scientific data

OPEN

DATA DESCRIPTOR

The changing face of floodplains in the Mississippi River Basin detected by a 60-year land use change dataset

Adnan Rajib[1]✉, Qianjin Zheng[1], Heather E. Golden[2], Qiusheng Wu[3], Charles R. Lane[4], Jay R. Christensen[2], Ryan R. Morrison[5], Antonio Annis[6] & Fernando Nardi[6,7]

Floodplains provide essential ecosystem functions, yet >80% of European and North American floodplains are substantially modified. Despite floodplain changes over the past century, comprehensive, long-term land use change data within large river basin floodplains are limited. Long-term land use data can be used to quantify floodplain functions and provide spatially explicit information for management, restoration, and flood-risk mitigation. We present a comprehensive dataset quantifying floodplain land use change along the 3.3 million km² Mississippi River Basin (MRB) covering 60 years (1941–2000) at 250-m resolution. We developed four unique products as part of this work, a(n): (i) Google Earth Engine interactive map visualization interface, (ii) Python code that runs in any internet browser, (iii) online tutorial with visualizations facilitating classroom code application, and (iv) instructional video demonstrating code application and database reproduction. Our data show that MRB's natural floodplain ecosystems have been substantially altered to agricultural and developed land uses. These products will support MRB resilience and sustainability goals by advancing data-driven decision making on floodplain restoration, buyout, and conservation scenarios.

Background & Summary

Riverine floodplains are vital and productive ecosystems that provide essential biological, geomorphic, and hydrologic functions[1,2]. Services provided by floodplains – including regulation of disturbances (e.g., flood attenuation), water supply, and waste treatment – are valued at approximately US$1.5 × 10^{12} yr^{-1} globally (in 2007 US$)[3]. Yet floodplains are continually threatened by human development and encroachment, including loss of floodplain-river connectivity due to channelization and levee construction[4], which exacerbates habitat loss[5] and hydrologic alteration[6].

Human modifications to floodplains include changes in land use from activities such as urbanization, agriculture, industry, and mining[7]. For instance, approximately 80–90% of floodplains across Europe have been intensively cultivated[8], and 90% of floodplains in North America are non-functional due to cultivation[9]. New developments in floodplains expose an increased population in the United States to flooding[10], and even a 1% chance of flooding can cause losses exceeding $78 billion per year in the US[11].

Flood-risk management efforts of the previous century have focused on minimizing flood impacts on humans through large and expensive infrastructure projects[12] at the expense of floodplain ecosystem health and resilience. However, programs such as floodplain buyouts and conservation can produce co-benefits for economies and

[1]Department of Environmental Engineering, Texas A&M University, Kingsville, Texas, USA. [2]U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, Ohio, USA. [3]Department of Geography, University of Tennessee, Knoxville, Tennessee, USA. [4]U.S. Environmental Protection Agency, Office of Research and Development, Athens, Georgia, USA. [5]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, Colorado, USA. [6]Water Resources Research and Documentation Center, University for Foreigners of Perugia, Perugia, Italy. [7]Institute of Environment and College of Arts, Sciences & Education, Florida International University, Miami, Florida, USA. ✉e-mail: adnan.rajib@tamuk.edu

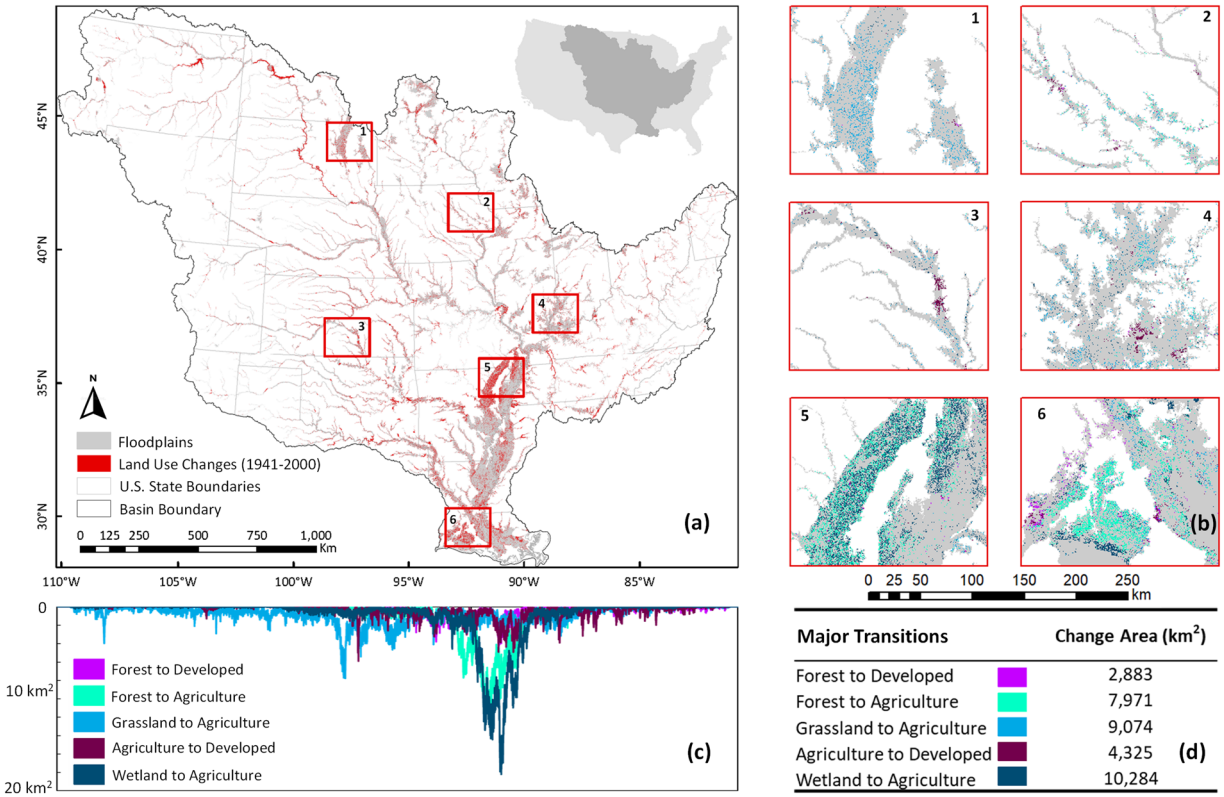

Fig. 1 Land use change in the Mississippi River Basin (MRB) floodplains between 1941 and 2000. (**a**) The "change" in this map is defined as the non-uniqueness of individual land use grid cells between the two end years. (**b**) The maps 1–6 correspond to six objectively chosen domains across different geophysical settings and stream orders in South Dakota (1), Iowa (2), Kansas (3), Indiana (4), Arkansas (5), and Louisiana (6). These maps show five major land transitions in MRB floodplains that are potentially irreversible (e.g., wetland → agriculture), posing negative impacts on floodplain ecohydrology and resilience. Plot (**c**) graphically shows how the five major potentially irreversible land transitions vary along the latitude at every 250-m horizontal resolution. Plot (**d**) summarizes the areal extent (km^2) of change between 1941 and 2000. The changes demonstrated in Fig. 1 can be further visualized at this interactive map interface: https://gishub.org/mrb-floodplain.

floodplain ecosystems[13]. Yet they require a comprehensive understanding of the history of floodplain changes along the full river continuum to ensure sustainable and effective floodplain and flood-risk management[14,15].

Despite the human-induced changes in floodplains over the past century, comprehensive data of long-term land use change within floodplains of large river basins are limited[16]. A recent large-scale study assessed floodplain conditions across England from 1900 to 2015 using land use data[17]. Others have focused on changes across smaller geographic expanses and shorter time scales, such as floodplain losses in Dhaka, Bangladesh[18] and Kumasi, Ghana[19]. No studies, to our knowledge, have integrated long-term (> 30 year) data to examine changes in floodplain land use across a large river basin. Data of long-term and large-scale floodplain land use are required (1) to effectively quantify floodplain functions and development trajectories, and (2) for a holistic perspective on the future of floodplain management and restoration and concomitantly flood-risk mitigation.

Here, we present the first available dataset to our knowledge that quantifies land use change along the floodplains of the Mississippi River Basin (MRB) covering 60 years (1941–2000) at 250-m resolution. The MRB is the fourth largest river basin in the world (3,288,000 km^2) comprising 41% of the United States and draining into the Gulf of Mexico, an area with an annually expanding and contracting hypoxic zone resulting from basin-wide over-enrichment of nutrients[20]. The basin represents one of the most engineered systems in the world, and includes a complex web of dams, levees, floodplains, and dikes. This new dataset reveals the heterogenous spatial extent of land use transitions in MRB floodplains. The floodplains have transitioned from natural ecosystems to predominantly agricultural land use (e.g., more than 10,000 km^2 of wetland loss due to agricultural expansion; Fig. 1). Developed land use within floodplain has also steadily increased (Fig. 2). These irreversible transitions in floodplain composition reduce storage[21] and conveyance[22] of natural flow, amplify flood risks posed by climate change[23,24], and hinder both ecosystems and human well-being[25].

To maximize the reuse of this dataset, we also include four unique products: (i) a Google Earth Engine interactive interface mapping MRB floodplain land use change over 60 years, (ii) a Google-based Python code that runs in any internet browser, (iii) an online tutorial with visualizations facilitating classroom application of the code, and (iv) an instructional video showing how to run the code and partially reproduce the dataset. We share all data through HydroShare: https://doi.org/10.4211/hs.41a3a9a9d8e54cc68f131b9a9c6c8c54.

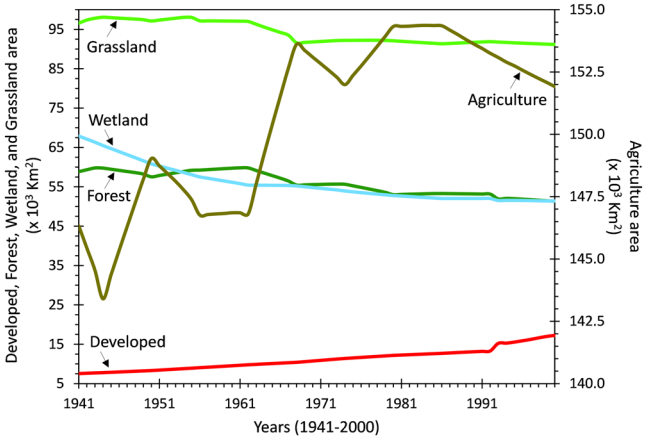

Fig. 2 Time series graphs showing 60 years (1941–2000) of continuous changes in different land use classes. While values in this figure are aggregated for the entire Mississippi River Basin floodplains, time-series of changes for each of the six domains shown in Fig. 1 are separately plotted in Supplementary Fig. 1.

The 60 years of spatially explicit floodplain land use change data produced herein are usable for flood-risk and nutrient management and research across the 31 US states that drain the MRB. A recent strategic plan of the Upper Mississippi River Restoration partnership, representing 0.5 million km^2 of the MRB, envisions "a healthier and more resilient Upper Mississippi River ecosystem that sustains the river's multiple use"[26]. This data will help achieve these goals by providing foundational information for data-driven decision making on floodplain restoration, buyouts, and conservation. Importantly, the data and associated materials can be the template for developing similar datasets for other river basins across the globe.

Methods

Input Data Sources. We derived the 60-year MRB floodplain land use change dataset from two input data sources: (i) the high-resolution global floodplain extent dataset GFPLAIN250m developed by Nardi *et al.*[27], and (ii) the annual continental United States land use data developed by USGS[28,29]. GFPLAIN250m is based on a geomorphic analysis of the Digital Terrain Model (DTM) identifying riparian areas underlying maximum flood levels. The GFPLAIN algorithm estimates distributed flood energy gradients, at the river basin scale, with a simplified hydrologic model that assigns every channel cell a maximum flood depth using the drainage area as a scaling variable[30,31]. Conceptually, this algorithm dissects floodplains from surrounding hillslopes as those low-lying landscape features that have been naturally shaped by accumulated geomorphic effects of past flood events. Therefore, the GFPLAIN250m dataset is built on the principle that a flood-prone area is implicitly contained in the DTM, decoupling the floodplain extent zoning from the need to preliminarily define a design flood event. The outcome is a DTM-based morphometric indexing of floodplain domains rather than a simulation associated to a specific return period, e.g., 100-year[32]. The dataset is publicly available at 250-m spatial resolution gridded GeoTIFF format, with coordinates set by World Geodetic System 1984 (WGS84).

The USGS land use data is based on a spatially explicit modeling framework which reconstructed a temporally continuous land use from widely acknowledged baselines including the National Land Cover Databases (NLCD)[33–35], more than 100 years of agricultural census information[36], and three decades of representative satellite imagery[37]. The dataset is publicly available at 250-m spatial resolution in gridded GeoTIFF format, with coordinates set by USA Contiguous Albers Equal Area Conic USGS version projected system. This land use dataset is divided into two parts: a 14-class historical land use for each year from 1938 to 1992[28] and a 17-class recent land use for each year from 1992 to 2005[29]. We included the land use from 1941 to 2000 in our approach to develop the 60-year MRB floodplain land use change dataset.

Procedure. Our methodology followed six consecutive steps (Fig. 3): (i) reprojection of floodplain and land use data to a consistent coordinate system, (ii) land use reclassification, (iii) extraction of floodplain land use, (iv) land use change detection, (v) formation of inter-class land transition matrix, and (vi) technical validation. These steps are discussed in detail below. All the associated tasks were performed in ArcGIS 10.5 and ENVI 5.1 geospatial analysis platforms.

(i) *Reprojection of coordinate systems*: The two primary inputs used in our approach, i.e., the global floodplain and annual USGS land use, were originally developed in two different coordinate systems. Because non-identical coordinate systems across corresponding datasets induce positional error in floodplain geospatial analysis especially across large river basins[38,39], we reprojected the coordinate system of the global floodplain to that of the USGS land use such that the two datasets become interoperable.

(ii) *Land use reclassification*: Commonly used land use datasets follow classification schemes (i.e., categorizing the intended purpose of a landscape parcel[40]) with multiple levels of nested hierarchy[41]. While a detailed land use classification offers critical insights to environmental monitoring and restoration research[42,43], the large semantic variability of land use classifications across disciplines often complicates their practical

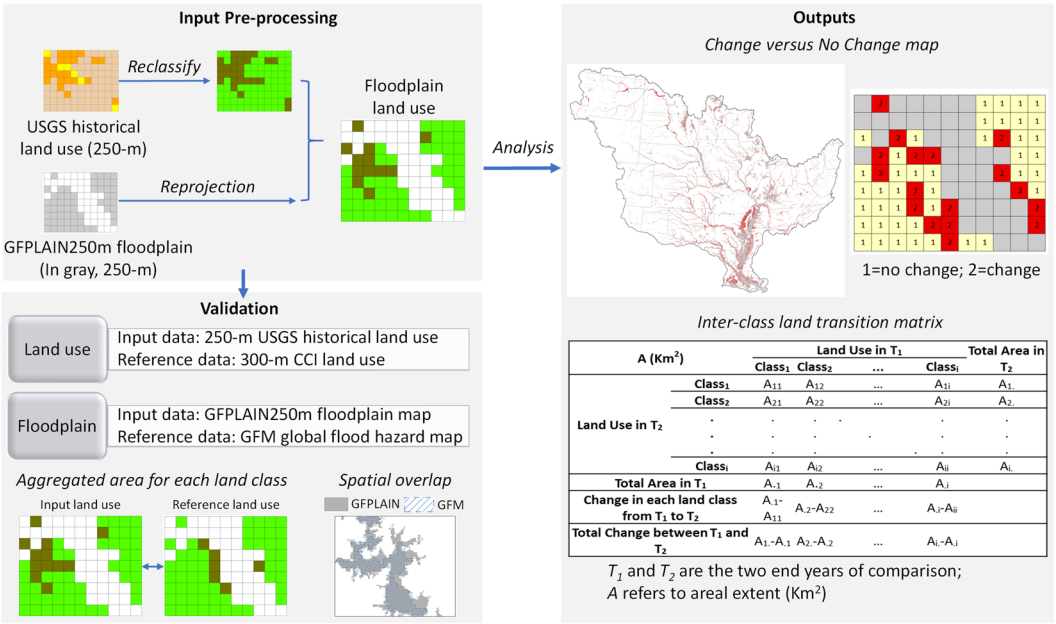

Fig. 3 Schematic of the methodological design.

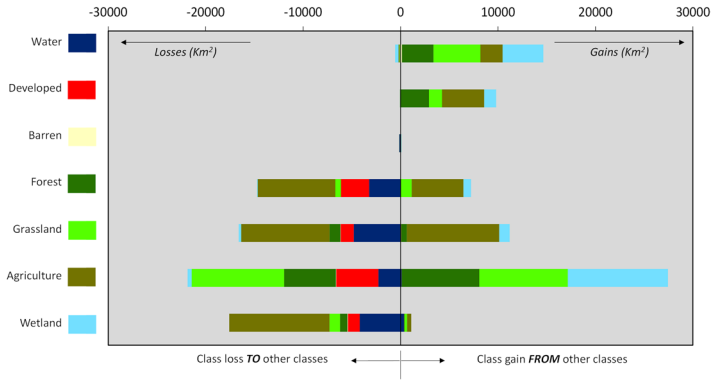

Fig. 4 Inter-class land loss versus gain between 1941 and 2000 for the entire MRB floodplains.

applications[44]. To allow easy integration of our land use change dataset with cross-disciplinary research and decision-making tools, we simplified the original classification scheme of the USGS land use data to produce seven generic classes. The new land use classes included: 1) open water, 2) developed area, 3) barren land, 4) forest, 5) grassland, 6) agriculture, and 7) wetland.

(iii) *Extraction of 60-year floodplain land use*: We used the MRB boundary polygon as a mask and clipped the MRB portion of the global floodplain (hereafter, the MRB floodplain). Following the same approach in a subsequent step, we used the MRB floodplain as a mask on the USGS land use and extracted a series of floodplain land use maps for each of the years from 1941 to 2000. We then calculated the areal extents of different land use classes by multiplying their corresponding total number of grid-cells with the spatial resolution of a single grid-cell (250*250 m^2), thus creating 60-year time-series of floodplain land use in the MRB (e.g., Fig. 2 and Supplementary Fig. 1). Since the USGS land use dataset was developed only for the continental United States, the small upstream portion of the MRB that drains two Canadian provinces (~1% of the basin's drainage area) was excluded from our analysis.

(iv) *Land use change detection between two end-years (1941 and 2000)*: We applied a statistical approach to detect the difference/non-uniqueness in land use between the two end-years of comparison (i.e., 1941 and 2000). Specifically, we calculated the number of unique grid-cell values between the two land use maps on a cell-by-cell basis. The outcome was a new map with only two possible values in the grid-cells. Value "1" indicated one unique value of a target grid-cell across the two input land use maps, meaning "no change" of land use between two points in time. Conversely, value "2" indicated that a target grid-cell had two non-unique values and hence a "change" of land use between two points in time.

(v) *Formation of inter-class land transition matrix*: To demonstrate the "nature of change"[45] in the MRB floodplains, we quantified how the land use therein transitioned from one class to the other(s) between two

		Area in year T_1*							
		$Class_1$	**$Class_2$**	**$Class_3$**	**$Class_4$**	**$Class_5$**	**$Class_6$**	**$Class_7$**	**Total area in T_2****
Area in year T_2*	**$Class_1$**	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}	A_{17}	$A_{1.}$
	$Class_2$	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	A_{26}	A_{27}	$A_{2.}$
	$Class_3$	A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	A_{36}	A_{37}	$A_{3.}$
	$Class_4$	A_{41}	A_{42}	A_{43}	A_{44}	A_{45}	A_{46}	A_{47}	$A_{4.}$
	$Class_5$	A_{51}	A_{52}	A_{53}	A_{54}	A_{55}	A_{56}	A_{57}	$A_{5.}$
	$Class_6$	A_{61}	A_{62}	A_{63}	A_{64}	A_{65}	A_{66}	A_{67}	$A_{6.}$
	$Class_7$	A_{71}	A_{72}	A_{73}	A_{74}	A_{75}	A_{76}	A_{77}	$A_{7.}$
Total area in T_1* **		$A_{.1}$	$A_{.2}$	$A_{.3}$	$A_{.4}$	$A_{.5}$	$A_{.6}$	$A_{.7}$	
Change in each land class between T_1 to T_2		$A_{.1}$-A_{11}	$A_{.2}$-A_{22}	$A_{.3}$-A_{33}	$A_{.4}$-A_{44}	$A_{.5}$-A_{55}	$A_{.6}$-A_{66}	$A_{.7}$-A_{77}	
Total change between T_1 and T_2		$A_{1.}$-$A_{.1}$	$A_{2.}$-$A_{.2}$	$A_{3.}$-$A_{.3}$	$A_{4.}$-$A_{.4}$	$A_{5.}$-$A_{.5}$	$A_{6.}$-$A_{.6}$	$A_{7.}$-$A_{.7}$	

Table 1. Schematic of the inter-class land transition matrix. *T_1 and T_2 are the two end-years of comparison; A refers to areal extent for a land use class (Km^2); diagonal cells in the table indicate "no change" between T_1 and T_2; **Sum of values in a row; ***Sum of values in a column.

Folder Name: Input Data				
ID	**Subfolder/File Name**	**File Type**	**Content Description**	**Provenance**
1	MRB_floodplain	GIS raster*	Riverine floodplain extents within the Mississippi River Basin (MRB), clipped from the GFPLAIN250m floodplain dataset • 250-m grid, GeoTIFF format • Albers Equal Area Conic projected coordinate	Nardi *et al.*[27].
2	MRB_floodplain_LU	GIS raster*	Input land use dataset for change detection • Clipped for the MRB floodplain extent • Modified to have seven generic land use classes • One corresponding dataset for each of the years from 1941 to 2000	Sohl *et al.*[28,29]; Supplementary Table 2
3	Reference_LU	GIS raster*	Reference land use dataset used for validation purposes (see items #10 and 11 in Table 3) • Clipped for the MRB floodplain extent • Modified to have same land use classes as in item #2 • One corresponding dataset for each of the years from 1992 to 2000	European Space Agency[58]; Supplementary Table 3
4	Reference_floodplain	GIS raster	Riverine floodplain extents within the MRB, clipped from the European Commission Joint Research Centre Global Flood Maps (GFM) • 1000-m grid, GeoTIFF format	Dottori *et al.*[56]

Table 2. Input dataset file descriptions. Note: All GIS raster and shapefile datasets are in the Albers Equal Area Conic projected coordinate system. *The raster datasets are in GeoTIFF format at 250-m spatial resolution (except item # 4).

end-years (1941 and 2000). We conducted this task using a widely acknowledged approach called *Transition Matrix* Analysis[33,46–53]. Table 1 schematically shows the resultant transition matrix across the seven land use classes of the MRB floodplains. Here, T_1 and T_2 respectively indicate the two end-years of comparison, while A_{ij} is the areal extent [L^2] that transitioned from class i at the initial year to class j at the final year. The last row in the transition matrix represents the net gain or loss of areal extent between T_1 and T_2 in every land use class. The inter-class land use transitions in the MRB floodplains between 1941 and 2000 are graphically presented in Fig. 4, while the corresponding calculations are provided in Supplementary Table 1.

Data Reco rds

The MRB floodplain land use change dataset is made available through an open-access geospatial data sharing platform HydroShare. Our archive also includes all corresponding input data, intermediate calculations, and supporting information. Tables 2 and 3 below provide an overview of the file contents. The entire archive can be downloaded as a single zip file from this web address: https://doi.org/10.4211/hs.41a3a9a9d8e54cc68f131b9a9c6c8c54[54].

Technical Validation

To ensure the technical quality and reliability of our MRB floodplain land use change dataset, we validated both GFPLAIN250m floodplain and USGS land use datasets with respect to the best available references.

Although the GFPLAIN250m floodplain[27] has been previously validated across different scales[55], we further compared it to the Global Flood Maps (GFM)[56] across the MRB Hydrologic Unit Codes (HUCs). We used critical success index (CSI) and true positive rate (TP rate) metrics to confirm the spatial consistency between GFPLAIN250m and GFM floodplain extents (Fig. 5). When compared to GFM, GFPLAIN tends to produce larger floodplain delineations in the headwater regions of the MRB (Fig. 5a). This was also evident in Fig. 5b,which

Folder Name: Output Data				
ID	**Subfolder/File Name**	**File Type**	**Content Description**	**Output Figure/Table**
5	ChangeMap	GIS raster*	A map showing the changes in MRB floodplain land use between 1941 and 2000; "change" is defined as the non-uniqueness of individual land use grid-cells • Grid value 1 indicates "no change" and 2 indicates "change"	Fig. 1a
6	ClassTransitionMaps	GIS raster*; shapefile	A map showing major land use class-transitions in MRB floodplains between 1941 and 2000 • Five irreversible transitions (e.g., wetland → agriculture) that would negatively impact floodplain ecohydrology and resilience Also includes boundary polygons for six objectively chosen domains to facilitate focused assessments	Fig. 1b
7	ClassTransition_ LatLong	MS Excel	Total area (km^2) for each of the five irreversible transitions mentioned in item #6 • Separate calculations along the latitudinal and longitudinal directions at every 250-m spacing	Fig. 1c
8	ClassTransitionMatrix	MS Excel	Inter-class transitions (km^2) across all seven generic land use classes between 1941 and 2000; based on the schematic presented in Table 1	Supplementary Table 1; Fig. 1d; Fig. 4
9	Timeseries	MS Excel	Timeseries of total area (km^2) for each of the seven generic land use classes • 60-year timeseries with corresponding values for every year from 1941 to 2000 • Seven datasets: the entire MRB floodplains and the six domains mentioned in item #6	Fig. 2; Supplementary Fig. 1
10	ValidationMaps	MS Excel; GIS Shapefile	Measures of fit comparing the input and reference floodplain extents (Table 2 – items #1 and 4 respectively) across the MRB • An Excel file showing 8-digit Hydrologic Unit Codes (HUC) and corresponding TP and CSI values • HUC-8 boundary polygon shapefile	Fig. 5b,c
		GIS raster*	Validation maps showing the spatial consistency between input and reference land use datasets (Table 2 – items #2 and 3 respectively) • 12 validation domains • Maps correspond to different years between 1992 and 2000 • File nomenclature: *x_y_z*; *x* = USGS (input) or Remotely sensed (reference) land use, *y* = US state where the validation domain is located; *z* = year	Fig. 6; Supplementary Figs. 2–4
11	ValidationResults	MS Excel	Total area (km^2) for different land use classes at each of the 12 validation domains, calculated separately from the two contrasting land use datasets (item #10) • Includes calculation of correlations (R^2) between two land use datasets	

Table 3. Output dataset file descriptions. Note: All GIS raster and shapefile datasets are in the Albers Equal Area Conic projected coordinate system. *The raster datasets are in GeoTIFF format at 250-m spatial resolution (except Table 2 item# 4).

showed lower CSI values in the headwater HUCs. Values of CSI were within acceptable ranges demonstrated by previous studies[27,57], however. TP rates (Fig. 5c) were greater than 0.8 for most of the HUCs, indicating a high overlap between GFPLAIN250m and GFM datasets. These assessments demonstrate the applicability of using the GFPLAIN250m data to identify continuous floodplains.

We validated the USGS land use using the European Space Agency's (ESA) Climate Change Initiative (CCI) data. The CCI data include a time-series of consistent global land use maps at 300-m spatial resolution on an annual basis from 1992 to 2019[58]. These global land use maps were derived from multiple satellite data sources, including Envisat Medium Resolution Imaging Spectrometer (MERIS) (2003–2012), Advanced Very High Resolution Radiometer (AVHRR) (1992–1999), SPOT-VGT (1999–2013), and PROBA-V (2013–2015)[59] (hereafter, we refer CCI land use as the *remotely sensed land use* for simplicity). In contrast, the 250-m spatial resolution USGS land use maps were based on a hindcast modeling (1938–2005), derived from the NLCD, Landsat satellite, and county-level agricultural census. We chose the CCI/remotely sensed dataset as the reference because it was developed by a different agency with different data sources, which ensured an independent validation of our land use change estimates (see Fig. 6).

The USGS land use contains 17 classes (Supplementary Table 2), whereas the remotely sensed land use contains 37 classes (Supplementary Table 3). To make these two datasets comparable, we reclassified them into seven generic classes, including open water, developed area, barren land, forest, grassland, agriculture, and wetland (see Supplementary Tables 2, 3). In addition, we reprojected the CCI coordinates from World Geodetic System 84 (WGS84) to USA Contiguous Albers Equal Area Conic projected system to enable a uniform comparison with the USGS data. After the reclassification and reprojection steps, we selected 12 sites to validate our MRB floodplain land use change dataset. These validation sites were chosen objectively to represent different geophysical settings across the MRB as well as different stream orders, including both major rivers and lower order tributaries. During the common period of data availability between USGS (input) and remotely sensed (reference) datasets, we conducted validations for the 12 selected sites from 1992 to 2000, with a different year randomly assigned to each validation site. It should be noted that the comparison was conducted at an aggregated level for each site. We did not evaluate the cell-by-cell correlations for two reasons. First, the two datasets were developed at two

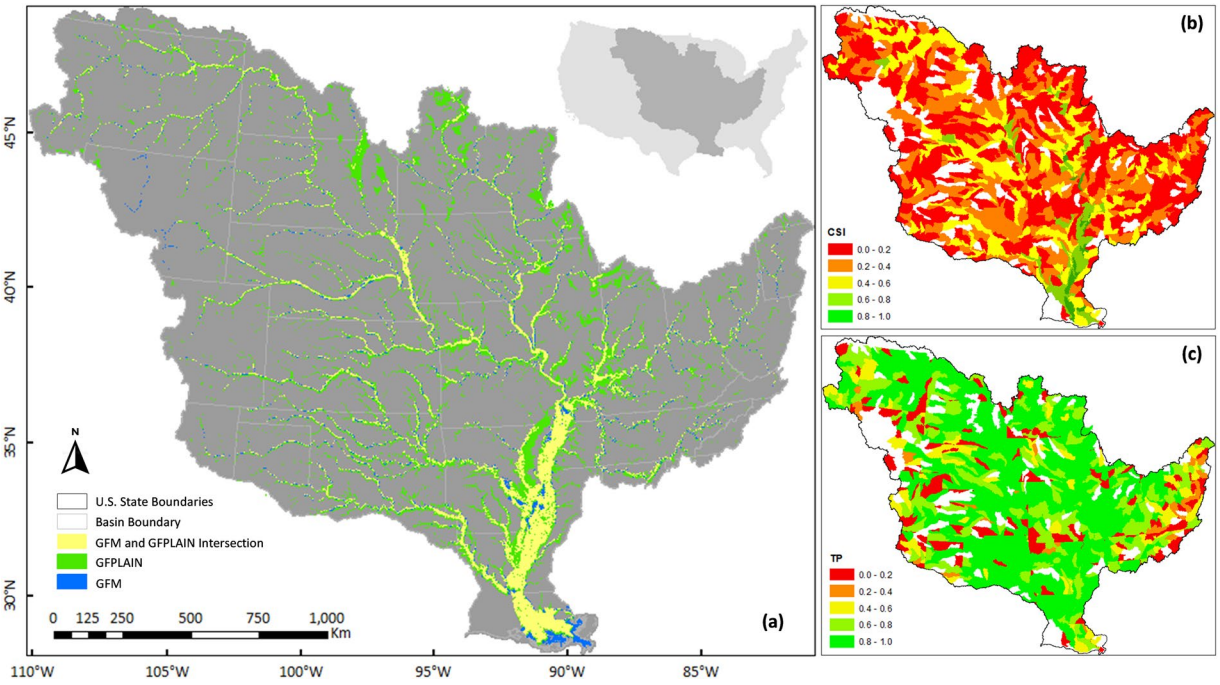

Fig. 5 Consistency between GFPLAIN250m (input) and GFM (reference) floodplain extents. (**a**) Yellow areas show the intersection between the two datasets[27,56]. (**b**,**c**) Map of the Critical Success Index (CSI) and True Positive Rate (TP rate) for each Hydrologic Unit watershed where the two contrasting floodplain datasets are intersected. CSI = A/(A + B + C), TP rate = A/(A + C); here A, B, and C – the yellow, green, and blue areas in map (**a**), respectively, indicate the matches, areas covered by the GFPLAIN250m only, and areas covered by the GFM only.

	Product	Access Link and Intended Objectives
1	Web interface	https://gishub.org/mrb-floodplain A Google Earth Engine interface facilitating the interactive visualization of major land use class-transitions in MRB floodplains (see item #6 in Table 3).
2	Semi-automatic coding framework	https://colab.research.google.com/drive/1vmIaUCkL66CoTv4rNRIWpJXYXp4TlAKd?usp = sharing A ready-to-use python code that operates entirely in Google's web-based high-performance programming platform called *Google Colaboratory*. • Allows users to reproduce the MRB floodplain land use change dataset (up to the item #5 listed in Table 3); users can run the code simply in a web browser without requiring to write any new code or setting up a programming environment in users' local computers. • Not limited to change detection only; an end-to-end workflow that performs all data discovery, download, and pre-processing tasks in a semi-automatic manner. • Scalable to the floodplains in any of the world's major river basins. In addition to the specific land use input[28,29] used in our work to generate the MRB floodplain land use change dataset, the current version of our code can also assimilate land use input from two different sources: 30-m National Land Cover Database (United States)[33–35] and 300-m Climate Change Initiative Land Cover Database (global)[58].
3	Classroom tutorial	https://serc.carleton.edu/hydromodules/steps/241489.html A web-based tutorial offering step-by-step instructions to run the Google Colaboratory python code, partially reproduce the MRB floodplain land use change dataset, and perform offline visualization of the output dataset in ArcGIS desktop; built to fit classroom instruction modules.
4	Instructional video	https://youtu.be/wH0gif_y15A A You Tube video that shows how to run the Google Colaboratory python code following our web-based tutorial; facilitates more efficient reproduction of the dataset.

Table 4. Software solutions and educational materials for enhancing the FAIR data properties.

different spatial resolutions (250-m and 300-m). Second, the respective definitions of land classes are not identical across the two datasets, although we reclassified them to bring some degree of consistency. Three validation sites are shown in Fig. 6, and the other nine validation sites are shown in Supplementary Figs. 2–4.

The validation results for the 12 selected sites show high correlations between the USGS and remotely sensed data across all land use classes, with R^2 ranging from 0.90 to 0.99. Agricultural land and grassland appear to show the largest discrepancy between the USGS and remotely sensed data among the seven land use classes, which is largely due to the potential inconsistency in respective land class definition schemes between these two datasets. The original USGS classification of hay/pasture was designated as grassland while in the remotely sensed dataset, there is no hay/pasture class but a cropland with herbaceous cover which was designated as agriculture in our simplified

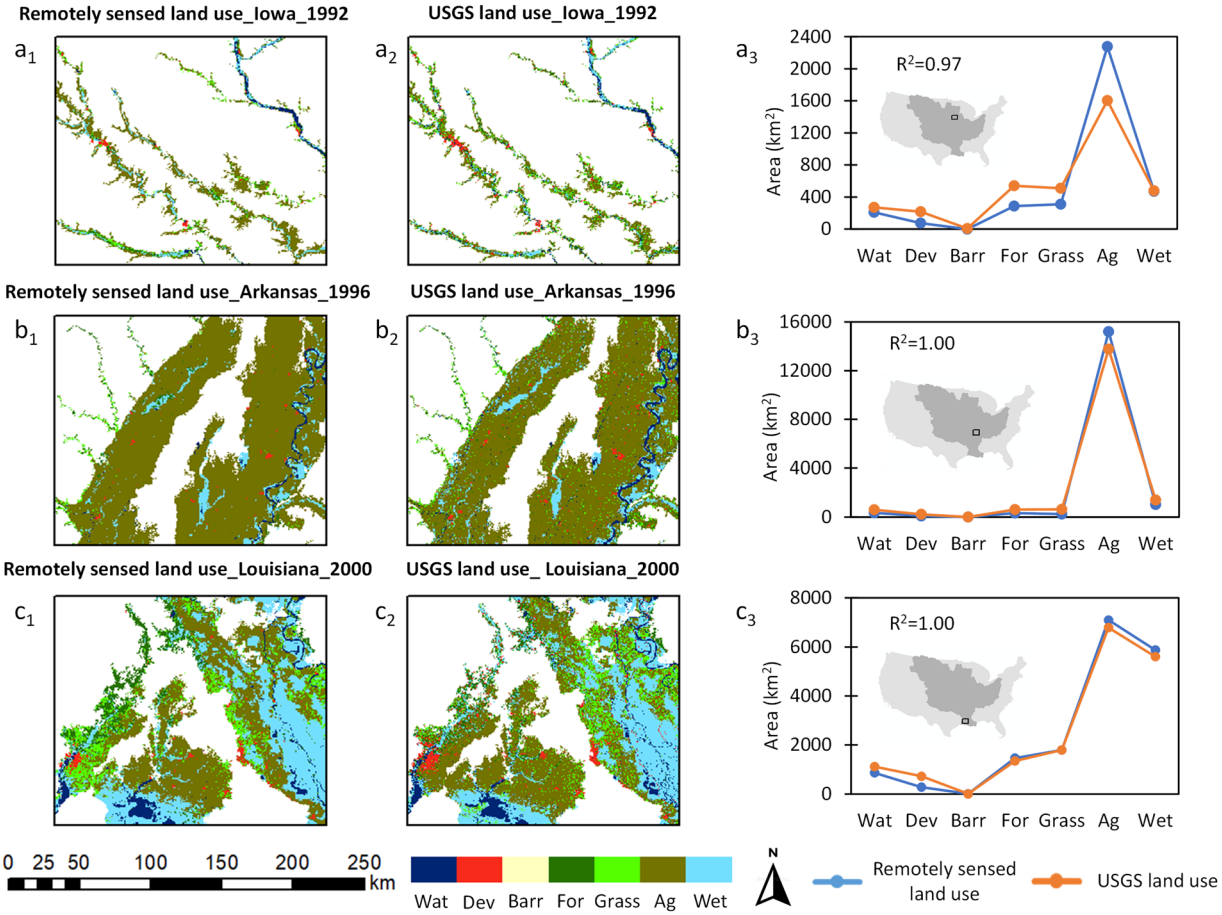

Fig. 6 Spatial comparison between USGS (input) and remotely sensed (reference) land use datasets in three different years. The subplots (**a**–**c**) correspond to zoomed-in portions of the Mississippi River Basin floodplains in Iowa, Arkansas, and Louisiana, respectively. a_1, b_1, c_1 show ESA's Climate Change Initiative (CCI) land use maps based on satellite observations (hereafter, the remotely sensed land use)[58], while a_2, b_2, c_2 show the land use data obtained from USGS land modeling framework[28,29]. The remotely sensed land use ($a_1 - c_1$) was our reference to validate the spatial consistency of the USGS land use ($a_2 - c_2$; the input land use data in our methodology). Subplots $a_3 - c_3$ show the correlation between remotely sensed and USGS datasets across different land use classes within the given spatial domains. The generic land use classes include water, developed, barren, forest, grassland, agriculture, and wetland (abbreviated as Wat, Dev, Barr, For, Grass, Ag, and Wet, respectively).

classifications (Supplementary Tables 2, 3). The other five land use classes are highly consistent across all validation sites. Overall, these validation results indicate that our input and accordingly our output datasets are sufficiently reliable.

Usage Notes

To ensure that the MRB floodplain land use change dataset, relevant inputs, and underlying methodology are Findable, Accessible, Interoperable, and Reproducible (FAIR)[60], we developed four software solutions and educational products (Table 4). These products, besides assisting other researchers with the reuse of our dataset, will also foster new research on floodplain resilience by allowing efficient analysis of floodplain land use change in any of the world's major river basins.

Code availability

The MRB floodplain land use change dataset is derived entirely through ArcGIS 10.5 and ENVI 5.1 geospatial analysis platforms (see *Methods* section for details). We developed additional open-access codes and visualization interfaces, however, to promote reproducibility and widespread application of the dataset. The python code is accessible at: https://colab.research.google.com/drive/1vmIaUCkL66CoTv4rNRIWpJXYXp4TlAKd?usp=sharing. The visualization interface is available online at: https://gishub.org/mrb-floodplain. See *Usage Notes* section for details.

Received: 25 February 2021; Accepted: 31 August 2021;
Published online: 15 October 2021

References

1. Junk, W. J., Bayley, P. B. & Sparks, R. E. The flood pulse concept in river-floodplain systems. In D. P. Dodge [ed.] *Proceedings of the International Large River Symposium*, *Canadian Special Publication of Fisheries and Aquatic Sciences* **106**, 110-127 https://www.waterboards.ca.gov/waterrights//water_issues/programs/bay_delta/docs/cmnt081712/sldmwa/junketal1989.pdf (1989).
2. Karpack, M. N., Morrison, R. R. & McManamay, R. A. Quantitative assessment of floodplain functionality using an index of integrity. *Ecological Indicators* **111**, 106051, https://doi.org/10.1016/j.ecolind.2019.106051 (2020).
3. Costanza, R. *et al.* Changes in the global value of ecosystem services. *Global Environmental Change* **26**, 152–158, https://doi.org/10.1016/j.gloenvcha.2014.04.002 (2014).
4. Wohl, E., Lane, S. N. & Wilcox, A. C. The science and practice of river restoration. *Water Resources Research* **51**, 5974–5997, https://doi.org/10.1002/2014WR016874 (2015).
5. Hamilton, S. K. Wetlands of Large Rivers: Flood plains. *Encyclopedia of Inland Waters* 607-610 https://doi.org/10.1016/B978-012370626-3.00065-X (2009).
6. Opperman, J. J., Luster, R., McKenney, B. A., Roberts, M. & Meadows, A. W. Ecologically functional floodplains: connectivity, flow regime, and scale. *Journal of the American Water Resources Association* **46**, 211–226, https://doi.org/10.1111/j.1752-1688.2010.00426.x (2010).
7. Waltham, N. J. *et al.* Lost floodplain wetland environments and efforts to restore connectivity, habitat, and water quality settings on the great barrier reef. *Front. Mar. Sci.* **6**, 71, https://doi.org/10.3389/fmars.2019.00071 (2019).
8. Tockner, K. & Stanford, J. A. Review of: riverine flood plains: present state and future trends. *Biological Sciences Faculty Publications* **29**, 166 https://scholarworks.umt.edu/biosci_pubs/166 (2002).
9. Erwin, K. L. Wetlands and global climate change: the role of wetland restoration in a changing world. *Wetlands Ecology and Management* **17**, 71, https://doi.org/10.1007/s11273-008-9119-1 (2009).
10. Johnson, K. A. *et al.* A benefit-cost analysis of floodplain land acquisition for US flood damage reduction. *Nat Sustain* **3**, 56–62, https://doi.org/10.1038/s41893-019-0437-5 (2019).
11. Quinn, N. *et al.* The spatial dependence of flood hazard and risk in the United States. *Water Resources Research* **55**, 1890–1911, https://doi.org/10.1029/2018WR024205 (2019).
12. Pinter, N. One step forward, two steps back on U.S. floodplains. *Science* **308**(5719), 207–208 https://science.sciencemag.org/content/308/5719/207 (2005).
13. Kousky, C. & Walls, M. Floodplain conservation as a flood mitigation strategy: examining costs and benefits. *Ecological Economics* **104**, 119–128, https://doi.org/10.1016/j.ecolecon.2014.05.001 (2014).
14. Tullos, D. Opinion: how to achieve better flood-risk governance in the United States. *Proceedings of the National Academy of Sciences of the United States of America* **115**(15), 3731–3734, https://doi.org/10.1073/pnas.1722412115 (2018).
15. Kundzewicz, Z. W., Hegger, D. L. T., Matczak, P. & Driessen, P. P. J. Opinion: flood-risk reduction: structural measures and diverse strategies. *Proceedings of the National Academy of Sciences of the United States of America* **115**(49), 12321–12325, https://doi.org/10.1073/pnas.1818227115 (2018).
16. Lambin, E. F., Geist, H. J. & Lepers, E. Dynamics of land-use and land-cover change in tropical regions. *Annual Review of Environment and Resources* **28**, 205–241, https://doi.org/10.1146/annurev.energy.28.050302.105459 (2003).
17. Entwistle, N. S., Heritage, G. L., Schofield, L. A. & Williamson, R. J. Recent changes to floodplain character and functionality in England. *Catena* **174**, 490–498, https://doi.org/10.1016/j.catena.2018.11.018 (2019).
18. Dewan, A. M. & Yamaguchi, Y. Land use and land cover change in Greater Dhaka, Bangladesh: using remote sensing to promote sustainable urbanization. *Applied Geography* **29**, 390–401, https://doi.org/10.1016/j.apgeog.2008.12.005 (2009).
19. Amoateng, P., Finlayson, C. M., Howard, J. & Wilson, B. Dwindling rivers and floodplains in Kumasi, Ghana: a socio-spatial analysis of the extent and trend. *Applied Geography* **90**, 82–95, https://doi.org/10.1016/j.apgeog.2017.11.007 (2018).
20. Rabalais, N. N., Turner, R. E. & Wiseman, W. J. Jr. Gulf of Mexico hypoxia, a.k.a. "the dead zone. *Annual Review of Ecology and Systematics* **33**, 235–263, https://doi.org/10.1146/annurev.ecolsys.33.010802.150513 (2002).
21. Wohl, E. An integrative conceptualization of floodplain storage. *Reviews of Geophysics* **59**, e2020RG000724, https://doi.org/10.1029/2020RG000724 (2021).
22. Scott, D. T., Gomez-Velez, J. D., Jones, C. N. & Harvey, J. W. Floodplain inundation spectrum across the United States. *Nat. Commun.* **10**, 5194, https://doi.org/10.1038/s41467-019-13184-4 (2019).
23. Hattermann, F. F. *et al.* Climatological drivers of changes in flood hazard in Germany. *Acta Geophysica* **61**, 463–477, https://doi.org/10.2478/s11600-012-0070-4 (2013).
24. Mallakpour, I. & Villarini, G. The changing nature of flooding across the central United States. *Nat. Clim. Change* **5**, 250–254, https://doi.org/10.1038/nclimate2516 (2015).
25. Corvalán, C., Hales, S., McMichael, A. J., Millennium Ecosystem Assessment (Program), & World Health Organization (Eds.). *Ecosystems and human well-being: Health synthesis* (World Health Organization, 2005).
26. Enhancing Restoration and advancing knowledge of the upper Mississippi river: a strategic plan for the upper Mississippi river restoration program 2015-2025. https://www.umesc.usgs.gov/ltrmp/documents/umrr_strategic_plan_jan2015.pdf (USGS, 2015).
27. Nardi, F., Annis, A., Di Baldassarre, G., Vivoni, E. R. & Grimaldi, S. GFPLAIN250m, a global high-resolution dataset of Earth's floodplains. *Scientific Data* **6**, 180309, https://doi.org/10.1038/sdata.2018.309 (2019).
28. Sohl, T. L. *et al.* Modeled historical land use and land cover for the conterminous United States: 1938-1992. *U.S. Geological Survey data release* https://doi.org/10.5066/F7KK99RR (2018).
29. Sohl, T.L. *et al.* Conterminous United States land cover projections - 1992 to 2100. *U.S. Geological Survey data release* https://doi.org/10.5066/P95AK9HP (2018).
30. Leopold, L. B., & Maddock, T. *The hydraulic geometry of stream channels and some physiographic implications*. (U.S. Geological Survey, 1953)
31. Nardi, F., Vivoni, E. R. & Grimaldi, S. Investigating a floodplain scaling relation using a hydrogeomorphic delineation method. *Water Resources Research* **42**(9), https://doi.org/10.1029/2005WR004155 (2006).
32. Di Baldassarre, G. *et al.* Brief communication: comparing hydrological and hydrogeomorphic paradigms for global flood hazard mapping. *Nat. Hazards Earth Syst. Sci.* **20**, 1415–1419, https://doi.org/10.5194/nhess-20-1415-2020 (2020).
33. Homer, C. *et al.* Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database. *ISPRS Journal of Photogrammetry and Remote Sensing* **162**, 184–199, https://doi.org/10.1016/j.isprsjprs.2020.02.019 (2020).
34. Yang, L. *et al.* A new generation of the United States national land cover database: requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing* **146**, 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006 (2018).
35. Jin, S. *et al.* Overall methodology design for the United States National Land Cover Database 2016 products. *Remote Sensing* **11**, 2971, https://doi.org/10.3390/rs11242971 (2019).
36. *USDA Census of Agriculture Historical Archive* http://agcensus.mannlib.cornell.edu/AgCensus/homepage.do;jsessionid=17C0132051BEB31DF79D01B0D07300F2 (US Department of Agriculture, 2007).
37. Sleeter, B. M. *et al.* Land-cover change in the conterminous United States from 1973 to 2000. *Global Environmental Change* **23**(4), 733–748, https://doi.org/10.1016/j.gloenvcha.2013.03.006 (2013).
38. Cao, Y. *et al.* Analysis of errors introduced by geographic coordinate systems on weather numeric prediction modeling. *Geosci. Model Dev.* **10**(9), 3425–3440, https://doi.org/10.5194/gmd-10-3425-2017 (2017).

39. Piwowar, J. M., Ledrew, E. F. & Dudycha, D. J. Integration of spatial data in vector and raster formats in a geographic information system environment. *International Journal of Geographical Information Systems* **4**, 429–444, https://doi.org/10.1080/02693799008941557 (2007).
40. Croissant, C. Landscape patterns and parcel boundaries: an analysis of composition and configuration of land use and land cover in south-central Indiana. *Agriculture Ecosystems and Environment* **101**, 219–232, https://doi.org/10.1016/j.agee.2003.09.006 (2004).
41. LaGro Jr., J. A. *Land-use Classification* (Elsevier Press, 2005).
42. Kutcher T. E. *et al*. *Habitat and Land Cover Classification Scheme for the National Estuarine Research Reserve System*. (National Estuarine Research Reserve System, 2008).
43. Buskey, E. J. *et al*. in *System-wide monitoring program of the national estuarine research reserve System: research and monitoring to address coastal management issues* Chapter 21 (Academic Press, 2015).
44. Feng, C.-C. & Flewelling, D. M. Assessment of semantic similarity between land use/land cover classification systems. *Computers, Environment and Urban Systems* **28**(3), 229–246, https://doi.org/10.1016/S0198-9715(03)00020-6 (2004).
45. Foufoula-Georgiou, E., Takbiri, Z., Czuba, J. A. & Schwenk, J. The change of nature and the nature of change in agricultural landscapes: Hydrologic regime shifts modulate ecological transitions. *Water Resources Research* **51**, 6649–6671, https://doi.org/10.1002/2015WR017637 (2015).
46. Biondini, M. & Kandus, P. Transition matrix analysis of land-cover change in the accretion area of the Lower Delta of the Paraná River (Argentina) reveals two succession pathways. *Wetlands* **26**, 981–991, https://link.springer.com/article/10.1672/0277-5212(2006)26[981:TMAOLC]2.0.CO;2#citeas (2006).
47. Hu, Y., Batunacun, Zhen, L. & Zhuang, D. Assessment of land-use and land-cover change in Guangxi, China. *Sci Rep.* **9**, 2189, https://doi.org/10.1038/s41598-019-38487-w (2019).
48. Liu, X. *et al*. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability* **3**, 564–570, https://doi.org/10.1038/s41893-020-0521-x (2020).
49. Teferi, E., Bewket, W., Uhlenbrook, S. & Wenninger, J. Understanding recent land use and land cover dynamics in the source region of the Upper Blue Nile, Ethiopia: spatially explicit statistical modeling of systematic transitions. *Agriculture, ecosystems & environment* **165**(15), 98–117, https://doi.org/10.1016/j.agee.2012.11.007 (2013).
50. Yu, Z., Guo, X., Zeng, Y., Koga, M. & Vejre, H. Variations in land surface temperature and cooling efficiency of green space in rapid urbanization: the case of Fuzhou city, China. *Urban forestry & urban greening* **29**, 113–121, https://doi.org/10.1016/j.ufug.2017.11.008 (2018).
51. Yuan, F., Sawaya, K. E., Loeffelholz, B. C. & Bauer, M. E. Land cover classification and change analysis of the twin cities (Minnesota) metropolitan area by multitemporal Landsat remote sensing. *Remote Sensing of Environment* **98**, 317–328, https://doi.org/10.1016/j.rse.2005.08.006 (2005).
52. Yuh, Y. G. *et al*. Effects of land cover change on great apes distribution at the Lobéké National Park and its surrounding forest management units, south-east Cameroon. A 13 year time series analysis. *Sci. Rep.* **9**, 1445, https://doi.org/10.1038/s41598-018-36225-2 (2019).
53. Zhao, J., Yang, Y., Zhao, Q. & Zhao, Z. Effects of ecological restoration projects on changes in land cover: a case study on the Loess Plateau in China. *Sci. Rep.* **7**, 44496, https://doi.org/10.1038/srep44496 (2017).
54. Rajib, A. *et al*. Land Use Changes in The Mississippi River Basin Floodplains: 1941 to 2000 (version 1). *HydroShare* https://doi.org/10.4211/hs.41a3a9a9d8e54cc68f131b9a9c6c8c54 (2021).
55. Annis, A., Nardi, F., Morrison, R. R. & Castelli, F. Investigating hydrogeomorphic floodplain mapping performance with varying DTM resolution and stream order. *Hydrological Sciences Journal* **64**(5), 525–538, https://doi.org/10.1080/02626667.2019.1591623 (2019).
56. Dottori, F. *et al*. Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources* **94**, 87–102, https://doi.org/10.1016/j.advwatres.2016.05.002 (2016).
57. Scheel, K., Morrison, R. R., Annis, A. & Nardi, F. Understanding the large-scale influence of levees on floodplain connectivity using a hydrogeomorphic approach. *Journal of the American Water Resources Association* **55**(2), 413–429, https://doi.org/10.1111/1752-1688.12717 (2019).
58. *Climate Change Initiative (CCI) Land Cover products* http://maps.elie.ucl.ac.be/CCI/viewer/download.php (2018).
59. *Land Cover CCI Product User Guide Version 2*. Tech. Rep. http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf (European Space Agency, 2017).
60. Wilkinson, M. D. *et al*. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018, https://doi.org/10.1038/sdata.2016.18 (2016).

Acknowledgements

A.R., Q.Z., and Q.W. were partially funded through the Department of Defense (award number FAIN # W912HZ2020050). R.R.M. was partially funded through the Colorado Water Center and the National Science Foundation (award number CBET-1916780). F.N. was partially funded through the National Science Foundation (award number HRD-1547798) as part of the Centers for Research Excellence in Science and Technology (CREST) Program at the Florida International University. The research presented was not performed or funded by the U.S. Environmental Protection Agency (EPA) and was not subject to EPA's quality system requirements. The views expressed in this article are those of the authors and do not necessarily represent the views or the policies of the EPA and the funding agencies.

Author contributions

A.R. and Q.Z. designed the research, performed the computational work, generated the graphics, and outlined the manuscript. H.E.G., C.R.L., and R.R.M. contextualized this work with the current state of science and assisted in manuscript development. Q.W. developed the map interface. J.R.C., Q.W., F.N., and A.A. performed technical validation of the data.

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/s41597-021-01048-w.

Correspondence and requests for materials should be addressed to A.R.

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