nature sustainability

Article

https://doi.org/10.1038/s41893-023-01202-9

Reversal of the levee effect towards sustainable floodplain management

Received: 26 November 2022

Accepted: 19 July 2023

Published online: 21 August 2023

Check for updates

Meng Ding[1,2,9], Peirong Lin[1,3,9] ✉, Shang Gao[4], Jida Wang[2,5], Zhenzhong Zeng[6], Kaihao Zheng[1], Xudong Zhou[7], Dai Yamazaki[7], Yige Gao[8] & Yu Liu[1]

Levees constrain roaring floodwater but are blamed for reducing people's perception of flood risks and promoting floodplain human settlements unprepared for high-consequence flood events. Yet the interplay between levee construction and floodplain development remains poorly quantified, obscuring an objective assessment of human–water relations. Here, to quantitatively assess how floodplain urban expansion is linked to levee construction, we develop a multiscale composite analysis framework leveraging a national levee database and decades of annual land-cover maps. We find that in the contiguous United States, levee construction is associated with a 62% acceleration in floodplain urban expansion, outpacing that of the county (29%), highlighting a clear change in risk perception after levees are built. Regions historically lacking strong momentum for population growth while experiencing frequent floods tend to rely more strongly on levees and we suggest these areas to develop a more diversified portfolio to cope with floods. Temporally, the positive levee effect is found to have weakened and then reversed since the late 1970s, reflecting the role of legislative regulations to suppress floodplain urban expansion. Our quantitative framework sheds light on how structural and non-structural measures jointly influence floodplain urban growth patterns. It also provides a viable framework to objectively assess the floodplain management strategies currently in place, which may provide useful guidance for managing flood risks towards sustainable development goals.

Flooding is one of the most devastating natural hazards, causing US$25.5 billion of economic losses and 6,570 fatalities worldwide annually on average between 1970 and 2020[1]. The property and life losses related to flooding have accelerated at a rate of 6.3% and 1.5% per year during the past five decades[2,3]. Yet despite the devastating consequences of floods and the increasing awareness of changing flood risks under a warming climate, floodplain encroachment is still inevitable largely due to increasing population pressure[4–7], which has posed grand challenges to flood risk management[8].

Levee construction is a well-known yet poorly quantified driving force for promoting floodplain development. As one of the oldest and least costly hydraulic engineering infrastructures, levees function

[1]Institute of Remote Sensing and Geographic Information Systems, School of Earth and Space Sciences, Peking University, Beijing, China. [2]Department of Geography and Geospatial Sciences, Kansas State University, Manhattan, KS, USA. [3]International Research Center of Big Data for Sustainable Development Goals, Beijing, China. [4]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, USA. [5]Department of Geography and Geographic Information Science, University of Illinois Urbana–Champaign, Urbana, IL, USA. [6]Department of Civil and Environmental Engineering, Southern University of Science and Technology, Shenzhen, China. [7]Global Hydrological Prediction Center, Institute of Industrial Science, The University of Tokyo, Tokyo, Japan. [8]Gwinnett County Department of Water Resources, Lawrenceville, GA, USA. [9]These authors contributed equally: Meng Ding, Peirong Lin. ✉e-mail: peironglinlin@pku.edu.cn

to confine the lateral distribution of roaring floodwater. However, levees can also reduce people's awareness of flood risks or the physical boundaries of a potential floodplain[9] and subsequently promote floodplain settlement (Fig. 1a,b); such a phenomenon is often known as the 'levee effect' (hereafter referred to as LE) and has been widely discussed among scholars since the 1940s[10–15] but rarely quantified across scales. The paradox of LE is that levees introduce a 'false sense of security', which may lead to unintended consequences despite the original goal of levee building to reduce flood risks. Although levees are designed to reduce the probability of flood hazards, studies have increasingly alerted people about the residual risks of levees, that is, levees can only protect areas from floods below the designed protection standards, but they cannot eliminate the occurrence of low-probability but high-consequence flood events[15–17]. Indeed, over the years, prominent cases of levee breaching have been reported, and due to the reduced awareness/preparedness, such events were often associated with high vulnerability and unexpected socio-economic losses, including but not limited to the New Orleans levee failures in 2005 after Hurricane Katrina[18], the Mississippi River levee failure in the 1927 and 2008 floods[19,20], and the UK levee breaches in the 1953 North Sea flood[21]. Currently, levees remain important structural measures to cope with flood risks[22]. Under this context, it is crucial to better understand the changing threat of flooding people may face even in areas protected by levees[23].

In this study, we are motivated to gain an in-depth understanding of how floodplain urban expansion can be linked with levee construction. Intuitively, a tight binding between levee construction and floodplain urban expansion indicates a strong socio-economic reliance on levees, which should call for attention given the underlying LE and its ramification for flood risk management. However, in the past, studies have reported difficulties in deriving a clear-cut linkage between levees and floodplain urban expansion patterns[12,13], partly due to a dearth of levee geodatabases and the highly complex nature of local decisions or policies compounding the problem[24–26]. Here we develop a multiscale composite analysis framework that leverages a national levee inventory[27] documenting key attributes such as levee completion year and levee-protected floodplain areas, which allows for an exclusive analysis on this factor. Building on the national levee database (NLD), we extract and analyse the floodplain and county urban expansion curves with seven decades of annual land-use land-cover change (LULCC) maps[28,29] for over a thousand locations. Then we introduce an LE index, which quantifies positive and negative LE using composite analysis and the similar concept of econometric methods (for example, difference-in-difference, synthetic control[30]) that compares the rates of accelerated urban expansion at the floodplain and the county levels (Methods); this helps to average out the compounding effects from other factors, allowing for an exclusive analysis of the LE. Ultimately, we provide a quantitative framework to assess the interplay between levee construction and floodplain urban expansion, and more importantly, an improved understanding of factors regulating this interplay. As climate change continues to intensify the hydrometeorological extremes[31–34], a deeper understanding into these inner dynamics is expected to inform wiser management strategies, which is key to achieving sustainable development goals for countries facing varying degrees of flood risk and financial/governance constraints.

Conceptualization to quantitative measures

Levees remain the least mapped features[25] for the river–floodplain system despite recent progress in mapping rivers, lakes and reservoirs/dams worldwide[35–38], which precludes a global-scale analysis of LE. Here we use the US NLD[39] as it is one of the most comprehensive and openly accessible levee geodatabases covering key attributes crucial to our analyses (Methods). Despite studies suggesting that the archived 39,445 km of levees only took up ~30% of the total US levee length[24,26], the 1,129 levee systems (Supplementary Fig. 1) cover the major ones and the number used is larger than in existing studies[12,13]. Documented levee constructions can date back to over 150 yr with a boom in the 1940–1960s, slightly earlier than but largely coinciding with the peak time of damming in the United States (Supplementary Fig. 2).

To assess whether there is an association (not necessarily a causal relationship) between levee construction and change in people's perceptional of flood risk, we first use the seven decades of annual LULCC maps from the US Geological Survey (USGS) to derive the urban expansion curves at two critical spatial scales, that is, the levee-protected floodplains and the counties that intersect the floodplains (see Methods for the pairing process). The former reflects urban developments that are more incentivized by the perceived flood protection from levees, whereas the latter also includes other socio-economic and biophysical drivers for urban growth. The raw urban expansion curves exhibit multiple breakpoints (Supplementary Fig. 3), suggesting the presence of many factors jointly promoting urban growth[40]. Thus, we introduce a composite analysis around the levee construction year (T_0 denotes its completion year) to exclusively focus on levees (Methods, equations 1 and 2). The high linearity of the curves suggests the effective averaging of other compounding factors, which allows for the calculation of the multiplicative changes in the urban expansion rates before and after T_0 (see the red and blue curves for the observed and the predicted urban expansion rates, respectively, assuming no perturbations; Fig. 1c). Then we define an LE index e that subtracts the accelerated urban expansion percentage at the county level from that at the floodplain level (Methods, equation 3) to quantify LE – $e > 0$ denotes an outpaced floodplain urban expansion compared with the county urban expansion after T_0, while $e \leq 0$ suggests a negative or non-prominent LE due to constantly faster urban expansion in other parts of the county. We find that the urban expansion rate in the levee-protected floodplain has accelerated by 62% (Fig. 1c) after T_0. This acceleration exceeds and is 2.1 times that of the county (29% acceleration; Fig. 1d), hence the $e = 0.33$ at the national scale offers strong evidence of urban expansion shifting towards places close to rivers, as well as quantitative confirmation of risk perception change after T_0.

We also calculate the ratio of urban area in the levee-protected floodplain ($U_p(t)$) to that in the county ($U_c(t)$) to investigate how the floodplain urban area ratio changes over time (Methods, equation 4). We find that the $\frac{U_p(t)}{U_c(t)}$ ratio first decreases linearly before T_0 (red line in Supplementary Fig. 4a), implying more allocation of urban expansion to the portion of counties beyond levee-protected floodplains, probably attributable to people's tendency to avoid flood risks. The ratio begins to increase after T_0, suggesting an outpaced urban expansion inside the floodplain, again highlighting a clear perceptional shift in treating floodplains as more habitable land than other parts of the county after levee construction. Subtracting the predicted ratio from the observed ratio and then dividing by the original $\frac{U_p(t)}{U_c(t)}$ ratio (Methods, equation 5), $E(t)$ represents the relative percentage of urban growth in the floodplain exclusively induced by levee construction, and a net of ~4% more urban areas have been promoted 10 yr after T_0 (Supplementary Fig. 4b). Interestingly, $E(t)$ starts to become positive1–2 yr before T_0, suggesting that an accelerated floodplain urban expansion occurred in parallel with the near-completion phase of levee construction, during which people's flood risk perception may already be shifting.

Strong regional levee effects

We further break down the positive LE at the national scale by probing into the spatio-temporal patterns, and investigate factors that may explain the varying strengths of LE. To address the limitations of the non-smoothed urban expansion curves at each individual levee level, we aggregate levees within different spatial units defined by the state (that is, administrative) and watershed (that is, natural hydroclimate) boundaries, such that spatial patterns of positive/negative LE can be

https://doi.org/10.1038/s41893-023-01202-9

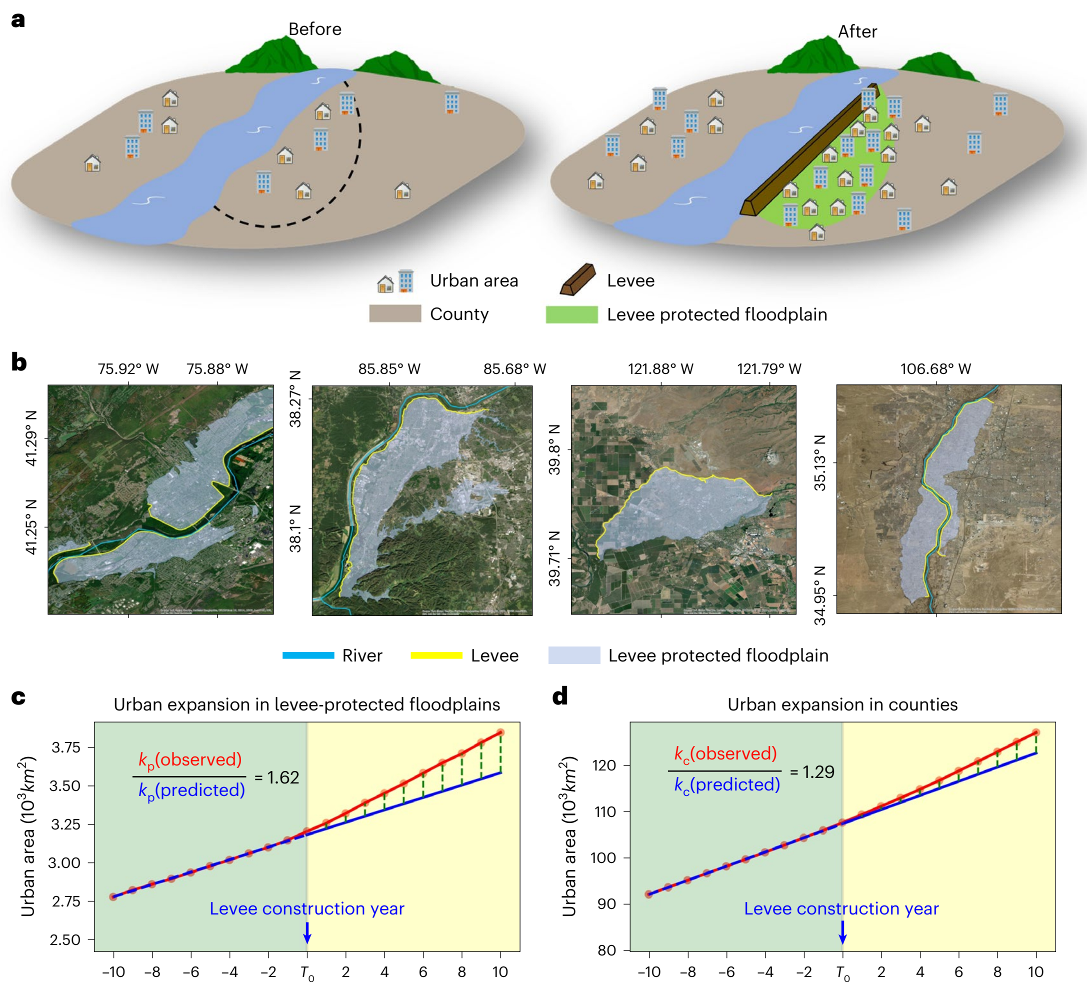

Fig. 1 | Composite analysis of the LE, lumping over the contiguous United States. a, Conceptual illustration of LE. **b**, Satellite images of four selected sites hinting urban development within the floodplain (from left to right: Larksville, Pennsylvania; Louisville, Kentucky; Chico, California; Albuquerque, New Mexico). **c**,**d**, Composite analysis of the urban area (10^3 km^2) in the levee-protected floodplains (U_p, **c**) and that in the counties (U_c, **d**). *X* axes separate the urban expansion time series into years before and after the levee construction year T_0, where all leveed locations in the United States are summed up for composite analysis; *k* shows the linear urban expansion rate. Background World Imagery Map source credits: Esri, Maxer, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN and the GIS User Community.

revealed. In Fig. 2a,b, the spatial composite index $e \le 0$ (blue) points to either limited floodplain urban expansion before and after T_0, or floodplain urban settlements already saturated before T_0. The former scenario implies the diversity of levee services, as exemplified by locations 1 and 2 (Fig. 2c) that function to protect agricultural lands rather than urban development, whereas the latter corresponds to high historical socio-economic pressures that had attracted floodplain urban growth long before levees were in place (locations 3 and 4 in Fig. 2c). These contrast with cases of $e > 0$ (red in Fig. 2a,b), which corresponds to noticeable urban sprawl towards rivers after levees were built.

We find that in general, regions historically lacking strong momentum for population growth while suffering from frequent flooding (for example, Lower Mississippi, Missouri, Arkansas-White-Red, Lower Colorado, the Pacific Northwest and Florida) tend to rely more strongly on levees (that is, $e > 0$; red in Fig. 2a,b and Supplementary Fig. 5a,c). For example, floods occur frequently along the Missouri River in eastern Nebraska, where several major floods in Omaha were reported[41]. For this location (location 7 in Fig. 2d), people considered the adjacency to the Missouri River meanders non-habitable without levees (urban area in grey), but their floodplain settlement strategy markedly changed after levees were built (urban area in red). South Florida presents another prominent example, where urban development relied heavily on flood control systems due to excessive surface water and prominent coastal flooding[42] (location 8 in Fig. 2d). Similarly, cases of $e > 0$ are found in Montana and New Mexico, where unpleasant natural conditions (for example, arid/semi-arid climates or mountainous areas) drove people to inhabit levee-protected areas due to the need for water resources. These are opposite to the eastern United States where no clear shifts in floodplain urban expansion patterns were found before and after T_0 ($e < 0$; blue in Fig. 2a,b and Supplementary Fig. 5b,d), prominently in the Ohio River Basin. This low reliance on levees was historically inevitable as the high population pressure since the colonial times drove floodplain urban expansion long before structural protections were in place. To further corroborate this finding, we replicate the composite analysis and quantify how LE varies at different levels of terrain variability and major city population[43] (Supplementary Fig. 6 and Table 1). Despite a certain degree of complexity implying the impacts of other exogenous factors, the general pattern shows that the signal of *e* tends to weaken with increasing population and decreasing terrain variability, which is consistent with our expectation. Meanwhile, urban expansion rates have accelerated in the levee-protected floodplains under all population and terrain conditions, confirming the overall prevalence of LE.

Areas with low reliance on levees also tend to have higher Federal Emergency Management Agency (FEMA) flood hazard map coverages[44]

 https://doi.org/10.1038/s41893-023-01202-9

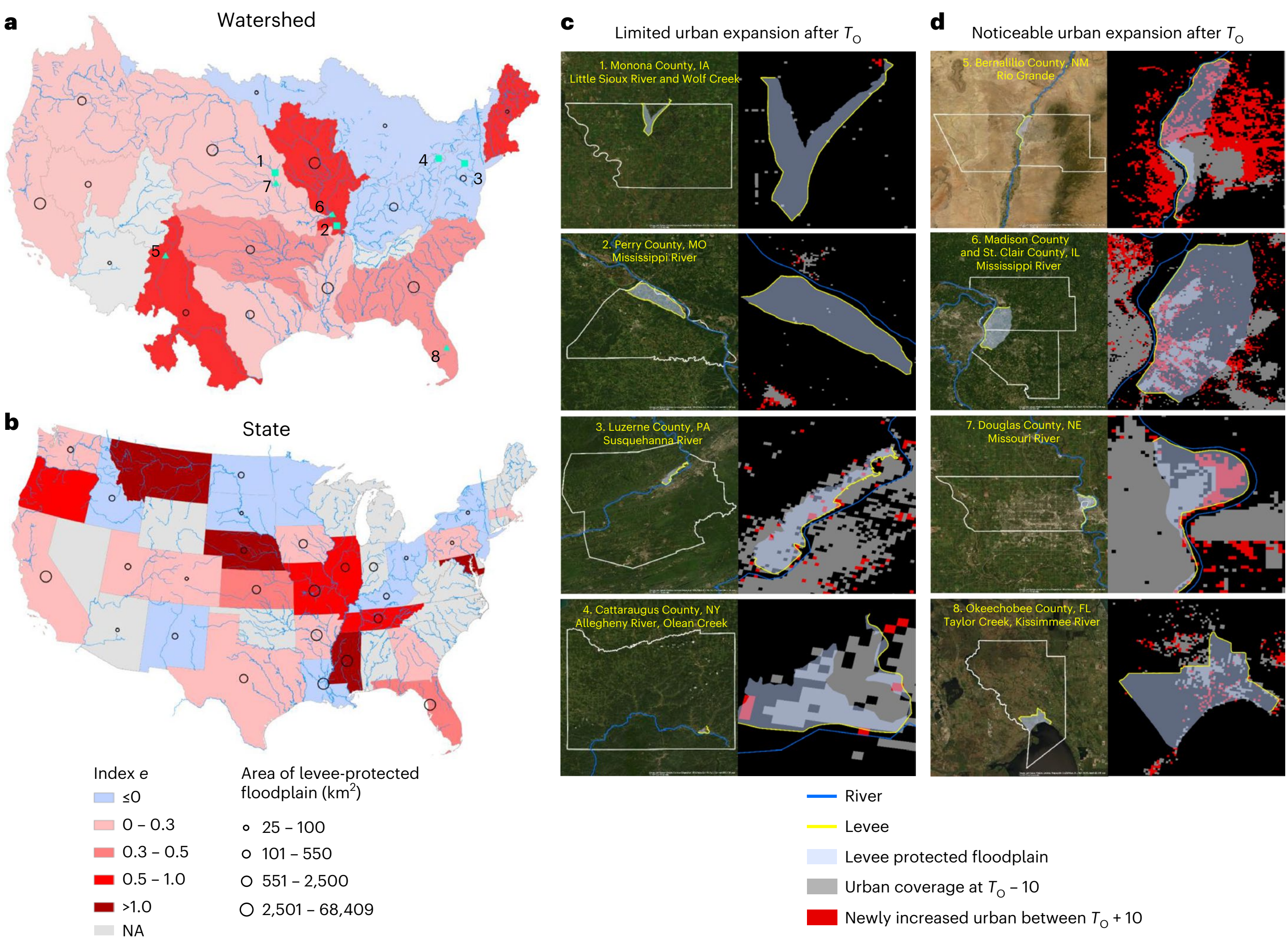

Fig. 2 | Spatial pattern of the LE. a,**b**, LE (*e*) at the watershed (HUC2, **a**) and state levels (**b**). **c**, Example cases of limited urban expansion after levee construction (locations 1–4, marked as green squares in **a**). **d**, The opposite cases (locations 5–8, marked as green triangles in **a**). In **c** and **d**, the left figure shows the county and the adjacent river where levee is located; the right figure shows the urban change at $T_0 - 10$ and $T_0 + 10$ (red denotes the newly increased urban area). Circles in **a** and **b** denote the total area of the levee-protected floodplains; the larger the area, the smoother the urban expansion curves for the validity of the spatial composite analysis. NA indicates a lack of either levee completion year or sufficient data for calculation. At locations 1 and 2, levees were built to protect local agricultural areas; at locations 3 and 4, the floodplain was fully urbanized before levees were built. It is reasonable to expect inconsistencies between **a** and **b** due to the modifiable areal unit problem and the scattered distribution of levees; here we focus more on the prominent consistencies for interpreting the results.

and vice versa. This may be explained by the fact that regions with high urbanization pressure and low reliance on levees (locations 3 and 4 in Fig. 2c) tend to possess better perceptions of flood risks, which often corresponds to more diversified ways of coping with floods, for example, funding to develop flood hazard maps, more comprehensive and well-firmed public alert systems. When levee systems were constructed later there, they were also designed with relatively higher protection standards as an ad hoc solution for protecting the already settled floodplains (for example, 2000-yr levees in the Ohio River Basin; Supplementary Fig. 7). By contrast, the majority of places that show a strong reliance on levees (red in Fig. 2; more cases in Supplementary Fig. 8) are typically devoid of FEMA flood hazard information and thus lack FEMA regulations to limit development, and their constructed levees also tend to have relatively lower protection standards in general (Supplementary Fig. 7). While not all high LE regions correspond to low flood protection standards, the tendency for such a distinct pattern provides an alarming picture of how areas exhibiting strong LE may probably have less preparedness for the residual risks of levee breaching or overtopping and thus high vulnerability to flooding, cautioning policymakers to pay special attention to these regions. Therefore, we suggest that such contrasting spatial patterns of LE be better considered in future flood risk management.

Temporal dynamics and the US floodplain management practices

Finally, we also analyse the temporal dynamics of LE to complement the above analysis that blurs out the years. To do that, we conduct a temporal composite analysis that groups the levees constructed 3, 5, 7 and 9 yr centred around the levee construction year to ensure that the results are consistent regardless of the examined timespans. While slight differences are found in the 3-yr composite (a likely result of the smaller sample size compared with other groups), all results point towards the positive LE before the late 1970s, suggesting the positive association between levee construction and the accelerated floodplain urban expansion primarily before the late 1970s (red in Fig. 3). Prominently, variations in the positive LE were observed before 1970, but LE then started to decline from the early 1970s until it came to a reversal after the late 1970s (blue in Fig. 3). The consistent temporal

https://doi.org/10.1038/s41893-023-01202-9

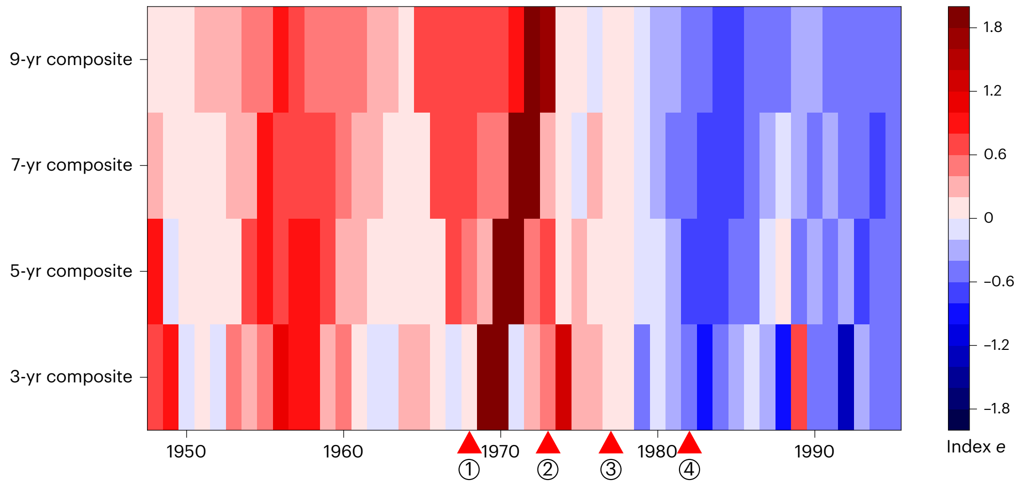

Fig. 3 | Temporal dynamics of the LE. The *X* axis shows the years of interest and the *Y* axis shows the analysis that groups different locations of levees constructed in the same years. 3-, 5-, 7- and 9-yr composite means that years centred around the years of interest are taken into account in the composite analysis. Triangles mark the years in which major legislative floodplain management policies were implemented; from 1 to 4: the 1968 NFIP; the 1973 Flood Disaster Protection Act; the 1977 Executive Order 11988: Floodplain Management; and the 1982 A Levee Policy for the NFIP.

patterns revealed interesting LE dynamics, highlighting distinct floodplain management strategies that have become effective. It is worth emphasizing that as our definition of the index *e* always compares the changes in urban expansion rate at two spatial scales, the natural slowdown of urban growth in the United States does not affect this result. The sensitivity analysis also suggests that the urban saturation level of the floodplain does not affect the decreased LE and its prominent reversal (Supplementary Fig. 9).

We further investigate the factors that can explain the observed LE dynamics and find the progress of the National Flood Insurance Program (NFIP) (triangle 1 in Fig. 3) as the key in shifting the US floodplain management policies from relying on levees alone to a joint use of structural and non-structural measures[45]. Because of NFIP, floodplain urban expansion started to face higher costs and more rigorous standards than before; thus, suppressed new developments were observed after initial rebounds in floodplain urban growth, which was faithfully captured by the index immediately after 1968. Then, the 1973 Flood Disaster Protection Act (triangle 2) added a mandatory insurance purchase requirement to NFIP, and the 1977 Executive Order 11988 (triangle 3) required that federal government activities in floodplains must strictly comply with NFIP; these collectively contributed to the LE weakening. Later, the 1982 'A Levee Policy' (triangle 4) specified detailed requirements of flood insurance purchase within levee-protected areas, which continued to strengthen the control of urban expansion in flood-prone areas and may explain the continuous decline and the eventual reversal in LE (Fig. 3). Overall, the LE dynamics reflected how each legislative non-structural measure has come into play. Our quantitative framework not only captures these detailed dynamics, but also highlights the timepoint when non-structural measures have become truly effective in suppressing floodplain urban growth in the United States.

Discussion

The interesting temporal dynamics offers a clear picture of the more sustainable floodplain management route that the United States has taken, which has been successful in suppressing 'risky growth'[46], in accordance with the call by the Intergovernmental Panel on Climate Change Sixth Assessment Report promoting joint use of structural versus non-structural measures for managing floodplains[47]. Thus, as a natural follow up, one might be interested in assessing LE dynamics more widely to offer insights into the floodplain management practices currently in place in other countries. Unfortunately, data limitations have precluded such an analysis; therefore, we instead surveyed ten countries frequently facing flood hazards to reveal their diverse structural and non-structural measures and the varying degrees of flood risk concerns (Supplementary Table 2). In general, the choice of adaptation measures varies considerably across regions due to their specific hydroclimatic and socio-economic contexts, precluding a simplified form of the expected LE and/or how it could affect risk management.

To offer a conceptual protocol, three possible pathways by which levees could affect flood risks are summarized, involving three determinants of risk: hazard, exposure and vulnerability[44] (right boxes in Fig. 4). The pathways are: (1) levees can reduce flood hazard under the designed flood protection standards and thus reduce flood risks[44]; (2) levees can reduce people's awareness or preparedness for floods, thus increasing flood vulnerability[11]; and (3) levees can stimulate urban expansion and flood exposure, thus increasing fatalities and economic losses with levee failure[48]. Given the complexity of these aggregated linkages, it remains challenging to directly associate flood risk with LE.

By developing a framework to assess LE, our study not only offers quantitative evidence of the existence of the last pathway, but also clarifies the interplay between levee construction and urban expansion, which can be jointly regulated by hydroclimatic and socio-economic factors as well as non-structural measures (left boxes in Fig. 4). This non-uniform LE dynamics in space and time adds a layer of uncertainty in risk management, as further confounded by the sign of changes in flood hazards under climate change[31] (top box in Fig. 4). As climate change continues to exacerbate weather extremes and the designed protection standard of hydraulic engineering infrastructures are increasingly challenged[4,8,49,50], it seems that directly advocating for similar non-structural measures (for example, those applied in the United States) for regions exhibiting a close link between levee construction and floodplain urban growth (that is, positive LE) should be prioritized, but the advocacy can be complicated by the specific hydroclimate and socio-economic background. Thus, should data become available, we argue that similar quantitative analyses need to be performed to facilitate an objective and in-depth assessment of spatio-temporal LE patterns. To achieve the sustainable development goals set by the United Nations that aim to reduce flood fatality worldwide, such assessments can foster a better rethinking of human–water relations to emphasize 'soft path solutions'[51] to complement physical infrastructures (for example, better legislation for floodplain management, raising awareness of flood risks and early-warning systems) towards better-informed strategic planning for sustainable floodplain management. Lastly, as such analyses would require data on hydraulic

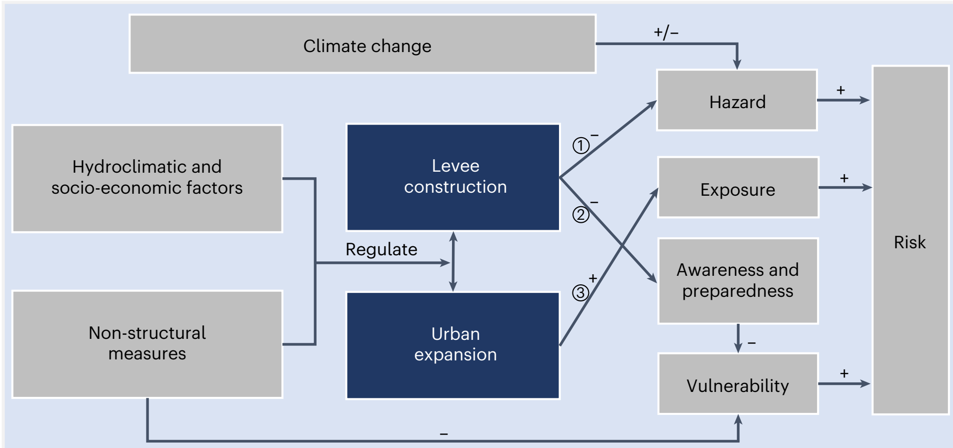

Fig. 4 | Conceptual diagram illustrating the role of levees in flood risk management. Numbers 1, 2 and 3 refers to the links of LE to three determinants of flood risk: hazard, exposure and vulnerability. '+' and '−' denote positive/negative correlation, respectively.

engineering infrastructures, our study also advocates for the development or open sharing of levee databases, in line with several recent studies[26,52], which will be useful for assessing the changing flood risks of the global river–floodplain system under increasing anthropogenic pressures.

Methods

Geodatabase of levee systems

The US NLD (https://levees.sec.usace.army.mil/#/; accessed November 2021) developed by the US Army Corps of Engineers (USACE) was used in our study. NLD records 6,969 levee systems in the United States with a total levee length of 39,445 km. NLD also provides important attributes such as the levee completion year, levee performance and the protection standard (as flood return period), among many others. The inventoried levee systems are represented by polylines and their protected areas (that is, the floodplain extent that a levee protects) are represented by polygons (Supplementary Fig. 1). The levee length ranges from 4.6 m to 578.27 km (median 1.63 km, mean 6.24 km), and the levee-protected area ranges from 140 m^2 to 19,511.71 km^2 (median 0.55 km^2, mean 31.26 km^2). Among the levees, 1,750 have documented construction completion years (from 1882 to 2021), but only 1,129 were constructed between 1948 and 1995, which satisfies our calculation criteria for the LE. Thus, these 1,129 levees were eventually used for the analysis.

Multitemporal urban land-cover maps at annual time scales

Many LULCC maps can be used to derive urban expansion curves useful for our analyses. Here we chose the USGS LULCC maps[28,29] that cover the period from 1938 to 2005 annually, because the levee constructions in the United States peaked in the 1940s–1960s (Supplementary Fig. 2). Despite a suboptimal spatial resolution of 250 m, this LULCC dataset included the peak levee construction period, thus allowing inclusion of as many levees as possible. This dataset was developed by combining a diverse set of data sources with a spatially explicit modelling framework, and it contained 14 LULC classes for the years 1938–1992 and 17 classes from 1992 to 2005. For both timespans, we extracted Class 2 (Urban/Developed) data for our analyses.

County, watershed and state shapefiles

We used the 1:500,000 2018 USA 116th Congressional Districts dataset as the county boundary (https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2018.html; accessed November 2021). We also used the watershed boundary dataset[53] (https://apps.nationalmap.gov/downloader/#/; accessed May 2021) and the 1:5,000,000 2018 US state shapefile (https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2018.html; accessed November 2021) for presenting the spatial distribution of LE.

Reprojection, rasterization and zonal averaging

The data layers have different projection systems and spatial extent/resolutions (Supplementary Table 3), and they were first reprojected to the USA Contiguous Albers Equal Area Conic USGS version (EPSG: 5070) projection system before other processing steps. The reprojected data files were then consistently rasterized and resampled to the same spatial resolution (250 m) and the same spatial extent (top: 3175292.633, left: −2357953.1839, right: 2282796.8161, bottom: 238542.633) as the USGS LULCC data with the Geospatial Data Abstraction Library. We then calculated the average urban percentages within each floodplain and county for time-series analysis (1938–2005) using zonal averages performed with the Python Pandas 'groupby' function (the Python scripts are openly available at https://github.com/peironglinlin/leveeRS).

Pairing levees with counties and protected floodplains

Each levee was paired up with a levee-protected area polygon, which comes with the NLD dataset and was digitized by a variety of methods (for example, hydraulic modelling, flood fill method, projected profile method, flow path method, manual digitizing) employed by USACE personnel (P. F. Kline, personal communication). We directly used this NLD polygon dataset for the floodplain-scale analyses. The minimum polygon in NLD is 0.000138 km^2, which is smaller than one LULCC pixel (0.0625 km^2), so we eliminated 270 levees whose protected areas are smaller than one pixel. In the remaining dataset, ~21.37% of the levee systems and their protected floodplain polygons intersect with multiple counties. For levees completely within the boundary of a single county, the pairing relationship was determined simply by using the 'completely within' spatial join function in ArcGIS. For levees intersecting multiple counties, we assumed that its floodplain urban expansion can be jointly influenced by all counties intersecting with it (meaning joint influence from the county policy or local government decisions). Accordingly, we visually inspected each of the 'one-to-many' spatial relationships and excluded cases where the intersection was caused by slivers or other mapping artefacts, retaining only the valid 'one-to-many' cases after visual inspection. The national-scale results presented in Fig. 1c,d are based on the correspondence of 1,129 unique levees and 601 unique counties.

 https://doi.org/10.1038/s41893-023-01202-9

Constraints applied in quantifying the LE

To focus our analyses on levees and to rule out other compounding factors, we applied two key layers of constraints for developing our quantitative framework, namely (1) 'spatial constraints' defined by two spatial scales, that is, the levee-protected floodplains (where the close proximity to rivers makes the levee protection a direct effect) and the counties (where the socio-economic drivers and policies within the administrative boundaries play a key role), and (2) 'temporal constraints' that separated our analysis critically by the levee construction year (T_0). Below, we introduce how we used them specifically.

Composite analysis

It is recognized early in the analysis that one cannot directly apply the floodplain-level LULCC calculation to assess LE due to the complex factors conjointly promoting urban growth (for example, multiple breakpoints not corresponding to T_0 in Supplementary Fig. 3), data limitations (for example, the suboptimal spatial resolution of the USGS LULCC data at 250 m due to the need for long historical data before 1985) and the scattered distributions of levees that can compromise a clear interpretation of LE. Therefore, we introduced the composite analysis that grouped levees at different spatial and temporal scales to derive the national-scale LE. More specifically, in equations (1 and 2), $u_{p,i}$ and $u_{c,i}$ denote the annual urban area within the levee-protected floodplain and within the county boundary, respectively; subscript i denotes the ith floodplain–county pair and t ranges from 1938 to 2005 (see example time series in Supplementary Fig. 3). Then the temporal constraints were applied by relabelling the LULCC as $t \in [T_0 - 10, T_0 + 10]$, and for the same year t, the urban area was summed up to derive the urban expansion curves (Fig. 1c,d); here, $U_p(t)$ and $U_c(t)$ denote the total urban area at the protected floodplain and county levels, respectively; n_p and n_c denote the total number of floodplains and counties, respectively.

$$U_p(t) = \sum_{i=1}^{n_p} u_{p,i}(t),\ t \in [T_0 - 10,\ T_0 + 10] \quad (1)$$

$$U_c(t) = \sum_{i=1}^{n_c} u_{c,i}(t),\ t \in [T_0 - 10,\ T_0 + 10] \quad (2)$$

Defining the LE index e

On the basis of equations (1 and 2) and Fig. 1c,d, factors other than levees were properly averaged out. We then borrowed the principles of the econometric methods to quantify LE, where it is crucial to choose a good control group. We selected county-scale urban expansion as the control group, as it shared the same socio-economic, hydroclimatic and other biophysical drivers for urban growth with the floodplain, except for levees. Building on this, a positive LE was logically derived if the urban expansion acceleration in the protected floodplain (treatment group) exceeded that of the county (control group). Therefore, an index e to quantify LE was constructed by calculating the difference in slope change of urban expansion between the levee-protected floodplains and counties (equation 3, where k_p and k_c denote the linear slope of the derived urban expansion curves in equations 1 and 2). In particular, the linear slope of the time-series urban area data from $T_0 - 10$ to T_0 can be used to calculate $k_p(\text{predicted})$ and $k_c(\text{predicted})$ assuming no perturbations from levee construction, and the linear slope from T_0 to $T_0 + 10$ can be used to calculate $k_p(\text{observed})$ and $k_c(\text{observed})$. In this analysis, we used the 10-yr timespan but our sensitivity analysis shows that the results are consistent under different timespans—we tested 4, 6, 8, 10, 12 and 14-yr timespans, and only 6, 10 and 14 yr are presented in Supplementary Fig. 10 for clarity. Specifically, e can be interpreted as the degree of urban growth in the protected floodplain exceeding that in the county, even in cases when county urban expansion is slowing down. Note that only ~0.5% of the levee-protected floodplains take up a percentage of larger than 20% in its corresponding county (the largest one is ~51%); this means that majority of the levee-protected floodplains are much smaller than the counties, thus making the index valid for our purpose.

$$e = \frac{k_p(\text{observed})}{k_p(\text{predicted})} - \frac{k_c(\text{observed})}{k_c(\text{predicted})} \quad (3)$$

In addition, to provide more information, we also calculated the ratio of urban area within the levee-protected floodplain to that within the county ($R_{p/c}(t)$, equation 4). Results show that the urban area in the floodplain takes up ~3% of the urban area in the county, which decreased before T_0; by extrapolating the linearly decreasing $R_{p/c}(t)$ to predict the ratio of urban area in floodplains ($R'_{p/c}(t)$) after T_0 by assuming no perturbations from levees, we derived the red and blue curves in Supplementary Fig. 4a. The relative difference between the observed and the predicted ratio ($R_{p/c}(t) - R'_{p/c}(t)$) can then be viewed as the net effect of levee-associated changes and divided by the original $R_{p/c}(t)$ to give $E(t)$ (equation 5), which denotes the percentage change in floodplain urban area exclusively associated with levee construction (Supplementary Fig. 4b).

$$R_{p/c}(t) = \frac{U_p(t)}{U_c(t)} \times 100\% \quad (4)$$

$$E(t) = [R_{p/c}(\text{t}) - R'_{p/c}(\text{t})]/R_{p/c}(\text{t}) \quad (5)$$

Spatial composite analysis

To break down the national-scale LE in the contiguous United States both spatially and temporally, we further introduced the spatial composite for state/watershed level analysis and the temporal composite for assessing the temporal dynamics. Again, as the urban expansion rate is difficult to directly assess at the level of individual levees, we used two grouping scales, that is, the Hydrologic Unit Code-2 (HUC2) watershed and state levels, and performed similar composite analysis as the national-scale composite to present a more visual and continuous distribution of the LE (Fig. 2). This spatial composite derived more-smoothed urban expansion curves for each spatial unit (see Supplementary Fig. 5 for examples) that allowed for the valid calculation of e. The same equations (1 and 3) were applied to each state and watershed that contains levees satisfying the criteria. It is worth noting that due to the imbalanced spatial distribution of the levee systems, some spatial units might not have had sufficient data sample size to derive well-smoothed curves; thus, we further used the summed area of the levee-protected floodplains (shown as circles in Fig. 2a,d) as the first-order approximation of the validity of our index e.

Temporal composite analysis

Finally, we also grouped levees constructed in the same years but in different locations to focus on the LE temporal dynamics. Similarly, to avoid the limitations due to insufficient data samples (locations g, h and i in Supplementary Fig. 3), we grouped the levee constructed in 3, 5, 7 and 9 yr centred around the years of interest for this analysis. This also helped to ensure that our results are robust irrespective of the grouping method. The derived temporal patterns of LE are presented in Fig. 3.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in this study were obtained from openly accessible data sources. The NLD[27] shapefiles were downloaded from https://levees.sec.usace.army.mil/#/, accessed November 2021. Historical maps of the USGS LULCC datasets (1938–2005)[28,29] were obtained from https://www.sciencebase.gov/catalog/item/59d3c73de4b05fe04cc3d1d1 and https://www.sciencebase.gov/catalog/item/5b96c2f9e4b0702d0e826f6d. The shapefiles of the Watershed Boundary Dataset (WBD)[53] were downloaded from https://apps.nationalmap.gov/downloader/#/. The dam data were obtained from the US National Inventory of Dams (USNID) dataset[39] (https://nid.sec.usace.army.mil). Data on the fatalities and economic losses due to flooding in each country are from the EM-DAT database[1] at https://public.emdat.be/data. Major cities[43]: https://hub.arcgis.com/datasets/esri::usa-major-cities/about. Our organized data are available from GitHub at https://github.com/peironglinlin/leveeRS/blob/main/data/COMID_systemID_intersection.csv and https://github.com/peironglinlin/leveeRS/blob/main/processed_data/Processed_data_for_Fig1c_and_Fig1d.csv. Source data are provided with this paper.

Code availability

Codes for data processing and analyses are openly available via GitHub at https://github.com/peironglinlin/leveeRS.

References

1. *EM-DAT: The International Disaster Database* (Centre for Research on the Epidemiology of Disasters, 2008); http://www.emdat.be/Database/Trends/trends.html
2. Tanoue, M., Hirabayashi, Y. & Ikeuchi, H. Global-scale river flood vulnerability in the last 50 years. *Sci. Rep.* **6**, 36021 (2016).
3. Willner, S. N., Otto, C. & Levermann, A. Global economic response to river floods. *Nat. Clim. Change* **8**, 594–598 (2018).
4. Andreadis, K. et al. Urbanizing the floodplain: global changes of imperviousness in flood-prone areas. *Environ. Res. Lett.* **17**, 104024 (2022).
5. Liu, X. et al. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nat. Sustain.* **3**, 564–570 (2020).
6. Rajib, A. et al. The changing face of floodplains in the Mississippi River Basin detected by a 60-year land use change dataset. *Sci. Data* **8**, 271 (2021).
7. Mård, J., Di Baldassarre, G. & Mazzoleni, M. Nighttime light data reveal how flood protection shapes human proximity to rivers. *Sci. Adv.* **4**, eaar5779 (2018).
8. Kreibich, H. et al. The challenge of unprecedented floods and droughts in risk management. *Nature* **608**, 80–86 (2022).
9. Montz, B. E. & Tobin, G. A. Livin'large with levees: lessons learned and lost. *Nat. Hazards Rev.* **9**, 150–157 (2008).
10. White, G. F. *Human Adjustment to Floods* Research Paper No. 29 (Department of Geography, Univ. Chicago, 1945).
11. Di Baldassarre, G. et al. Hess opinions: an interdisciplinary research agenda to explore the unintended consequences of structural flood protection. *Hydrol. Earth Syst. Sci.* **22**, 5629–5637 (2018).
12. Hutton, N. S., Tobin, G. A. & Montz, B. E. The levee effect revisited: processes and policies enabling development in Yuba County, California. *J. Flood Risk Manage.* **12**, e12469 (2019).
13. Ferdous, M. R., Wesselink, A., Brandimarte, L., Di Baldassarre, G. & Rahman, M. M. The levee effect along the Jamuna River in Bangladesh. *Water Int.* **44**, 496–519 (2019).
14. Hino, M., Field, C. B. & Mach, K. J. Managed retreat as a response to natural hazard risk. *Nat. Clim. Change* **7**, 364–370 (2017).
15. Di Baldassarre, G. et al. Sociohydrology: scientific challenges in addressing the sustainable development goals. *Water Resour. Res.* **55**, 6327–6355 (2019).
16. Di Baldassarre, G. et al. Water shortages worsened by reservoir effects. *Nat. Sustain.* **1**, 617–622 (2018).
17. François, B., Schlef, K. E., Wi, S. & Brown, C. M. Design considerations for riverine floods in a changing climate—a review. *J. Hydrol.* **574**, 557–573 (2019).
18. Sills, G. L., Vroman, N. D., Wahl, R. E. & Schwanz, N. T. Overview of New Orleans levee failures: lessons learned and their impact on national levee design and assessment. *J. Geotech. Geoenviron. Eng.* **134**, 556–565 (2008).
19. Daniel, P. *Deep'n as It Come: The 1927 Mississippi River Flood* (Univ. Arkansas Press, 1977).
20. Bernhardt, M. et al. Mississippi river levee failures: June 2008 flood. *ISSMGE Int. J. Geoeng. Case Histories* **2**, 127–162 (2011).
21. Vellinga, P. & Aerts, J. in *Natural Disasters and Adaptation to Climate Change* (eds Boulter, S. et al.) 136–146 (Cambridge Univ. Press, 2013).
22. Ward, P. J. et al. A global framework for future costs and benefits of river-flood protection in urban areas. *Nat. Clim. Change* **7**, 642–646 (2017).
23. Best, J., Ashmore, P. & Darby, S. E. Beyond just floodwater. *Nat. Sustain.* 5,811–813 (2022).
24. Wing, O. E. et al. A new automated method for improved flood defense representation in large-scale hydraulic models. *Water Resour. Res.* **55**, 11007–11034 (2019).
25. Scussolini, P. et al. FLOPROS: an evolving global database of flood protection standards. *Nat. Hazards Earth Syst. Sci.* **16**, 1049–1061 (2016).
26. Knox, R. L., Morrison, R. R. & Wohl, E. E. Identification of artificial levees in the contiguous United States. *Water Resour. Res.* **58**, e2021WR031308 (2022).
27. *National Levee Database (NLD)* (US Army Corps of Engineers (USACE), 2021); https://levees.sec.usace.army.mil/#/
28. Sohl, T. et al. Modeled historical land use and land cover for the conterminous United States. *J. Land Use Sci.* **11**, 476–499 (2016).
29. Sohl, T. L. et al. *Conterminous United States Land Cover Projections—1992 to 2100* (US Geological Survey, 2018).
30. Abadie, A., Diamond, A. & Hainmueller, J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J. Am. Stat. Assoc.* **105**, 493–505 (2010).
31. Zhan, W., He, X., Sheffield, J. & Wood, E. F. Projected seasonal changes in large-scale global precipitation and temperature extremes based on the CMIP5 ensemble. *J. Clim.* **33**, 5651–5671 (2020).
32. Blöschl, G. et al. Changing climate both increases and decreases European river floods. *Nature* **573**, 108–111 (2019).
33. Merz, B., Hall, J., Disse, M. & Schumann, A. Fluvial flood risk management in a changing world. *Nat. Hazards Earth Syst. Sci.* **10**, 509–527 (2010).
34. Reynard, N. S., Prudhomme, C. & Crooks, S. M. The flood characteristics of large UK rivers: potential effects of changing climate and land use. *Clim. Change* **48**, 343–359 (2001).
35. Allen, G. H. & Pavelsky, T. M. Global extent of rivers and streams. *Science* **361**, 585–588 (2018).
36. Wang, J. et al. GeoDAR: georeferenced global dams and reservoirs dataset for bridging attributes and geolocations. *Earth Syst. Sci. Data* **14**, 1869–1899 (2022).
37. Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nat. Commun.* **7**, 13603 (2016).
38. Lehner, B. et al. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* **9**, 494–502 (2011).
39. *US National Inventory of Dams (USNID) Dataset* (US Army Corps of Engineers (USACE), accessed 2022); https://nid.sec.usace.army.mil
40. Collenteur, R. A., De Moel, H., Jongman, B. & Di Baldassarre, G. The failed-levee effect: do societies learn from flood disasters? *Nat. Hazards* **76**, 373–388 (2015).

41. Holbrook, E. Historic floods of the big muddy. *Risk Manage.* **58**, 18–20 (2011).
42. Bueno, J. A., Tsihrintzis, V. A. & Alvarez, L. South Florida greenways: a conceptual framework for the ecological reconnectivity of the region. *Landsc. Urban Plan.* **33**, 247–266 (1995).
43. *USA Major Cities* (Esri Data and Maps, accessed 2023); https://hub.arcgis.com/datasets/esri::usa-major-cities/about
44. Qiang, Y., Lam, N. S., Cai, H. & Zou, L. Changes in exposure to flood hazards in the United States. *Ann. Am. Assoc. Geogr.* **107**, 1332–1350 (2017).
45. *National Flood Insurance Program (NFIP) Floodplain Management Requirements. A Study Guide and Desk Reference for Local Officials* (FEMA, 2005); https://www.fema.gov/sites/default/files/documents/fema-480_floodplain-management-study-guide_local-officials.pdf
46. Rentschler, J. et al. Global evidence of rapid urban growth in flood zones since 1985. Preprint at *Res. Square* https://doi.org/10.21203/rs.3.rs-1460344/v1 (2022).
47. IPCC *Climate Change 2022: Impacts, Adaptation, and Vulnerability* (eds Pörtner, H.-O. et al.) (Cambridge Univ. Press, 2022).
48. Johnson, B. A. et al. High-resolution urban change modeling and flood exposure estimation at a national scale using open geospatial data: a case study of the Philippines. *Comput. Environ. Urban Syst.* **90**, 101704 (2021).
49. Tellman, B. et al. Satellite imaging reveals increased proportion of population exposed to floods. *Nature* **596**, 80–86 (2021).
50. Brunner, M. I. et al. An extremeness threshold determines the regional response of floods to changes in rainfall extremes. *Commun. Earth Environ.* **2**, 173 (2021).
51. Gleick, P. H. Global freshwater resources: soft-path solutions for the 21st century. *Science* **302**, 1524–1528 (2003).
52. Zhao, G., Bates, D. P., Neal, J. & Yamazaki, D. Flood defense standard estimation using machine learning and its representation in large-scale flood hazard modelling. *Water Resour. Res.* https://doi.org/10.1029/2022WR032395 (2023).
53. *Watershed Boundary Dataset* (USGS, accessed 2022); https://apps.nationalmap.gov/downloader/#/

Acknowledgements

This study was supported by the Open Research Program of the International Research Center of Big Data for Sustainable Development Goals, Grant No. CBAS2022ORP05. We acknowledge the funding support from the Fundamental Research Funds for the Central Universities, Peking University on 'Numerical modelling and remote sensing of global river discharge' (no. 7100604136). M.D. acknowledges the travel funding supported by the Department of Geography and Geospatial Sciences, Graduate School, and the College of Arts and Sciences at Kansas State University. We thank P. F. Kline from USACE for providing detailed information about NLD, and X. He for helpful discussions.

Author contributions

P.L. conceived the study. P.L. and M.D. designed the methodology. M.D. and P.L. performed the analysis and conducted the validation. P.L. and M.D. drafted the original paper with inputs from S.G., J.W., Z.Z., D.Y., X.Z., Y.G. and Y.L. M.D., P.L. and K.Z. produced the figures and tables. M.D., P.L. and K.Z. revised the paper with inputs from all co-authors. All authors contributed to the interpretation of results, writing and revision of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41893-023-01202-9.

Correspondence and requests for materials should be addressed to Peirong Lin.

Peer review information *Nature Sustainability* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Peirong Lin

Last updated by author(s): 2023/07/13

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
x	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	✓	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
x	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	✓	A description of all covariates tested
☐	✓	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	✓	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
x	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	✓	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	✓	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
x	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data acquisition was manually done by accessing the links provided in the Data section.
Data analysis	Our codes for data processing and analyses are made openly available via GitHub https://github.com/peironglinlin/leveeRS.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data used in this study are obtained from openly accessible data sources.
The NLD shapefiles were downloaded from https://levees.sec.usace.army.mil/#/.
Historical annual maps of the USGS LULCC datasets (1938-2005) were obtained from https://www.sciencebase.gov/catalog/item/59d3c73de4b05fe04cc3d1d1 and https://www.sciencebase.gov/catalog/item/5b96c2f9e4b0702d0e826f6d.
The shapefiles of the Watershed Boundary Dataset (WBD) were downloaded from https://apps.nationalmap.gov/downloader/#/.
The dam data were obtained from the US National Inventory of Dams (USNID) dataset (https://nid.sec.usace.army.mil).
Data on the fatalities and economic losses due to flooding in each country are from the EM-DAT database: https://public.emdat.be/data.
Major cities: https://hub.arcgis.com/datasets/esri::usa-major-cities/explore?location=37.154357%2C-113.479736%2C4.38.
Our organized data are available from GitHub https://github.com/peironglinlin/leveeRS/blob/main/data/COMID_systemID_intersection.csv and https://github.com/peironglinlin/leveeRS/blob/main/processed_data/Processed_data_for_Fig1c_and_Fig1d.csv.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☑ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n/a
Data exclusions	n/a
Replication	n/a
Randomization	n/a
Blinding	n/a

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	n/a
Research sample	n/a
Sampling strategy	n/a
Data collection	n/a
Timing	n/a
Data exclusions	n/a
Non-participation	n/a
Randomization	n/a

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	A large-scale study covering the contiguous US
Research sample	1279 levee systems
Sampling strategy	All available data meeting the calculation criteria.
Data collection	All available data meeting the research purpose.
Timing and spatial scale	1938-2005, Contiguous US
Data exclusions	If levee protected area is smaller than one pixel
Reproducibility	Reproducible.
Randomization	n/a
Blinding	n/a

Did the study involve field work? ☐ Yes ☑ No

Field work, collection and transport

Field conditions	n/a
Location	n/a
Access & import/export	n/a
Disturbance	n/a

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
X	☐	Antibodies
X	☐	Eukaryotic cell lines
X	☐	Palaeontology and archaeology
X	☐	Animals and other organisms
X	☐	Clinical data
X	☐	Dual use research of concern
X	☐	Plants

Methods

n/a	Involved in the study	
X	☐	ChIP-seq
X	☐	Flow cytometry
X	☐	MRI-based neuroimaging

Antibodies

Antibodies used	n/a
Validation	n/a

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	n/a
Authentication	n/a
Mycoplasma contamination	n/a
Commonly misidentified lines (See ICLAC register)	n/a

Palaeontology and Archaeology

Specimen provenance	n/a
Specimen deposition	n/a
Dating methods	n/a

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	n/a
Wild animals	n/a
Reporting on sex	n/a
Field-collected samples	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	n/a
Study protocol	n/a
Data collection	n/a
Outcomes	n/a

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
X	☐	Public health
X	☐	National security
X	☐	Crops and/or livestock
X	☐	Ecosystems
X	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
X	☐	Demonstrate how to render a vaccine ineffective
X	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
X	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
X	☐	Increase transmissibility of a pathogen
X	☐	Alter the host range of a pathogen
X	☐	Enable evasion of diagnostic/detection modalities
X	☐	Enable the weaponization of a biological agent or toxin
X	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a

ChIP-seq

Data deposition

n/a ☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

n/a ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	n/a
Files in database submission	n/a
Genome browser session (e.g. UCSC)	n/a

Methodology

Replicates	n/a
Sequencing depth	n/a
Antibodies	n/a
Peak calling parameters	n/a
Data quality	n/a
Software	n/a

Flow Cytometry

Plots

Confirm that:

n/a ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

n/a ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

n/a ☐ All plots are contour plots with outliers or pseudocolor plots.

n/a ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	n/a
Instrument	n/a
Software	n/a
Cell population abundance	n/a
Gating strategy	n/a

n/a ☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	n/a
Design specifications	n/a
Behavioral performance measures	n/a

Imaging type(s)	n/a
Field strength	n/a
Sequence & imaging parameters	n/a
Area of acquisition	n/a

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	n/a
Normalization	n/a
Normalization template	n/a
Noise and artifact removal	n/a
Volume censoring	n/a

Statistical modeling & inference

Model type and settings	n/a
Effect(s) tested	n/a

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

n/a

Correction

n/a

Models & analysis

n/a	Involved in the study
x	Functional and/or effective connectivity
x	Graph analysis
x	Multivariate modeling or predictive analysis

Functional and/or effective connectivity

n/a

Graph analysis

n/a

Multivariate modeling and predictive analysis

n/a