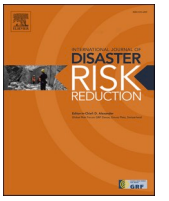
International Journal of Disaster Risk Reduction 72 (2022) 102835

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr



A statistical comparison of flood-related economic damage in Indian states with reflections on policy implications

Sourav Das [a], Arshay Nimish Sheth [b], Priya Bansal [c], Joon Chuah [d], Robert Wasson [a,e,*]

[a] College of Science and Engineering, James Cook University, McGregor Road, Smithfiled, Queensland, 4878, Australia
[b] Mathematics Institute, Zeeman Building, University of Warwick, Coventry, CV4 7AL, UK
[c] Economics Division, Competition Commission of India, 9th Floor, Office Block-1, Opposite Ring Road, East Kidwai Nagar, New Delhi, 110023, India
[d] Tembusu Residential College, National University of Singapore, 28 College Avenue E, #B1-01, 138598, Singapore
[e] Fenner School of Environment and Society, Australian National University, Canberra, ACT, 0200, Australia

ARTICLE INFO

Keywords:
Floods
Power law
Principal components analysis

ABSTRACT

Economic damage caused by floods in India is a serious problem that has disrupted development and the fight against poverty in some parts of the country. It is therefore important to mitigate the effects of floods as effectively as possible. An analysis of aggregated economic damage in the categories of crops, housing and utilities in six Indian States from 1953 to 2011 is presented using rigorous statistical methods. The main result is that increasing mitigation efforts since Independence have not produced a monotically decreasing long term trend in damage, although the magnitude of reduction cannot be known precisely, based on publicly available data. Andhra Pradesh has the highest total damage followed by Bihar, West Bengal, Orissa, Assam, and Gujarat. The reasons for this ranking are unclear because the only publicly available proxy for potential damage, has a weak statistical effect. Similarities in damage between some States suggest that inter-State learning may be of value despite differences in hydrology, landscape and economic activity. Despite the deficiencies in the data and its spatial resolution, there is sufficient evidence to inspire new thinking and action about flood mitigation in India, for which District and Taluk level data will be essential.

1. Introduction

The economic risk and social vulnerability to riverine floods in India are among the highest in the world, with millions of people exposed, and billions of rupees (Indian currency) worth of property and infrastructure at stake [1–3]. Between 1953 and 2011, data from the [1–4] indicates that the total economic cost of floods in India was over US$69 billion (at 2017 prices). Recurring damage by floods does not only produce immediate costs; it may also be setting back the development of the country [5,6] In addition, recurring flood disasters can produce and maintain poverty traps [7] and force outmigration [8]. A recent call for a new approach to an effective flood policy in India by Ref. [9] was a response to recent and historical flooding.

Flood mitigation in India has mainly been by building embankments along rivers [10] since the devastating floods of 1954 [11]

* Corresponding author. College of Science and Engineering, James Cook University, McGregor Road, Smithfiled, Queensland, 4878, Australia.
E-mail addresses: Sourav.das@jcu.edu.au (S. Das), arshay.sheth@warwick.ac.uk (A.N. Sheth), bansalpriya1994@gmail.com (P. Bansal), joonchuah@gmail.com (J. Chuah), Wasson.robertj@gmail.com, Robert.wasson@jcu.edu.au (R. Wasson).

https://doi.org/10.1016/j.ijdrr.2022.102835
Received 6 November 2020; Received in revised form 20 January 2022; Accepted 31 January 2022
Available online 16 February 2022
2212-4209/© 2022 Elsevier Ltd. All rights reserved.

although drainage improvement on floodplains has also been attempted and there are warning systems, flood detention basins and at least mention of floodplain zoning to reduce human vulnerability to floods ([8,12]. It is also claimed that reservoirs provide flood protection ([13,14] although this is contested [e.g.,15]. Embankments can be considered a form of Disaster Risk Reduction (DRR) which is a systematic assessment of risk and ways of reducing it. In the case considered here, embankments are 'flood control' measures in contrast to evacuation, relief payments, provision of food and other amenities that occur during or after a flood [16]. point to the importance of assessing flood-related economic damage to assess the effectiveness of the 'flood control approach' that is manifest in embankments, in contrast with a 'flood risk management' approach that takes a wider view of mitigation.

Ideally such an assessment should be catchment-specific and would take into account land use, hydrology, geomorphology and policy interventions including embankments. Unfortunately, such an approach is not possible for most of the country and the only publicly available data for economic damages from floods comes from the Central Water Commission aggregated at the State level from data collected in lower level administrative units. While not ideal, it is nonetheless important to extract as much information from these data as possible.

There have been several analyses of the CWC data, two of which are now briefly descibed [14]. posited the simple hypothesis that as

Fig. 1. This map of India is an open source image of the political boundaries of Republic of India, courtesy https://www.mapsofindia.com/. It shows the federal states, Assam, Andhra Pradesh (AP, in this paper), Bihar, Gujarat, Orissa (Odhisha, new name), and West Bengal (WB, in this paper).

flood protection increases, economic damage should decrease, although they used the number of people affected by floods as the response variable in a simple regression analysis. Also, they used the area protected by flood mitigation strategies as the independent variable, thereby aggregating all forms of protection, and their detailed analysis focused only on the states of Uttar Pradesh, Bihar and West Bengal for the relatively short period of 1971–1996. They concluded that the area affected by floods is an insignificant variable in accounting for the number of people affected by floods and therefore, vulnerability to floods has not decreased despite increasing allocation of funds to protection. Such a simple analysis is unlikely to account for variations in the maintenance and failure of embankments, hydrologic and river channel changes, and the ways in which people's behaviour and settlement patterns respond to the construction of embankments. An analysis by Ref. [10] who used aggregated data for all of India from 1953 to 2016 and also concluded that, despite enormous effort at protection, flood-related economic damage has increased. These and other studies have not extracted the greatest amount of information from the CWC data, because of inadequate statistical methods, in some cases use of less than all of the available data series, and use of either India-wide data or data for only a few States.

The approach adopted here is to use the longest records available for as many States as possible, and apply a range of rigorous statistical methods to maximize the amount and quality of information. In the interests of providing information for better flood mitigation, and considering the possible future stresses that climate change may bring [17] it is essential that differences between States are highlighted because of the spatial differences (e.g. culture, policies, practices) for a country as large as India. Comparisons between States may be of value to design spatially nuanced mitigation and to provide a case for more central government funding to the worst affected states. Keeping in mind what is possible from the available data (see below), the following questions are posed:

Research Questions.

1. Are there monotonic (increasing or decreasing) trends in the damage data by State, given that according to Ref. [14] there should be a declining trend as mitigation strategies are affected?
2. Which States have the highest damages and why?
3. Which States are most similar in terms of damage?
4. What is the best probability distribution for the data and what information do the properties of the best distribution provide?
5. What are the lessons for policy makers from this analysis?

To help with geographic context in Fig. 1 we show the political map of India that includes the six Indian federal states which are investigated in this manuscript.

A key contribution of the paper is that the analysis did not find any monotonic trend in the macro-economic flood damages – increasing and certainly not decreasing – for any of the states, though we observe that the time series display clustering in time but also by states, despite the presence of several missing values (with unknown reasons) and poorly known methods of sampling. Statements more nuanced than this would require granular data at lower levels of administration - Districts and Taluks.

The rest of the paper is organized as follows. A brief description of all econometric and statistical methods are given in Section 2. Section 2 is largely self-contained but for readers interested in more details we provide additional references. Results and discussions, including relevant plots (Figs. 2–8)and tables are presented in Section 3. We give concluding remarks in Section 4, including limitations.

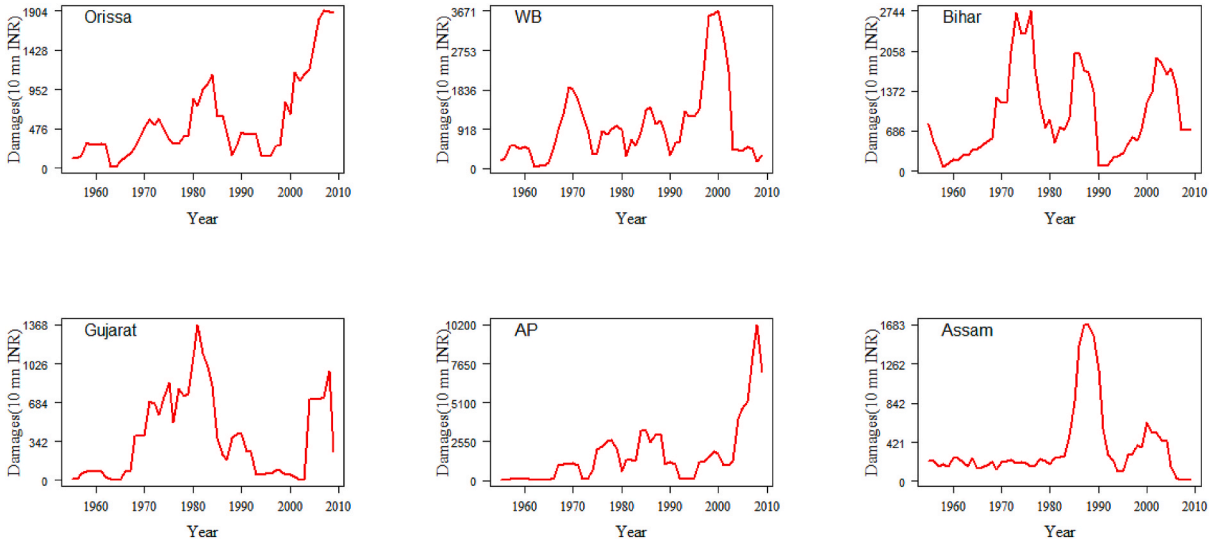


Fig. 2. Panel plots show five year moving averages for total crop damages from 1953 to 2011 in the Indian states of Andhra Pradesh (AP), Assam, Bihar, Gujarat, Orissa and West Bengal (WB). Damage estimates are given in 10 mn Indian Rupees (INR) in 2011 prices. Gaps in the plots highlight missing observations.

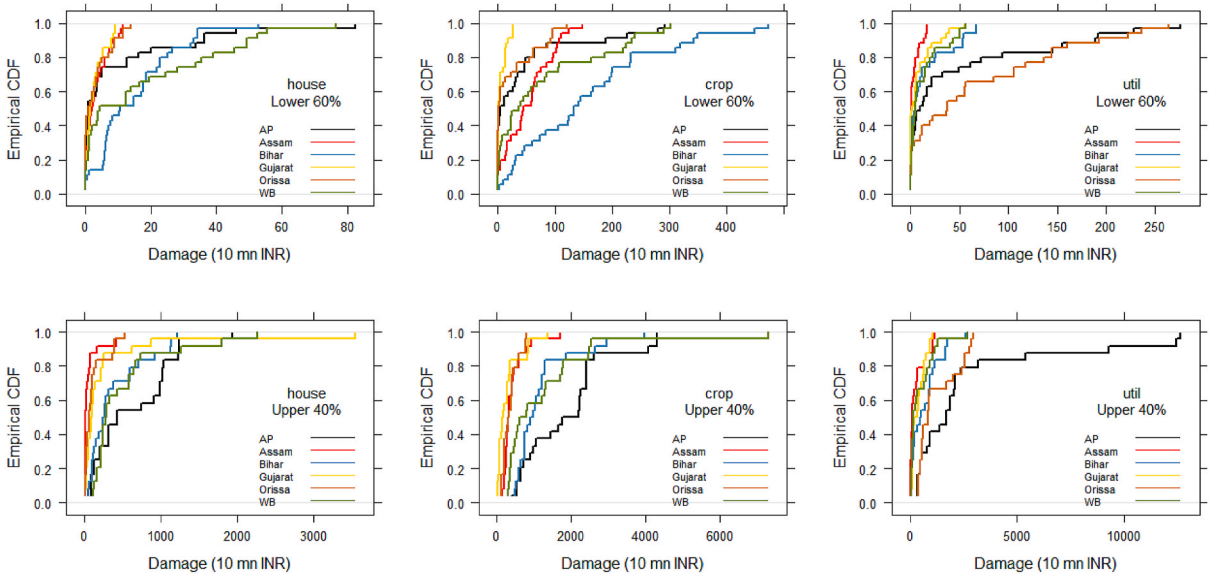


Fig. 3. The panels compare the empirical distribution function component damages, by states, as specified in the legends. Top row shows the lower sixty percentile while the bottom row shows the upper forty percentile.

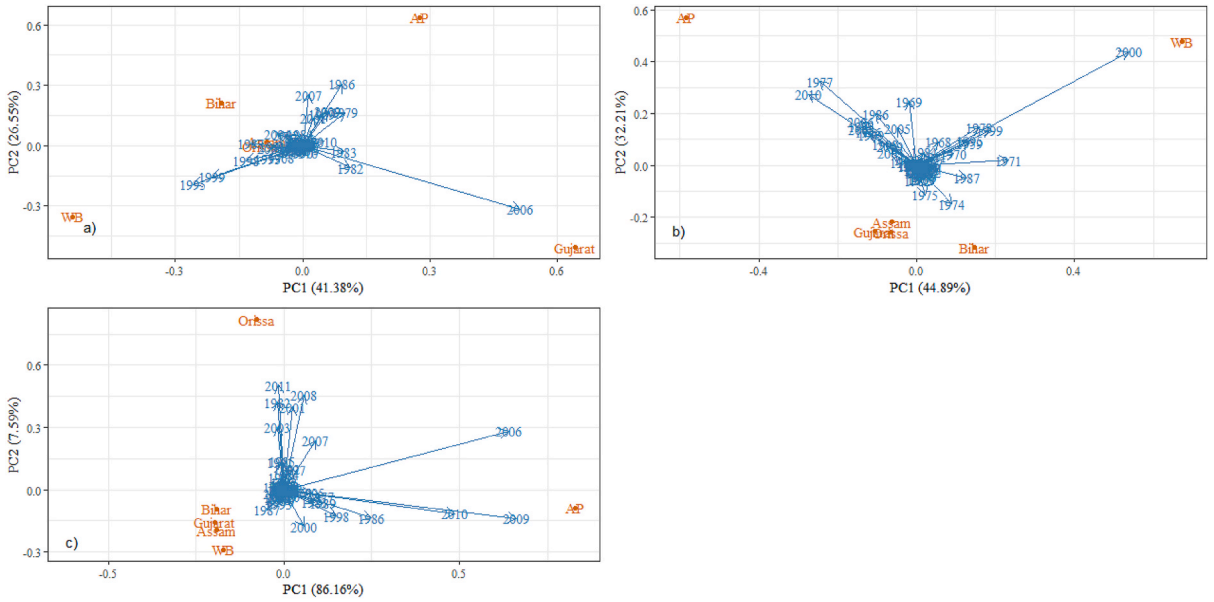


Fig. 4. The panels in the figure show the biplot of first and second principal components of the damages to housing (a), crops (b) and utility (c), respectively, for six Indian States. The loading vectors depict the most influential years in terms of variance. The biplot also reveals clustering of states based the first and second principal component features.

2. Materials and methods

2.1. Economic damage data

Annual data from 1953 to 2011 are available from the Central Water Commission (CWC: http://cwc.gov.in/sites/default/files/statewiseflooddatadamagestatistics.pdf)) for damage to crops, houses and utilities in Indian National Rupees (INR) by State. These data are collected in villages and then aggregated to District and State. They appear to be collected using the same method but it has not been possible to validate this and there has been no analysis of their uncertainties by for example replication of data in the same villages. We initially obtained annual time series on damages for ten Indian states. However, due to the presence of a large number of missing observations – mostly with unknown causes - we have analysed data on six states that had approximately 80% observations available on average (across components). These states are Assam, Andhra Pradesh, Bihar, Gujarat, Orissa, and West Bengal.

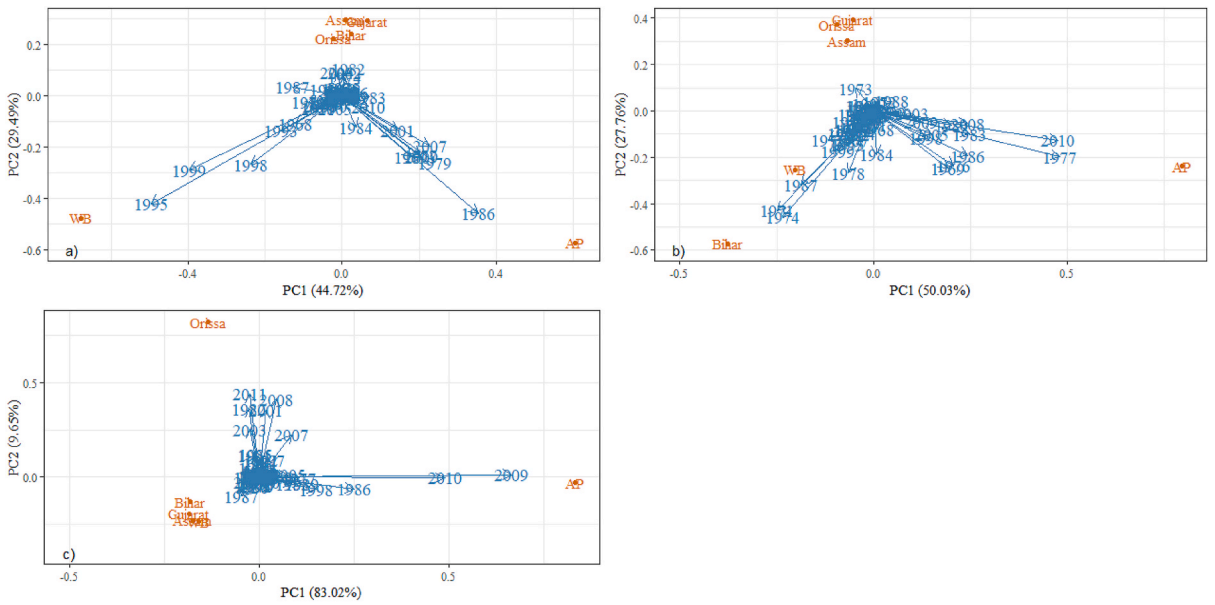


Fig. 5. The panels show the biplot of first and second principal components of the damages to housing (a), crops (b) and utility (c), respectively, after removing the effects of the years 2000 and 2006.

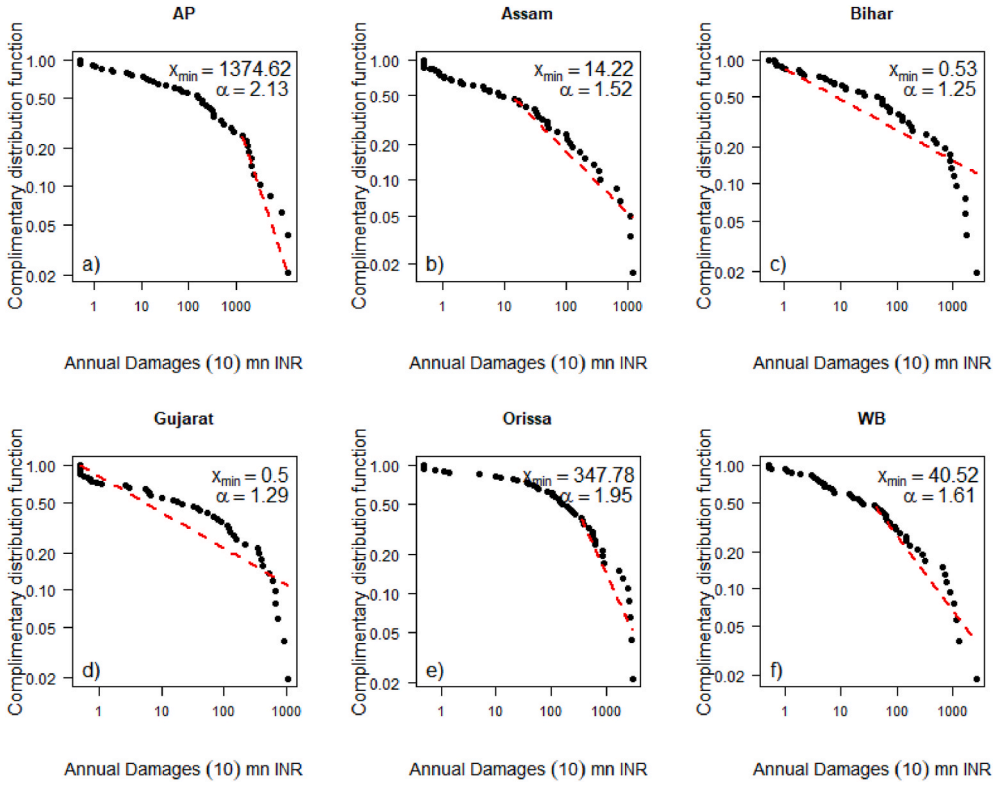


Fig. 6. Panels in the figure depict empirical complimentary distribution function of **utility damage** with estimates of model parameters, $\widehat{\alpha}$ and x_{min} in inset, for six Indian States. The slope of the red dashed line is $\widehat{\alpha}$ and the x axis intercept of the line is x_{min}.

Proportion of missing observations are given in supplementary Table 5. The state of Uttarakhand was carved out of the Uttar Pradesh (UP) introducing issues regarding unreported inconsistencies for the data on these two states, that need further investigation. This includes lack of distinction between missing values and zeros. Hence, we have removed these two states from our analyses, despite

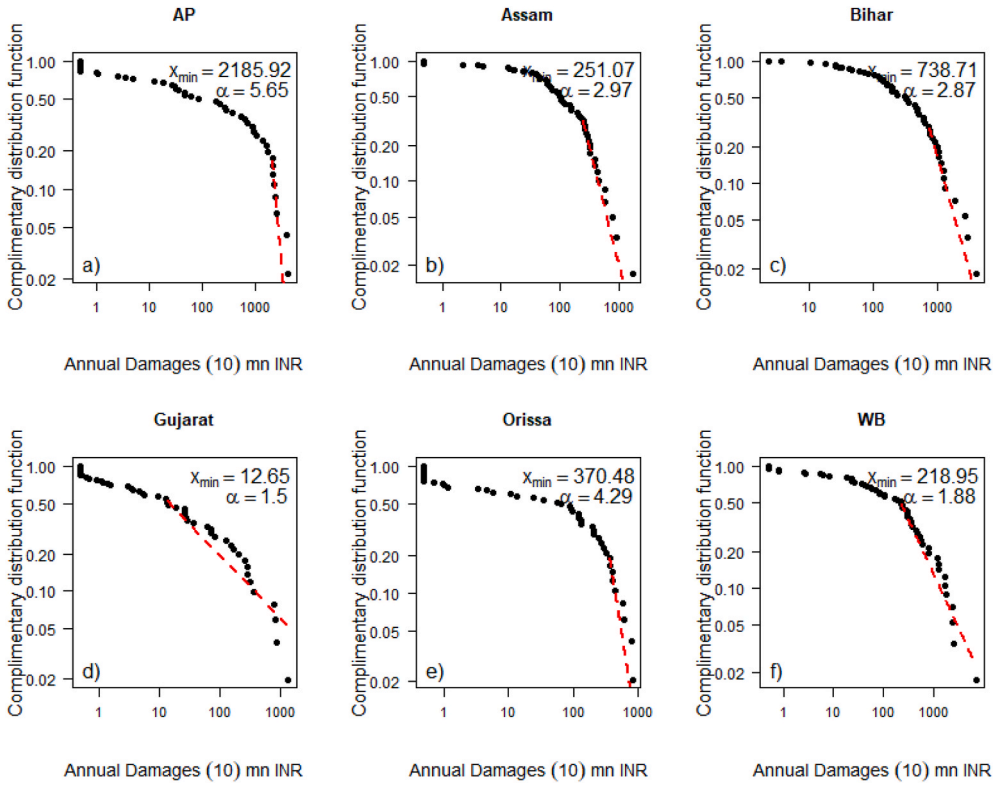


Fig. 7. Power law empirical complimentary distribution functions for **crop damage** with estimates of the model parameters $\hat{\alpha}$ and x_{min} in the insets, for six Indian States. The slope of the red dashed line is $\hat{\alpha}$ and the x axis intercept of the line is x_{min}.

reporting much smaller missing proportions.

2.2. Normalization

The data were normalized [see [18] for a discussion] for a discussion of the need for normalization), to remove the effect of increasing wealth and therefore damage potential, in the following way.

It would also be useful to normalize further by using population [19] but for most of the record it is not possible to determine the numbers of people affected by floods year by year. While the flood-related damage data are clearly from areas affected by floods, publicly available population data for the entire period are for whole states.

2.3. A measure of damage potential

To determine why some States have higher damage than others (Research question #2) it has been hypothesised that higher economic activity will result in more damage, although this will be modulated by changes in hydrology. The testing of this hypothesis is limited by the lack of publicly available catchment-specific data. The best that can be done is to use Gross State Domestic Product (GSDP) for the six States for the period 1980–2012, retrieved from the website of the Ministry of Statistics and Programme Implementation (http://mospi.nic.in/data). These data are in current prices and not constant prices The annual time series of damage data have been normalized using Gross National Income (GNI) deflators. The data on Gross National Income (GNI) for the period 1980–2019 was retrieved from the GOI (Govt. Of India) India Budget website (https://www.indiabudget.gov.in/economicsurvey/doc/Statistical-Appendix-in-English.pdf). The GNI estimates are in two categories: Nominal GNI and Real GNI. The Real GNI accounts for inflation and is measured in constant prices Nominal GNI is measured in current prices. The GNI deflator is given by GNI Deflator = (Nominal GNI*100)/Real GNI and has been calculated for each year from 1980–81 to 2018–19. The GSDP has been normalized in terms of 2018–19 prices. Therefore, to calculate the multiplication factor, the GNP deflator for the year 2018–19 (i.e. the normalization year) is divided by the GNP deflator for each year from 1980–81 to 2011–12. Then the annual GSDP for a State is adjusted by using each year's the multiplication factor. This exercise is carried out for each of the six States to obtain the normalized GSDP values as per 2018–19 prices.

2.4. Statistical methods

We now provide a brief description of the statistical methodology used to investigate the research questions set out in the Introduction. These methods are Power law distribution and it estimation, empirical distribution function and the principal components analysis (PCA).

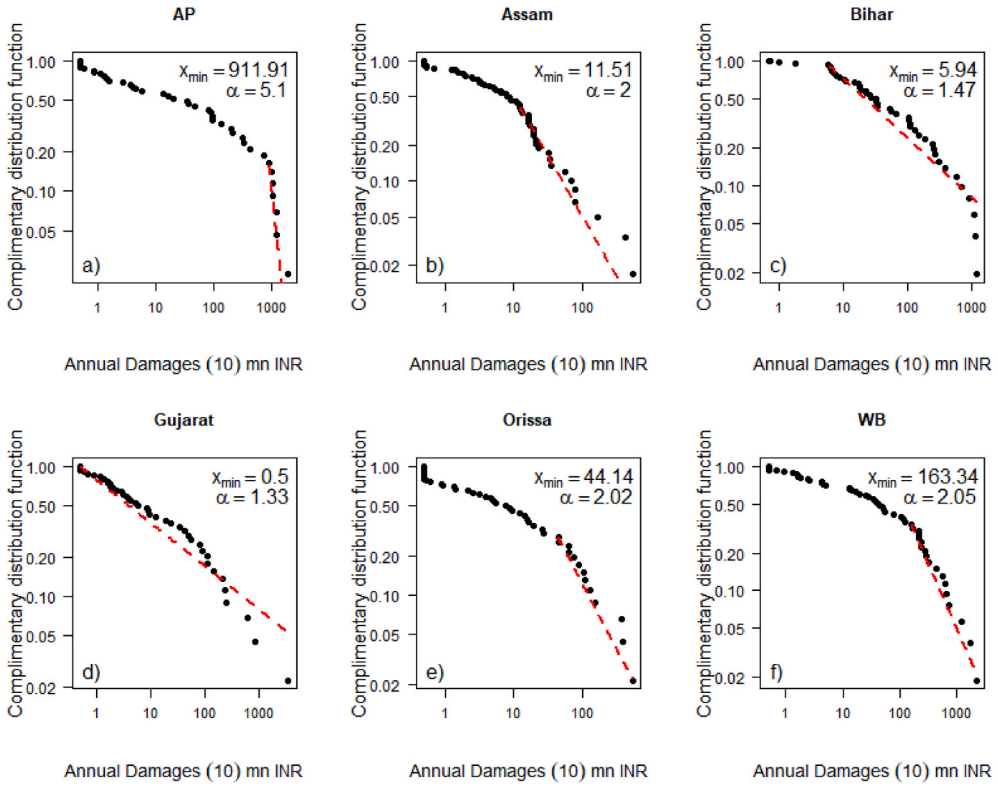


Fig. 8. Power law empirical complimentary distribution functions for **housing damage** with estimates of model parameters, $\widehat{\alpha}$ and x_{min} in the inset, for six Indian States. The slope of the red dashed line is $\widehat{\alpha}$ and the x axis intercept of the line is x_{min}.

2.4.1. Power law distribution

Power laws distributions have been widely applied to explain the statistical variation of various extreme natural events. We investigated if the flood related damages can be modelled using this distribution as layed out in Research Question 4. This is also known as parametric modelling. A satisfactory paramateric model can be projected into the future or past for forecasting or hind-casting. In this section we provide a self-contained introduction to this statistical distribution. In the below we use the phrase random process (X), interchangeably, with flood damages. A random process (X) is said follow a power law distribution if it has the following probability density function:

$$p_X(x) = C\,x^{-\alpha},\ x > 0,\ \alpha > 0 \tag{1}$$

where α is the scale parameter and C is the constant such that $\int_0^\infty p_X(x)\,dx = 1$. Equation (1) also shows that α can be interpreted as the gradient of a distribution on a log-linear scale. However, the full range of observations rarely displays a monotonically decreasing linear trajectory. Instead a threshold x_{min} is estimated such that

$$p_X(x) = C\,x^{-\alpha},\ x > x_{min},\ \alpha > 0 \tag{2}$$

The estimation procedure has the following two steps-

Table 1
Descriptive statistics of total annual damage caused by floods in six Indian States for 1953–2011. Total damage is a composite of housing damage, crop damage and utility damage.

Descriptive statistics						
Total damage	Orissa	WB	Bihar	Gujarat	AP	Assam
Min.	0.00	0.00	0.00	0.00	0.00	0.00
1st Qu.	0.67	41.35	132.40	1.01	0.53	63.65
Median	235.26	279.12	326.00	23.87	77.41	153.04
Mean	604.84	949.98	954.90	361.14	1807.16	352.08
3rd Qu.	751.94	1209.47	1074.60	314.83	1779.59	361.63
Max.	3961.87	10651.38	6374.10	3544.58	15593.24	3394.84

Based on the observed data, use either least squares or the Kolmogorov-Smirnov statistic to estimate x_{min}.

Implement the principle of maximum [20,21]. To estimate α and obtain the standard error of estimates and goodness-of-fit statistics.

Note that often power laws are of interest when $\alpha > 2$. Power laws with $\alpha < 2$ are extremely heavy tailed distributions for which no statistical moments exist . Thus for $\alpha < 2$ one can not obtain probability estimates of long term mean effects or its variance. Also reliable likelihood estimates of α is intertwined with estimation of x_{min}. This makes estimates of $\alpha < 2$ particularly prone to systematic irregularities in sampling of the data, that affects the current study. These include including missing observations (and imputation). We have demonstrated this further using parametric bootstrap (see Supplementary). Thus in this paper we have focused our discussion of power law for $\alpha > 2$, while also reporting estimates with $\alpha < 2$. The maximum likelihood estimates of parameters for the component damages are given in Tables 2 and 3.

For Research Question # 2 a comparison is made between the six States. For each damage component – crop, housing and utility - the minimum value of the bootstrap mean estimate of x_{min} (across all States) is used as a baseline threshold x_{min} and the scale parameter α for each State is re-estimated for comparison. Results are given in Tables 3a, 3b and 3c.

2.4.2. Empirical distribution function (ECDF)

The ECDF of a random variable X (say) is the relative cumulative frequency of events. It provides a complete sample based description of its univariate statistical distribution. In this paper we use ECDF as a non-parametric alternative to the power-law distribution described in Section 2.4.1. This estimation procedure is particular useful for this data due to issues surrounding missingness and irregularity of sampling. Formally, the empirical distribution function $F_X(x)$ is defined as

$$F_X(x) = \frac{Frequency\ of\ events \leq x}{Total\ sample\ size}.$$

In Section 3.2 we show the ECDF of flood related damages across the six Indian states. A key finding from the plots given in Section 3.2 are the prevalent similarity (and separation) between the distribution of annual flood damage among the states. To investigate this further we use principal component analysis, that we describe next.

2.4.3. Principal components analysis (PCA)

PCA was introduced by Karl Pearson [22] and is a standard unsupervised learning method often used to draw inferences from multivariate data when the dataset consists of several correlated covariates. A common use of PCA is to estimate clustering in the data and potential factors affecting such similarity. In this article we have used PCA to detect clustering of Indian states and similarity of annual damages, separately, by each component of damage. This would answer the research Questions 2,3 and 5 set out in the introduction.

The broad intuition PCA is as follows. Suppose in a dataset we have n observations each consisting of p features as shown the following matrix,

$$X = \begin{matrix} X_{11} & X_{12}\ldots & X_{1p} \\ X_{21} & X_{22}\ldots & X_{2p} \\ X_{n1} & X_{n2}\ldots & X_{np} \end{matrix} \tag{3}$$

such that several of these features are correlated. PCA is a multivariate statistical method that constructs $m < p$ secondary features $\{Z_1, Z_2, \ldots, Z_m\}$ by linearly combining the primary features such that the secondary features retain most of the variance in the data.

That is,

$$Z_1 = \max Var\left(\sum_{i=1}^{p} \alpha_{i1} X_i\right)\ such\ that\ \sum_{i=1}^{p} {\alpha_{i1}}^2 = 1, \tag{4}$$

has the highest variance of all possible linear combinations $\sum_{i=1}^{p} \alpha_{i1} X_i$ given in equation (4). Likewise Z_2 has the second highest, Z_3 is a secondary feature that has the third highest variance of all such linear combination. Further, the secondary features $\{Z_1, Z_2, \ldots, Z_m\}$ are

Table 2
Maximum likelihood estimates and standard errors of the scale parameter $\hat{\alpha}$ for a power-law distribution for three component flood damages, across six Indian States.

Estimated parameters $\hat{\alpha}$ – by damage types						
States	Crop		Housing		Utility	
	alpha	Standard Error	alpha	Standard Error	alpha	Standard Error
AP	5.646	0.685	5.104	0.626	2.127	0.163
Assam	2.970	0.256	1.997	0.130	1.519	0.068
Bihar	2.868	0.252	1.473	0.066	1.248	0.034
Gujarat	1.501	0.070	1.332	0.050	1.288	0.040
Orissa	4.291	0.475	2.018	0.150	1.947	0.140
WB	1.884	0.117	2.046	0.144	1.606	0.083

Table 3
The estimated scale parameter x_{min} for a power-law distribution for three component flood damages, across six Indian States.

States	Crop	Housing	Utility
AP	2185.92	911.91	1374.62
Assam	251.07	11.51	14.22
Bihar	738.72	5.94	0.53
Gujarat	12.65	0.5	0.5
Orissa	370.48	44.14	347.78
WB	218.95	163.34	40.52

uncorrelated. A primary purpose of PCA is to reduce the dimension of a large number of correlated primary features $\{X_1, X_2, \ldots, X_p\}$ into a smaller set of uncorrelated secondary features, $\{Z_1, Z_2, \ldots, Z_m\}$ that collectively retain most of the variation in the data, especially when the sample size (n) is much smaller than the number of features, p. These features can then be used for causal modelling such as regression or unsupervised methods such as cluster analysis.

The flood damages data are multiple annual time series data obtained across six Indian states. Based on equation (3) we can organize these time series as a data with 6(n) samples and 63(p) potentially correlated features. A common approach in time series modelling and inference is to investigate prevalent serial correlatedness using parametric models, such as the auto-regressive moving average models, and then project these models to obtain predictions in the future and their past. Analysts can also multivariate time series to models to estimate association between multiple time series. However, due to the various inconsistencies in the data, including sampling and missingness, conventional time series modelling – and hence investigation of temporal correlation -is difficult. Due to this we have considered a novel implementation of PCA. We have two primary objectives –

1. The empirical distribution functions indicate presence of clustering in time series. We wish to obtain visual depictions of inter-state clustering, and
2. Associate them with particular year.

To this end we have performed PCA on the flood damages with annual damage components as feature vectors. Findings are presented in as principal component *biplots* (see for example [23] showing inter-state clustering and years that largely drive them. The loading vectors are indicated using red arrows for each year. State labels are superimposed on the plots to show inter-state similarity (and dissimilarity) across the entire period of observation. In future work we would investigate the cause behind these clustering and correlation.

2.5. Missing values

The CWC data has many missing values, the cause of which are unknown. This has prevented the choice of a rigorous probabilistic model. Hence for the present analysis, missing observations have been imputed using the sample median for that State. The outcomes of the PCA were also compared for mean and third quartile imputed data. Substantial differences were not found in the distribution of variation of the principal components or the factor loadings. Hence, the median was used since it is less affected by extreme values.

2.6. GSDP vs. damage

Two Pearson correlation analyses were performed using univariate imputation by median and removal of missing values from the analysis. The results are in Table 4 in the Supplementary Material and show low to moderate correlations.

3. Results and discussion

All computations and results were obtained using the statistical programming language R [22]. Power laws were fit using the R package PowerLaw [21]. For principal components analysis plots we used the routines prcomp() and autoplot() from the package ggfortify [25].

3.1. Descriptive statistics

For Research Question #1, time series plots of 5-year moving averages for component damages are given in Fig. 2 and descriptive statistics of damages are given in Table 1. Andhra Pradesh (AP) followed by Bihar and West Bengal (WB) have consistently had the highest damages. Fig. 2 and Tables 1, 1.2 (see supplementary) and 1.3 (see supplementary) show that housing and crop damages are greater than utility damages, but there are nuances. The median damage values of the States is quite similar but the means and maxima of damages for AP and WB are substantially higher. These results have several possible explanations including a higher frequency of extreme weather events or being less prepared in terms of natural hazards management. Later in this section the canonical years, which account for the largest amount of damage, will be identified.

The panels in Fig. 2 do not suggest monotonically increasing or decreasing trends, however substantially higher damages have occurred during the late 90's and early 2000's in Orissa, WB, Bihar, and AP with high values at other times in other States except for AP. It is also apparent that the distribution of damages is both skewed and heavy tailed.

Descriptive statistics, histograms and trend plots stratified by component damages are given in the Supplementary showing the variation in damage statistics across States and by components.

3.2. Empirical distribution function

The empirical cumulative distribution functions (ECDF) of the lower sixty percentile (top row) and upper forty percentile (bottom row) of component damages for the six Indian States are plotted in Fig. 3. We wanted to investigate if the distribution and clustering of the states varied between the extreme and moderate The ECDF trajectories of the bottom row of plots of the upper forty percentile-damages depict the clustering of the States by damages. It is apparent that some States have relatively more negatively skewed housing and crop damages than others. But clustering of the ECDF curves among states is also apparent. Notably, Assam, Gujarat and Orissa have similar and overlapping ECDF trajectories for the component damages. Also, the ECDF of the upper forty percentile damages of Andhra Pradesh, West Bengal and Bihar stochastically dominate [26] the other three States. But there are also important heterogeneities especially in the clustering of States between the top and bottom percentile damage distribution. Andhra Pradesh dominates all states in all negative skewness and heavy-tailed component damages. West Bengal had significantly greater crop damage in the year 2000 while Gujarat had large housing damage in 2006. These descriptive observations led to the following two alternative approaches for statistical inference.

3.3. Principal components analyses (PCA)

To analyze heterogeneities in the time series of annual damages between States, and their association with years, a PCA [24] on the annual damages was performed. For each component damage the States are treated as samples and annual time series damages as features. Conventional PCA requires balanced data; that is, for each State the same length of time series. Hence missing observations were imputed as discussed in Section 3.5. The results of PCA are depicted using the biplots— scatterplots between the first and second principal components—in the panels of Fig. 3.

Observations from the PCA are as follows. The factor loadings (indicated by the red arrows) represent damages in the labelled years. The biplots (PC2 vs PC1) in Fig. 4 include arrows, the length of which is proportional to the magnitude of impact a year has in explaining the maximum variation in the data. The panels also show clustering in the feature space of years. Alignment in the orientation of arrows (and their length) indicate that damages in those years have similar impact across all States. For example, in housing damages we find that the years 1995, 1996 and 1998 have similar effects across all states while 2006 has a significantly different impact on variation across the States. 2006 is an outlying year for housing and utility damages, largely driven by observations in Gujarat. In 2000 WB witnessed much larger crop damages than other states. Other prominent clustering of yearly damages in the biplot for utility damages are expressed in the vertical alignment of 1982, 2001, 2003, 2007, 2008 and 2011, indicating that these are the years that account for 7% of overall variation in utility damages, across all States.

From this analysis the following similarities occur among States for the magnitude of damage-

a. Housing- Assam and Orissa.
b. Crops – Assam, Gujarat, Orissa and Bihar.
c. Utility – Assam, Bihar, Gujarat and West Bengal.

When the years 2000 and 2006 are removed, because they are outliers, the overall variation in the data decreases and there are some changes in the clustering patterns for housing damage. The first two principal components collectively include a higher proportion of overall variation in all damage data. The plots are shown in Fig. 5 and the (most reliable) clusters become-

d. Housing- Assam, Gujarat, Bihar and Orissa.
e. Crops – Assam, Gujarat and Orissa.
f. Utility – Assam, Bihar, Gujarat and West Bengal.

3.4. Modelling a power law

The ECDFs of natural events are often compared against a power-law distribution (see Section 2.4.1). Values of $\alpha > 2$ indicate that the events can be fit by a power law and are prone to random occurrences, particularly among the extremes. Such a distribution allows the estimation of the annual probability of extreme events for disaster preparedness, enables an inter-State comparison, and correlation with GSDP using values of $\alpha < 2$ for values $> x_{min}$. The PCA and the ECDF show that there is substantial heterogeneity in the pattern of damages between the States, stratified by the various component damages. Hence, power-law density functions were fit separately for each State and component. The parameter estimates $\widehat{\alpha}$ and x_{min} are given in Tables 2 and 3 Figs. 6–8 show the estimate distribution. More information about x_{min} is provided in the Supplementary Material.

Table 2 shows that data for AP, Assam, Bihar and Orissa can be fit by a power-law ($\widehat{\alpha} > 2$) for crop damage, only AP and WB can be fit by a power law for Housing and only AP for Utilities. This could partly be explained by weather events for crops and with less such effect for housing and utilities where human decision making plays a larger role in the creation of exposure and vulnerability to floods. While plausible, spatially resolved data at District and/or river catchment level will be required for a more robust conclusion.

3.5. Limitations of the power law and future research

The fitting of a power law in the data used here is extremely sensitive to the precise estimation of the inflection point x_{min} in the ECDF using a least squares method (refer to the paper). It is well known that least squares minimizers are affected by outlying or missing observations. Ironically, power laws are used to model extreme observations which might be considered to be outliers. More research is needed for other estimation methods for x_{min}. Further, the present data has several missing observations. Any imputation would likely affect the least squares estimates of x_{min} and hence estimates, $\widehat{\alpha}$. For these reasons parametric bootstrapping [29,30] was used to assess the reliability of the current estimates. The results are given in the Supplementary Section. Large bootstrap standard

errors of estimation were found for x_{min} and hence $\widehat{\alpha}$.

There are several potential alternatives that could be investigated to analyze this data. Such methods include extreme value distributions, robust statistics and change point detection. The methodological problems highlighted in this paper should be investigated separately.

3.6. GSDP vs. damage

The low to moderate correlations (supplementary Table 4) are probably because State-wide values of GSDP are poor indicators of the potential for damage in the vicinity of rivers, and also because linear relationships between such complex variables may not be expected. This is another reason for at least District level or catchment level analyses. The highest correlation is for damage to utilities in AP and Orissa, followed by crops and housing in AP. These results suggest that utilities are particularly vulnerable in AP and Orissa and crops and housing in AP. In Gujarat there is a moderately high negative correlation with GSDP, suggesting that the more utilities contribute to GSDP, the lower the damage to them. This puzzling result needs further investigation.

4. Conclusions

The primary objective of this paper was to conduct a temporal survey of flood related damages in some modern Indian states since the formation of the republic. During this period India has seen major transitions in economy, agriculture and urbanisation, thanks to a burgeoning population. We were particularly interested to investigate if flood damages were completely random 'black swan' like phenomena or did they offer signs of systematic events. Due to the paucity of data, we have restricted our analysis to six states. Five of these states (Assam, Bihar, Orissa and West Bengal) have been affected by recurring floods while the western state of Gujarat was largely kept as control. The results of power law modelling (Section 3.4), but further inter-state clustering of damages as observed through empirical distribution function (ECDF) (Section 3.2) and principal components analysis (PCA) (Section 3.3) indicate that component damages are quite likely influenced by structural and policy issues affecting different states in heterogeneous ways. For instance, we observe that the same event (such as flooding in years 2000, 2006 or 2009) have different adverse effects on the three components for each state. A full causal inference would require a granular investigation involving economic and policy-based analysis. We are considering analyses based on data from the lowest local government hierarchies – state districts. These would be reported separately. We conclude with categorical response to the specific research questions set out in the Introduction.

With respect to the research questions, the following conclusions have been reached:

1. There are no trends in any of the categories of damage for the period from 1953 to 2011. Therefore, the efforts at flood mitigation since Independence in the six States have not reduced damage to low levels, although damage may have been higher without mitigation. Governments should adopt policies which eventually lead to distribution of components' damages that are positively skewed and not heavy tailed. If hydrologic data and embankment records were made available for all of the States, probability models can be used to help evolve such policy that will account for the relationship of damages with flood magnitude. PCA and ECDF analyses demonstrate that strategies would need to reflect the differences that affect each state. For example in the case of Assam, for which peak flows are available for the Brahmaputra River, there is no evidence of such a change in peak flow (M. Mirza, pers. comm., 2020).
2. Total damage is ranked in the following way, from highest to lowest: AP, Bihar, WB, Orissa, Assam, Assam, Gujarat. The reason for this ranking is not clear based on a correlation with Gross State Domestic Product (GSDP) by damage category, suggesting that further investigation is required.
3. Similarities between the damage profiles of States (Section 3.3) suggest that inter-State learning could be valuable. The clusters of States are surprising given the differences between them hydrologically and economically, under scoring the value of PCA.
4. The power law is the best distribution for most of the crop damage data, but even less so for housing and utilities. While by no means certain, it is possible that random weather events have most impact through floods on crops and less so on housing and utilities the loactions of which are a result of human decisions that increase exposure and vulnerability to floods. More spatially resolved data is necessary to test this idea but, if substantiated, suggests that for housing and utilities human decisions can reduce damage; perhaps a self-evidently true statement.
5. The most reliable information for policy makers from this analysis is that the effort so far in flood mitigation has not markedly reduced flood damage in any of the three categories. Therefore, a new approach is required that is evidence-based and probably requires experimentation [27] in a few different locations before being widely adopted. Such experimentation should involve at least non-structural interventions such as floodplain zoning and regulation [28]. Another conclusion is that inter-State learning could be of value between the clusters of States identified here. AP appears to be particularly vulnerable to flood damage for reasons that are not apparent from this analysis. There is an urgent need for more spatially resolved analysis, ideally at the scale of river catchments or at least Districts to test some of the ideas presented in this and other papers. Finally, there is a large body of scientific literature, only some of which has been referred to here, which points to the inadequacy of current flood mitigation strategies. This literature does not appear to impinge on those making decisions that affect millions of people, often with disastrous consequences.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The initial work for this paper was carried out at the National University of Singapore from which funding was received. Dr. Monirul Mirza is thanked for the Brahmaputra peak flow data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijdrr.2022.102835.

References

[1] H.C. Winsemius, J.C.J.H. Aerts, L.P.H.V. Beek, M.F.P. Bierkens, A. Bouwman, B. Jongman, J.C.J. Kwadijk, W. Ligtvoet, P.L. Lucas, D.P. van Vuuren, P.J. Ward, Global drivers of flood risk, Nat. Clim. Change 6 (2015) 381–385.
[2] World Resources Institute, World's 15 Countries with the Most People Exposed to Floods, 2015 https://www.wri.org/blog/2015/03/world's-15-countries-most-people-exposed-river-floods accessed 9/2/18). Written by Tianyi Luo, Andrew Maddocks, Charles Iceland, Philip Ward and Hessel Winsemius.
[3] P.S. Ward, G.E. Shively, Disaster risk, social vulnerability, and economic development, Disasters 41 (2) (2017) 324–351.
[4] Central Water Commission, State Wise Flood Damage Statistics, 2012 accessed 6/2/18: and previously at: http://www.cwc.gov.in/main/downloads/FFM.2200-2251.27112012.pdf, http://www.cwc.nic.in/main/downloads/Water_Data_Complete_Book_2005.pdf.
[5] D.S. Bisht, C. Chatterjee, S. Kalakoti, P. Upadhyay, M. Sahoo, A. Panda, Modeling urban floods and drainage using SWMM and MIKE URBAN: a case study, Nat. Hazards 84 (2) (2016) 749–776.
[6] Y. Parida, S. Saini, J.R. Chowdhury, Economic growth in the aftermath of floods in Indian states. Environment, Develop. Sustain. (2020), https://doi.org/10.1007/s10668-020-00595-3.
[7] G.P. Salazar, A.O. Diaz, G.G. López, Natural disasters and poverty: understanding the systemic complexity, in: Proceedings of the 34th Conference of the System Dynamics Society, July 2016, Delft, The Netherlands, 2016.
[8] S. Coelho, Assam and the Brahmaputra: Recurrent Flooding and Internal Displacement, 2012 accessed 21/3/18, http://labos.ulg.ac.be/hugo/wpcontent/uploads/sites/38/2017/11/The-State-of-Environmental-Migration-2013-63-73.pdf.
[9] Y. Jameel, M. Stahl, S. Ahmad, A. Kumar, G. Perrier, India needs an effective flood policy, Science 369 (6511) (2020) 1575.
[10] M.P. Mohanty, S. Mudgil, S. Karmakar, Flood management in India: a focussed review on the current status and future challenges, Int. J. Disaster Risk Reduc. 49 (2020) 101660.
[11] K.B. Ray, Flood prevention in the rivers of Bihar, north bengal and Assam, October 9, 1954, The Economic Weekly 6 (41) (1954) 1121–1126.
[12] I. Pal, S. Singh, Disaster risk reduction and response management for flood: a case study of Assam, India, in: I. Pal, R. Shaw (Eds.), Disaster Risk Governance in India and Cross Cutting Issues, Springer Nature Singapore, 2018.
[13] P.K. Mohapatra, R.D. Singh, Flood management in India, in: Flood Problem and Management in South Asia, Springer, Dordrecht, 2003, pp. 131–143.
[14] S. Gupta, A. Javed, D. Datt, Economics of flood protection in India, Nat. Hazards 28 (2003) 199–210.
[15] G. Sadoff, N.R. Harshadeep, D. Blackmore, X. Wu, A. O'Donnell, M. Jeuland, S. Lee, D. Whittington, Ten fundamental questions for water resources development in the Ganges: myths and realities, Water Pol. 14 (2013) 147–164.
[16] B. Merz, H. Kreibich, R. Schwarze, A. Thieken, Assessment of economic flood damage, Nat. Hazards Earth Syst. Sci. 10 (2010) 1697–1724, https://doi.org/10.5194/nhess-10-1697-2010.
[17] H. Ali, P. Modi, V. Mishra, Increased flood risk in Indian sub-continent under the warming climate, Weather Clim. Extremes 25 (2019) 100212.
[18] W. Kron, M. Steuer, A. Löw Wirtz, How to deal properly with a natural catastrophe database-analysis of flood losses, Nat. Hazards Earth Syst. Sci. 12 (2012) 535–550.
[19] R.A. Pielke Jr., J. Gratz, C.W. Landsea, D. Collins, M.A. Saunders, R. Musulin, Normalized hurricane damage in the unites states: 1900-2005, Nat. Hazards Rev. 9 (1) (2008) 29–42.
[20] A. Clauset, C.R. Shalizi, M.E. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
[21] C. Gillespie, Package 'poweRlaw', 2020.
[22] J.C. Gower, D.J. Hand, Biplots, Chapman & Hall, 1996.
[23] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014. http://www.R-project.org/.
[24] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, vol. 112, Springer, New York, 2013, p. 18.
[25] Y. Tang, M. Horikoshi, W. Li, ggfortify: unified interface to visualize statistical results of popular R packages, R J 8 (2) (2016) 474.
[26] P. Billingsley, Probability and Measure, John Wiley & Sons, 2008.
[27] D. Huitema, A. Jordan, S. Munaretto, M. Hildén, Policy experimentation: core concepts, political dynamics, governance and impacts, Pol. Sci. 51 (2018) 143–159, https://doi.org/10.1007/s11077-018-9321-9.
[28] S. Modak, P. Kapuria, From Policy to Practice: Charting a Path for Floodplain Zoning in India, ORF (Observer Research Foundation) Occasional Paper, May 2020, 2020, p. 39.
[29] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, CRC pPess, 1994.
[30] P. Hall, Theoretical comparison of bootstrap confidence intervals, Ann. Stat. 927–95 (29) (1988). Pearson, K. (1901). Principal components analysis. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 6(2), 559.