Hydrological Sciences Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/thsj20

Development of a global flood risk index based on natural and socio-economic factors

Yuko Okazawa [a], Pat J.-F. Yeh [b], Shinjiro Kanae [c] & Taikan Oki [b]

[a] Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Megoru-ku Tokyo, Japan

[b] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Megoru-ku, Tokyo, Japan

[c] Department of Mechanical and Environmental Informatics, Tokyo Institute of Technology, 2-12-1-W8-4 O-okayama, Megoru-ku, Tokyo, Japan

To cite this article: Yuko Okazawa, Pat J.-F. Yeh, Shinjiro Kanae & Taikan Oki (2011): Development of a global flood risk index based on natural and socio-economic factors, Hydrological Sciences Journal, 56:5, 789-804

To link to this article: http://dx.doi.org/10.1080/02626667.2011.583249

Development of a global flood risk index based on natural and socio-economic factors

Yuko Okazawa[1], Pat J.-F. Yeh[2], Shinjiro Kanae[3] & Taikan Oki[2]

[1] *Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Megoru-ku Tokyo, Japan now at: ALMEC Corporation, 1-19-14 Aobadai, Megoru-ku, Tokyo, Japan*

[2] *Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Megoru-ku Tokyo, Japan*
taikan@iis.u-tokyo.ac.jp

[3] *Department of Mechanical and Environmental Informatics, Tokyo Institute of Technology, 2-12-1-W8-4 O-okayama, Megoru-ku, Tokyo, Japan*

Received 6 October 2008; accepted 2 February 2011; open for discussion until 1 January 2012

Citation Okazawa, Y., Yeh, P., Kanae, S. & Oki, T. (2011) Development of a global flood risk index based on natural and socio-economic factors. *Hydrol. Sci. J.* **56**(5), 789–804.

Abstract A global flood risk index (FRI) is established, based on both natural and social factors. The advanced flood risk index (AFRI) is the expectation of damage in the case of a single flood occurrence, estimated by a linear regression-based approach as a function of hazard and vulnerability metrics. The resulting equations are used to predict potential flood damage given gridded global data for independent variables. It is new in the aspect that it targets floods by units of events, instead of a long-term trend. Moreover, the value of the AFRI is that it can express relative potential flood risk with the process of flood damage occurrence considered. The significance of this study is that not only the hazard parameters which contribute directly to flood occurrence, but vulnerability parameters which reflect the conditions of the region where flood occurred, including its residential and social characteristics, were shown quantitatively to affect flood damage.

Key words flood risk; natural factors; social factors

Développement d'un indice global de risque d'inondation fondé sur des facteurs naturels et socio-économiques

Résumé Un indice global de risque d'inondation (FRI) est établi, basé sur des facteurs naturels et sociaux. L'indice avancé AFRI correspond à l'espérance des dommages dans le cas de l'occurrence d'une inondation unique, estimée à partir d'une approche basée sur une régression linéaire, fonction de mesures de l'aléa et de la vulnérabilité. Les équations qui en résultent sont utilisées pour prévoir les dommages potentiels à partir d'une grille de données globales de variables indépendantes. L'innovation relève du fait que cela considère des inondations par unités d'événement, au lieu d'une tendance à long terme. De plus, la valeur ajoutée de AFRI est que cet indicateur peut exprimer un risque d'inondation potentiel relatif, compte tenu du processus d'occurrence des dommages liés aux inondations. L'importance de cette étude réside dans le fait qu'elle montre quantitativement que les paramètres qui influencent les dommages sont non seulement ceux qui régissent l'aléa et donc l'occurrence de l'inondation, mais aussi ceux qui régissent la vulnérabilité et donc les propriétés de la région impactée par l'inondation, y compris ses caractéristiques résidentielles et sociales.

Mots clefs risque d'inondation; facteurs naturels; facteurs sociaux

1 INTRODUCTION

Flood is one of the most serious natural disasters today. According to the World Bank (2005), regions affected by floods during 1985–2003 comprise more than one third of the Earth's surface, inhabited by over 82% of the world's population. The fact that the Gross Domestic Product (GDP) and the total agricultural production in these regions are three times the world's average suggests that these flood-prone regions are also resource-concentrated areas.

Floods can be caused by various events, such as intense precipitation resulting in a drastic increase

ISSN 0262-6667 print/ISSN 2150-3435 online

doi: 10.1080/02626667.2011.583249
http://www.informaworld.com

in river discharge, snowmelt, ice-jam, glacial lake outburst, etc. However, the degree of damage caused by floods in a specific region is dependent on many natural and socio-economic factors, such as the density of population and assets, land use, infrastructure development (e.g. dikes and dams), and the speed and accuracy of information transmission (e.g. early warning systems). However, the relationships between these factors and associated flood risk have not been fully investigated. Here, flood risk is defined as the possibility of the degree of damage from an occurrence of flooding. Quantifying flood risk from various natural and socio-economic factors will allow us to assess how flood risk would change corresponding to the changes in population, climate and land-use conditions, and also how the policy of flood damage mitigation can potentially reduce the flood risk.

There are numerous definitions existent for "risk". Davidson (1997) explained risk as the result of hazard, exposure, vulnerability, capacity and measures, while Villagran de Léon *et al.* (2006) defined it as a four-dimensional concept composed of hazard, vulnerability, preparedness and capacity. Moreover, Wisner *et al.* (2004) argued that risk is a multiplied result of hazard and vulnerability, elaborating this concept by defining risk as the result of hazard and vulnerability divided (reduced) by capacity. Indeed, the appropriate way to define risk is still a controversial issue. Given that the assessment of flood risk is still in its infancy, however, we adopt in this study the simplest concept that "hazard" refers to the natural disasters (floods) themselves; "vulnerability" refers to whether the society can cope with these natural disasters; and "risk" refers to the integrated outcome of the magnitude of "hazard" and the degree of "vulnerability". Therefore, the "risk" is defined by the hazard multiplied by the vulnerability, and it is also equivalent to the "expected damage" from the hazard (natural disasters).

2 MOTIVATION AND OBJECTIVE

Previous studies on flood risk have been conducted by over 20 institutions worldwide. Here, a brief review is given.

2.1 The United Nations Development Programme (UNDP)

The UNDP (2004) has developed an event-based Disaster Risk Index (DRI) based on the regression analysis of past floods. Regression equations were constructed for each disaster type in each country, with the exposure, GDP per capita, and population density as independent variables, and the number of deaths as the dependent variable. Analysis of the data showed that countries with low GDP per capita, low density of population and high physical exposure are associated with high levels of flood risk. Relative vulnerability is defined as the result of the comparison between exposure and deaths, i.e. the lower the number of deaths to the same degree of exposure, the lower the vulnerability. Therefore, the DRI represents the risk of disaster as the degree of exposure and vulnerability.

2.2 The World Bank

The World Bank (2005) has calculated Natural Disaster Hotspots, defined as the top three deciles of the number of cumulative deaths from floods according to the Emergency Events Database (EM-DAT, http://www.emdat.be/). The data represent the average damage during the study period of 1985–2003. The hotspots were calculated globally on a grid scale (1° and 2.5°). The calculated hotspots simply reflect the degree of actual damage; for example, the mortality risk from floods was determined in regions that have a large cumulative number of deaths over the study period, including eastern China, the Korean Peninsula, India, Bangladesh, Latin America and the Caribbean.

2.3 Munich Re

The Munich Re Group (2004) has developed the Hazard Index for Mega-cities (HIM), calculated by multiplying the hazard, vulnerability and exposure on a city basis. This index aims at allowing a comparison of flood risks between mega-cities, and is geared to the risk of material losses. Various natural hazards were objectively weighted by allocating their average annual losses. A catastrophic loss with a low occurrence probability was then calculated. A uniform basis of a 1000-year loss (i.e. probable maximum loss) was used. The results showed that the index is most significantly influenced by the degree of exposure, followed by the hazard, whereas the vulnerability only plays a minor role. This is possibly a result of considerably larger spread in the adopted indices of exposure and hazard than the index of vulnerability. The estimated HIM of flood is found to be particularly high in Calcutta, India and Dhaka, Bangladesh.

2.4 National Institute for Land and Infrastructure Management (Japan)

Hara *et al.* (2009) developed the Flood Vulnerability Index (FVI) for assessing flood risks. The FVI is an index for assessing the vulnerability to flood disasters that can be applied at the river-basin scale. It consists of a precipitation factor and three components: hydrogeographic and socio-economic factors, and countermeasures. These major components were divided into 11 indicators (i.e. sub-components), which were selected based on factor diagram analysis in terms of flood disasters. Then, the FVI values were estimated using multiple linear regression analysis for the major river basins around the world.

2.5 International Centre for Water Hazard and Risk Management (ICHARM)

The ICHARM (2007) of Japan has developed flood risk maps with hazard, vulnerability and resilience as dependent factors. Data were obtained at the country, river-basin and grid scales, based on which three flood risk maps were produced. One indicator was selected for each of the hazard, vulnerability and resilience factors. For example, the Reciprocal of Forest Area (vulnerability factor) is multiplied by the number of floods occurred (hazard factor), and then divided by the Digital Access Index (resilience factor) to derive the flood risk. The Digital Access Index was developed by the International Telecommunications Union, and characterizes a country's ability to access information and communication technology (ICT) based on infrastructure, affordability, knowledge and the quality and actual usage of ICT.

2.6 Motivation and objective of this study

Nevertheless, the aforementioned studies on flood (disaster) risks may still have the following limitations: (a) the resolution of these analyses, such as the country level, is too coarse for certain regional-scale applications; (b) most of them considered the resulting damage as the risk, but the sensitivities of contributing factors such as hazard, vulnerability, exposure and resilience were not analysed in a systematic framework; and (c) most of these studies focused on the cumulative or average damage; thus it is difficult to analyse the damage caused by a single catastrophic event, because of the averaging effect due to more frequently occurring floods with a relatively small degree of damage.

The present study aims to improve the limitations of previous flood risk studies by developing a new global flood risk index that incorporates both natural and socio-economic factors. The newly developed index is referred to as the Advanced Flood Risk Index (AFRI), which quantifies the expected value of damage caused by a single flood occurrence, as it focuses on the event scale instead of the long-term statistical trend of floods. The AFRI is a function of the metrics of flood hazard and vulnerability stratified by different flood generating mechanisms (i.e. flood types), estimated from a simple regression approach based on available global gridded data sets of influencing factors. It can be used to predict potential future flood damage, and the derived regression relationship between the AFRI and dependent factors can also be used to test the sensitivity of flood damage to the change in population, land cover and urbanization.

3 DATA

The globally-distributed data on flood hazard, vulnerability and damage have been collected from various sources for the study period of 1985–2000 on a 0.5° × 0.5° grid resolution. A brief introduction on the data used in this study is provided below.

3.1 Flood hazard data

Although many natural factors influence flood occurrence, for simplicity only the most important indicator, precipitation, is selected to represent flood hazard here. The global 0.5° × 0.5° gridded precipitation data set developed by Ngo-Duc *et al.* (2005) is utilized herein in view of its global coverage and relatively high accuracy.

3.1.1 Anomaly of weekly moving-average precipitation Since the influence of precipitation on flood occurrence depends on the flow concentration time of a basin (i.e. the accumulated time needed for a falling raindrop to reach the river and then flow downstream), precipitation that occurred within three days both before and after a recorded flood occurrence date is considered by using the following weekly moving average:

$$\overline{X_{\text{week}}} = \frac{\sum_{i=t-3}^{t+3} X_i}{7} \tag{1}$$

and the following standardized precipitation anomaly (A_{week}) is used in order to facilitate global comparison:

$$A_{\text{week}} = \frac{\overline{X_{\text{week}}} - \overline{\overline{X_{\text{week}}}}}{\sigma_{\text{week}}} \tag{2}$$

where σ_{week} represents the standard deviation of $\overline{X_{\text{week}}}$.

3.1.2 Return period The return period is the average recurrence time of a hydrological phenomenon above or below a certain threshold intensity. In this study, the return period of recorded flood events, assumed to follow a Gumbel distribution, is calculated based on daily precipitation data (Ngo-Duc *et al.* 2005) as an indicator of the severity of precipitation for a specific region.

3.2 Flood vulnerability data

The vulnerability parameters are selected from a group of parameters covering a wide range of attributes such as economy, health, land cover, population, river and vegetation. The 48 candidate parameters (see Table 1 for the list of parameters and data sources) are chosen based on their global availability and data consistency. For example, some parameters such as flood dike length are not included since they are not available globally at present. Although most countries may have their own data on flood dikes, the definition of flood dike as well as the accuracy and specification of the data, among other aspects, are not consistent among countries. Data are obtained from relatively more reliable sources, such as UN databases, and they are collected at the most commonly available frequency (e.g. yearly) for all the time-varying parameters, such as GDP, population and forest cover during the same period as flood damage data. Thus, the socio-economic conditions at the flood occurrence time can be reflected in the analysis of flood risk.

The following five-step screening procedure was conducted on the selected 48 candidate parameters (Table 1) in order to reduce the total number of parameters and improve their appropriateness to represent flood vulnerability:

(1) Minimize the dependence among selected parameters: if several candidate parameters are highly correlated (i.e. correlation coefficient > 0.8), only one of them is kept for further testing. This step is necessary to prevent biased results, since the redundancy among inter-correlated vulnerability parameters may affect the regressed relationship and their sensitivity to flood damage parameters.
(2) Spatial coverage of parameters: the available data of parameters must cover at least 80% of the world.
(3) Temporal coverage of parameters: the available data of parameters must cover the target period (1985–2000).
(4) Rationality of parameters: the selected parameters need to be those whose relationship to flood can be explained logically.
(5) Utility for political implications: the selected parameters need to be useful for the policy-making of flood damage mitigation.

While the above steps (1)–(3) can be tested objectively, steps (4) and (5) have to be judged by reference to related documents and reports. If a certain parameter has been verified by previous research to have a logical relation with floods, it is treated here as an appropriate parameter in Step (4). Similarly, if a certain parameter has been considered as useful for making flood mitigation policies, then it is treated as an appropriate parameter in Step (5). Finally, those parameters which fulfil the test in Step (1) and at least three of the remaining four tests are selected as final candidate parameters as highlighted in Table 1, by bold font.

3.3 Flood damage data

Several data sets on the flood damage are available globally, as summarized below:

3.3.1 EM-DAT EM-DAT is the flood damage database operated by the World Health Organization (WHO) in collaboration with the Centre for Research on the Epidemiology of Disasters (CRED). It largely categorizes disasters into natural disasters and technological disasters, further classifies into more detail, and also provides the damage data. The data are recorded on the country level (e.g. a typhoon event affecting several countries is recorded for each country), and include the information on damage for the cities affected. Although it provides long-term data (e.g. since the 1960s for Japan), some previous studies have reported that they are not always reliable (ICHARM, 2005).

Table 1 List of the 48 candidate vulnerability parameters and the details in the screening process for selecting final vulnerability parameters. The final selected vulnerability parameters are highlighted in **bold**.

ID	Data name	Test 1	Test 2		Test 3	Test 4	Test 5	Total	Data source
1	**Population within 100 km of coast**	**1**	**85.00**	**1**	**1**	**1**	**1**	**5**	UNEP/DEWA/GRID-Europe
2	NOAA/Global Vegetation Index 8-year maximum	1	98.84	1	0	0	0	2	DEWA/GRID-Europe
3	**Agricultural area—% of land area**	**1**	**91.14**	**1**	**1**	**1**	**1**	**5**	FAO
4	% of agricultural production to GDP	1	89.59	1	1	0	0	3	World Bank
5	Aid dependency ratio	1	30.45	0	1	1	0	3	World Bank
6	Average altitude	1	100.00	1	0	0	0	2	GTOPO30
7	National budget per capita	0	88.75	1	1	Null	Null	Null	Central Intelligence Agency
8	**Number of cities**	**1**	**100.00**	**1**	**1**	**1**	**1**	**5**	ESRI
9	% of population living on less than US$1 per day	1	63.35	0	1	1	0	3	World Bank
10	% of population living on less than US$2 per day	0	63.93	0	1	Null	Null	Null	World Bank
11	**Number of dams**	**1**	**100.00**	**1**	**0**	**1**	**1**	**4**	University of Yamanashi
12	Population density	1	91.60	1	1	0	0	3	UN Population Division
13	Digital access index	0	88.49	1	0	Null	Null	Null	International Telecommunication Union
14	Area per discharge telemetry station	0	42.21	0	0	Null	Null	Null	INFOHYDRO
15	% durable structure	1	28.64	0	0	1	1	3	World Bank
16	Economically active population	0	90.82	1	1	Null	Null	Null	ILO
17	Education index	0	88.43	1	1	Null	Null	Null	UNDP
18	Number of fixed telephones per 1000 people	0	90.50	1	1	Null	Null	Null	World Bank
19	Government final consumption expenditure	0	89.08	1	1	Null	Null	Null	World Bank
20	GDP	0	87.98	1	1	Null	Null	Null	World Bank
21	**GDP per capita**	**1**	**87.98**	**1**	**1**	**1**	**0**	**4**	World Bank
22	Household final consumption expenditure	0	83.39	1	1	Null	Null	Null	World Bank
23	Foreign direct investment	0	88.62	1	1	Null	Null	Null	World Bank

(*Continued*)

Table 1 (Continued).

ID	Data name	Test 1	Test 2		Test 3	Test 4	Test 5	Total	Data source
24	Adult literacy rate	0	67.36	0	1	Null	Null	Null	World Bank
25	**Number of mobile telephones per 1000 people**	**1**	**90.50**	**1**	**1**	**1**	**1**	**5**	World Bank
26	Volume of nonlife insurance	0	72.07	0	0	Null	Null	Null	Swiss Re Company
27	Official development assistance and official aid	1	71.43	0	1	1	0	3	World Bank
28	**Population increase—grid based**	**1**	**90.30**	**1**	**1**	**1**	**0**	**4**	IPCC/SRES
29	**Paddy area**	**1**	**100.00**	**1**	**0**	**1**	**1**	**4**	Leff, 2004
30	**Population density—grid based**	**1**	**99.16**	**1**	**1**	**1**	**0**	**4**	CIESEN, Columbia University, and CIAT
31	Area per precipitation telemetry station	0	40.47	0	0	Null	Null	Null	INFOHYDRO
32	Net enrolment Ratios in primary education	1	72.53	0	1	0	1	3	UNESCO
33	Number of radios per 1000 people	0	90.50	1	1	Null	Null	Null	UNESCO Institute for Statistics
34	**Maximum reservoir storage**	**1**	**100.00**	**1**	**0**	**1**	**1**	**4**	University of Yamanashi
35	Reservoir area	1	100.00	1	0	0	1	3	University of Yamanashi
36	Improved sanitation coverage—rural	0	84.62	1	1	Null	Null	Null	WHO/UN Children's Fund
37	Improved sanitation coverage—total	0	84.62	1	1	Null	Null	Null	WHO/UN Children's Fund
38	Improved sanitation coverage—urban	0	89.08	1	1	Null	Null	Null	WHO/UN Children's Fund
39	Net enrolment ratios in secondary education	0	58.89	0	1	Null	Null	Null	UNESCO
40	Total population	0	91.14	1	1	Null	Null	Null	UN Population Division
41	Number of televisions per 1000 people	0	84.68	1	1	Null	Null	Null	UNESCO Institute for Statistics
42	Total unemployment—% of total labour force	1	67.61	0	1	0	0	2	ILO
43	Urban population—% of total population	0	91.14	1	1	Null	Null	Null	UN Population Division
44	Improved drinking water coverage – rural	0	89.08	1	1	Null	Null	Null	WHO/UN Children's Fund
45	Improved drinking water coverage – total	0	89.08	1	1	Null	Null	Null	WHO/UN Children's Fund
46	Improved drinking water coverage – urban	0	90.50	1	1	Null	Null	Null	WHO/UN Children's Fund
47	Area per water level telemetry station	1	35.88	0	0	1	1	3	INFOHYDRO
48	**% of forest cover**	**1**	**91.01**	**1**	**1**	**1**	**1**	**5**	FAO/FAOSTAT

UNEP: United Nations Environment Programme; DEWA: Division of Early Warning and Assessment; GRID: Global Research Information Database; FAO: Food and Agriculture Organization; GTOPO30: Global 30 Arc Second Elevation Data Set; ESRI: Environmental Systems Research Institute Inc.; UN: United Nations; INFOHYDRO: The Hydrological Information Referral Service; ILO: International Labour Organization; UNDP: United Nations Development Programme; IPCC SRES: Intergovernmental Panel on Climate Change Special Report on Emissions Scenarios; CIESEN: Center for International Earth Science Information Network; CIAT: International Center for Tropical Agriculture; UNESCO: United Nations Educational, Scientific and Cultural Organization; WHO: World Health Organization; FAOSTAT: Statistics Division of the Food and Agriculture Organization

3.3.2 Dartmouth Flood Observatory The Dartmouth Flood Observatory operated by the Dartmouth University, Boulder, USA, has compiled a flood database (http://floodobservatory.colorado.edu/), which emphasizes flood disasters and records data according to flood type. Each flood event is recorded with its geophysical information, i.e. the name of country and city, longitude and latitude, and categorized by 11 flood types (heavy rain, tropical cyclone, extra-tropical cyclone, monsoonal rain, snowmelt, rain and snowmelt, ice jam/break-up, dam/levee break or release, brief torrential rain, tidal surge and avalanche-related). The flood type is determined by the Dartmouth Flood Observatory through expert judgment based on the season, location and duration of floods, and supplemented by the information collected from local institutions or media (see Fig. 1 for the geographical distribution of flood occurrence for each flood type from 1985 to 2000). These data are available from 1985 until recently. The location (region, city, river basin, etc.) in which a flood occurred is recorded in the Dartmouth Flood Observatory database by each flood type, and then geo-referenced to a certain grid by the authors, and the numbers of flood occurrence in each grid are added up to generate the global maps, as presented in Fig. 1. Incomplete records are eliminated from the data, and this gives a total of 1547 complete flood events for the study period 1985–2000.

3.3.3 ADRC Disaster Information Archive Since 1995, the Asian Disaster Reduction Center (ADRC, http://www.adrc.asia/top_aca.php) has provided the Disaster Information Archive, which collects and archives flood data primarily from newspaper sources. Detailed information, such as the causes of death and economic damage, can be attained from this database. However, there are fewer records than in the above two databases; also, it only provides the data for Asian regions.

To select appropriate flood damage data, the following criteria must be fulfilled: (a) the data are provided on a global basis; (b) the geospatial

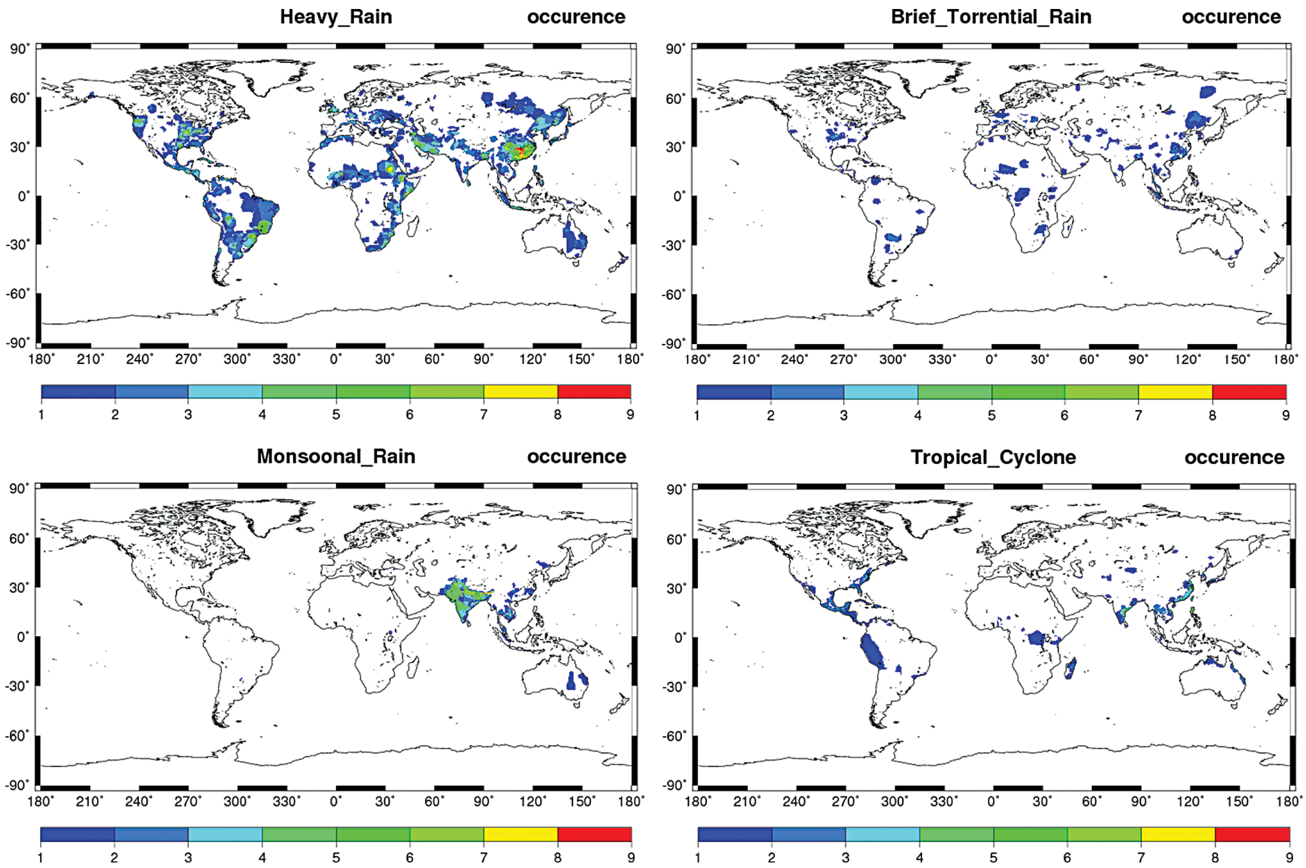

Fig. 1 Results of the chi-squared test for the measure of association between the damage parameters and hazard/vulnerability parameters. All statistics over 100 have been truncated as 100. The degree of freedom is either 16 or 20, depending on whether the null data exists or not for each parameter.

Table 2 List of selected damage, hazard and vulnerability parameters. ("000 pop" denotes a thousand of population; and "no./ 000 pop" denotes the number of dead/displaced people per thousand of population).

Damage parameters		Vulnerability parameters		Hazard parameters	
D-1-1	Fatalities – absolute (no.)	V-1	Population within 100 km of coast (000 pop.)	H-1	Standardized relative precipitation (mm)
D-1-2	Fatalities – relative (no./ 000 pop.)	V-2	Agricultural area (% of total land area)	H-2	Return period of precipitation (year)
D-2-1	Displaced – absolute (no.)	V-3	GDP per capita (USD/ capita)		
D-2-2	Displaced – relative (no./ 000 pop.)	V-4	No. of mobile telephones (no./ 000 pop.)		
D-3-1	Damage – absolute (USD)	V-5	Forest cover (% of land area)		
D-3-2	Damage – relative (% to GDP)	V-6	No. of cities (no.)		
D-4	Flooded area (ha)	V-7	No. of dams (no.)		
D-5	Affected region (km^2)	V-8	Maximum reservoir storage (000 m^3)		
		V-9	Population increase (%/ year)		
		V-10	Population density (no./ km^2)		
		V-11	Paddy area (m^2)		

gridded data are available; and (c) the types of floods are provided in order to differentiate them by their causes. After careful evaluation, the only database that fulfils all of these criteria is the Dartmouth Flood Observatory Database, based on which the AFRI will be derived in the following analysis.

Table 2 summarizes the selected flood damage, hazard and vulnerability parameters. The relative damage is calculated so that the influence of flood damage of different regions can be compared objectively. For example, the (grid-based) parameter Fatalities – relative (D-1-2) is the ratio of the number of deaths to the total population of a region, and the parameter Displaced – relative (D-2-2) is the ratio of the number of displaced people to the total population. Damage – relative (D-3-2) is the ratio of estimated economic losses to the GDP in US dollars. Each recorded flood is geo-referenced to all hazards and vulnerability parameters in order to produce a new database that includes both natural and socio-economic factors during the recorded flood event. That is, not only can one know the degree of damage for a certain flood event (e.g. the number of deaths, the degree of economic losses), the new database also provides information on the intensity of precipitation, population, geographical features, telecommunication levels, and so on, for the region and the period of flood occurrence. Figure 2 shows the results of the chi-squared test for the statistical significance between two hazard parameters (H-1 and H-2), 11 vulnerability parameters (V-1 to V-11), and eight damage parameters (D-1-1 to D-5), in which a total of 104 combinations is being tested (i.e. eight damage parameters are tested for their occurrence in relation to the total of 13 hazard and vulnerability parameters). The null hypothesis is that "hazard and vulnerability parameters are independent from damage parameters". The result of the chi-square test indicates that 65% of all 104 parameter combinations are within the acceptance region at the significance level of 5%, and, for certain combinations, even at the 1% significance level. Therefore, it can be concluded that the 13 selected hazard and vulnerability parameters have strong relationships with the eight selected damage parameters; hence these are suitable for subsequent use in the regression analysis, as presented below.

4 RESULTS

Each of the damage, hazard and vulnerability parameters is categorized into five classes so that each class has an equal number of flood events. For comparison among parameters of different units, the five classes are scored from 1 to 5, since the scale and unit of each parameter are different and also they were obtained from different sources with varying accuracy, duration and collecting methods.

The flood risk index (AFRI) is calculated by conducting a step-wise regression with a confidence interval of 95% to determine the regression coefficients for each hazard and vulnerability parameter, namely, the degree of contribution of each parameter to the variance of damage parameters.

The following formula considers the risk (damage) as the product of hazard and vulnerability parameters, and the relative influence of contributing parameters is represented by their respective exponent:

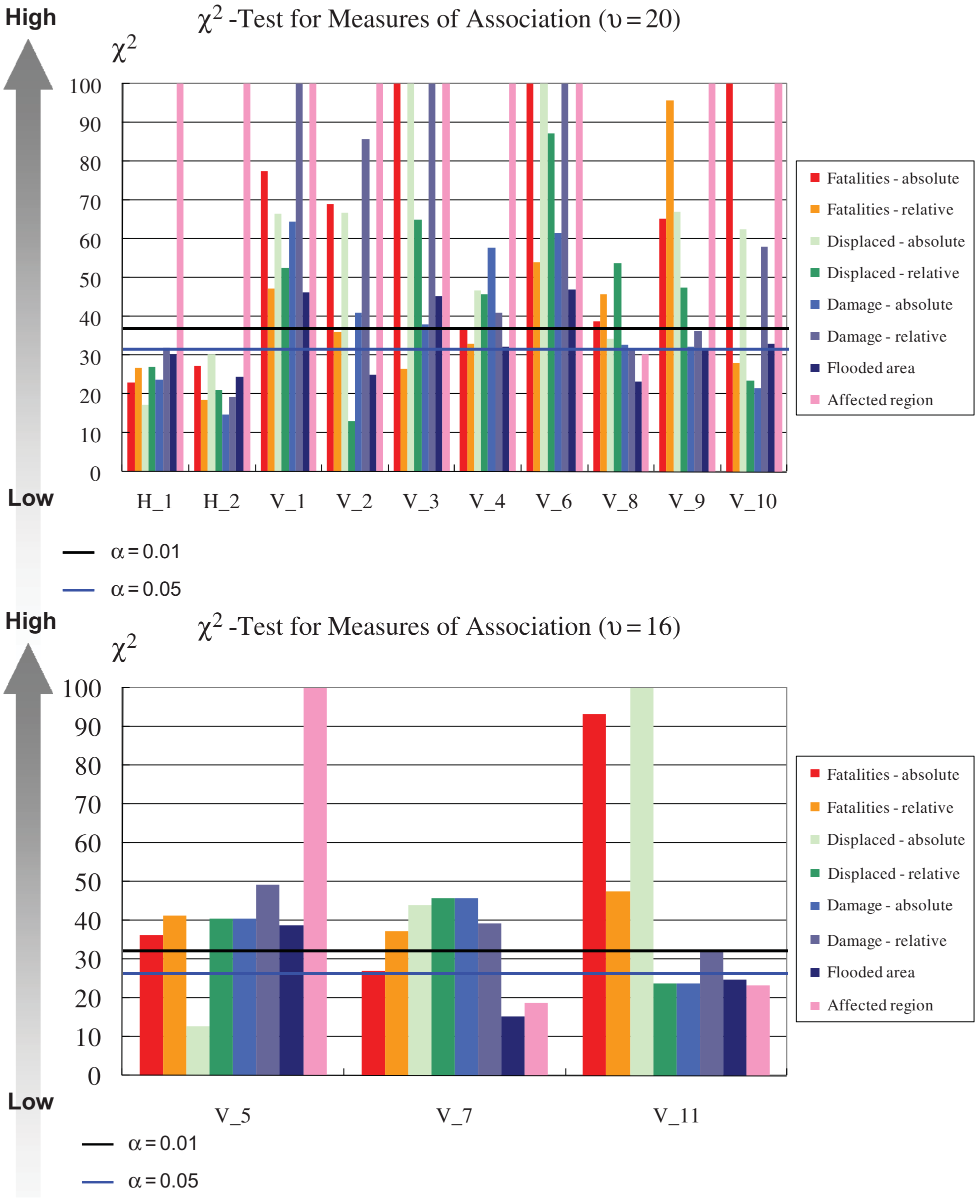

Fig. 2 Degree of flood occurrence in the target period 1985–2000 stratified by the four flood types (data from the Dartmouth Flood Observatory).

$$Damage = (H1)^a \times (H2)^b \times (V1)^c \times \ldots \times (V11)^m \times n$$

$$\ln Damage = a \ln (H1) + b \ln (H2) + c \ln (V1) + \ldots + m \ln (V11) + n \qquad (3)$$

With all parameters categorized into five classes (each class having an equal number of flood events, scored from 1 to 5), by conducting a stepwise regression to the original data set, the hazard and vulnerability parameters that best explain the variance of damage data are adopted, and this process ends when the inclusion of an additional parameter fails to increase the coefficient of determination (R^2) more than 1%. Then, each hazard and vulnerability parameter is substituted in the obtained regression equation by the assigned score of 1 to 5 (depending on the time and region of the flood occurrence), and thus the flood risk can be calculated from equation (3). Finally, the AFRI is derived by normalizing the damage obtained in equation (3) by the maximum value of damage of all grids so that all AFRI values range between 0 and 1.

The AFRI is calculated for each flood type categorized by the Dartmouth Flood Observatory, as plotted in Fig. 1. The total number of flood events recorded in the Dartmouth Flood Observatory

Database is 1547, of which the most frequently occurring four flood types are HR (heavy rain, 60%), BR (brief torrential rain, 13%), TC (tropical cyclone, 12%) and MR (monsoonal rain, 7%).

Figure 3(a) shows a plot of the calculated AFRI for the number of displaced people ((D-2-1); flood type: heavy rain). and presents a comparison of potential flood damage caused by a certain flood type calculated based on the AFRI. As may be seen in Fig. 3(a), the AFRI is high over areas such as the eastern Indian subcontinent and western China, while it is relatively low in arid and cold areas. Most flood-prone areas and areas near major rivers (such as the Mississippi, Amazon, and Yangtze rivers) have relatively high AFRI, apart from the European continent. This can be explained by the countermeasures towards flood disaster reduction undertaken in Europe historically, that affect the calculated AFRI in this region. In Fig. 3(b), the record from the Dartmouth Flood Observatory is plotted for the same damage parameter and flood type

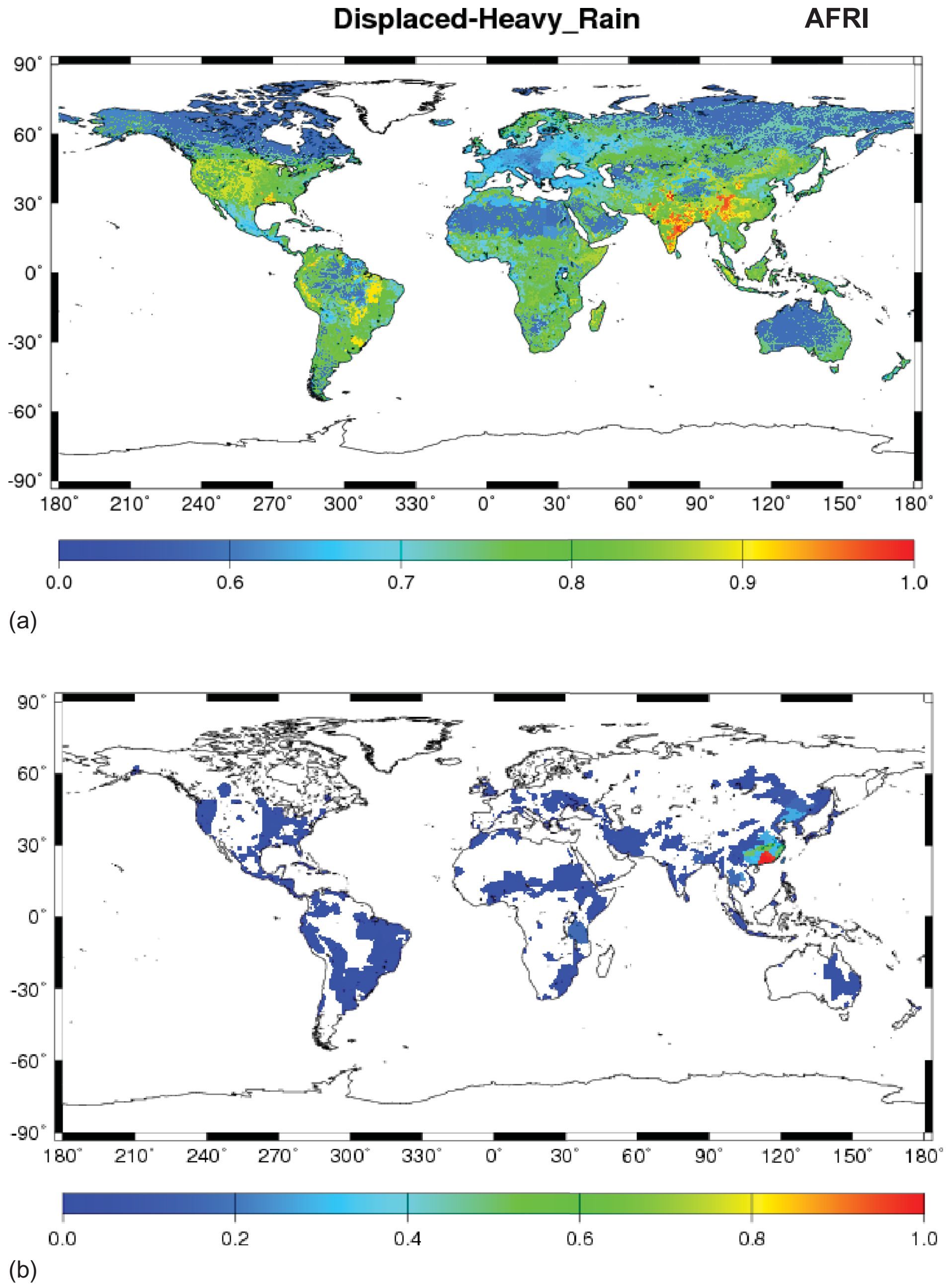

Fig. 3 (a) The global map of calculated AFRI and (b) the number of displaced people from the record of Dartmouth Flood Observatory, for the parameters Displaced number of people (D-2-1) and Heavy rain floods. For comparison, both plots have been normalized to have the maximum grid value as 1.

Table 3 Summary of the coefficients of determination (R^2) for the calculation of AFRI corresponding to different flood types and damage parameters.

	D-1-1	D-1-2	D-2-1	D-2-2	D-3-1	D-3-2	D-4	D-5
Heavy rain	0.19	0.12	0.27	0.17	0.17	0.27	0.11	0.20
Brief torrential rain	0.16	0.10	0.28	0.12	0.00	0.52	0	0.14
Monsoonal rain	0.26	0.05	0.30	0.15	0.36	0.55	0.30	0.24
Tropical cyclone	0.16	0.10	0.24	0.07	0.07	0.41	0	0.24

((D-2-1); heavy rain), and comparison with Fig. 3(a) shows that, for regions where the calculated AFRI is low, no flood events were recorded in the Dartmouth Flood Observatory database either.

As summarized in Table 3, the R^2 values for the total 32 regressed AFRI expressions corresponding to different combinations of flood type and damage parameter range from 0 to 0.55. As an example, Fig. 4 shows the scatter plots between the calculated AFRI ((D-2-1); heavy rain) and the damage data from the Dartmouth Flood Observatory; the correlation coefficient between them is 0.52 (i.e. $R^2 = 0.27$, see Table 3). Although the correlation in Fig. 4 is merely moderate, an increasing trend of the displaced number of people can be observed when AFRI exceeds 0.7, indicating that the AFRI for this specific damage parameter and flood type is more accurate in areas with high flood risks.

As mentioned earlier, in this study the risk is defined by the hazard parameters multiplied by the vulnerability parameters, and it is also equal to the expected damage from the hazard. However, the risk defined in this way still does not account for the empirical characteristics of past flood occurrence in a specific region. Thus, the concept of exposure is considered in the following. First, a parameter EX, is defined as the number of flood events that occurred in a grid during the target period. The AFRI is calculated only for areas where actual flood damage has occurred previously (if EX = 0, AFRI = 0), and it is related to the actual characteristics of floods in a region. The advantage of this approach is that all four types of flood can be considered separately according to their frequency of occurrence, which is unique for each region, and the risk (the product of hazard and vulnerability) is more in accord with the past flood events owing to the incorporation of the exposure concept.

For each grid, a percentage weight, w_i is determined according to the past occurrence of four major flood types:

$$w_i = \frac{\mathrm{EX}_i}{\max \mathrm{EX}} \tag{4}$$

and the AFRI is calculated as the weighted sum as follows:

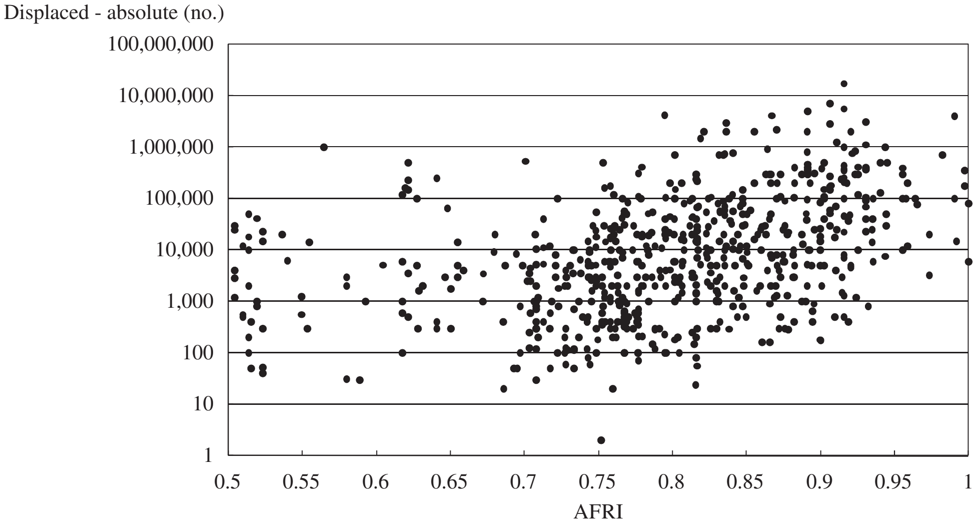

Fig. 4 Scatter plot of the calculated AFRI *vs* the number of displaced people in the Dartmouth Flood Observatory records, both for the flood events caused by heavy rain. (Note that the scale of y-axis is logarithmic).

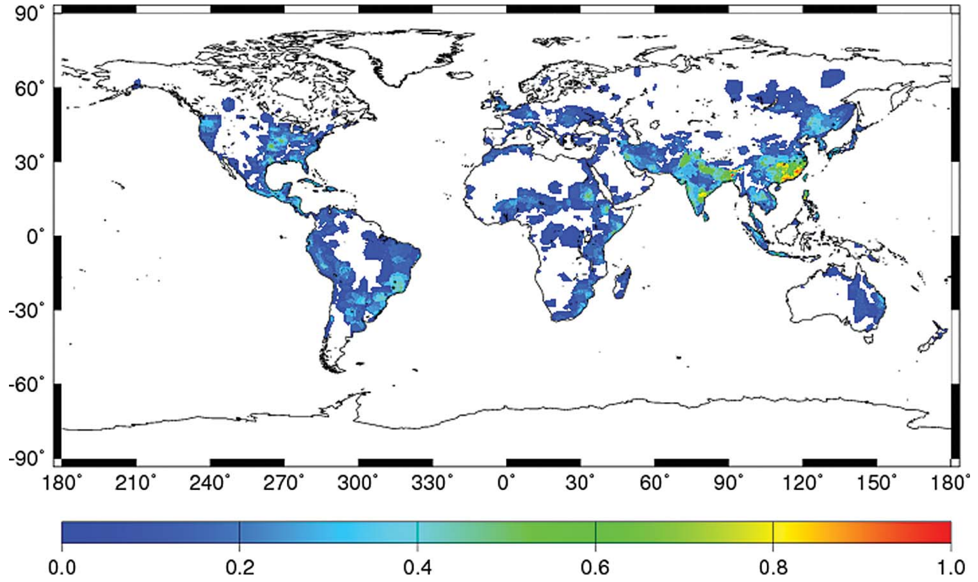

Fig. 5 AFRI and Exposure by Dartmouth Flood Observatory Records (the colours indicate the AFRI of Fig. 3 only for areas where actual flood events took place).

$$\mathrm{AFRI} = w_{\mathrm{HR}}\mathrm{AFRI}_{\mathrm{HR}} + w_{\mathrm{BR}}\mathrm{AFRI}_{\mathrm{BR}} + w_{\mathrm{MR}}\mathrm{AFRI}_{\mathrm{MR}} + w_{\mathrm{TC}}\mathrm{AFRI}_{\mathrm{TC}} \quad (5)$$

The global distribution of the AFRI derived from equation (5) is plotted in Fig. 5 and is discussed in the next section.

5 DISCUSSION

One of the most notable advantages of the developed AFRI index is that the factors which contribute to floods can be analysed individually. Table 4 summarizes whether each hazard and vulnerability parameter indicates an increase (shown with arrow pointing up) or decrease (shown with arrow pointing down) in flood damage. From this table, the GDP per capita (V-3) for floods caused by monsoonal rain and tropical cyclones, and number of cities (V-6) for all types of floods are the parameters that lead to reduced flood damage, as well as the population increase (V-9) for the floods caused by heavy rain. This suggests that, overall, a higher density of population and assets could lead to reduced flood damage, which is consistent with the finding in the study by UNDP (2004). One reason could be that, in regions of higher population density with most assets concentrated, disaster mitigation measures are likely to be implemented more effectively prior to disaster occurrence. This is in line with the concept of compact city development – infrastructure investment is concentrated and cost-effective. However, population increase (V-9) intensifies the damage in the case of monsoon floods, with the most prominent regions being the Indian subcontinent where the population has been growing rapidly and massive flood damage has often occurred (as shown in Fig. 3).

Information technology factors (e.g. the number of mobile phones per population) also play a significant role in the interpretation of flood risk. Previous studies (e.g. WB 2005, UNDP 2004) have shown the relationship between flood damage and socio-economic factors, such as population or GDP, but it is worthwhile to note that our present study indicates that information technology has a close relation with flood damage, resulting in decreased flood damage for floods caused by heavy rain. Note that this factor is not proportional to GDP, as the inter-correlation between various vulnerability factors has been screened out prior to the calculation of the AFRI.

Another interesting finding is the relationship between flood damage and paddy area (V-11). While this factor shows a positive relation with damage parameters, such as the displaced and affected regions for floods caused by heavy rain and brief torrential rain, a negative relation is found with the economic damage factor (i.e. damage USD) for floods caused by monsoonal rain. A reasonable explanation is that paddy areas in general function as a reservoir that is more likely to be inundated. However, the resulting economic damage is relatively small since paddy areas are seldom located in urbanized regions.

Table 4 The contribution of hazard and vulnerability parameters to flood damage for each of four major flood types. (see Table 2 for description of parameters: ↑ indicates an increase, ↓ indicates a decrease).

	D-1-1	D-1-2	D-2-1	D-2-2	D-3-1	D-3-2	D-4	D-5
Heavy rain:								
H-1								↑
H-2								
V-1	↑		↑			↓		↑
V-2								
V-3	↑							
V-4	↓		↓	↓	↓	↓	↓	↓
V-5			↓				↓	↓
V-6	↓	↓	↓	↓	↑	↓		↓
V-7								
V-8								
V-9		↓		↓				↓
V-10		↑	↓				↓	↓
V-11			↑	↑				↑
Brief torrential rain:								
H-1								
H-2								↑
V-1						↓		
V-2								
V-3								
V-4								
V-5								↓
V-6	↓		↓	↓		↓		↓
V-7								
V-8								
V-9	↑	↓						
V-10								
V-11			↑					
Monsoonal rain:								
H-1								↑
H-2								
V-1						↓		
V-2								
V-3	↓				↓	↓		↓
V-4						↓		
V-5			↑	↑				
V-6			↓					
V-7								
V-8								
V-9		↓	↑			↑	↑	
V-10								↓
V-11					↓	↓		
Tropical cyclone:								
H-1								
H-2								
V-1		↓			↑	↓		
V-2								
V-3						↓		↓
V-4			↑					
V-5								
V-6	↓		↓	↓				
V-7		↓						
V-8								
V-9								↓
V-10			↑					
V-11								

The potential flood risk around the globe, as shown in Fig. 3, is calculated without considering the actual occurrence of flood events. In Fig. 5, the AFRI (displaced number of people: D-2-1) plotted against damage with consideration of exposure (i.e. regions where floods have occurred in the target period of 1985–2000). Notice that Fig. 5 shows only the areas where actual flood events took place during 1985–2000; for those regions not assigned an AFRI value in Fig. 5, this does not necessarily mean that the possibility of flood occurrence is zero. It should be emphasized that the proposed AFRI does not represent the likelihood of flood occurrence, but rather the *expected damage from floods* conditioned upon *the past flood event occurrence*. It should also be emphasized that the global pattern shown in Fig. 5 is rather consistent with that derived from the Dartmouth Flood Observatory Database (Fig. 3(b)), whereas this is not the case between the two maps in Fig. 3, suggesting the significance of the concept of “exposure” in characterizing flood risk.

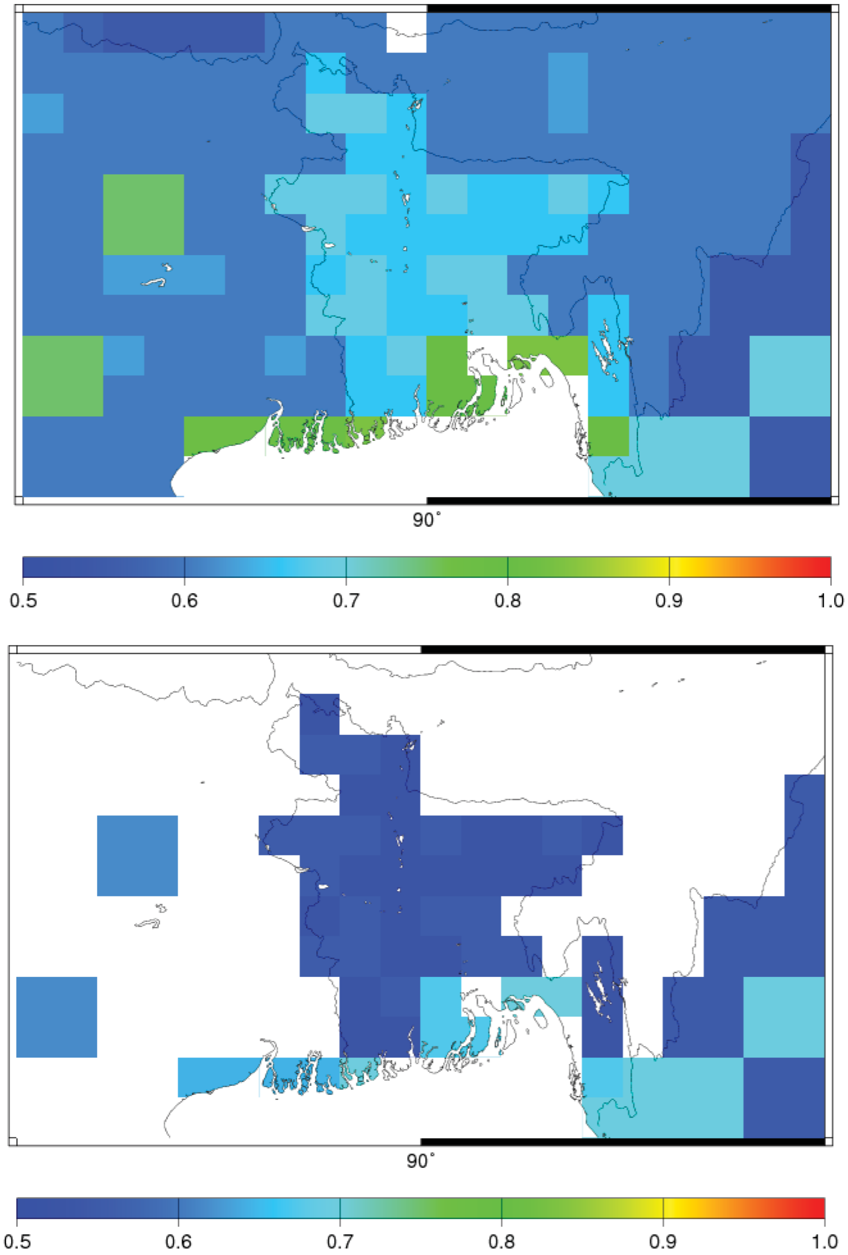

Fig. 6 Comparison of relative damage (for monsoon rain floods) in Bangladesh. The upper part shows the relative damage in a flood caused by monsoon rain given the current GDP level in Bangladesh, while the lower part shows the case if the GDP level in Bangladesh were to rise to the same level as current Japan.

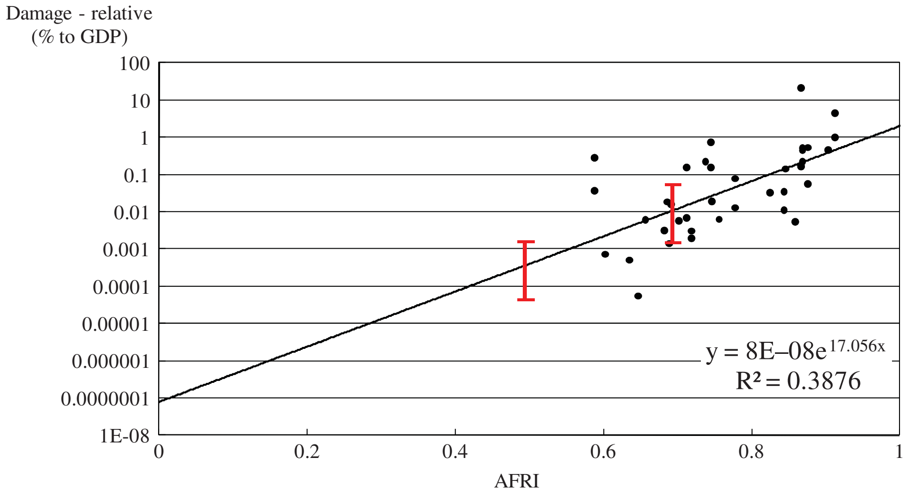

Fig. 7 Interpretation of AFRI to relative damage (for monsoon rain floods) in Bangladesh. The *x*-axis indicates AFRI, while the *y*-axis indicates the relative monetary damage from floods, i.e. the percentage of economic damage in US dollars to the GDP of the country. In the test case shown, AFRI decreases from 0.7 to approximately 0.5, which means that the expected relative monetary damage decreases from 0.01% to approximately 0.0005% of Bangladesh's total GDP.

Regarding the hazard parameters, our results indicate their limited contribution to flood damage relative to the vulnerability parameters (except for that of affected regions: D-5). This is consistent with the result of Yoshimura *et al.* (2008), who demonstrated the relationship between hazard and damage by using affected region as an indicator to measure the extent of damage from floods. However, the fact that the relationship between damage and hazard is only seen for the affected regions could be a result of the hazard data used (i.e. global precipitation in this study). It is commonly recognized that most of the currently available global precipitation data sets contain large uncertainty, in particular in their accuracy of representing extreme rainfall events. Similar analyses related to the risk of hazards as presented in this study call for further improvement of global precipitation data set. Moreover, the hazards can be better calculated by using river discharge data, provided that an accurate global river discharge data set will be available in the future.

Figure 6 plots the AFRI in Bangladesh (D-3-2; monsoonal rain). Note that Bangladesh is selected here over India as a demonstration example, because the unit for relative damage is based on the country GDP, which is too diverse in India. This damage and flood type combination was selected for conducting the test case because the R^2 of AFRI calculation shown in Table 3 is the highest among all 32 combinations. The changed parameter for the example was GDP per capita (V-3). If the current GDP per capita in Bangladesh were to rise to the same level as that in current Japan, the AFRI would decrease from an average of 0.7 to 0.5, as shown in Fig. 7. The estimated current economic damage in Bangladesh, represented by an AFRI of 0.7, is interpreted to be around 0.01% of the current GDP in Bangladesh. Therefore, if the GDP per capita in Bangladesh were to rise to the same level as in current Japan, it can be estimated that the economic damage from monsoonal rain floods would be approximately 0.0005% of the future GDP in Bangladesh (see Fig. 7).

Finally, although the number of samples used in this study is relatively small, it should be emphasized that global flood damage data, collected and compiled in a standardized format, have become accessible only in recent years. Therefore, the results presented here can be considered as preliminary, and represent an initial effort toward more in-depth analysis and calculation of potential flood risks in the near future.

6 CONCLUSIONS

A new global flood risk index, the AFRI, based on both natural and socio-economic factors was developed in this study. The AFRI is an event-based index that indicates the expected damage from a single flood event, and, in contrast to previous studies, it targets floods by the unit of events instead of the long-term statistical trend. Moreover, the AFRI can express the relative potential flood risk, namely the relative degree of expected damage, with the flood damage occurrence being considered such that it enables comparison among different regions and periods. Although it is not interpreted by the absolute amount of damage, it is sufficient to be used as a simulation tool for policy making in urban planning and land-use policies. For example, it may be used in predicting the effects of population density or paddy area change on flood damage, or estimating the effect of dam construction on the flood damage. In addition, the research presented herein can be extended to investigate the change in flood risk due to global warming (Hirabayashi *et al.*, 2008).

The uniqueness of the new Advanced Flood Risk Index is that not only the hazard parameters that directly influence flood occurrence, but also the vulnerability parameters that reflect the socio-economic characteristics of a region, can be quantitatively represented in the evaluation of flood risk. Moreover, the AFRI can also be applied as an objective tool for assessing flood adaptation policies. For example, the change in expected flood damage due to alteration in land use can be predicted by the AFRI, and subsequently the results can be considered as guidelines for future urban planning. Another example is for policy makers to foresee the relationships between socio-economic change (e.g. population and economic growth) and flood damage, which will benefit the estimation of expected damage from future floods and also allow the evaluation between potential economic losses and the needed investments for the reduction of losses. This in turn can lead to more accurate cost–benefit analyses and more appropriate budget allocation.

Acknowledgements This work was supported by the Global Environment Research Fund (GERF: S5) from the Ministry of Environment, JSPS KAKENHI, Grants-in-Aid for Scientific Research (S)(19106008), and Innovative Program of Climate Change Projection for the 21st Century from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

REFERENCES

Davidson, R., 1997. An Urban Earthquake Disaster Risk Index. The John A. Blume Earthquake Engineering Center. Report no. 121.

Hara, Y., Umemura K., Kato K., Connor R. and Sato, Y., 2009. The Development of Flood Vulnerability Index applied to 114 major river basins around the world. *Journal of the Japan Socciety Hydrology and Water Resources*, 22 (1), 10–23.

Hirabayashi, Y., Kanae, S., Emori, S., Oki, T. and Kimoto, M., 2008. Global projections of changing risks of floods and droughts in a changing climate. *Hydrological Sciences Journal*, 53 (4), 754–772.

ICHARM (International Centre for Water Hazard and Risk Management), 2005. *Technical Report on Global Trends of Water-related Disasters*. Tokyo: ICHARM.

ICHARM (International Centre for Water Hazard and Risk Management), 2007. *The Establishment of Risk Maps of Global Flood*. Tokyo: ICHARM.

Munich Re Group, 2004. *Megacities – Megarisks Trends and challenges for insurance and risk management*.

Ngo-Duc, T., Polcher, J. and Laval, K., 2005. A 53-year forcing data set for land surface models. *Journal Geophysica Research*, 110, D06116, doi:10.1029/2004JD005434.

UNDP (United Nations Development Programme), 2004. *Reducing Disaster Risk*. Available from: www.undp.org/cpr/whats_new/rdr_english.pdf.

Villagran de León, J. C., Bogardi, J., Dannenmann, S. and Basher, R., 2006. Early Warning Systems in the context of Disaster Risk Management. *Entwicklung & Ländlicher Raum*. 2, 23–25.

Wisner, B., Blaikie, P., Cannon, T. and Davis, I., 2004. *At Risk: Natural Hazards, People's Vulnerability, and Disasters*. Oxford: Routledge.

World Bank, 2005. *Natural Disaster Hotspots: A Global Risk Analysis*. Washington DC: World Bank Publications.

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