Moftakhari H and AghaKouchak A 2019 Increasing exposure of energy infrastructure to compound hazards: cascading wildfires and extreme rainfall *Environ. Res. Lett.* **14** 104018

Mori N, Shimura T, Yoshida K, Mizuta R, Okada Y, Fujita M, Khujanazarov T and Nakakita E 2019 Future changes in extreme storm surges based on mega-ensemble projection using 60-km resolution atmospheric global circulation model *Coast. Eng. J.* **61** 295–307

Nakajo S, Umeda J and Mori N 2020 Applicability of D4PDF dataset to Global Stochastic Tropical Cyclone Model *Int. Conf. Coastal. Eng.* **36** 26

Neal J, Keef C, Bates P, Beven K and Leedal D 2013 Probabilistic flood risk mapping including spatial dependence *Hydrol. Process.* **27** 1349–63

Newman M E J and Girvan M 2004 Finding and evaluating community structure in networks *Phys. Rev.* E **69** 026113

Piani C, Weedon G P, Best M, Gomes S M, Viterbo P, Hagemann S and Haerter J O 2010 Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models *J. Hydrol.* **395** 199–215

Quinn N, Bates P D, Neal J, Smith A, Wing O, Sampson C, Smith J and Heffernan J 2019 The spatial dependence of flood hazard and risk in the United States *Water Resour. Res.* **55** 1890–911

Saleh F, Ramaswamy V, Wang Y, Georgas N, Blumberg A and Pullen J 2017 A multi-scale ensemble-based framework for forecasting compound coastal-riverine flooding: the Hackensack-Passaic watershed and Newark Bay *Adv. Water Resour.* **110** 371–86

Sasaki H, Kurihara K, Takayabu I and Uchiyama T 2008 Preliminary experiments of reproducing the present climate using the non-hydrostatic regional climate model *SOLA* **4** 25–8

Shakti P 2017 Quantitative precipitation estimation and hydrological modeling in Japan *J. Japan. Soc. Hydrol. and Water Resour.* **30** 6–17

Swiss Re Institute 2018 Natural catastrophes and man-made disasters in 2017: a year of record-breaking losses, Sigma vol 1 (Swiss Re Institute) pp 4–5 (https://www.swissre.com/institute/research/sigma-research/sigma-2018-01.html)

Tanaka T, Kobayashi K and Tachikawa Y 2020 Analyzing return period of number of flooded rivers by Typhoon Hagibis and its future change *J. Japan Soc. Civil Eng. B1* accepted (https://doi.org/10.2208/jscejhe.76.1_159)

Tanaka T and Tachikawa Y 2015 Testing the applicability of a kinematic wave-based distributed hydrological model in two climatically contrasting catchments *Hydrol. Sci. J.* **60** 1361–73

Wahl T, Jain S, Bender J, Meyers S D and Luther M E 2015 Increasing risk of compound flooding from storm surge and rainfall for major US cities *Nat. Clim. Change* **5** 1093–7

Winter B, Schneeberger K, Förster K and Vorogushyn S 2020 Event generation for probabilistic flood risk modelling: multi-site peak flow dependence model vs. weather-generator-based approach *Nat. Hazards Earth Syst. Sci.* **20** 1689–703

Yang J A, Kim S, Mori N and Mase H 2018 Assessment of long-term impact of storm surges around the Korean

Peninsula based on a large ensemble of climate projections *Coast. Eng.* **142** 1–8

Yoshida K, Sugi M, Mizuta R, Murakami H and Ishii M 2017 Future changes in tropical cyclone activity in high-resolution large-ensemble simulations *Geophys. Res. Lett.* **44** 9910–917

Zhou Y, Wang Z, Jin J, Cheng L and Zhou P 2017 Uncertainty analysis of designed flood on Bayesian MCMC algorithm: a case study of the Panjiakou reservoir in China *Environ. Earth Sci.* **76** 788

Zscheischler J *et al* 2018 Future climate risk from compound events *Nat. Clim. Change* **8** 469–77