A River Ran Through It: Floodplains as America's Newest Relict Landform

Richard Knox[1], Ellen Wohl[2], and Ryan Morrison[1]

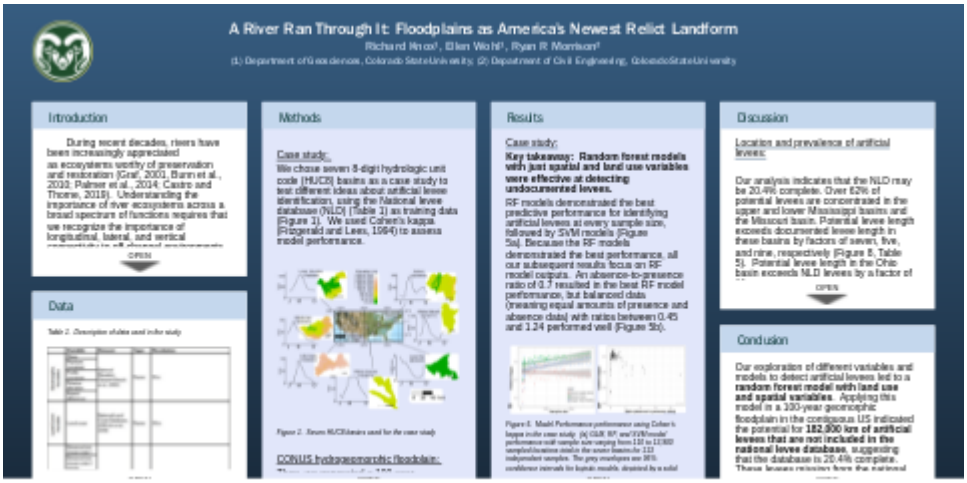
[1]Colorado State University
[2]Colorado State University Fort Collins

November 24, 2022

Abstract

Recent advances in Earth observation data and computing ability create exciting opportunities for national and global studies of human impacts to water resources. But, with a lack of complete databases of artificial levees, there remains a need to better understand how artificial levees impact floodplain extent at regional and larger scales. Here, we estimate river-floodplain disconnection in the contiguous United States using an incomplete artificial levee database, machine learning algorithms, and hydrogeomorphic floodplain delineation models. We tested different topographic, land use, and spatial variables with different machine learning techniques in a case study of seven geographically diverse HUC8 basins before applying the technique at the national scale. We found that a parsimonious random forest model without topographic variables was 97% accurate. When applied to areas within a national 100-year hydrogeomorphic floodplain, the model indicated the potential for more than 180,000 km of undocumented artificial levees, meaning that the National Levee Database (NLD) is about 20% complete. More than 62% of potential levees are concentrated in the Upper and Lower Mississippi and Missouri basins. The stream order distribution of potential and NLD levees are similar; however, potential levees are primarily located along stream orders 3 and 6 while the NLD locations are along stream orders 2, 3 and 4. Using this, we explored the national impacts of artificial levees on floodplain extent by comparing two hydrogeomorphic floodplains based on (1) an unmodified USGS 1 arc second DEM and (2) a modified DEM with known and potential levees erased from the topography. We found that the overall impact of artificial levee removal was to shift the location of flooding. Over 30% of the CONUS 100-year floodplain was cultivated or developed land use.

A River Ran Through It: Floodplains as America's Newest Relict Landform



Richard Knox[1], Ellen Wohl[1], Ryan R Morrison[2]

(1) Department of Geosciences, Colorado State University; (2) Department of Civil Engineering, Colorado State University

PRESENTED AT:

AGU FALL MEETING
New Orleans, LA & Online Everywhere
13–17 December 2021

Poster Gallery
brought to you by
WILEY

INTRODUCTION

During recent decades, rivers have been increasingly appreciated as ecosystems worthy of preservation and restoration (Graf, 2001, Bunn et al., 2010; Palmer et al., 2014; Castro and Thorne, 2019). Understanding the importance of river ecosystems across a broad spectrum of functions requires that we recognize the importance of longitudinal, lateral, and vertical connectivity to off-channel environments (Ward, 1989; Kondolf et al., 2006; Harvey and Gooseff, 2015). One anthropogenic feature that adversely impacts lateral connectivity is artificial levees, which can be defined as raised linear features built between active channels and floodplains to contain peak flows in the channel (Tobin, 1995). The length of artificial levees in the U.S. is unknown but estimates range between 48,000 and 167,000 km, corresponding to coverage of roughly 1% and 3% of total estimated river km in the contiguous US (Heine and Pinter, 2012; ASCE, 2017). The USACE started a national levee inventory in 2006, which resulted in the National Levee Database (NLD). The NLD is currently estimated to be 30% complete (ASCE, 2017), but a comprehensive evaluation of the NLD's thoroughness has not been completed (Wing et al., 2017). Consequently, there is no national-scale assessment of how artificial levees have altered lateral connectivity on U.S. rivers (Wohl, 2017) analogous to Graf's national-scale assessments of the effects of dams on river longitudinal connectivity (1999, 2001). As a first step toward creating such a national levee assessment, we explore methods to remotely identify the presence of artificial levees. Nearly every study on the identification of artificial levees has exclusively used topography or topographic-derived geomorphic variables with the exception of two studies that used spectral signatures (Steinfeld and Kingsford, 2013; Steinfeld et al., 2013).

In the 1900s, American floodplain development kept pace with flood protection efforts, resulting in the constant rise of average flood-related economic losses (White, 2000). Worldwide, the restoration, rehabilitation, and conservation of large floodplain rivers are increasingly in conflict with development (Sparks, 1995; Wohl et al., 2015). Managing these conflicts requires an understanding of floodplain location and extent, as well as the water and sediment interactions between floodplain and channel (Wohl et al., 2015; Nardi et al., 2018). An explosion in availability of Earth observation datasets and computational power has created new opportunities for the evaluation of floodplain mapping models (Annis et al., 2019), including hydrodynamic models at the continental scale (Wing et al., 2017) and hydrogeomorphic models at basin, continental, and global scales (Nardi et al., 2018; Annis et al., 2019; Nardi et al., 2019; Scheel et al., 2019). Surprisingly, there are few studies that evaluate the impact of artificial levees on floodplain extent at large watershed scales (Scheel et al., 2019). One example of such an evaluation employed the hydrogeomorphic GFPLAIN flood model (Nardi et al., 2019) on two versions of a DEM (digital elevation model), one with artificial levees removed, in the 100,000 km^2 four-digit hydrologic unit code (HUC) (Table 1) Wabash basin (Scheel et al., 2019). At the continental scale, however, it remains unknown to what extent floodplains have been disconnected from channels in the USA or elsewhere in the world.

We improve upon previous artificial levee studies by employing and testing different categories of data (i.e., geomorphic, land use type, and spatial) to the specific problem of identifying artificial levees. Our primary objective is to estimate the locations and spatial distribution of artificial levees across the contiguous U.S., especially as they relate to the completeness of the NLD. We then use these results to explore the spatial extent of lateral disconnectivity caused by artificial levees in the CONUS. We apply a GFPLAIN flood model calibrated with FEMA flood-hazard

maps to two digital elevation models: one unmodified and one with artificial levees removed. We use these analyses to determine the spatial distribution and stream order patterns of floodplain disconnection by artificial levees in the CONUS.

DATA

Table 1. Description of data used in the study

	Variable	Dataset	Type	Resolution
Geomorphic Variables	Slope	National Elevation Dataset (Gesch et al., 2002)	Raster	10 m
	Planform curvature			
	Profile curvature			
	Relative elevation			
	Aspect difference			
Land Cover Variable	Land cover	National Land Cover Database, 2016 (Jin et al., 2019)	Raster	30 m
Spatial Variables	Distance from stream order 1	National Hydrology Dataset Plus High Resolution (Buto & Anderson, 2020)	Vector	-
	Distance from stream order 2 stream			
	Distance from stream order 3 stream			
	Distance from stream order 4 stream			
	Distance from stream order 5 stream			
	Distance from stream order 6 stream			
	Basin boundaries	USGS Hydrologic Unit Maps (Seaber et al., 1987)	Vector	-
	"A" and "AE" flood zones	FEMA Flood maps	Vector	-

METHODS

Case study:

We chose seven 8-digit hydrologic unit code (HUC8) basins as a case study to test different ideas about artificial levee identification, using the National levee database (NLD) (Table 1) as training data (Figure 1). We used Cohen's kappa (Fitzgerald and Lees, 1994) to assess model performance.

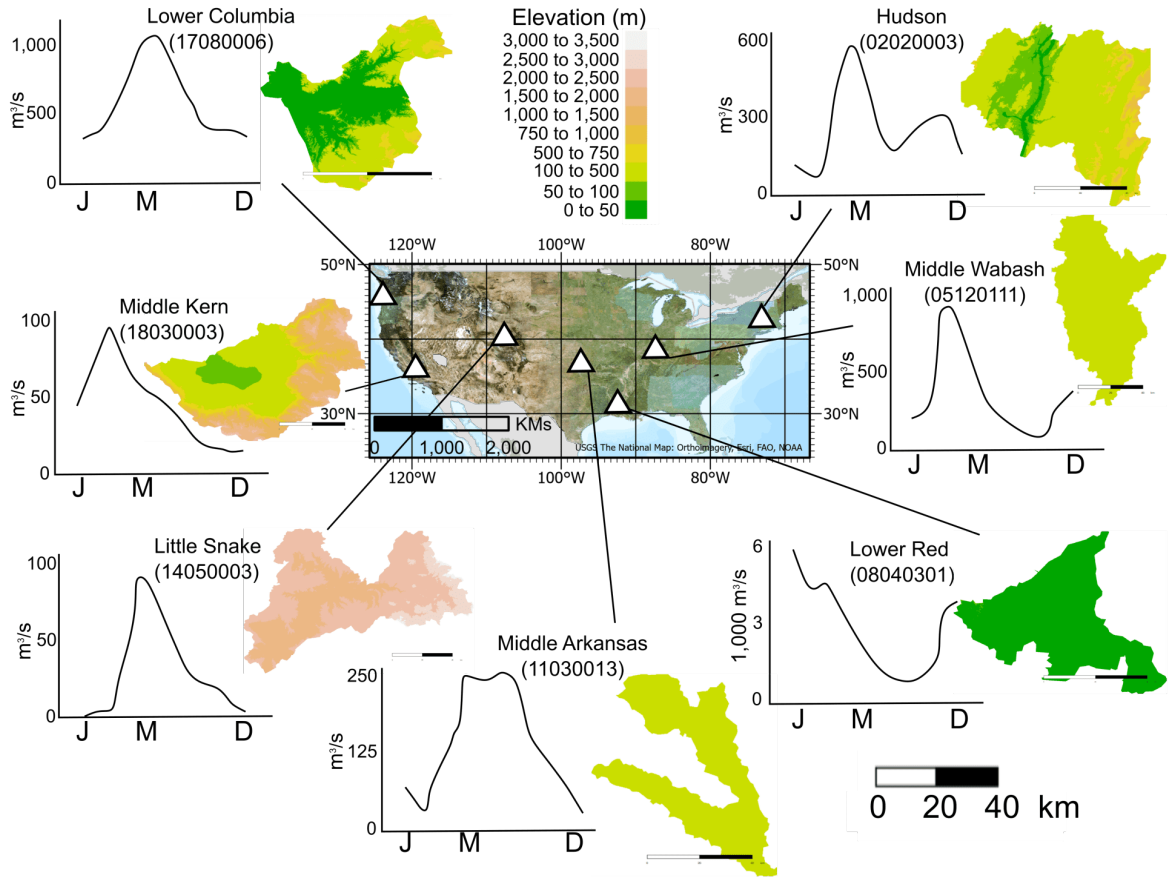


Figure 1. Seven HUC8 basins used for the case study

CONUS hydrogeomorphic floodplain:

Then, we generated a 100-year hydrogeomorphic floodplain for the continental US using the hydrogeomorphic floodplain algorithm GFPLAIN (Nardi et al., 2006, Nardi et al., 2013) and calibrated it to FEMA special flood zones A and AE (Table 1) in each 2-digit HUC basin (HUC2) using streams of order one through six (Figure 2).

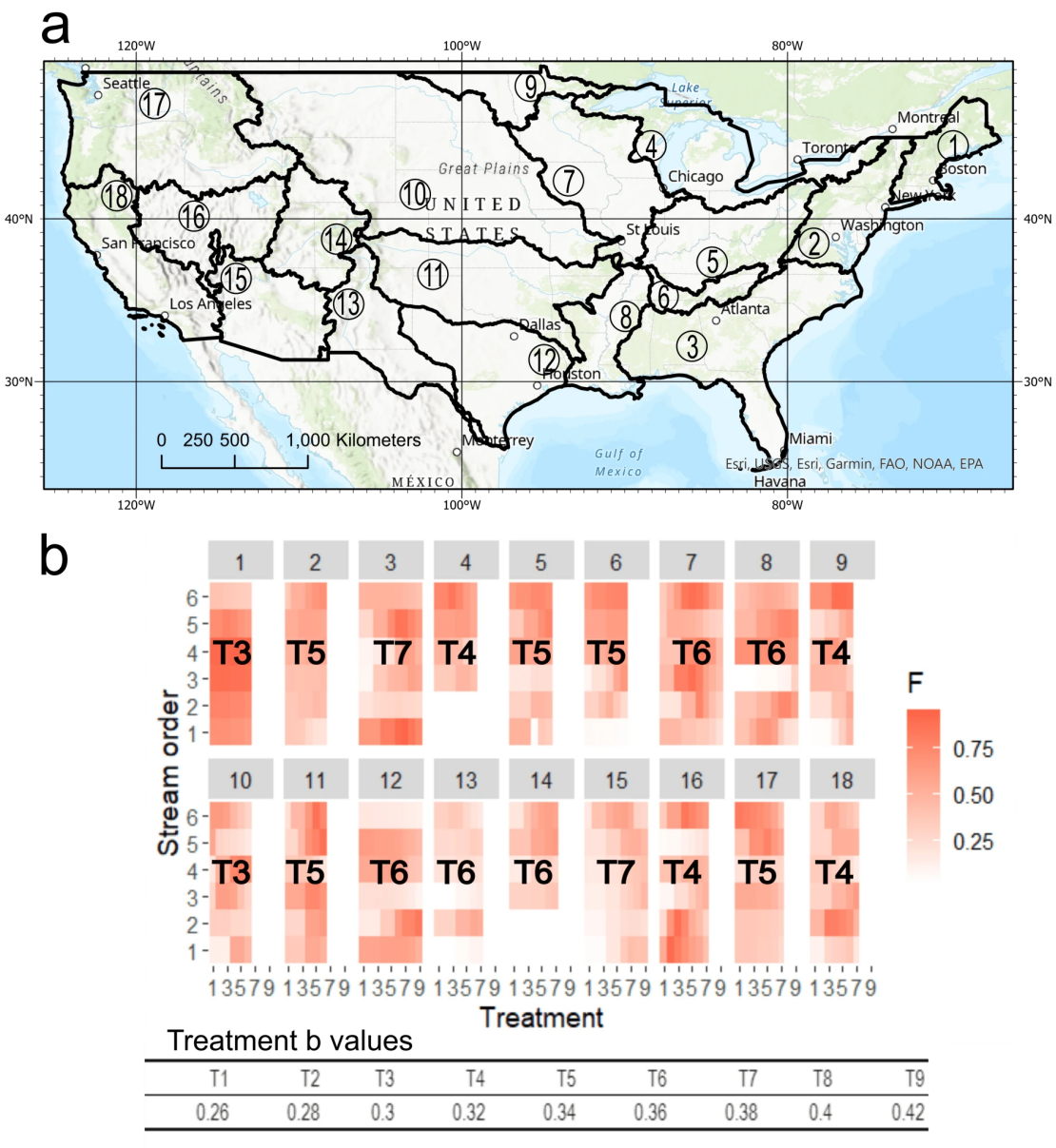


Figure 2. Calibration of the 100-year hydrogeomorphic floodplain. (a) The location of calibration in each HUC2 basin is annotated by a digit, which also corresponds to the HUC number. (b) Calibration results from each basin and the value of b chosen for each basin.

The GFPLAIN algorithm identifies geomorphic floodplains in two main steps: (1) terrain analysis of a DEM for basin drainage extraction and (2) floodplain delineation. It uses an adaption of a scaling regression from Leopold and Maddock (1953) to relate stage to upstream contributing area:

$$FH = {}_{a}A^{b}$$

where FH_i is the maximum flow depth for the recurrence interval i, a and b are dimensionless scaling parameters, and A is the contributing area (Scheel et al., 2019). This floodplain was used as the studies geographic extent.

Leave one-out-cross validation and national model:

After testing different variables, machine learners, and sampling strategies in the case study (Figure 3, Table 2) using R (R Core Team, 2020) and ArcGIS Pro (ESRI Inc., 2020), we conducted a leave-one-out cross validation (Stone, 1974) in the lower Mississippi River basin and then we tested different models at the CONUS level using R, ArcGIS Pro, and Google Earth Engine (Gorelick et al., 2017) We applied a high performing random forest model to the CONUS resulting in a prediction surface. This surface was segmented and analyzed to determine potential levee location, length, and stream order association (Figure 3).

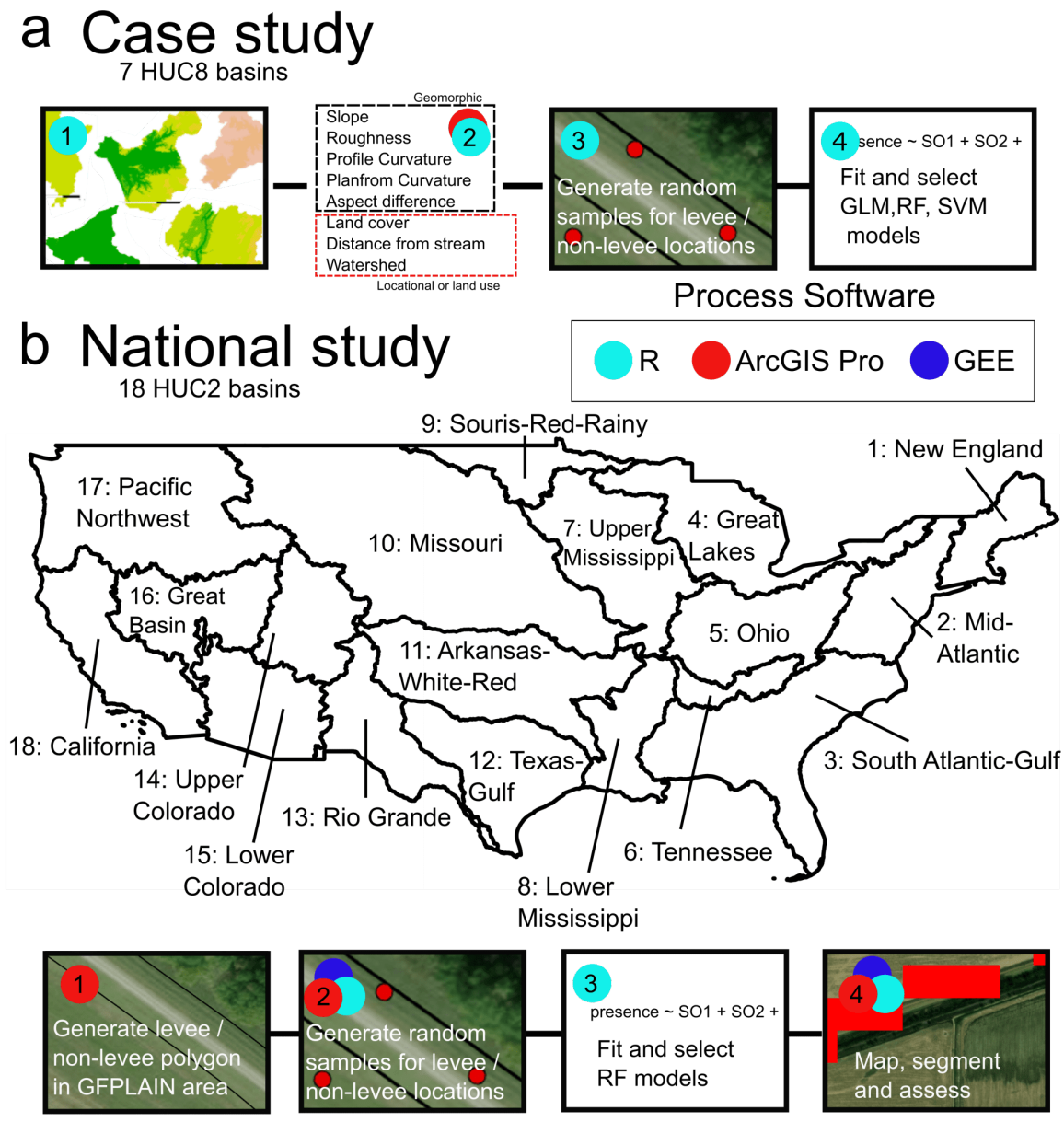


Figure 3. Workflow for case and national study.

Table 2. Model names and variables used in the case study.

Model	Total variables	Variables
1	13	Geomorphic variables (slope, profile curvature, planform curvature, relative elevation, aspect difference), NLCD, basin, distance from stream order 1-6
2	12	Model 1 without slope
3	12	Model 1 without profile curvature
4	12	Model 1 without planform curvature
5	12	Model 1 without relative elevation
6	12	Model 1 without NLCD
7	12	Model 1 without aspect difference
8	12	Model 1 without basin
9	7	Model 1 without distance from stream order variables
10	6	Distance from stream order variables only
11	11	Model 1 without distance from stream order variables and aspect difference
12	8	NLCD, basin, distance from stream order 1-6

Erasing artificial levees:

Then we modified the 1 arc second (30m resolution) USGS EDNA CONUS DEM (Table 1) to "erase" artificial levees from it (Figure 4).

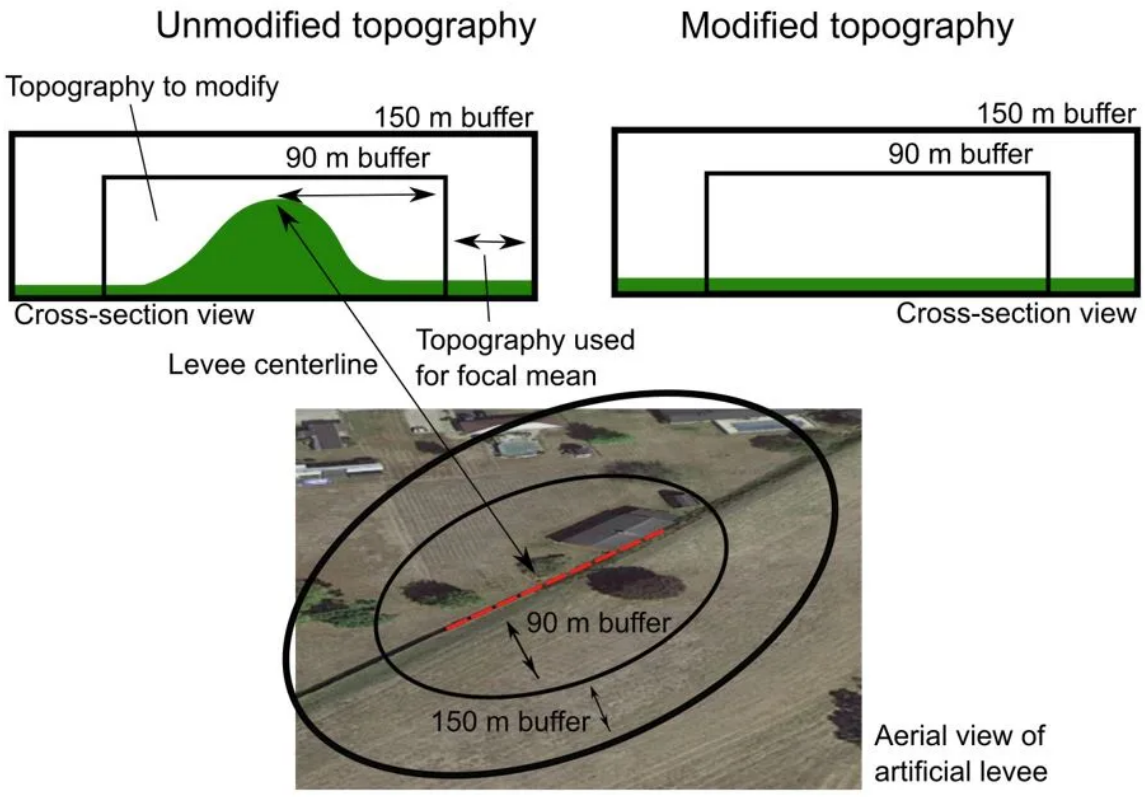


Figure 4. Topography modification for NLD and potential levees. The topography within 90 m of the levee centerline is modified by applying a focal mean with a 120 m radius using only the topography between the 90 m and 150 m buffers.

We applied GFPLAIN again to the modified topography and developed custom ArcGIS Pro and R scripts to analyze the differences between floodplain extent from unmodified and modified topography. We worked by HUC 2 basin and identified areas of agreement and disagreement. Our analysis focuses mainly on the latter because areas of disagreement are created solely by the presence or removal of artificial levees.

Analyzing disconnected floodplains and artificially flooded areas:

Areas of disagreement between the two floodplain are classified as either disconnected floodplain or artificially flooded and are analysed using ArcGIS Pro. To be clear, disconnected floodplain areas are those disconnected from streamflow by the installation of artificial levees. Artificially flooded areas are those that are caused to flood by the installation of artificial levees. These areas were measured in terms of square kilometers and their coverage in the 2016 National Land Cover Database (NLCD) (Table 1) was determined in ArcGIS Pro. We determined the largest stream order associated with each floodplain segment by searching in ArcGIS Pro within 500 m of each segment for every stream segment in the National Hydrology Dataset (Table 1).

RESULTS

Case study:

Key takeaway: Random forest models with just spatial and land use variables were effective at detecting undocumented levees.

RF models demonstrated the best predictive performance for identifying artificial levees at every sample size, followed by SVM models (Figure 5a). Because the RF models demonstrated the best performance, all our subsequent results focus on RF model outputs. An absence-to-presence ratio of 0.7 resulted in the best RF model performance, but balanced data (meaning equal amounts of presence and absence data) with ratios between 0.45 and 1.24 performed well (Figure 5b).

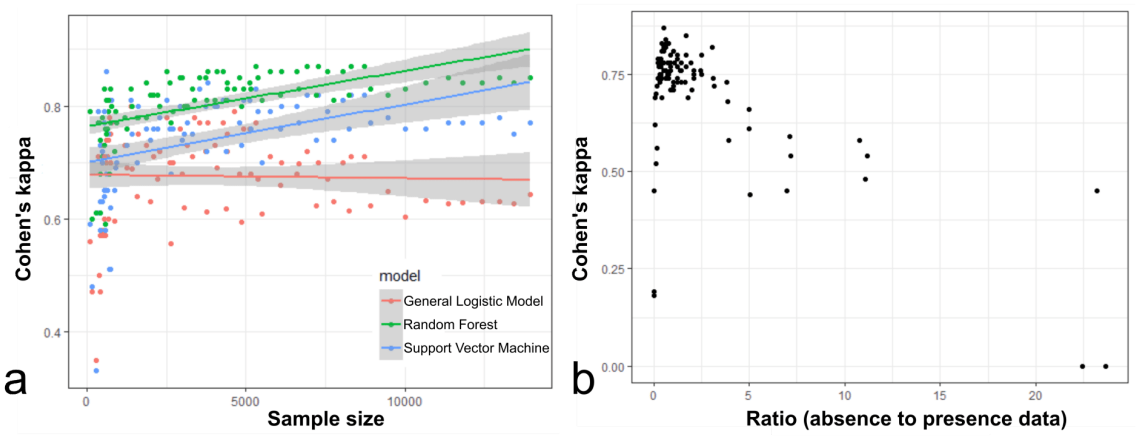


Figure 5. Model Performance performance using Cohen's kappa in the case study. (a) GLM, RF, and SVM model performance with sample size varying from 110 to 13,900 sampled locations total in the seven basins for 113 independent samples. The grey envelopes are 95% confidence intervals for logistic models, depicted by a solid line, fit to the data (b) Performance of 93 independent RF models by varying absence/presence ratio of sampled locations while controlling for sample size (n ~ 832 sampled locations).

Different RF models, each with 100 trees of three variables sampled at each node, were applied to 50 different random samples from 1,000 sampled locations and a 0.7 absence/presence ratio (Figure 6). Model 1, with all variables, only slightly outperformed models with one less variable with kappas in the 0.75-0.8 range. A model without any geomorphic variables (model 12) performed almost as well as the full model.

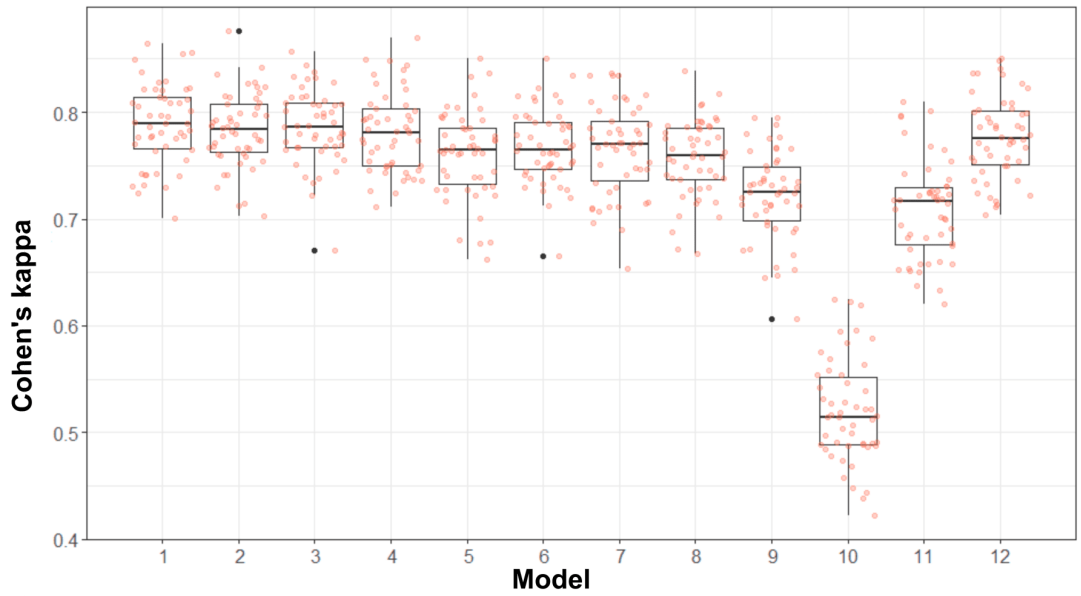


Figure 6. RF model performance by Cohen's kappa and variables for 50 data sets, each with a 0.7 ratio of absence to presence data and ~ 1,000 sampled locations. Boxplots are plotted along with individual model values. The model number on the x-axis corresponds to models listed in table 2.

An RF model using model 12 detected 61% of levees when they were left out of the training dataset. Detected levees were longer than undetected levees such that sum of the length of detected levees (7,473 km) represented 94% of total levee length (7,910 km) (Figure 7a). Levees were close together, with 74% of levees within 5 km of each other and 94% of levees within 25 km (Figure 7b).

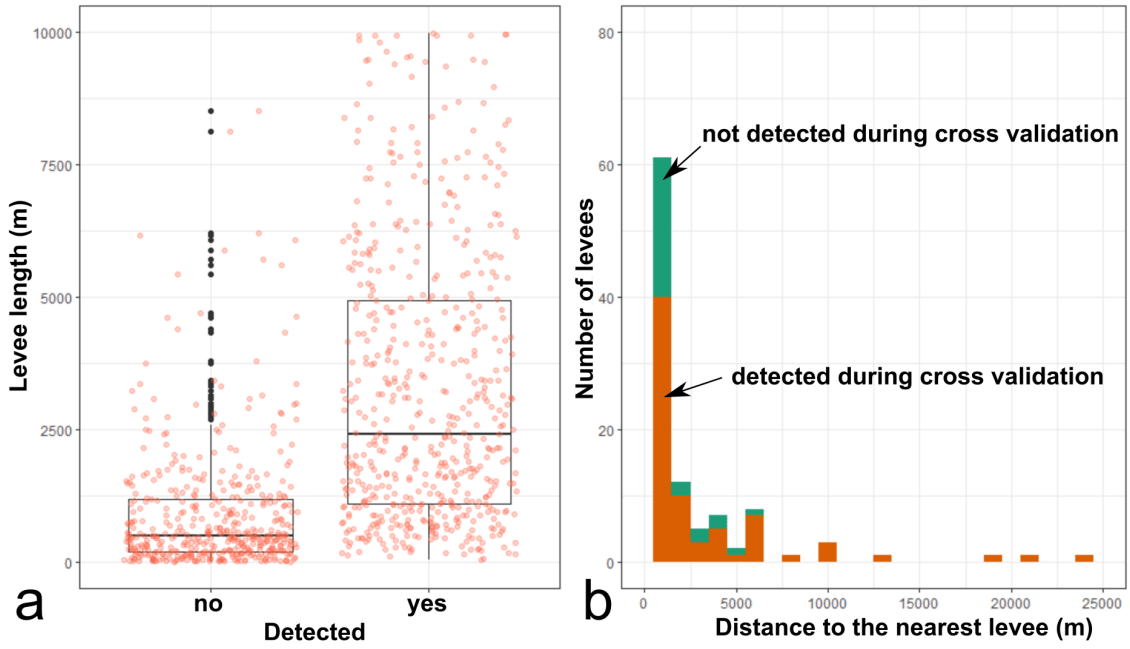


Figure 7. Results from the leave-one-out cross-validation. (a) Longer levees were detected more often than shorter levees so that detected levees represent 94% of total levee length. (b) Levees are close together, with 74% of levees withine 5 km of each other and 94% within 25 km.

National study:

Key takeaway: We detected 182,000 km of potential levees, mainly in the Mississippi and Missouri basins.

We tested different variables, model types, sample sizes and absence/presence ratios at a national scale (Table 3).

Table 3. Model performance in the national study

Model[a]	ML[b]	Size[c]	Ratio[d]	Result (kappa)[e]
1	RF	1700	0.7	0.69
1	RF	1700	1	0.69
1	GLM	1700	1	0.47
1	SVM	1700	1	0.55
12	RF	1700	0.7	0.65
1	RF	170,000	0.7	0.84
12	RF	170,000	0.7	0.94
12 + relative elevation	RF	170,000	0.7	0.89
12 + profile curvature	RF	170,000	0.7	0.9
12 + aspect difference	RF	170,000	0.7	0.87
12 + slope	RF	170,000	0.7	0.89

Note. [a]"Model" corresponds to the variables listed in table 2. [b]The "ML" column denotes the machine learning or statistical model used. [c]"Size" denotes the total sample size taken from each HUC2 basin for both model training and testing. [d]"Ratio" denotes the ratio of absence to presence in the sample. [e]The result denotes the Cohen's kappa of the model on the testing sample, where we used a 70/30 random split for training and validation in all models.

Potential levees were concentrated in the upper and lower Mississippi and the Missouri basins (basins 7,8,10 in Table 4 and Figure 8). Potential levees were also concentrated along streams of order 2 to 6, constituting 75% of total levee length (Figure 9). There were 146,404 potential levees identified constituting a total length of 182,213 km (Table 4). Normalized artificial levee length by stream order length increases by stream order, approaching 0.20 for stream order 10 (Figure 9). Potential levees and those documented in the NLD represent coverage of 2% of the total length of streams in the contiguous United States.

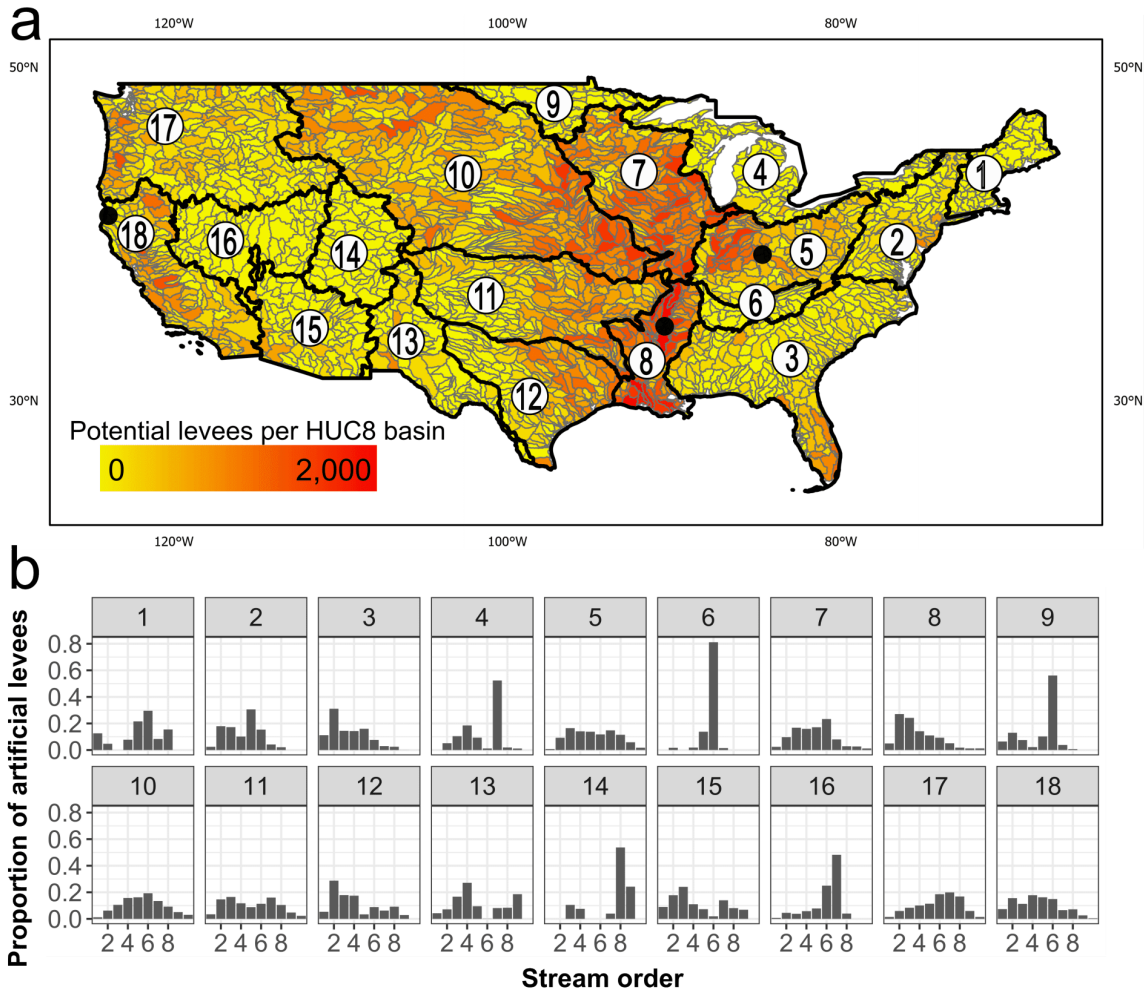


Figure 8. A spatial and stream order representation of potential artificial levees by HUC8 and HUC2 basin. (a) The number of potential levees per HUC8 basin. HUC2 basin boundaries, in bold, are denoted by number. Three black dots indicate potential levees examined in Figure 10. (b) The proportion of artificial levee length along each stream order in 18 HUC2 basins.

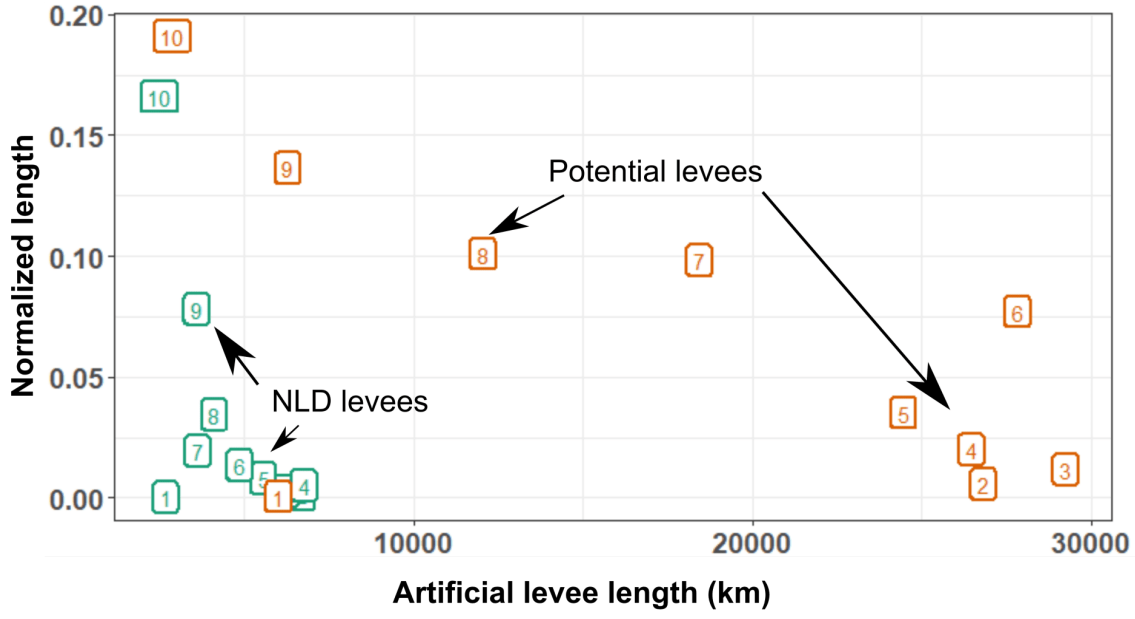


Figure 9. Potential and NLD artificial levee stream orders by normalized length and the sum of levee length for that order.

To illustrate a few locations where we identified levees not present in the NLD, we highlight three potential levees that we were able to ascertain are definitely levees (Figure 10).

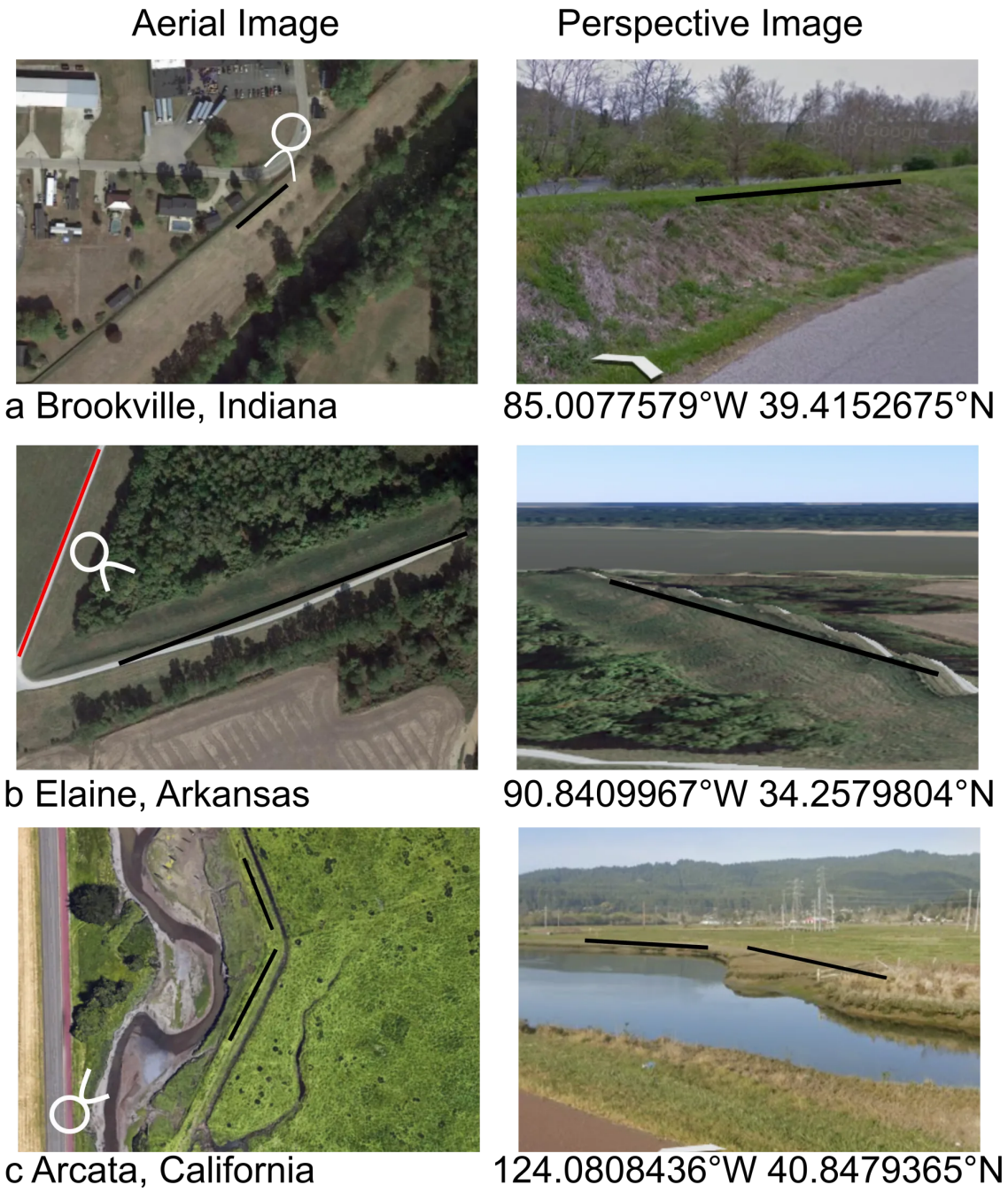


Figure 10. Aerial and perspective images of potential artificial levees discovered during this process. The white circle and arrow on the aerial image indicates the perspective location for the perspective image. Black lines correlate locations between the two images. (a) Brookville, Indiana levee is visible as a long linear feature in aerial imagery, on Google Street view, and written about in an online news article (Norwood, 2020). (b) Elaine, Arkansas levee on the Mississippi River is connected to a NLD levee (indicated by a red line) but remains undocumented. (c) Arcata, California levee along the Gannon Slough and U.S. Route 101 is likely related to salt marsh reclamation for pasturage.

Table 5. NLD and potential levee lengths (km) by HUC2 basin.

HUC2 basin	NLD (km)	Potential levees (km)
New England (1)	89	25
Mid-Atlantic (2)	617	2,220
South Atlantic-Gulf (3)	2,410	5,921
Great Lakes (4)	41	277
Ohio (5)	1,148	12,216
Tennessee (6)	45	40
Upper Mississippi (7)	4,804	35,374
Lower Mississippi (8)	7,912	38,657
Souris-Red-Rainy (9)	466	459
Missouri (10)	4,438	39,221
Arkansas-White-Red (11)	2,939	16,073
Texas-Gulf (12)	2,403	9,230
Rio Grande (13)	1,074	965
Upper Colorado (14)	154	9
Lower Colorado (15)	1,582	1,471
Great Basin (16)	133	392
Pacific Northwest (17)	2,082	10,747
California (18)	14,306	8,916
Total Length (km)	46,643	182,213

Flooplain disconnection:

Key takeaways: Artificial levee removal shifted the location of flooding. Over 30% of the CONUS floodplain is either cultivated or developed land use.

Differences in floodplain extent are clustered together (Figure 11) and near known and potential artificial levees. The Lower Mississippi River (6,714 km2), California (2,043 km2), and Missouri Basins (2,016 km2) had the greatest total artificially flooded and disconnected floodplains (Table 6).

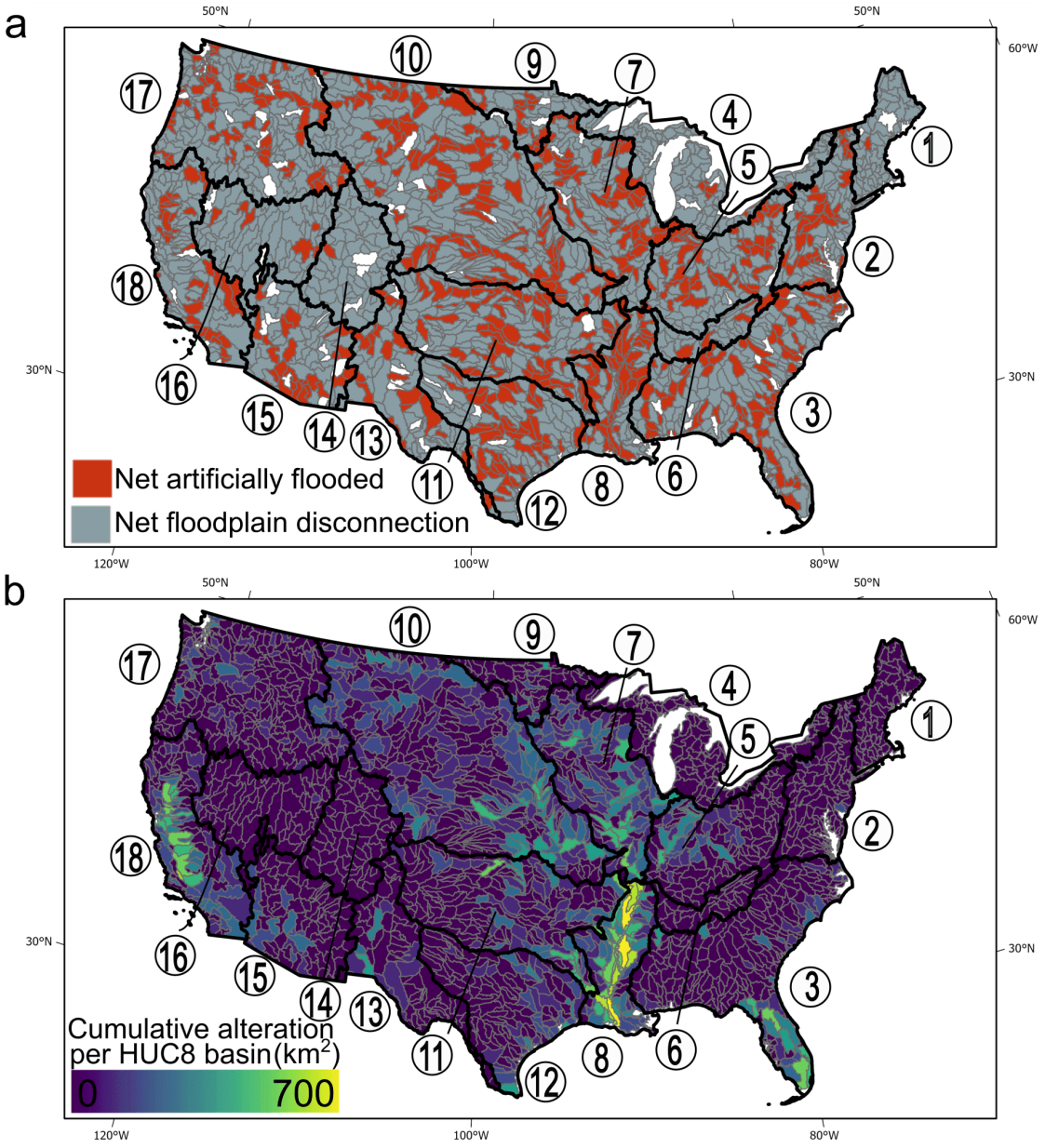


Figure 11. Net disconnection and cumulative alteration in the CONUS HUC8 basins. (a) Net disconnection compares the area of artificially flooded and floodplain disconnection in HUC8 basins. (b) Cumulative alteration by floodplain disconnection and artificial flooding. The 18 HUC2 basins are annotated in each figure with thick black lines and by numbers.

These basins have the greatest (46,569 km), fourth greatest (23,222 km), and second greatest (43,659 km) lengths, respectively, of known and potential artificial levees.

Table 6. Area in square kilometers of each type of area by HUC2 basin.

HUC2 basin	Agreement (km²)	Artificially flooded (km²)	Disconnected (km²)
New England (1)	13,463	1	2
Mid-Atlantic (2)	33,647	20	23
South Atlantic-Gulf (3)	223,063	378	575
Great Lakes (4)	25,378	10	11
Ohio (5)	32,939	273	244
Tennessee (6)	8,850	1	0
Upper Mississippi (7)	73,917	786	971
Lower Mississippi (8)	117,658	4,252	2,462
Souris-Red-Rainy (9)	22,767	44	83
Missouri (10)	81,588	992	1,024
Arkansas-White-Red (11)	59,114	850	803
Texas-Gulf (12)	84,760	190	273
Rio Grande (13)	29,765	70	68
Upper Colorado (14)	12,168	0	0
Lower Colorado (15)	28,366	82	86
Great Basin (16)	37,210	9	12
Pacific Northwest (17)	40,731	179	197
California (18)	31,730	776	1,267
Total (km²)	957,113	8,911	8,100

Land use patterns of artificially flooded and disconnected floodplains are similar but with some notable differences (Figure 12). By far, cultivated land uses (cultivated crops and hay/pasture) make up the largest proportion (55% for artificially flooded and 47% for disconnected only) of each type of floodplain. Wetlands (15% artificially flooded and 11% disconnected floodplain), forested (11% and 16%), and developed (11% and 12%) categories constitute progressively smaller proportions of land use (Table 7).

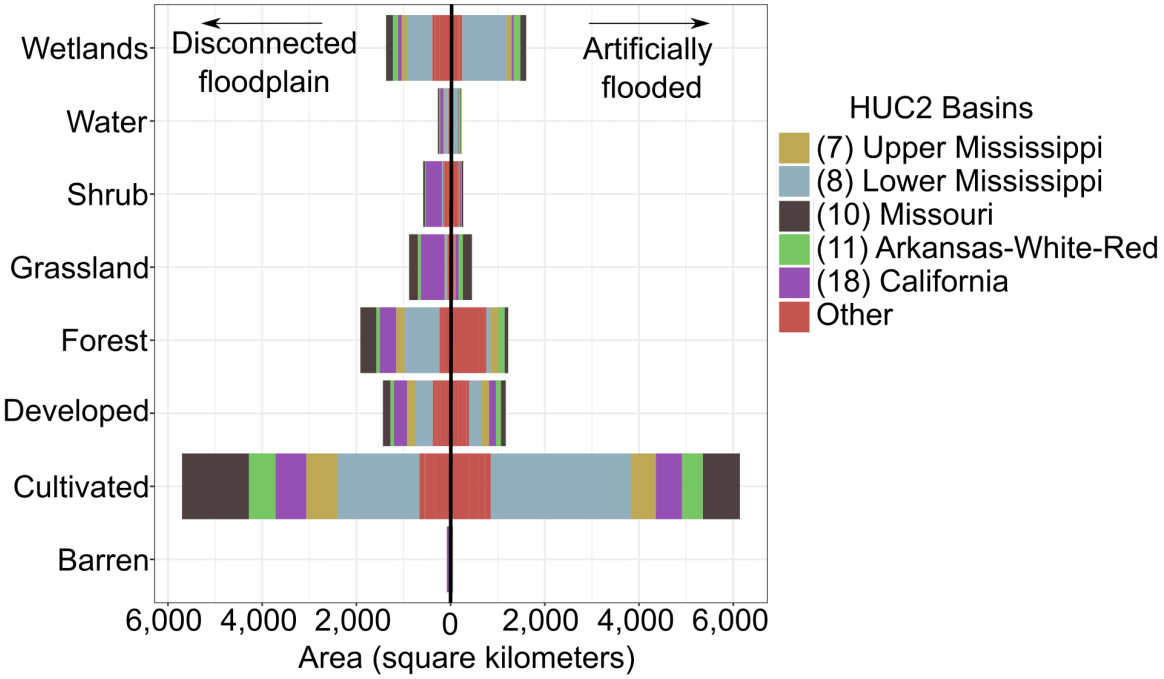


Figure 12. CONUS land cover area (square kilometers) of disconnected floodplain and artificially flooded areas with HUC2 basin contributions annotated by color.

There are several notable differences in the artificially flooded and disconnected floodplains (referred to as “disagreement areas” when discussed as a group) when compared to the agreement areas. Cultivated land uses constitute twice the area in disagreement areas (55-47%) when compared to agreement areas (24%). Forested and developed areas experience similar trends, with much less land use in agreement areas when compared to disagreement areas. Predictably, agreement areas include more wetlands, open water, and shrub cover.

Table 7. Percent of land use using the 2016 NLCD in artificially flooded, disconnected, and agreement floodplains for the CONUS.

Land use	Artificially flooded	Disconnected floodplain	Agreement
Barren land	0	0	3
Cultivated crops	47	36	18
Hay pasture	8	11	6
Deciduous forest	7	7	4
Evergreen forest	2	6	4
Mixed forest	2	3	1
Developed high intensity	1	1	0
Developed low intensity	3	4	2
Developed medium intensity	2	2	1
Developed open space	5	5	3
Emergent herbaceous wetlands	3	3	7
Woody wetlands	12	8	17
Herbaceous	4	7	7
Open water	2	2	17
Perennial snow ice	0	0	0
Shrub and scrub	2	5	10
Unclassified	0	0	0

Stream order is a metric used to classify streams: a first order stream has no tributaries, and stream order increases downstream from the confluence of two streams of equal order (Strahler, 1957). Artificial levees are more likely to disconnect floodplains in first to third order streams, whereas the levees are more likely to enhance floodplain inundation in streams of fourth and higher orders (Figure 13). Stream order contribution patterns vary widely by HUC2 basin.

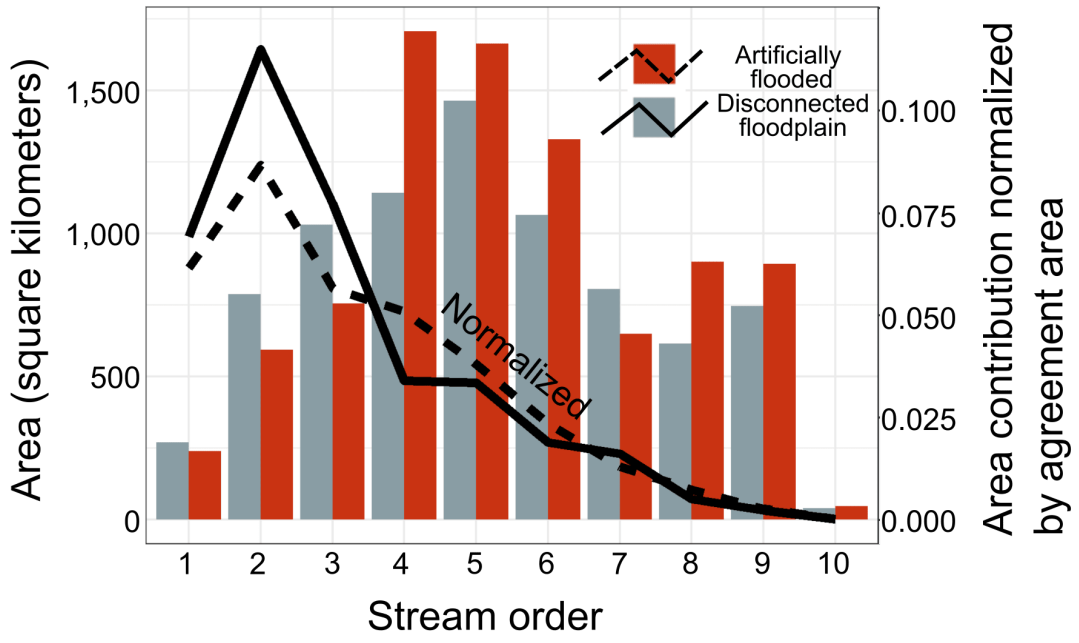


Figure 13. Actual and normalized areas of artificially flooded and disconnected floodplain in the CONUS, distinguished by stream order. Areas are normalized by stream order contributions to the agreement areas.

When normalized by the stream order contribution to agreement areas, disagreement areas peak in order two streams and then decrease with increasing stream order, indicating the effects of artificial levees on smaller order streams (Figure 13).

DISCUSSION

Location and prevalence of artificial levees:

Our analysis indicates that the NLD may be 20.4% complete. Over 62% of potential levees are concentrated in the upper and lower Mississippi basins and the Missouri basin. Potential levee length exceeds documented levee length in these basins by factors of seven, five, and nine, respectively (Figure 8, Table 5). Potential levee length in the Ohio basin exceeds NLD levees by a factor of 11.

Spatial and geographic implications of artificial levee findings:

We are interested in the causes of the difference in accuracy between models that employ land use and spatial variables (e.g., model 12) and those models using geomorphic variables (e.g., model 1). Detecting artificial levees presents significant technical challenges due to their small size, geographic ubiquity, and varied morphology (Steinfeld and Kingsford, 2013). Artificial levees can be massive structures or features nearly invisible to both the eye and topographically (Figure 14), with the height of some artificial levees less than the vertical error of topographic datasets (e.g., the mean relative vertical accuracy of the NED is 0.81 m with the accuracy of 95% of locations within 2.93 m (Gesch et al., 2014)). Furthermore, the resampling process of digital elevation models tends to smooth out topographic crests (such as those of levees) making the features more topographically stealthy or even invisible (Wing et al., 2019). Consequently, it is not surprising that spatial and land use patterns seem to be more useful than geomorphic patterns in a national study given the diverse geomorphic signatures of both documented artificial levees (such as those in figure 14, which can be used as training data) and undocumented levees.

Figure 14. Two artificial levees in the NLD. (a) Over 7 m high, a massive levee west of the overbank structure at the Old River Complex, Louisiana, USA. (b) Almost invisible, Fort Collins North- Cache La Poudre River, ~1 m high, Colorado, USA, indicated by two arrows.

Recent investigations have raised concerns over validation strategies for large scale modeling studies where the employment of spatially autocorrelated training and validation data leads to inflated estimates of model accuracy (e.g., Ploton et al., 2020), so we consider it appropriate to discuss the suitability of the validation techniques employed here. We consider the spatial patterns expressed by the distance from stream order and land use variables to be real patterns created by humans because land use and stream flow were primary factors in the decision

process that led to artificial levee construction. In addition, our method of mapping model 12 over the GFPLAIN floodplain is considered interpolation, not extrapolation, because we are applying the model in the same domain (i.e., the same geographic extent and variable domain) as that from which the training data are generated. Validation error of random samples is considered accurate in models with applications in similar geographic and variable domains (Roberts et al., 2017). Our training and validation samples (n ~ 3,060,000) are drawn from the same geographic and variable space as the model application area (the full 100-year GFPLAIN floodplain). We are not applying the model in a different geographic area. The detection of unknown levees representing 94% of total levee length in the leave-one-out cross-validation substantiates these claims.

Floodplain disconnection:

The finding that the artificially flooded extent was larger than the disconnected floodplain extent (Table 6) was unexpected, although the results are within 99.9% of each other. This corroborates other research illustrating the unintended upstream and downstream flooding caused by artificial levees (e.g., Tobin, 1995; Criss and Shock, 2001; Heine and Pinter, 2012; Czech et al., 2016).

Where artificial levees disconnect floodplains, their removal can increase active floodplain area through two processes; simple floodplain expansion and lateral flowline alteration (Figure 15).

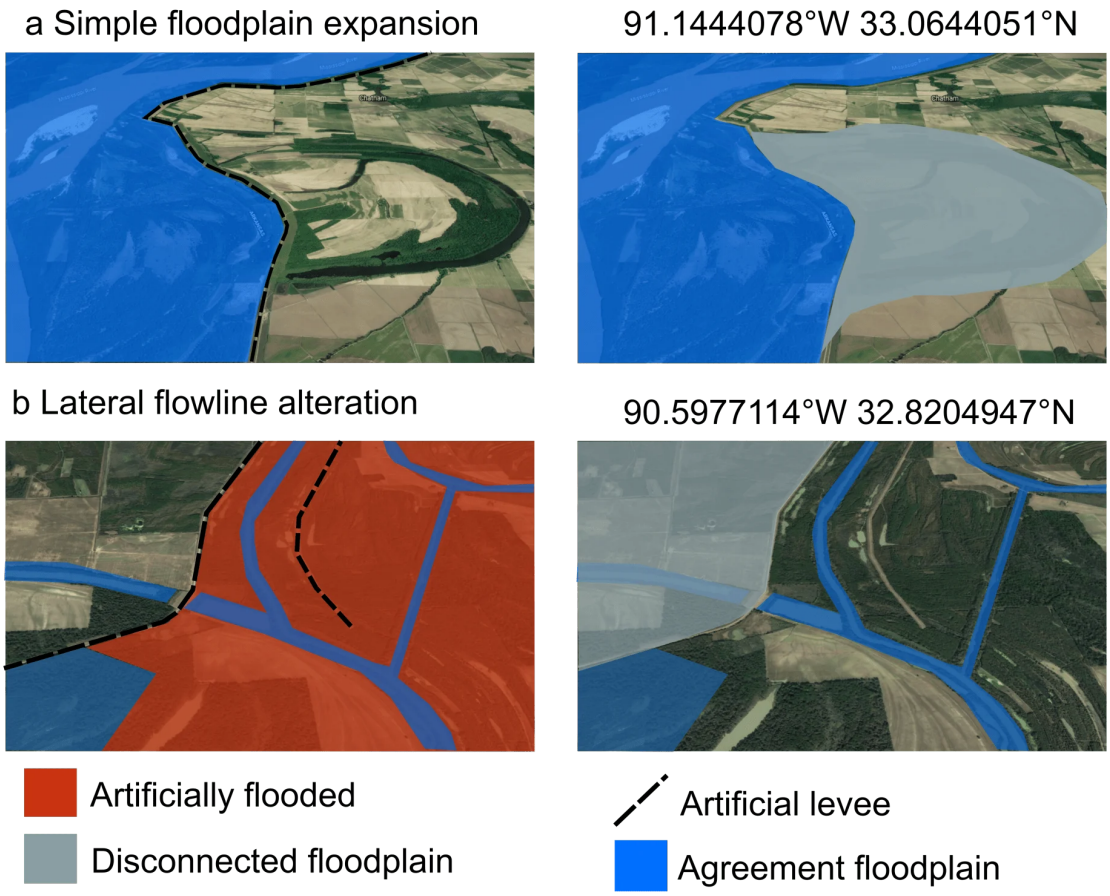


Fig. 15. Types of hydrologic alteration experienced from artificial levee (darker green color) removal along the Lower Mississippi River (a) and the Yazoo Delta (b). (a) Simple floodplain expansion occurs when floodplain extent increases after levee removal, in the absence of other effects. (b) Lateral flowline alteration occurs when removal influences flow direction and accumulation

of flood waters.

An example of simple floodplain expansion comes from the left bank of the Lower Mississippi River near Glen Allan, MS (Figure 15a). The removal of levees in the Yazoo delta provides an example of lateral flowline alteration (Figure 15b), which involves the adjustment of cell flow direction, cell accumulation, and river network identification so that floodplain extent is shifted from the south side to the north side of the artificial levees.

Gilbert White noted that the main policy aim of the last century was to minimize losses on floodplains instead of maximizing social benefits (White, 2000). In spite of that aim and the expenditure of billions of dollars on flood protection projects, flood losses in the US continued to rise and were 2.5 times higher during the period 1951-1985 than 1916-1950 (Tobin, 1995). What insight can this study provide to this problem? We found that if we considered cultivated (cultivated crops and hay pasture) and developed land uses as those susceptible to economic losses, those areas constitute 66%, 59%, and 30% of the artificially flooded, disconnected floodplain, and agreement area floodplains, respectively. This 297,794 km2 constitutes 30.6% of floodplain areas and 3.7% of the entire CONUS. That nearly one-third of floodplain areas in the CONUS are used for some sort of economic purpose likely explains at least one of the causes for the trend noted by Tobin (1995) and White (2000).

CONCLUSION

Our exploration of different variables and models to detect artificial levees led to a **random forest model with land use and spatial variables**. Applying this model in a 100-year geomorphic floodplain in the contiguous US indicated the potential for **182,000 km of artificial levees that are not included in the national levee database**, suggesting that the database is 20.4% complete. These levees missing from the national database were **concentrated in the lower and upper Mississippi and Missouri basins** and mostly along streams of order 2 through 6. When normalized for total stream length, larger stream orders were more impacted than smaller streams, with more than a third of stream order 10 streams impacted by NLD or potential levees.

We removed known and potential artificial levee locations from a modified 1 arc second DEM of the contiguous United States. We then generated two hydrogeomorphic floodplains using the modified and unmodified DEM and compared the location and area, land use, and the stream order of rivers associated with each floodplain segment. **The overall effect of artificial levee removal was not to just extend the floodplain, but rather to shift the location of flooding.** Disconnected floodplain (protected from flooding) and artificially flooded (induced to flood by artificial levees) areas each accounted for about 1% of the total CONUS floodplain, which was more than 960,000 km^2. More than 60% of the disagreement areas (mapped floodplain that differed with and without artificial levee presence) were cultivated, forested, wetland, or developed land use. **More than 30% of the CONUS floodplain was either cultivated or developed.** These results are indicative, but on a massive scale, of previous artificial levee investigations that illustrated the unintended consequences of artificial levee installation. Also, the finding that over 30% of the CONUS floodplain has a cultivated or developed land use seems to explain at least one of the causes of the troubling trend of increasing flood damage noted by Tobin (1995) and White (2000).

ABSTRACT

Recent advances in Earth observation data and computing ability create exciting opportunities for national and global studies of human impacts to water resources. But, with a lack of complete databases of artificial levees, there remains a need to better understand how artificial levees impact floodplain extent at regional and larger scales.

Here, we estimate river-floodplain disconnection in the contiguous United States using an incomplete artificial levee database, machine learning algorithms, and hydrogeomorphic floodplain delineation models. We tested different topographic, land use, and spatial variables with different machine learning techniques in a case study of seven geographically diverse HUC8 basins before applying the technique at the national scale. We found that a parsimonious random forest model without topographic variables was 97% accurate. When applied to areas within a national 100-year hydrogeomorphic floodplain, the model indicated the potential for more than 180,000 km of undocumented artificial levees, meaning that the National Levee Database (NLD) is about 20% complete. More than 62% of potential levees are concentrated in the Upper and Lower Mississippi and Missouri basins. The stream order distribution of potential and NLD levees are similar; however, potential levees are primarily located along stream orders 3 and 6 while the NLD locations are along stream orders 2, 3 and 4.

Using this, we explored the national impacts of artificial levees on floodplain extent by comparing two hydrogeomorphic floodplains based on (1) an unmodified USGS 1 arc second DEM and (2) a modified DEM with known and potential levees erased from the topography. We found that the overall impact of artificial levee removal was to shift the location of flooding. Over 30% of the CONUS 100-year floodplain was cultivated or developed land use.

REFERENCES

American Society of Civil Engineers (2017). 2017 Infrastructure Report Card: Levees, Reston, Va. [Available at https://www.infrastructurereportcard.org/cat-item/levees/ (https://www.infrastructurereportcard.org/cat-item/levees/).]

Annis, A., Nardi, F., Morrison, R. R., & Castelli, F. (2019). Investigating hydrogeomorphic floodplain mapping performance with varying DTM resolution and stream order. Hydrological Sciences Journal, 64(5), 525-538.

Bunn, S. E., Abal, E. G., Smith, M. J., Choy, S. C., Fellows, C. S., Harch, B. D., ... & Sheldon, F. (2010). Integration of science and monitoring of river ecosystem health to guide investments in catchment protection and rehabilitation. Freshwater Biology, 55, 223-240.

Buto, S. G., & Anderson, R. D. (2020). NHDPlus High Resolution (NHDPlus HR)---A hydrography framework for the Nation (No. 2020-3033). US Geological Survey.

Castro, J. M., & Thorne, C. R. (2019). The stream evolution triangle: Integrating geology, hydrology, and biology. River Research and Applications, 35(4), 315-326.

Criss, R. E. and Shock, E. L. (2001). Flood enhancement through flood control. Geology, 29(10), 875-878.

Czech, W., Radecki-Pawlik, A., Wyżga, B., & Hajdukiewicz, H. (2016). Modelling the flooding capacity of a Polish Carpathian river: a comparison of constrained and free channel conditions. Geomorphology, 272, 32-42.

Esri Inc. (2020). ArcGIS Pro (Version 2.7). Esri Inc. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Fitzgerald, R. W., & Lees, B. G. (1994). Assessing the classification accuracy of multisource remote sensing data. Remote sensing of Environment, 47(3), 362-368.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., & Tyler, D. (2002). The national elevation dataset. Photogrammetric engineering and remote sensing, 68(1), 5-32.

Gesch, D. B., Oimoen, M. J., & Evans, G. A. (2014). Accuracy assessment of the US Geological Survey National Elevation Dataset, and comparison with other large-area elevation datasets: SRTM and ASTER (Vol. 1008). US Department of the Interior, US Geological Survey.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.

Graf, W. L. (1999). Dam nation: A geographic census of American dams and their large-scale hydrologic impacts. Water resources research, 35(4), 1305-1311.doi:

10.1029/1999WR900016.

Graf, W. L. (2001). Dam age control: restoring the physical integrity of America's rivers. Annals of the Association of American Geographers, 91(1), 1-27.

Harvey, J., & Gooseff, M. (2015). River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resources Research, 51(9), 6893-6922. doi: 10.1002/2015WR017617.

Heine, R. A., & Pinter, N. (2012). Levee effects upon flood levels: an empirical assessment. Hydrological Processes, 26(21), 3225-3240.

Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., ... & Howard, D. (2019). Overall methodology design for the United States national land cover database 2016 products. Remote Sensing, 11(24), 2971.

Kondolf, G. M., Boulton, A. J., O'Daniel, S., Poole, G. C., Rahel, F. J., Stanley, E. H., ... & Nakamura, K. (2006). Process-based ecological river restoration: visualizing three-dimensional connectivity and dynamic vectors to recover lost linkages. Ecology and society, 11(2).

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., ... & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature communications, 11(1), 1-11.

Leopold, L. B., & Maddock, T. (1953). The hydraulic geometry of stream channels and some physiographic implications (Vol. 252). US Government Printing Office.

Nardi, F., Vivoni, E. R., & Grimaldi, S. (2006). Investigating a floodplain scaling relation using a hydrogeomorphic delineation method. Water Resources Research, 42(9).

Nardi, F., Biscarini, C., Di Francesco, S., Manciola, P., & Ubertini, L. (2013). Comparing a large-scale DEM-based floodplain delineation algorithm with standard flood maps: The Tiber River Basin case study. Irrigation and Drainage, 62(S2), 11-19.

Nardi, F., Morrison, R. R., Annis, A., & Grantham, T. E. (2018). Hydrologic scaling for hydrogeomorphic floodplain mapping: Insights into human-induced floodplain disconnectivity. River Research and Applications, 34(7), 675-685.

Nardi, F., Annis, A., Di Baldassarre, G., Vivoni, E. R., & Grimaldi, S. (2019). GFPLAIN250m, a global high-resolution dataset of Earth’s floodplains. Scientific data, 6(1), 1-6.

Palmer, M. A., Hondula, K. L., & Koch, B. J. (2014). Ecological restoration of streams and rivers: shifting strategies and shifting goals. Annual Review of Ecology, Evolution, and Systematics, 45, 247-269.

R Core Team (2020). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913-929.

Scheel, K., Morrison, R. R., Annis, A., & Nardi, F. (2019). Understanding the Large-Scale Influence of Levees on Floodplain Connectivity Using a Hydrogeomorphic Approach. JAWRA Journal of the American Water Resources Association, 55(2), 413-429.

Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). Hydrologic unit maps: US Geological Survey water supply paper 2294. US Geological Survey.

Scheel, K., Morrison, R. R., Annis, A., & Nardi, F. (2019). Understanding the Large-Scale Influence of Levees on Floodplain Connectivity Using a Hydrogeomorphic Approach. JAWRA Journal of the American Water Resources Association, 55(2), 413-429.

Sparks, R.E., Blodgett, K.D., Casper, A.F., Hagy, H.M., Lemke, M.J., Velho, L.F.M., & Rodrigues, L.C. (2017). Why experiment with success? Opportunities and risks in applying assessment and adaptive management to the Emiquon floodplain restoration project. Hydrobiologia, 804, 177-200.

Steinfeld, C. M. M., & Kingsford, R. T. (2013). Disconnecting the floodplain: earthworks and their ecological effect on a dryland floodplain in the Murray–Darling Basin, Australia. River Research and Applications, 29(2), 206-218.

Steinfeld, C. M., Kingsford, R. T., & Laffan, S. W. (2013). Semi-automated GIS techniques for detecting floodplain earthworks. Hydrological Processes, 27(4), 579-591.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society: Series B (Methodological), 36(2), 111-133.

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. Eos, Transactions American Geophysical Union, 38(6), 913-920.

Tobin, G. A. (1995). The levee love affair: a stormy relationship? 1. JAWRA Journal of the American Water Resources Association, 31(3), 359-367.

White, G. F. (2000). Water science and technology: some lessons from the 20th century. Environment: Science and Policy for Sustainable Development, 42(1), 30-38.

Wing, O. E., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous U nited S tates. Water Resources Research, 53(9), 7968-7986.

Wing, O. E., Bates, P. D., Neal, J. C., Sampson, C. C., Smith, A. M., Quinn, N., ... & Krajewski, W. F. (2019). A new automated method for improved flood defense representation in large-scale hydraulic models. Water Resources Research, 55(12), 11007-11034. DOI: 10.1029/2019WR02597.

Wohl, E., Bledsoe, B. P., Jacobson, R. B., Poff, N. L., Rathburn, S. L., Walters, D. M., & Wilcox, A. C. (2015). The natural sediment regime in rivers: Broadening the foundation for ecosystem management. BioScience, 65(4), 358-371.

Wohl, E. (2017). Connectivity in rivers. Progress in Physical Geography, 41(3), 345-362., DOI: 10.1177/0309133317714972.