ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/391735650

Predictability of Cottonwood Recruitment Along a Dynamic, Regulated River

Article *in* Ecohydrology · May 2025

DOI: 10.1002/eco.70048

CITATIONS
0

READS
53

6 authors, including:

Sierra Phillips
University of California, Davis
4 PUBLICATIONS **2** CITATIONS

SEE PROFILE

Romina Díaz Gómez
Stockholm Environment Institute (SEI)
33 PUBLICATIONS **78** CITATIONS

SEE PROFILE

Gregory Brian Pasternack
University of California, Davis
262 PUBLICATIONS **5,752** CITATIONS

SEE PROFILE

All content following this page was uploaded by Gregory Brian Pasternack on 14 May 2025.

The user has requested enhancement of the downloaded file.

Title: Predictability of cottonwood recruitment along a dynamic, regulated river

Running head: cottonwood recruitment predictability

Authors: Rachel E. Wright[1], Sierra J. Phillips[1], Romina Diaz-Gomez[1], Yufang Jin[1], Mary L. Cadenasso[2], Gregory B. Pasternack[1*]

[1]Department of Land, Air, and Water Resources, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

[2]Department of Plant Sciences, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

* Corresponding author: gpast@ucdavis.edu

ORCIDs

Wright: 0009-0008-9816-4412, Phillips: 0000-0001-8539-5037, Diaz-Gomez: 0000-0002-8919-554X, Pasternack: 0000-0002-1977-4175, Jin: 0000-0002-9049-9807, Cadenasso: 0000-0002-6521-067X

Accepted, Corrected

ABSTRACT

Riparian vegetation planting and management are vital to river engineering projects. To inform these activities, science needs to provide practitioners with a better understanding of influences on recruitment and where vegetation will most likely establish and survive. This study investigated whether the spatially explicit recruitment of *Populus fremontii* (Fremont cottonwood), a dominant riparian species in the western USA, could be predicted along a dynamic, alluvial regulated river. We used a ~ 34-km segment of the Yuba River in California, USA, which was mapped in 2017 after a large flood reset the terrain. Five years later from August through November 2022, a field campaign characterized precise locations of juvenile cottonwoods. We evaluated predictions from deterministic and statistical models. For the deterministic test, a spatially distributed riparian seedling recruitment model was used with expert-estimated parameters. The model was not accurate in this case but was informative. For the statistical approach, a supervised classification Random Forest (RF) algorithm, driven by 2017 hydrophysical and topographic variables, was trained and cross-validated using 2022 cottonwood presence and absence observations. The RF model had an overall accuracy of 87% and an AUC-ROC value of 94%, with the most important variables being the detrended DEM, channel proximity, and inundation survival. Topographic variables were much more significant than hydrophysical ones. Further developments to understand underlying governing equations and recruitment model parameters will draw on lessons from the RF model. Both deterministic and statistical models are recommended to evaluate riparian vegetation restoration designs, as each yields unique insights.

Keywords: Riparian seedling recruitment, ecohydraulics, machine learning, river revegetation, cottonwood

1 INTRODUCTION

The anthropogenically driven degradation of riverine ecosystems poses a serious problem in the arid and semiarid regions of western North America, where water resources are often intensely regulated for human needs by dams and diversions (Hauer & Lorang, 2004; Poff et al., 2003). Western U.S. riparian forests are now remnants of varying size and quality, and habitat-forming foundational species are severely impacted (Abell, 1989; Braatne et al., 1996; Patten, 1998). The current crisis is expected to be worsened by the growing effects of regional climate change, such as intensified disturbance regimes, declining water supply, and even more infrastructural development (e.g. Dettinger et al., 2015).

Cottonwoods (*Populus*) are a foundation (Ellison et al., 2005; Whitham et al., 2006) and dominant riparian tree species in this region (Braatne et al., 1996; Patten, 1998). Riparian trees stabilize channel banks, provide rich wildlife habitats, maintain biodiversity, create shade and shelter, and produce streamwood (Gregory et al., 1991; Naiman et al., 1993). River engineering projects and long-term management programs seek to restore and conserve cottonwood forests, and an inability to accurately predict cottonwood seedling recruitment poses a constraint on these efforts.

To untangle impediments on riparian vegetation conservation, this study tested the ability of existing scientific theory and methods to accurately predict locations of cottonwood recruitment throughout a regulated, dynamic river corridor by implementing the state of knowledge in two different modeling approaches, one deterministic and one statistical. The

Accepted Article

physical processes and environmental conditions included in the two models were further explored to examine the conditions needed for both natural recruitment and for accurately predicting recruitment. Prediction accuracy was evaluated using field observations of young cottonwoods, yet the novelty of this work lies in exploring outcomes of different prediction approaches and how those outcomes can inform riparian ecology and conservation.

1.1 Cottonwood Seedling Recruitment

Cottonwood recruitment occurs both sexually through seeds and asexually as clonal processes. Riparian regrowth by seedling recruitment is important to maintain, as it supports genetic diversity, offsets losses due to mortality, and is the primary method of colonization for this pioneer species (Braatne et al., 1996; Dixon et al., 2012; Mahoney & Rood, 1998; Stromberg, 1993). Seedling recruitment is dependent on several hydrophysical processes which occur seasonally over an annual growing season. However, suitable recruitment conditions may not occur every year.

Bare surfaces are created when flows high enough to induce sediment mobilization uproot existing vegetation, clear away ground cover and detritus, and/or bury young vegetation by depositing sediments. When high flows recede, cottonwoods may colonize suitable areas of newly deposited sediment and moist open ground (Friedman et al., 1995). These barren surfaces are important to pioneer species like cottonwoods, which are shade intolerant and poor competitors, making access to full sunlight critical for seedling growth and development (Braatne et al., 1996; Johnson 1994). Mature cottonwoods produce an abundant number of seeds every year that are primarily dispersed by water (Braatne et al., 1996; Karrenberg et al., 2002; Moggridge & Gurnell, 2010; Stromberg, 1993). With a limited dispersal period and a seed

Accepted Article

viability that quickly declines (Braatne et al., 1996; Mahoney & Rood, 1998; Stella et al., 2006), cottonwoods have adapted to dynamic riparian environments by using climatic signals to couple their seed release with spring snow melt pulses. Annual variability in the timing and magnitude of flows results in some years with more prolific recruitment than others (Dixon et al., 2012; Scott et al., 1997; Stromberg, 1998).

Cottonwoods are dependent on the groundwater table and the associated capillary fringe in the substrate for moisture. For germination to occur, substrate requires continual moisture for the first few weeks of establishment (Cooper et al., 1999; Fenner et al., 1984). After germination, surface moisture conditions and receding water table rates impact the success of seedling growth and development (Amlin & Rood, 2002; Stella et al., 2010). Seedlings must be able to grow sufficiently long roots to reach the receding water levels (Stromberg, 1993), with drought stress or mortality for seedlings where the water table recedes faster than their roots can grow (Amlin & Rood, 2002; Mahoney & Rood, 1991; Stella et al., 2010).

Areas where cottonwood seedlings colonized may then be vulnerable to high flows with sufficiently intense hydraulic forces that result in scouring or depositional processes, risking mortality (Politti et al., 2018). When these large flows result in areas flooded with slow moving or stationary water, erosional and sediment transport impacts are less (Amlin & Rood, 2001). However, prolonged inundation over multiple weeks can also be stressful or lethal to seedlings as it can lead to oxygen depletion in the root zone (anoxic conditions), root growth suppression, reduce transpiration, and root decay (Amlin & Rood, 2001; Auchincloss et al., 2012). Impacts to a seedling and the number of inundation days it can survive are dependent upon the age and size of the seedling, as well as the depth, clarity, and temperature of the water (Auchincloss et al., 2012; Friedman & Auble, 1999).

Accepted Article

1.2 Riparian Vegetation Modeling

Solari et al. (2016) and You et al. (2015) provided summaries of models analyzing vegetation within dynamic riparian environments. These models include effects of vegetation on hydro-morphodynamics by influencing flow resistance (Järvelä, 2004; Luhar & Nepf, 2013), sediment transport (Lopez & Garcia, 1998), or bank dynamics (Bertoldi et al., 2014; Zong & Nepf, 2011), and the reverse of hydro-morphodynamics on vegetation by impacting seed dispersal (Groves et al., 2009; Merritti & Wohl, 2016), recruitment (Mahoney & Rood, 1998), or mortality and woody debris inputs (Edmaier et al., 2011; Gregory et al., 2003; Haga et al., 2002; Villanueva et al., 2014). Mathematical models also vary in purpose to analyze systems at the individual (Scott et al., 1999), population (Clipperton et al., 2003; Phipps, 1979), or community (Camporeale & Ridolfi, 2006) level. In addition, these mathematical models may differ by being deterministic, statistical-empirical, or statistical-stochastic, or a combination thereof (Jajarmizadeh et al., 2012).

This study used the Riparian Seedling Recruitment Model (RSRM), a two-dimensional (2D) spatially distributed, deterministic algorithm designed to determine the theoretical suitability of locations for riparian seedling recruitment by predicting the potential success for a seedling to survive through its first year of life (Phillips & Pasternack, 2022). The conceptual basis of the RSRM is the 'recruitment box model' (Mahoney & Rood, 1998), which relates the timing and inter-annual pattern of stream stage to the physiological needs for cottonwood seedling recruitment. The approach is similar to other spatially distributed models for riparian tree seedling recruitment (Benjankar et al., 2014, 2020; Stella, 2005; Tranmer et al., 2023), but includes several novel developments that promote spatially-explicit mechanistic realism and practical utility for not only river assessment but also river ecological engineering design

Accepted, Corrected

(Phillips & Pasternack, 2022; Phillips et al., 2025). The RSRM is implemented in free, open-source software called River Architect (Schwindt et al., 2020; https://riverarchitect.github.io).

Time, finances, site accessibility, and other such local constraints may prevent *in situ* collection of environmental variables or measurements necessary to calibrate a numerical deterministic model. Nevertheless, the ways in which a given model is uncertain and its levels of accuracy and precision can serve as indicators of the state of understanding of a phenomenon. Alternatively, a small sample of data might be better used for training a statistical-empirical model, such as an artificial intelligence machine learning (AI/ML) model when there is an abundance of remote sensing data, especially environmental variables derived from airborne LiDAR (Diaz-Gomez et al., 2025; Guisan et al., 1999; Rew et al., 2005; Shoutis et al., 2010; Vogiatzakis & Griffiths, 2006).

One such ML procedure is the Random Forest (RF), which uses classification trees to repeatedly split the input data into more homogenous groups using different combinations of explanatory variables (Breiman, 2001). RF's allow for the exploration of prediction patterns and processes through the use of both discrete and continuous explanatory variables, variable importance measures, and graphical representations. A previous study by Diaz-Gomez et al. (2025) used a RF with topographic metrics derived from airborne LiDAR to predict where vegetation had successfully established on the same testbed river as this study. Presence and absence points of naturally established vegetation were randomly selected from LiDAR-derived data and used with 17 topographic predictor variables, ultimately achieving an accuracy metric (i.e., AUC) of 77% (Diaz-Gomez et al., 2025). The workflow created by Diaz-Gomez et al. (2025) was modified for this study to predict presence and absence of juvenile cottonwoods.

Accepted, Corrected

1.3 Research Questions

The goal of this study was to evaluate the state of cottonwood seedling recruitment predictability when addressing a real river in need of active conservation measures. To achieve this goal, both a deterministic and a statistical-empirical model were used to explore the potential explanatory power each approach offers. The testbed observational dataset was obtained by a labor-intensive effort to precisely locate and measure 2,957 young cottonwoods, which is notable compared to coarser polygon-level riparian zone evaluations in past studies. The deterministic model implemented mathematical equations describing hydrophysical processes and was coupled with empirically derived biophysical cottonwood metrics. The statistical model explored empirical relationships between environmental variables and cottonwood presence/absence. Each approach makes different assumptions and provides unique, valuable insights. We asked three research questions that focused on juvenile cottonwoods of five years old or younger, defined as < 5 m tall (Braatne et al., 1996; Nagler et al., 2005; Zamora-Arroyo et al., 2001).

First, how do predicted cottonwood seedling recruitment locations from a deterministic model compare to field locations of juvenile cottonwoods? We hypothesized that field locations of juvenile cottonwoods would occur in the more favorable and optimal recruitment locations predicted by the deterministic model, because this approach models hydrophysical processes (e.g. scouring flows, seed dispersal, inundation periods, and stream and water-table recession) important for cottonwood seedling recruitment. Second, how do the field locations of juvenile cottonwoods compare to presence/absence classification predictions by the statistical RF algorithm? Due to the use of both hydrophysical variables important for cottonwood seedling recruitment and topographic variables that capture the heterogeneity and small-scale variations in terrain needed for maintaining riparian vegetation diversity, enough spatial information should be available for successful model predictions. Third, do the most important hydrophysical and topographic variables ranked by the RF algorithm explain suitable conditions for recruitment and cottonwood presence and absence? Both topographic and hydrophysical variables were included as potential predictor variables, with the distance from and elevation above the wetted channel hypothesized to be the most important drivers in juvenile cottonwood presence. This is based on terrain-hydrology-cottonwood establishment relationships, as seeds are deposited in receding flood flows along the active channel margins to create recruitment bands (Braatne et al., 1996; Mahoney & Rood, 1998; Scott et al., 1997; Stromberg, 1993), while recruitment elevations are dependent on an access to moisture that does not result in scouring by high flows at lower elevations or drought stress at higher elevations (Mahoney & Rood, 1998; Scott et al., 1997). Given the set-up and analysis of two different types of models in one article to enable comparison and synthesis, many details are provided in the Supplementary Materials.

2 STUDY SETTING

The lower Yuba River (LYR) is a ~ 37.5-km-long, gravel-cobble regulated river in northern California's Central Valley. The Yuba catchment drains 3,480 km^2 of the western Sierra Nevada Mountains before reaching the confluence with the Feather River (**Figure 1**). This area has a subtropical climate, experiencing cool, wet winters and hot, dry summers. The LYR's hydrology is driven by winter rainstorms and spring snowmelt, with most of the annual precipitation occurring between November and March. Flow to the LYR is partially regulated by dams and diversions (YCWA, 2013), including 79-m-high Englebright Dam and 7.3-m-high Daguerre Point Dam (both mining sediment barrier dams), but to a less degree than for other rivers in the region (Escobar-Arias & Pasternack, 2011).

Accepted, Corrected

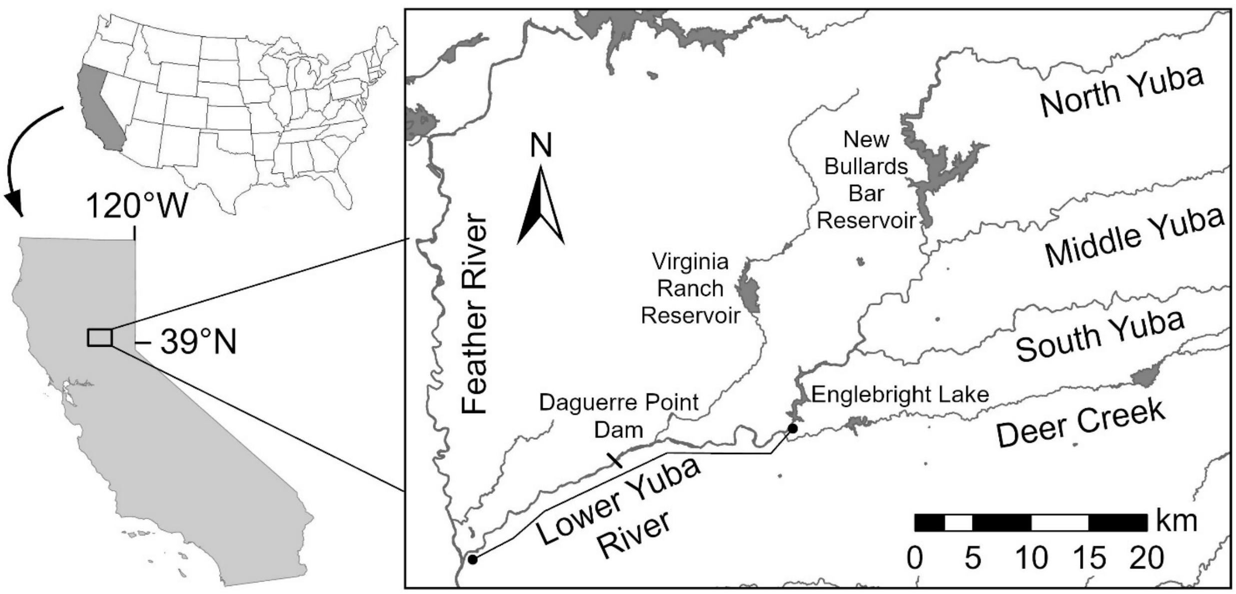

Figure 1. Lower Yuba River (LYR) map and location within northern California.

An estimated 280,209,355 m^3 of hydraulic mining sediment was created within the Yuba's catchment during and after California's Gold Rush between 1852 and 1906 (James, 2005), with almost 90% of it remaining there by the 1980's (Adler, 1980). Sedimentation aggraded the LYR's natural channel by 8-26 m (Adler, 1980; Gilbert, 1917) and changed the channel pattern. The deposition of unconsolidated mining sediment buried pre-existing riparian vegetation and may have altered riparian conditions by reducing the extent and diversity of vegetation, covering the existing soil with debris, and may have altered the capillary fringe and impacted soil moisture availability for roots (YCWA, 2013).

While the historic river valley underwent dramatic changes, the containment of flow into a smaller corridor where it has been for many decades has by now yielded a remarkably dynamic river responding to a dynamic flow regime (Gervasi et al., 2021). Aerial photographs from 1937 to 2010 present a cumulative increase in riparian vegetation along the LYR (YCWA, 2013). Many woody species are supported, including in order from more abundant to least, varying willow species (*Salix* spp.), Fremont cottonwood (*Populus fremontii*), blue elderberry (*Sambucus nigra* ssp. *caerulea*), black walnut (*Juglans hindsii*), Western sycamore (*Platanus racemosa*), Oregon ash (*Fraxinus latifolia*), white alder (*Alnus rhombifolia*), tree of heaven (*Ailanthus altissima*), and grey pine (*Pinus sabiniana*) (YCWA, 2013). However, even though riparian vegetation is increasing, the river still has large expanses of unshaded terrain.

3 METHODOLOGY

3.1 Experimental Design

The novelty of this research lies in exploring how the outcomes of different prediction approaches inform the science of riparian ecology and the practice of riparian conservation, especially with the difficult challenge of matching individual organism locations. To answer the questions posed, an experimental design was developed integrating high-resolution geospatial data and biological data (**Figure 2**). To answer the first question investigating how predicted seedling recruitment locations compared to field locations of juvenile cottonwoods, the Riparian Seedling Recruitment Model (RSRM) was used to predict seedling recruitment potential along the LYR for the years 2017-2021. Among deterministic models for this purpose, none had yet been tested against individual organism locations, let alone a large dataset of 2,957 locations spanning a long length of river. To answer the second question evaluating a statistical-empirical model's prediction accuracy for cottonwood presence/absence, a Random Forest (RF) algorithm was used with topographic variables and hydrophysical outputs from the RSRM as predictor variables. For the third question, a ranking of variable importance generated by the RF was used

Accepted, Corrected

to investigate the top ranked predictor variables and their biophysical sensibility for cottonwood recruitment.

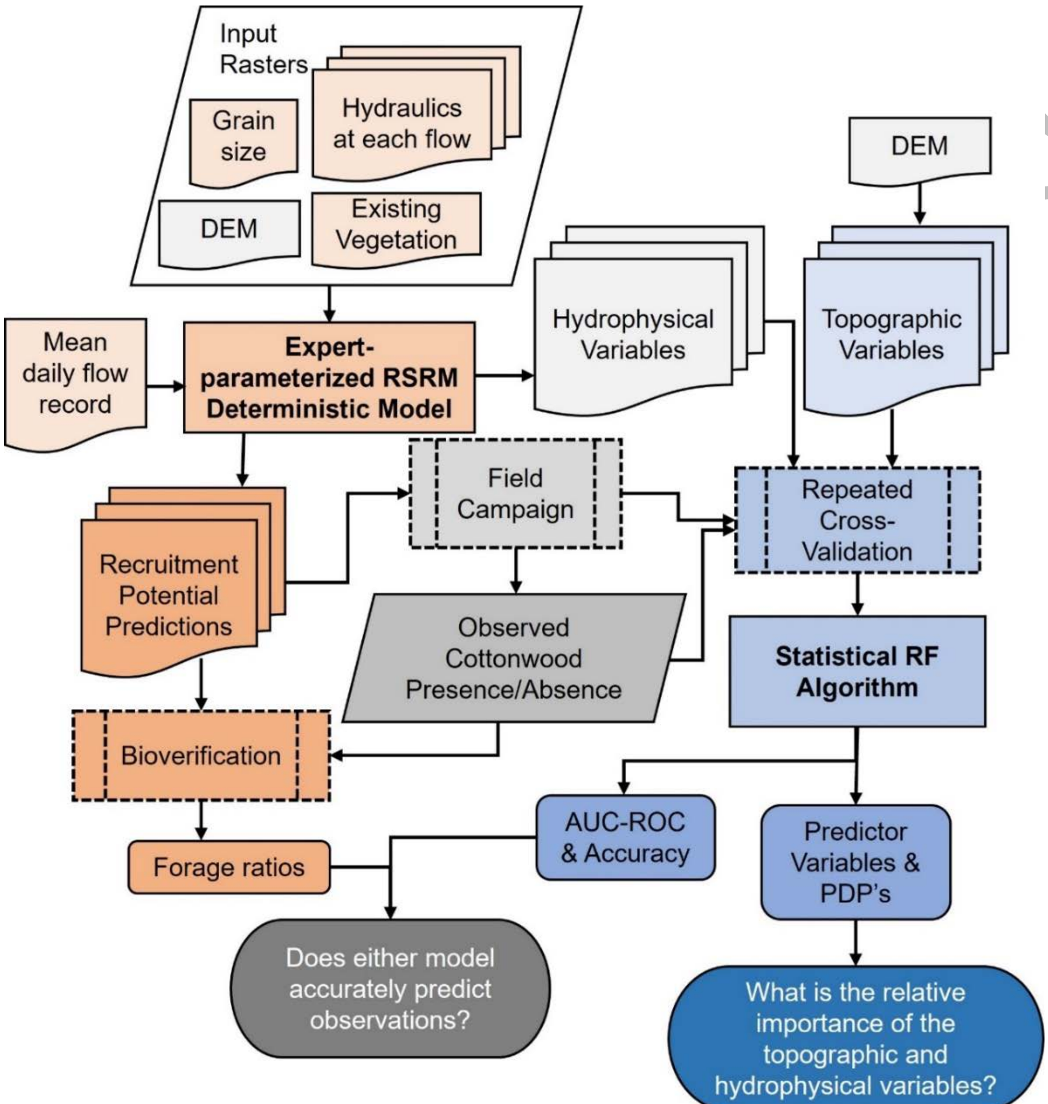

Figure 2. Experimental design and structure for the use of a deterministic and statistical model. The orange color follows the design for the Riparian Seedling Recruitment Model (RSRM), the blue follows the Random Forest (RF), and the grey represents overlapping data or questions. Methods used to explore model accuracy have a dashed outline.

3.2 Riparian Seedling Recruitment Model (RSRM)

As this was the first application of a novel model, there were no pre-existing model parameter value sets from past calibrated models to use as a starting point to parameterize this model and inform the field campaign. Further, in the absence of any pre-existing seedling observational data, model parameter sets could not be calibrated in advance. This component of the study did not seek to implement a *post hoc* framework in which the answers would be used to tune the model. Instead, the approach was consistent with how this type of model might be implemented in conservation practice for project planning and design in the absence of pre-existing data. Thus, physically realistic values were chosen on an expert basis with reference to the literature on Fremont cottonwoods and then newly collected observational data was used to test how well the model performed with the expert-based values. Commonly, models do not perform well without calibration, even when they are supposed to be physically realistic, but it is important to do that testing at the outset with a new model to help evaluate whether the scientific understanding underlying model structure and model parameterization is literally true and accurate enough for prediction.

3.2.1 Hydrophysical variables

Four hydrophysical processes were used in the RSRM to evaluate whether suitable site conditions for recruitment were met:

(1) Preparation of the bed through higher flows generating a large enough dimensionless bed shear stress to create new bare surfaces before seed dispersal,

(2) Desiccation or drought survival from stream stage and groundwater recession,

(3) Survival during prolonged inundation periods,

(4) Scour survival from flows with a high enough dimensionless bed shear stress for scouring effects after germination.

Threshold values were set in the RSRM to evaluate whether hydrophysical processes created suitable conditions needed for seedling recruitment and survival (**Table 1**). The RSRM uses the wetted area extents for the maximum and minimum flows during the seed dispersal period to create the spatial domain for areas of possible germination. An existing vegetation map was included to remove areas with established vegetation from analysis, as these are areas with competition for sunlight and moisture.

Table 1. Criteria set for the physical processes in the RSRM by Phillips & Pasternack, 2022, where the metric code was given depending on where a cell fell within the criteria for each condition and the bed shear stress was divided into the bed preparation phase and the scour survival phase. The seed dispersal period reflects the recruitment box in Mahoney & Rood, 1998.

Process	Criteria	Condition	Metric
Seed Dispersal Period	May 2 - July 4		
Bed Shear Stress (Bed preparation / scour survival)	0.047	Fully prepared / Fully disturbed	1 / 0
	0.030	Partially prepared / Partially disturbed	0.5 / 0.5
	0.000	Unprepared / Undisturbed	0 / 1
Mortality Coefficient (Desiccation survival)	< 20 days	Favorable	1
	20-30 days	Stressful	0.5
	>30 days	Lethal	0
Inundation	< 14 days	Favorable	1
	14	Stressful	0.5

Accepted, Corrected

28	Lethal	0

3.2.2 Model inputs

The mean daily flow record was collected at different points along the LYR (**Table S4**). The physical processes and parameter settings in the RSRM allow for user defined values based on the species of interest. Flow records were defined to start two years prior to the year of interest to evaluate if an area is prepared for germination through the sediment mobilizing flows that create bare surfaces. May 2nd was set as the beginning of cottonwood seed dispersal and marks the beginning of the year of interest for analysis. The years analyzed in this study were 2017 to 2021, requiring the mean daily flow record for the years of 2015 to 2022.

Spatial data inputs used in RSRM modeling included topography, hydraulics, sediment grain size, and vegetation raster data. Topographic data and derivative variables were available from a 0.91-m resolution 2017 digital elevation model (DEM) (**Table S1**). The DEM came from a point cloud integrating airborne LiDAR, boat-based multi-beam echosounder, and ground-based RTK GPS surveys (Silva & Pasternack, 2018; Gervasi et al., 2021). Steady-state, spatially explicit hydraulic rasters for velocity, depth, and water surface elevation (WSE) were available from a validated two-dimensional (2D) hydrodynamic model simulated using TUFLOW HPC with outstanding performance (Pasternack, 2023; **Table S2 & Table S3**). A total of 45 flows from 8.5-2,464 m^3/s were used, covering the range of discharges that occurred on the LYR for the years of interest. Grain size data for the LYR was available for 2017 from a previous study that used a RF algorithm based on LiDAR data and grain size samples from the field to create a sediment facies map (Díaz Gómez et al., 2022). The average grain size was approximated for

Accepted Article

regions not included within the mapped area. An existing 2017 vegetation map created from LiDAR data was used to identify areas of established vegetation (i.e., taller than 0.6 m).

As the DEM and vegetation map were created in 2017 and the study domain is too large to re-map topo-bathymetry and grain-size metrics annually, we applied a fixed topography and a fixed grain-size map in the RSRM for the study period. That choice is supported by the study years of 2017-2022 being a dry period of ecological recovery after geomorphically significant flooding in early 2017. Floods during the study period were of modest magnitude and short duration, with duration having been found to be very important to this river's morphodynamic volatility (Gervasi et al., 2021). Assessment of the validity of this assumption is presented in discussion section 5.1.

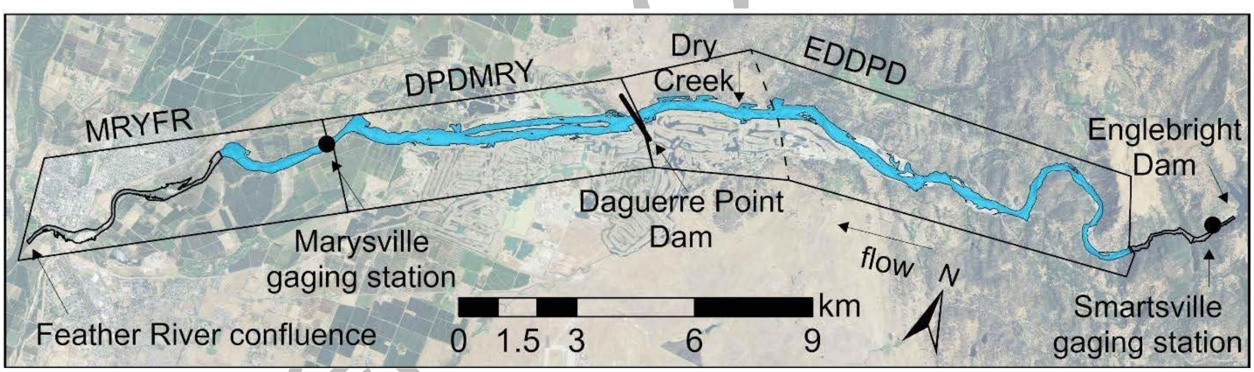

Figure 3. The LYR with the three modeling domains (MRYFR, DPDMRY, and EDDPD), gaging stations, and dams. MRYFR encompasses the segment immediately downstream of the Marysville gaging station to the confluence with the Feather River, DPDMRY extends from Daguerre Point Dam to the Marysville gaging station, and EDDPD is from the Englebright Dam to Daguerre Point Dam. The blue region of the river represents the section that was used for field data collection. The dashed line within EDDPD represents where this domain was split into two hydrological sections.

Accepted

The study area (**Figure 3**, blue) was split into three different RSRM modeling domains for computational and analysis efficiencies. Further, the Dry Creek tributary inflow and an irrigation diversion at Daguerre Point Dam (**Figure 3**) required use of three different discharge data sources. The inflow from Dry Creek required the upstream modeling domain to be split into two hydrological sections (**Table S4**).

3.2.3 Recruitment potential predictions

After applying constraints to the hydrophysical processes, the recruitment potential for a given cell is determined. The metric for each hydrophysical process is weighted by a coefficient and then the product of those terms is computed to create recruitment predictions at a 0.46-m^2 resolution (**Table 2**). This method is a common approach in environmental science and management (Leclerc et al., 1995; Renard et al., 1997). In the absence of any prior knowledge, all coefficients were given the same value of 1.0. Each hydrophysical process was weighted equally when computing the recruitment potential classes.

Table 2. Final recruitment potential classes from Phillips & Pasternack, 2022.

Description	Stressful Parameters	Combined Value
Optimal	0	1
Favorable	1	0.5
Stressful	2	0.25
Tolerable	3	0.125
Likely lethal	4	0.0625
Lethal	-	0

Accepted, Corrected

3.2.4 Field site selection

Site selection was made on an equal-effort basis (Johnson, 1980; Manly & Alberto, 2014) using a two-way stratification. Observation sites were stratified based on the RSRM-predicted recruitment potential class map and three longitudinal river sections used as 2D modeling domains. An equal number of randomly selected sites from within each class were surveyed (**Figure 4**). As a goal, ≥ 10 sites per class across the LYR were to be sampled, with 3-4 sites per class in each modeling domain (**Figure 3**). Sites that were inaccessible, unsafe, or highly impacted were excluded.

Accepted, Corrected

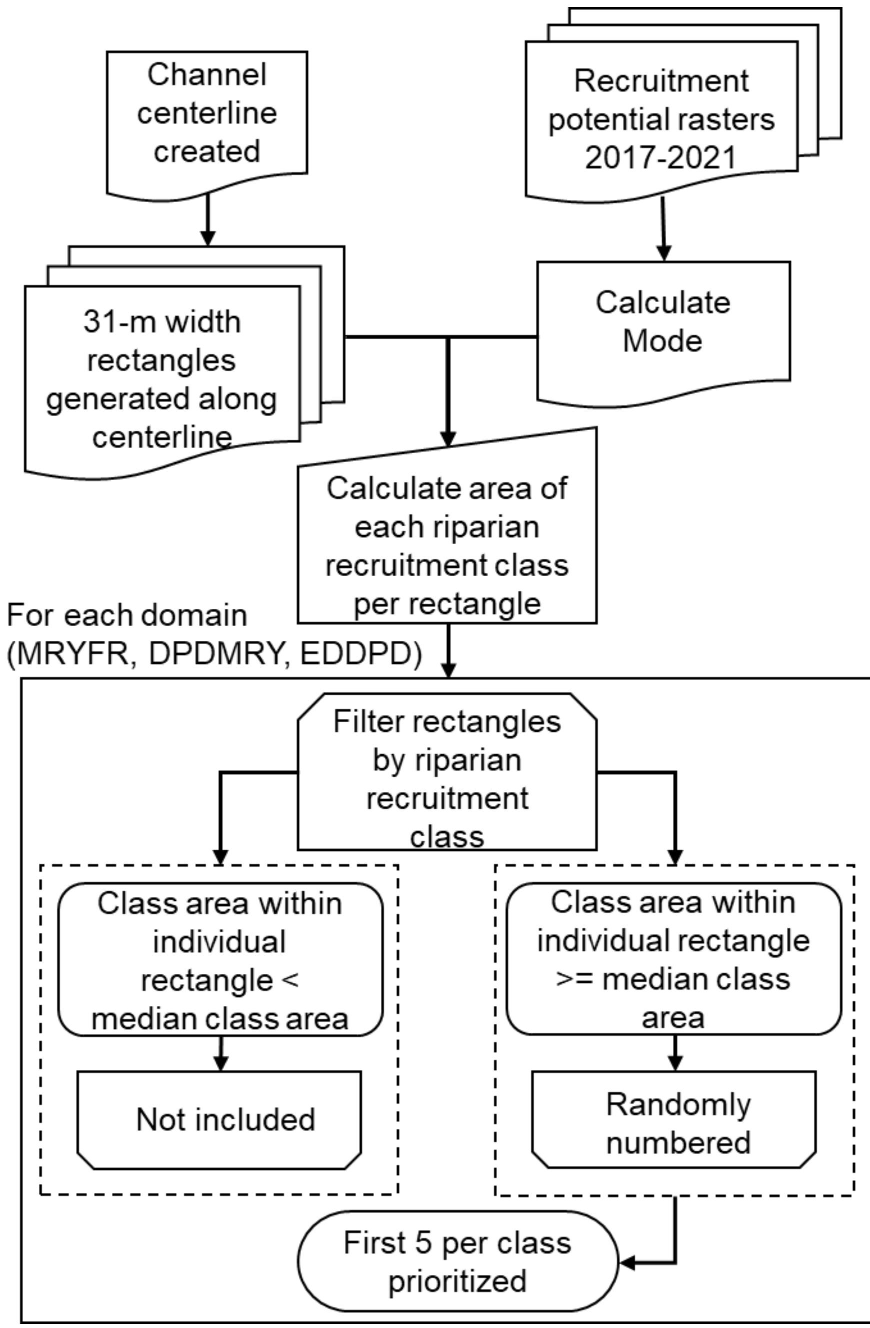

Figure 4. Workflow for field site selection. Mode refers to calculating that statistical variable for each raster. This process was repeated for each modeling domain illustrated in **Figure 3**.

The first step that had to be done was to divide the river domain into a population of potential sampling sites. Following the method of Wyrick and Pasternack (2014), the LYR's river-corridor centerline was stationed and sectioned into 31-m longitudinal rectangles, laterally spanning the wetted area of the highest discharge of 2,384-m^3 for 2015-2022. These rectangles were then split along the centerline to create 2253 sites. Due to meandering and topographic nonuniformity, site lateral extents and areas varied. While sites technically extended into the water, we only observed dry land beyond the low-flow water's edge present during field work. **Figure 5** illustrates sampling sites.

The RSRM has 6 values, or classes, for annual recruitment potential, so in the second step we sought to have an equal number of sampling sites for each class. One complication is that the study period spanned 5 years, so each river location has 5 classes. To resolve this, we calculated the mode (i.e. the most frequently occurring class) of annual recruitment potential to represent the recruitment potential over the study period in each modeled grid cell.

Another complication is that a potential sampling site may have more than one class in its area. To account for that, we then converted the raster of mode recruitment classes into polygons and "unioned" those polygons with the sampling sites from step one, so that every sampling site was now subdivided by areas of mode recruitment potential class.

We then developed a procedure to distinguish whether a class was sufficiently abundant in a site for that site to be appropriate to use as a sample of that class. Aligned with this concern was that consideration that it is not meaningful, logistically easy, cost-effective, or representative to sample within very small areas, so we sought to remove small class polygons from process of selecting field sites. To do that, the two-way polygons of recruitment class and sampling site were filtered and grouped by class in each model domain. A rectangle could be grouped multiple

Accepted, Corrected

times if it contained more than one class. The median area for each class within each model domain was then calculated and used as a size threshold, eliminating sampling sites with a class area below the median from selection. For example, imagine a site that has 3 m^2 of class 3 and 400 m^2 of class 4 in it. Such a site is not representative of class 3, so would have been eliminated from consideration for representing class 3 but kept for that purpose for class 4. Remaining sampling sites with a class area above the median were assigned random numbers. Sampling sites numbered one through five were given priority for field observation, with considerations for safety and accessibility sometimes necessitating going to the next lower ranked priority site.

3.2.5 LYR field data collection

Field sites were surveyed August through November 2022 (**Figure 5**). For sites that had cottonwoods, the height and diameter was collected from every individual present (**Figure 6)**. A survey-grade Trimble R8 RTK GPS receiving real-time corrections from a commercial regional benchmark network was used to record geographic coordinates (horizontal accuracy of ~ ± 3 cm) of every cottonwood observed within the site.

Accepted, Corrected

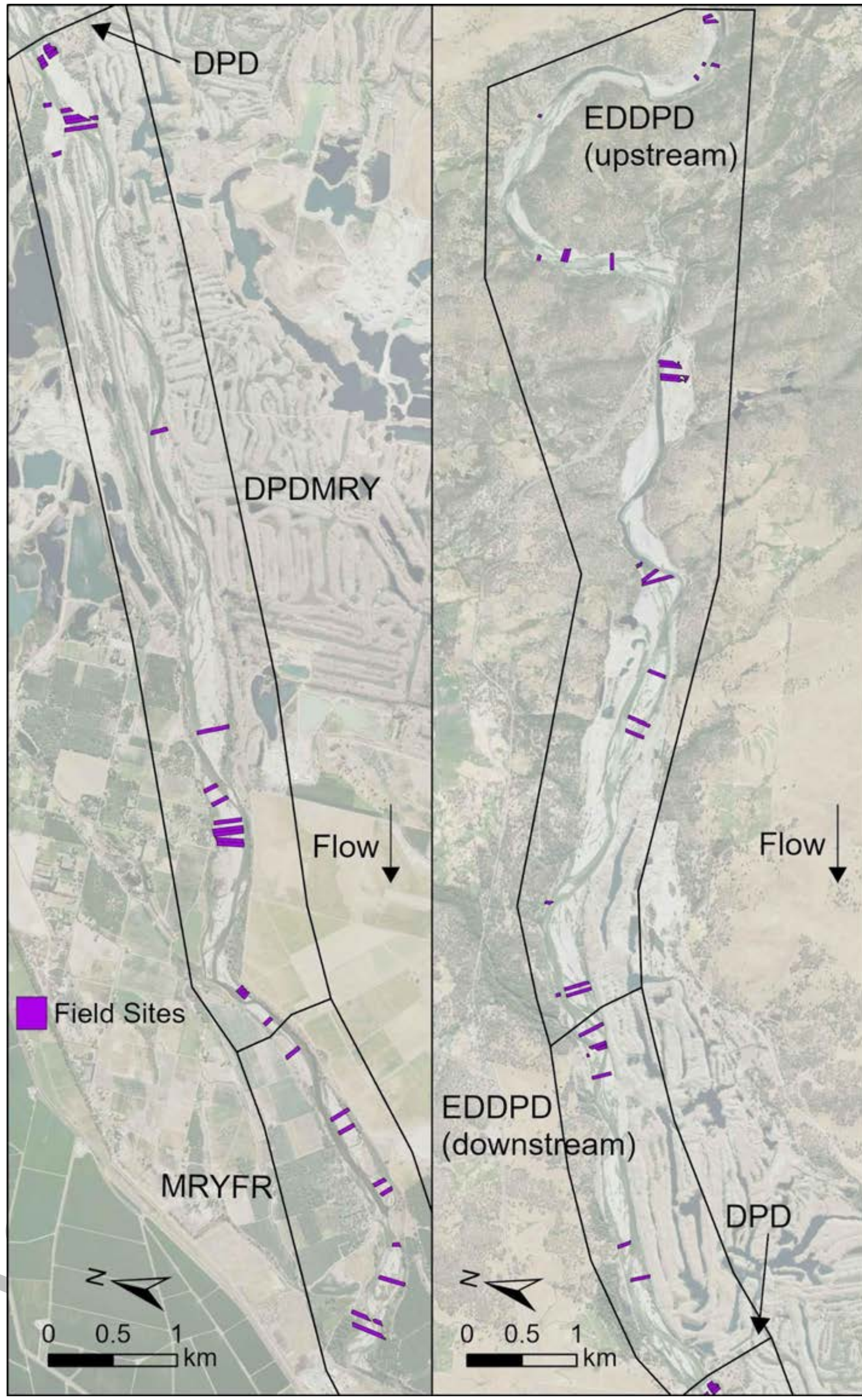

Figure 5. LYR field sites within the modeling domains. The map is split at Daguerre Point Dam (DPD). The downstream of DPD is the left image and upstream of DPD is the right.

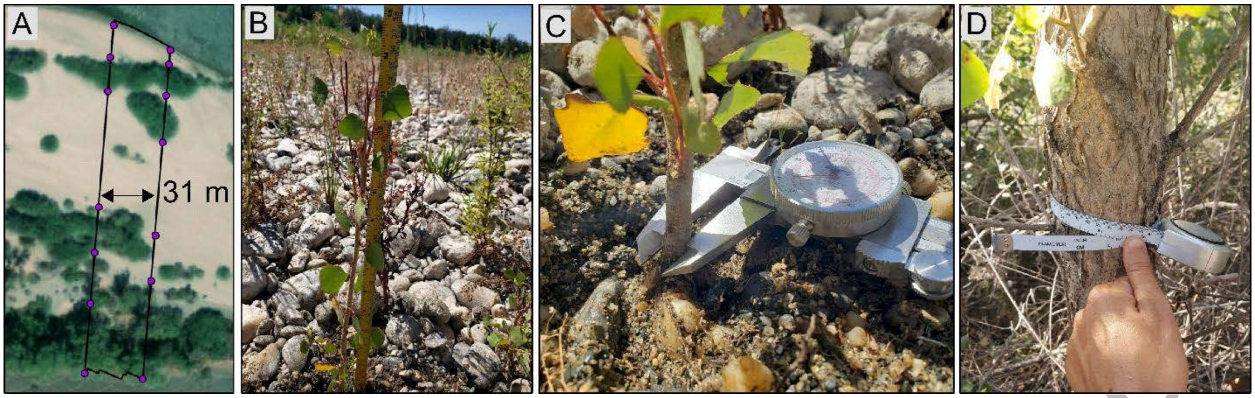

Figure 6. An example field site (A) and methods for data collection: B) measuring height with a tape measure, C) diameter above the root collar with a caliper, and D) DBH with a diameter tape.

A total of 2,957 juvenile cottonwood locations were recorded within the boundaries of 70 sampled sites. The recruitment class of 'Likely Lethal' had less sampled classes due to the way the RSRM calculates recruitment classes (**Table 3**). Most sites in DPDMRY were at the upstream and downstream ends, as the middle could only be reached by kayak. EDDPD was fully accessible.

Table 3. Number of recruitment classes sampled per domain.

	Recruitment Class						
Domain	Lethal	Likely Lethal	Tolerable	Stressful	Favorable	Optimal	Total # Sites
MRYFR	4	0	4	2	0	0	10
DPDMRY	5	0	7	4	6	5	27
EDDPD	6	4	6	6	6	5	33
Total # Sites	15	4	17	12	12	10	70

Accepted, Corrected

3.2.6 RSRM bioverification

An evaluation of RSRM prediction accuracy was performed through a method termed bioverification, comparing the number of recorded locations of juvenile cottonwoods within each area of recruitment potential classes modeled (Kammel et al., 2016; Moniz et al., 2020). Bioverification uses an electivity index to evaluate two criteria. First, there must be at least one recruitment potential class exhibiting preference and at least one exhibiting avoidance to demonstrate that the model can differentiate conditions. Second, the electivity index must increase as recruitment potential increases from class to class (Kammel et al., 2016). Many electivity indices exist, but given the abundance and simplicity of this data, the classic forage ratio (FR) was used. It was calculated as the ratio of the percent of cottonwood observations in the mode recruitment potential class (i.e., percent occurrence, aka utilization) to the percent area of that class (i.e., percent availability). A FR > 1 indicates an organism's preference for a habitat, while a FR < 1 indicates an avoidance. The further from 1.0 a FR value is, the more a habitat is preferred or avoided by the designated organism. A FR ≈ 1 for a class indicates behavior indistinguishable from random and cannot be attributed to a species showing preference or avoidance for that class in the model. A second set of FR values were also computed using the 2017-2021 maximum recruitment potential class values for a given cell (**Table S7; Figure S2**).

Accepted Article

3.3 Random Forest Algorithm

A RF supervised classification algorithm was used to address the second research question (**Figure 7**), modifying the RF model previously used to predict and analyze all riparian vegetation on the LYR (Diaz-Gomez et al., 2025). A two-step pre-processing was undertaken to prepare predictors and a binary response variable of presence or absence. The values of each predictor were then extracted at each binary response variable location and used in the *caret* package in R (Kuhn, 2008) to perform the RF and generate hypothesis-testing metrics. The number of trees used was 500 as the conservative default value needed to stabilize the prediction accuracy (Maxwell et al., 2018; Diaz-Gomez et al., 2025). The number of predictor variables randomly sampled at every node was defined by a grid-search method with a resolution of 1, with node values between 1 and the number of variables (20) (Probst et al., 2018, Zhang et al., 2020, Diaz-Gomez et al., 2025).

Accepted, Corrected

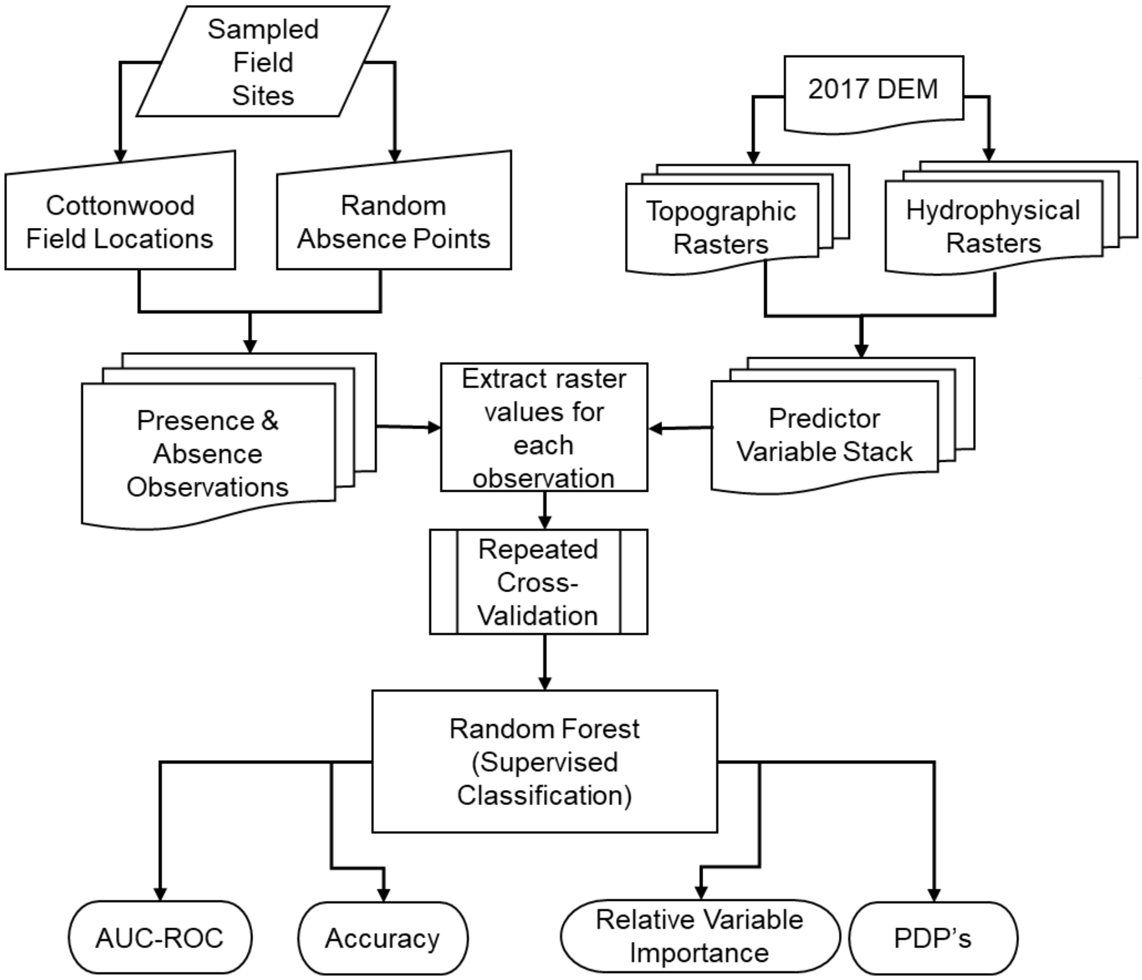

Figure 7. Data processing and RF machine learning model workflow.

3.3.1 Binary response variable

For the supervised classification, cottonwood presence and absence were used as the binary class variable. Field locations of juvenile cottonwoods were used to indicate presence cells at 0.91-m resolution. Cells that fell within cottonwood clusters were classified as presence, with no differentiation made between cells containing one or more cottonwoods. As a result, the 2,957 observed juvenile cottonwood locations were reduced to 1,349 presence cells. An equal number of absence cells were randomly created without duplication in any one cell within

Accepted

surveyed sites after excluding presence cells, resulting in a total of 2,698 samples for model training and testing.

3.3.2 Predictors

Twenty variables were chosen for predicting juvenile cottonwood presence or absence, using the four hydrophysical variables from the RSRM and 16 DEM-derived topographic features characterizing the terrain in 2017 (**Table 4**). The RSRM created 0.46-m resolution rasters for the four 2017 hydrophysical rasters, which were later resampled to the same 0.91-m resolution as the DEM and the subsequent topographic variables. Topographic variables were numerically continuous, while hydrophysical variables had three discrete values (**Table 4**). As a sensitivity analysis, the RF was run multiple times with varying categories of the predictor variables to compare performance metrics, with those results in Supplementary Materials.

Table 4. Predictor variables consisting of the hydrophysical processes rasters produced by the RSRM and topographic rasters produced from the 2017 LYR DEM.

Predictor	Indicator of	Source
Bed Preparation	Preparation of the ground before seed dispersal by scouring flows to clear existing vegetation and debris. 1 = fully prepped, 0.5 = partially prepped, 0 = unprepared	(Phillips & Pasternack, 2022)
Recession Rate	Rate of stream stage and water table decline compared to seedling root growth. 1 = favorable rate, 0.5 = stressful rate, 0 = lethal rate	(Phillips & Pasternack, 2022)

Accepted, Corrected

Inundation Survival	Inundation length impacts on seedlings. 1 = favorable inundation, 0.5 = stressful inundation, 0 = lethal inundation	(Phillips & Pasternack, 2022)
Scour Survival	Impact of scouring from higher flows after germination. 1 = undisturbed, 0.5 = partially disturbed, 0 = fully disturbed	(Phillips & Pasternack, 2022)
Detrended Elevation (m)	Removes the down-valley slope while preserving local topographic variations	(Pasternack et al., 2018)
Channel Proximity (m)	Distance of a cell to the 2017 wetted baseflow (~1,000 cfs). Influences the depth to the water table, inundation duration and depth, and distance to the river (Auchincloss et al., 2012).	ArcGIS Pro
Flow Direction	Direction of flow from every cell to its steepest downslope neighboring cell	ArcGIS Pro
Flow Accumulation	Accumulated flow to each cell	ArcGIS Pro
Topographic Profile	Direction of maximum slope by being upwardly convex (-), upwardly concave (+), or flat (0). Characterizes surface moisture accumulation, flow and speed of water by gravity, erosion and deposition of sediments.	(Evans & Murphy, 2023)
Topographic Planform	Perpendicular to the maximum slope by being laterally concave (-), laterally convex (+), or flat (0). Characterizes surface moisture accumulation, flow and speed of water by gravity, erosion and deposition of sediments.	(Evans & Murphy, 2023)
Topographic Curvature (McNab)	Surface curvature index based on features confining the view from the center of a 3x3 window (McNab, H.W. 1989).	(Evans & Murphy, 2023)

Accepted, Corrected

	Characterizes surface moisture accumulation, flow and speed of water by gravity, erosion and deposition of sediments.	
Topographic Slope	Rate of elevation change or steepness at each cell (degrees). Steeper terrain has a higher slope while flatter terrain has a smaller slope.	ArcGIS Pro
Lateral Relative Aspect	Pixel direction facing toward (-) or away (+) from the river	(Diaz-Gomez & Pasternack, 2021)
Longitudinal Relative Aspect	Pixel direction facing downstream or upstream	(Diaz-Gomez & Pasternack, 2021)
Topographic Position Index (TPI)	Difference in elevation between a cell's central point and the average elevation of its nearest neighbors. The central point is higher (+) or lower (-) than its average surroundings (De Reu et al., 2013). Describes local topographic heterogeneity	(Hijmans, 2023)
Terrain Ruggedness Index (TRI)	Ruggedness by calculating the sum elevation change between a cell and the 8 nearest neighbors. Smaller values are less rugged, greater values are more rugged. Provides an indicator of topographic heterogeneity (Riley et al., 1999).	(Hijmans, 2023)
Roughness	Difference between the maximum and minimum value of a cell and its 8 surrounding neighboring cells.	(Hijmans, 2023)
Vector Ruggedness Measure (VRM)	Terrain ruggedness by measuring 3-D vector dispersion. Ranges from 0 (flat) to 1 (rugged). Captures both aspect and slope for a measure of terrain heterogeneity (Sappington et al., 2007).	(Evans & Murphy, 2023)
Surface Relief Ratio (SRR)	Rugosity of a continuous surface within a specified window. Expresses topographic geometry (Pike & Wilson, 1971).	(Evans & Murphy, 2023)

Accepted

Pixel-Scale Topographic Variability	LiDAR point standard deviation within a pixel.	(Weber & Pasternack, 2017)

3.3.3 Implementation

The repeated k-fold cross-validation resampling method (Kuhn & Johnson, 2013) was used for RF implementation, where k indicates the number of groups the dataset is split into (10 herein). The observations at 0.91-m resolution were divided into 10 subsets (or folds) of equal size, each with 269 samples. One group is a holdout for validation and nine train the model, repeated 10 times until every group has trained and validated the model.

3.3.4 RF model performance

The RF model was tested for the same two performance indicators as the RSRM – differentiation of presence/absence locations and whether the directionality of the variables matched biophysical mechanistic sensibility. Differentiation was evaluated using several metrics, beginning with an averaged confusion matrix to portray the number of correct and incorrect predictions by the RF in terms of the overall accuracy, producer’s and user’s accuracy, and omission and commission errors (Fawcett, 2006; Sokolova & Lapalme, 2009). The next assessment used the area under the curve (AUC) for the receiver operator characteristic (ROC) (Fawcett, 2006).

Biophysical mechanistic sensibility was also evaluated with several methods. Permutation-based feature importance (Breiman, 2001) and partial dependence plots (PDPs) (Friedman, 2001) evaluated variable importance and helped interpret how each predictor affected the presence of cottonwood given everything else is the same. Then the median, upper quartile, and lower quantile values for the more important predictors were inspected to evaluate differences in presence and absence locations. The directionality of the predictor-cottonwood relationship and the predictor extent of presence and absence conditions were evaluated against biophysical reasoning.

4 RESULTS

4.1 Cottonwood Recruitment Patterns

A higher density of cottonwoods was located downstream of DPD than upstream of it (**Figure 8**). Downstream of DPD recruitment for seedlings that germinated in the 2022 summer was observed to be in sporadic dense patches, located mostly on lateral and point bars, islands, or along backwater and abandoned channels. Newly formed and dynamic islands were also found to have dense clustering of seedlings. Juvenile cottonwoods that were not in their first growing season were found in scattered stands among other riparian species, such as willows, or by themselves. There were dense and robust stands of mature cottonwoods observed in these regions.

Upstream of DPD, a dense seedling cluster was found on an active point bar. However, juvenile cottonwood locations were more scattered and individual in this section. Moving upstream, the surfaces became barer and had less vegetation present. A smaller number of mature cottonwoods was observed in this segment.

Accepted, Corrected

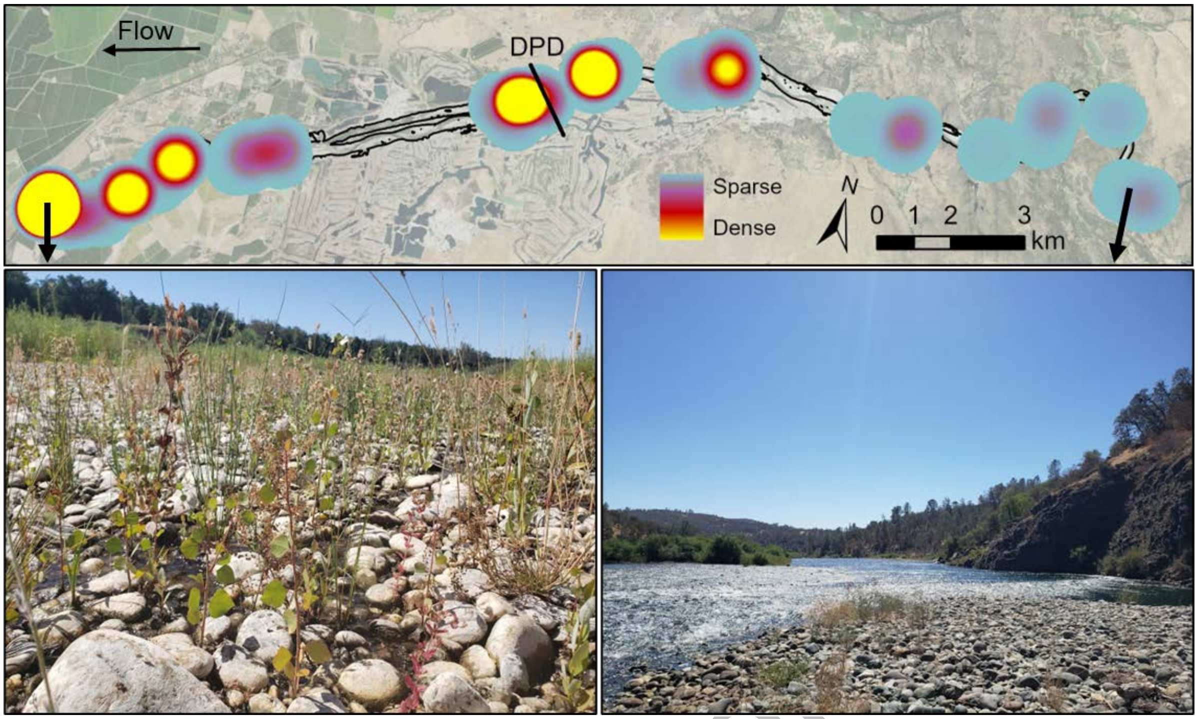

Figure 8. Heat map of surveyed cottonwoods along the LYR with two perspectives, one downstream of DPD and one above.

4.2 Question 1: Does the RSRM accurately predict cottonwood seedling presence?

Among the total of 2,957 young cottonwood presence locations used in the forage ratio (FR) calculations, 1,408 presence points were located within RSRM modeled results and 1,550 presence points were outside modeled areas (because the RSRM considers this area beyond the modeled wetted area during seed dispersal). Four recruitment potential classes had FR values indicating avoidance, while one ("tolerable") had a value indicating preference (**Table S5; Figure 9**). These results meet the first criteria that requires the occurrence of both preference and avoidance to be exhibited by separate classes, which can be seen in **Figure 9**. However, the second criteria which requires that the FR, the percent of cottonwood occurrence to the percent area available for each recruitment class, increase as the recruitment potential increases, was not met. The FR calculated using the 2017-2021 maximum recruitment potential class values, instead of the mode, did not change overall results (**Table S6; Figure S2**).

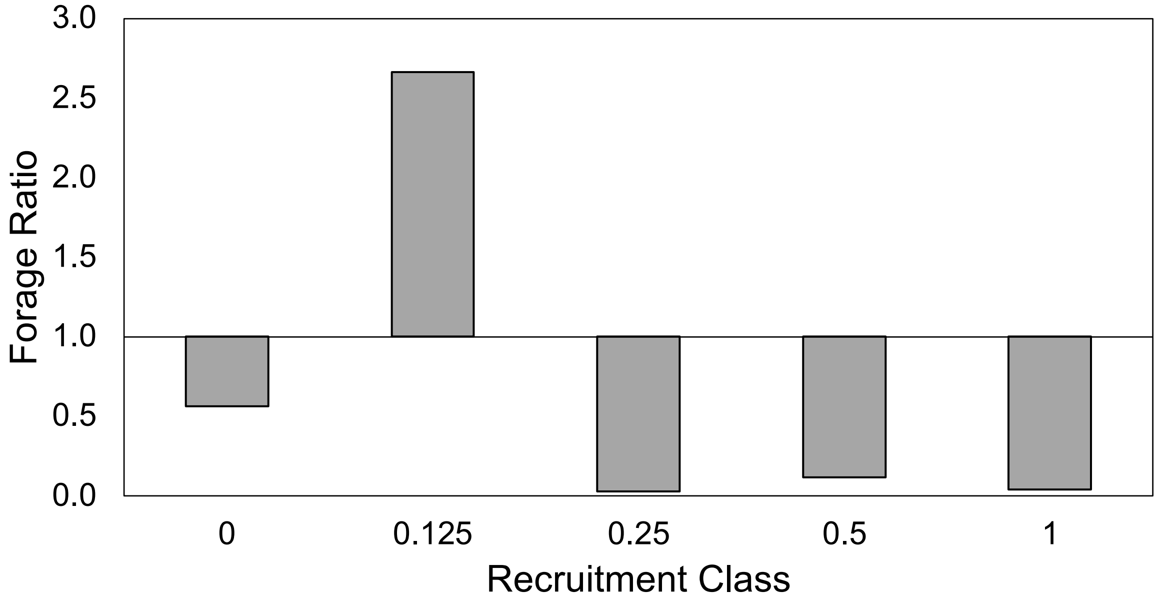

Figure 9. Forage ratio test results using the 2017-2021 mode recruitment class values for the expert-parameterized implementation of the RSRM. The class that represented "tolerable" recruitment potential was the only one that resulted in an FR indicating preference.

4.3 Question 2: Does the RF Model accurately predict presence and absence?

When compared with the testing data, the averaged confusion matrix of the RF prediction portrayed an overall accuracy of 87% of correctly predicting either an absence or presence point with a p-value < 2e-16 (**Table 5**). The RF model had a higher accuracy for predicting presence, while still obtaining the goal of having a good balance between the producer and user accuracies for presence and absence. The model performed well with an AUC-ROC of 94%, reaching this value with 8 variables available at each tree node (**Figure 10**). Remarkably, even a single variable produces an AUC-ROC > 90% and adding just three more variables increases the overall accuracy to 93%.

Table 5. The averaged confusion matrix for the RF repeated cross-validation scheme, with the bolded values representing the correctly classified observations.

Prediction	Reference			
	Absence	Presence	Total	User Accuracy
Absence	**1147**	151	1299	88%
Presence	202	**1198**	1400	86%
Total	1349	1349	2698	
Producer Accuracy	85%	89%		

Corrected

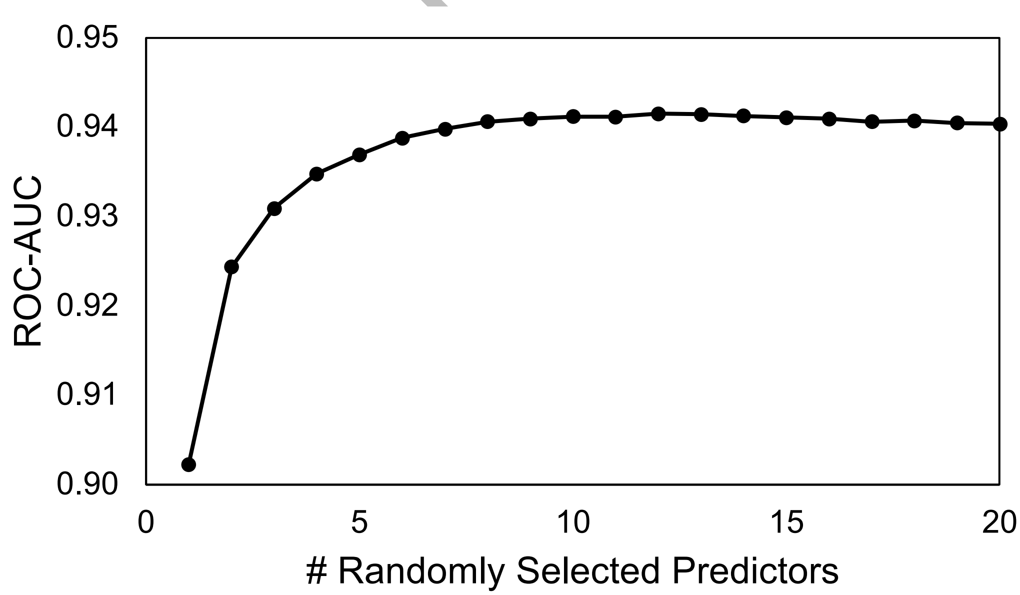

Figure 10. AUC-ROC value by number of randomly selected predictors.

4.4 Question 3: Drivers for cottonwood presence and absence

For predicting the presence/absence of young cottonwoods, the top four most important variables based on the RF-generated variable importance ranking include detrended DEM, channel proximity, inundation survival, lateral relative aspect, and the vector ruggedness measure (VRM), (**Figure 11**). Inundation survival was the fifth and the only hydrophysical variable from the RSRM in the top five.

Partial dependence plots (**Figure 12**) and statistical distribution metrics (**Table 6**) found that detrended elevation, channel proximity, lateral relative aspect, and VRM exhibited biophysically realistic directionality, while inundation survival did not. For example, the probability of cottonwood presence increased from detrended elevations of 0 to 2 m before a sharp peak, then decreased as the elevation increased further. For detrended elevation, presence cells were found at a lower median elevation than the absence points. The upper quantile value for presence was equivalent to the lower quantile value for the absence cells.

Table 6. Median, lower quantile (LQ), and upper quantile (UQ) values of the presence and absence points for the top five most important predictor variables.

	Presence			Absence		
Predictor	LQ	Median	UQ	LQ	Median	UQ
Detrended Elevation (m)	1.3	1.6	2.4	2.4	3.3	4.1
Channel Proximity (m)	0.0	5.4	19.9	14.1	36.4	65.5
Inundation Survival	0	0	1	1	1	1
Lateral Relative Aspect	-0.96	-0.09	0.97	-0.78	0.53	0.95
VRM	6E-05	2E-04	1E-03	8E-05	3E-04	1E-03

Accepted. Corrected

Based on these results, inundation survival was removed from the predictors and a new RF model was applied to compare model performance and the ranked variables of importance (**Figure S5**). Accuracy was similar, and the detrended elevation and channel proximity remained the two most important variables.

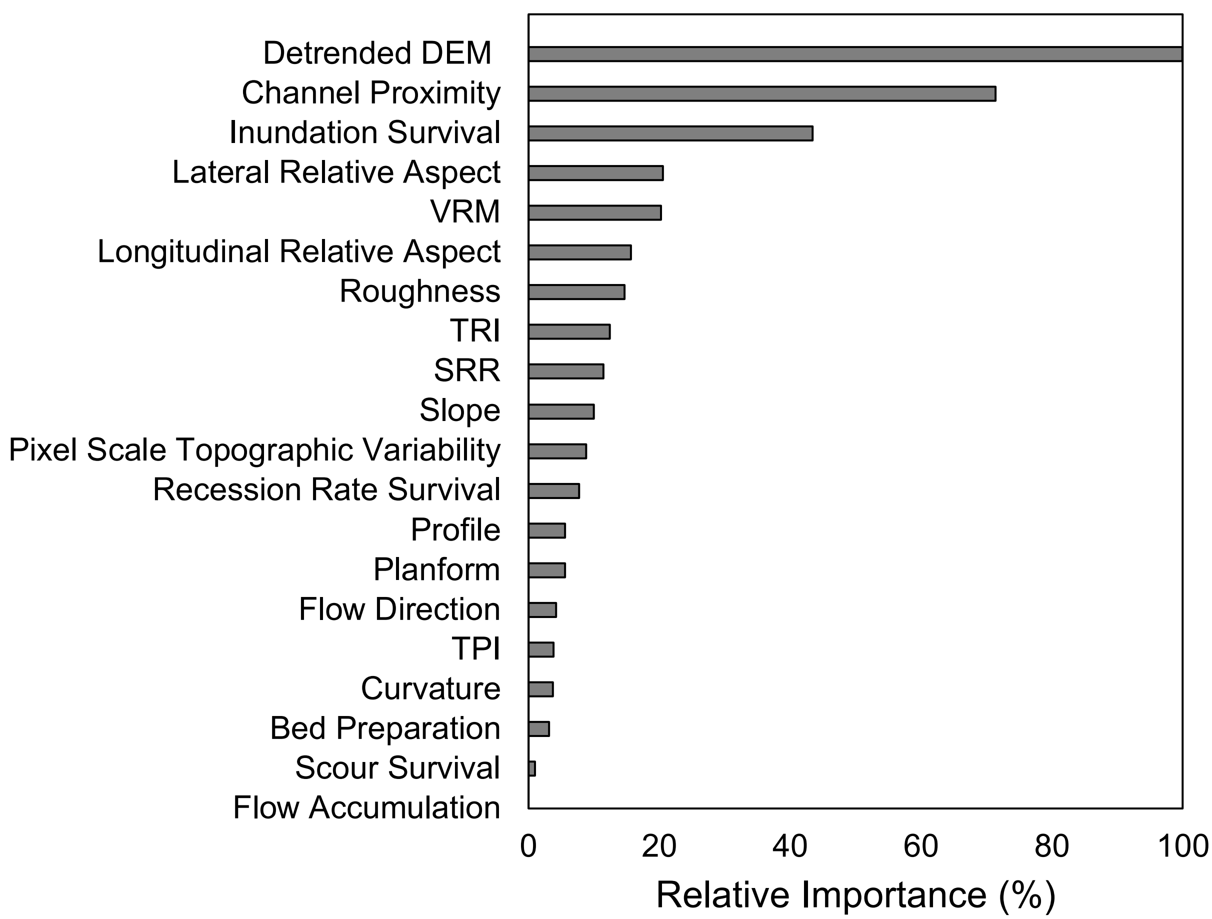

Figure 11. Relative contribution of variable importance for the RF model's predictions of cottonwood presence and absence. The most important variable is identified and assigned an

importance of 100%, with the other variables ranked relative to it. Each variable is explained in Table 4.

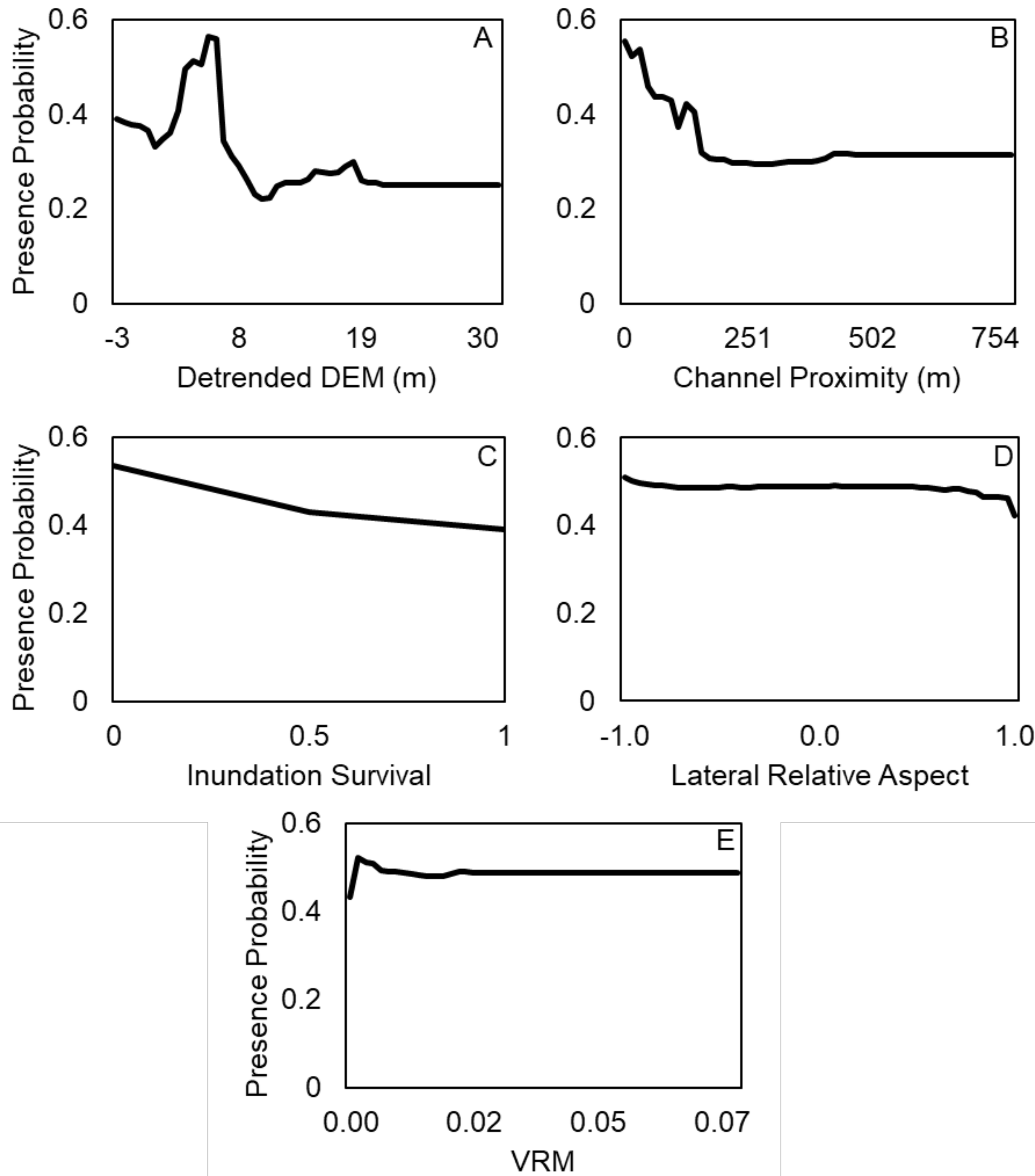

Figure 12. Partial dependence presence probability for the top four predictor variables: A) detrended DEM (m), B) channel proximity (m), C) inundation survival, D) lateral relative aspect,

and E) vector ruggedness measure (VRM). The black line represents the mean marginal response when the other predictor variables were kept constant.

5 DISCUSSION

5.1 Understanding RSRM Results

After comparing RSRM results with LYR locations of juvenile cottonwoods, the expert-based model did not bioverify. The test seemed to result in random results that did not necessarily indicate habitat preference or avoidance. Factors explaining this outcome include (1) a time lag between the years analyzed with the RSRM and when field surveys occurred; (2) unrecognized importance of local environmental factors in establishing model parameters and initial conditions; and (3) unrecognized differential sensitivities of the parameter criteria used.

While the RSRM produces seedling recruitment predictions for a seedling after its first year of life, it does not account for mortality that may have occurred after. Site selection for field data collection was based on the mode RSRM results for 2017-2021, so it is difficult to determine model accuracy if most seedlings died in earlier years. In addition, during the five years between the collection of data to create the 2017 DEM and the 2022 field season for this study, a few floods occurred and caused local geomorphic changes, potentially degrading model accuracy- though the same issue faced the RF model and it still yielded excellent performance. Many locations had minor to no changes, but particularly some depositional locations were highly prone to dynamism wherein the channel completely migrated and left behind abandoned or remnant channels. This resulted in the RSRM not being able to make predictions that reflected the current streambank or in areas with newly formed islands or land features. As erosional and

Accepted, Corrected

depositional processes are important for the creation or disappearance of new or bare surfaces and the location of the wetted channel for access to water are important for seedling recruitment, morphological changes need to be acknowledged. Without the monitoring of sites for each of these years the RSRM was used, a complete evaluation is difficult.

While the criteria and thresholds set for the hydrophysical processes in the RSRM (**Table 1**) was chosen based on existing scientific literature, site specific decisions based on local factors need to be considered (Stella et al., 2010). It is also difficult to compare criteria and results from differing studies due to variations in experimental designs and environmental conditions (Politti et al., 2018). An uncertainty exists in the criteria set for the RSRM, as chosen values may not apply well to the LYR or using the same threshold values for the entire extent of the LYR may have been too general.

The sensitivity for one or more parameters could be high, impacting the success of results even if the values chosen were close to being suitable for the LYR. Smaller sections that were carefully selected and studied may have been necessary for a more successful validation of the RSRM. A propagation of error from the 2D modeled hydraulic inputs, interpolation of the WLE, and the modeling of the RSRM itself may have also impacted results. The RF relative importance analysis suggests that different variables have unequal roles, and so the choice of equal weighting of hydrophysical variables in the RSRM may require re-assessment.

There could also be other factors or processes that were not considered in the development of the RSRM. Variables relating to the climate or location and number of mature cottonwoods that may release seeds relative to model predictions were not included. The RSRM considers areas within the floodplain inundation extent during the seed dispersal period as potential sites for germination, with no distinction between seeds dispersed by water, wind, or

Accepted, Corrected

the disruption of seed distribution due to dams. A larger density of juvenile cottonwoods was in the sampled sites located below Daguerre Point Dam when compared to the upstream sites. One factor that may be contributing to this is that more robust and expansive riparian forests with mature cottonwoods were observed in this area, potentially producing a larger number of seeds available to recruit on suitable surfaces. There could also be a role for the generally erosional setting upstream of DPD and depositional setting downstream of it (Carley et al., 2012), though that can change through time (Gervasi et al., 2021).

5.2 Cottonwood Presence and Absence

Despite some floods and morphodynamic changes from 2017 to 2022, the RF model was able to accurately predict juvenile cottonwood presence and absence in 2022 based on conditions in 2017, as indicated by performance assessment metrics. An accuracy of 87% was achieved for correctly predicting cottonwood presence or absence. Sensitivity was larger than the specificity, indicating the model was better at correctly predicting presence locations versus absence. AUC-ROC was high, reflecting optimal performance by the RF (Fawcett, 2006). The RF had a strong performance, similar to other binary RF classification models (Cutler et al., 2007; Maxwell et al., 2018), suggesting that the predictor variables provided enough useful environmental information to identify characteristics of cottonwood recruitment locations.

The two most important variables from the RF were detrended elevation and channel proximity, which are indicators for depth to the water table and flood inundation depth. The directionality of the statistical relations aligned with observations in the field and expectations from cottonwood literature and is observed in a visual representation of the predicted probability of cottonwood presence along a small section of the LYR (**Figure 13**). Juvenile cottonwoods

Accepted, Corrected

were found to be at lower elevations and closer to the baseflow channel when compared to the absence points, likely having been deposited on the moist substrate during receding flows. Seedlings that had recruited along the active margin of the channel were mostly located in dense clusters on the edges of point and lateral bars, which have the geomorphic surfaces and sediment processes needed to create suitable bare surfaces for cottonwood seedling recruitment (Braatne et al., 1996, 2007; Mahoney & Rood, 1998). In one location of dense clustering, the migration of the active channel had created large extents of new bare surfaces, allowing a large band of new recruitment.

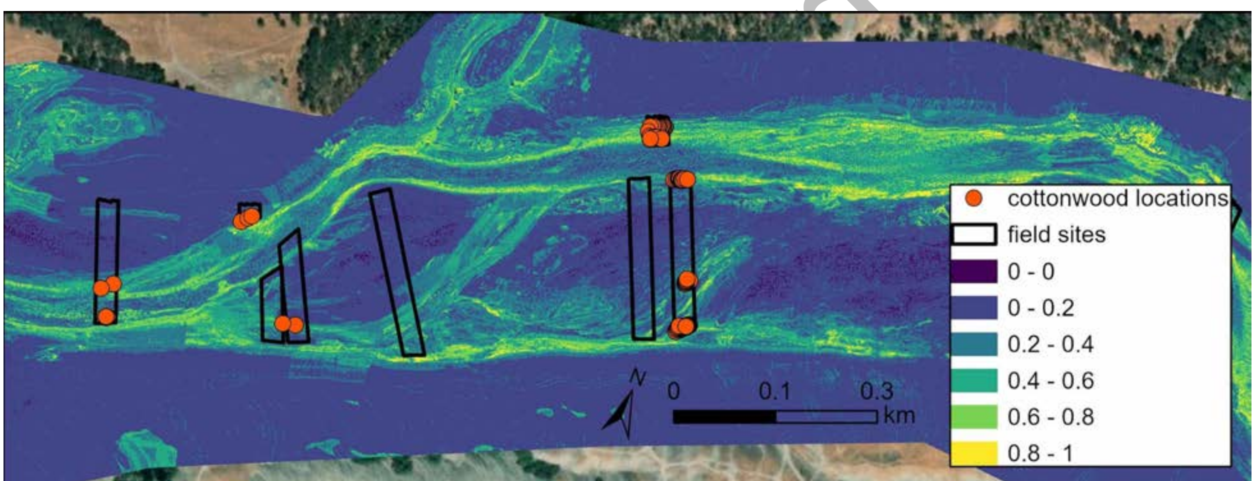

Figure 13. An example of the RF's predicted probability of cottonwood presence (1 is presence, while 0 is absence), with surveyed juvenile cottonwood locations within the sampled field sites.

Inundation survival was ranked as the 3rd most important variable, with presence points having a median value of lethal inundation and absence having favorable inundation value. This is opposite to what was initially hypothesized yet was analyzed for biophysical sensibility. Seedlings that recruited on new or bare surfaces close to the baseflow wetted area would experience longer durations of inundation when compared to those that recruited at higher elevations. New recruits in risky locations may have been modeled in the lethal inundation zones by the RSRM. In addition, inundation depth, duration, and a combination of both have significant impacts on seedling health and survival (Auchincloss et al., 2012), highlighting the importance of having the right parameterization for this variable. Changes that have occurred in active areas of the channel margin since 2017 may also have an influence. Recruitment since 2017 has occurred on newly created surfaces close to the channel that did not exist in 2017, which may have been modeled as inundated by the RSRM. The absence points were located further away from the active channel and may have experienced short or no periods of inundation. The ideal conditions of abandoned or remnant channels may have also been captured. Many abandoned channels below Daguerre Point Dam had been extensively colonized by juvenile cottonwoods, as the process of fine sediment deposition as the abandoned channel dewaters creates ideal moist surfaces and conditions for rapid colonization by pioneer species (Latella et al., 2024; Stella et al., 2011). Colonized abandoned channels above Daguerre Point Dam were not observed.

Long term survival for many of these recruited seedlings is not probable due to their location relative to the river water surface level during higher flow events, as they are likely to be scoured when sediment is mobilized. Cottonwood seedlings that had survived beyond their first few growing seasons and large cottonwood trees were observed in backwater areas or within willow and cottonwood bands on point bars and high on the riverbank far from the late summer stage position when the field sampling occurred.

Presence points were also more likely to face towards the river then away, as shown by the 4th most important variable being lateral relative aspect. The lateral relative aspect is linked

Accepted, Corrected

to hydraulic and sediment processes (Díaz Gómez et al., 2022), and may also be associated with the deposition of seeds that were transported by water. The vector ruggedness measure (VRM) ranked 5th in this study but was the most important influencer in another study on the 2017 riparian vegetation of the LYR (Diaz-Gomez et al., 2025). The median VRM for presence and absence points were similar, but the presence points had a larger average VRM than the absence points. VRM expresses heterogeneity in the surface by representing both aspect and slope, indicating that micro-variability in the terrain is needed for cottonwoods to recruit and establish.

5.3 Management Implications

The ability to accurately predict cottonwood seedling recruitment locations along a dynamic, regulated river is useful for informing riparian revegetation efforts and planting projects. Areas where seedlings naturally recruit indicate desirable locations and environmental characteristics that could be used to maximize recruitment opportunities for cottonwoods when managing river flows during varying water years. The identification of recruitment areas and their environmental characteristics can also help to inform manual plantings, which are used as a common cottonwood revegetation method (González et al., 2018). Plantings have the benefit of human site selection in areas determined to be favorable and may not be as vulnerable as a newly germinated seedling. The success of a planting does not first depend on disturbance flows to create new, bare surfaces, and larger plantings with already present roots may not be as vulnerable to receding water table levels or scouring flows. Yet a revegetation effort may fail due to unaddressed underlying factors (Briggs et al., 1994; Stromberg, 2001) and fluvial morphodynamics can cause site suitability to change on a frequency set by the disturbance regime. While many planting projects may report high mortality rates, is this necessarily a sign

Accepted, Corrected

of failure? Natural mortality occurs annually in seedling recruitment, so some mortality should be expected with plantings too. A realistic planting mortality threshold for "success" should be defined to achieve the desired result for a revegetation effort, with a successful planting effort attempting to replicate the prolific seedling recruitment that occurs during the successful years that support sustainable populations.

Although the RSRM was not calibrated, it was used in a way that is common in management practice, so study results have consequences for professional practice. Commonly, projects are done at sites lacking long-term monitoring data or the breadth of scientific investigations done over the last two decades on the LYR. As a result, practitioners rely on literature and their expert judgment for whatever models they are applying. The results of this study suggest that the underlying science to make a mechanistic predictive model is still missing key factors and the absence of model calibration introduces uncertainties in understanding variable importance and parameterization. It is particularly puzzling when the RF model yielded remarkably accurate results from just two very simple topographic inputs. Thus, how topography asserts itself through a "mechanistic chain" of cause and effect is highly significant and still elusive to simulate, necessitating further work. It may also be that the RSRM would be successful in a different setting than the LYR.

Like many rivers around the world, the LYR is dynamic, so it is important to have tools that can be effective as rivers change. Full morphodynamic modeling is plausible but it is still experimental (Camporeale et al., 2013) and, at meter resolution, very computationally expensive for long river segments. The DEM and hydraulic spatial data from 2017 were used in this study to model recruitment through 2021, which meant that the morphological changes to the LYR since 2017 were not accounted for. This is a realistic constraint as agencies or organizations

Accepted, Corrected

involved in river management efforts may be limited by money, making infeasible yearly monitoring and the frequent updating of large river datasets (i.e., high resolution DEM, vegetation, substrate, etc.). The ability to accurately model cottonwood seedling recruitment, or the recruitment of other pioneer species, using datasets that are not updated on a frequent basis is a valuable tool for the planning and implementation of river revegetation projects. At this time, machine-learning modeling outperforms deterministic modeling in this context.

6 CONCLUSIONS

This study found that the deterministic RSRM did not bioverify, which could be due to time lags between the years modeled and when field work occurred, uncertainty in the parameters due to local conditions, or sensitivity in the chosen criteria. While the RSRM did not bioverify, the RF model was successful in predicting the presence or absence of juvenile cottonwoods. This indicates that there is enough useful information available about the environmental characteristics of juvenile cottonwood locations needed to predict recruitment. Detrended elevation and channel proximity were ranked as the two most important predictor variables by the RF. The methods described in this study could be used to help inform revegetation efforts through natural recruitment or manual plantings, potentially resulting in more cost-effective and successful projects. Care should be taken to study the characteristics of a given site to make sure model criteria are suitable.

Accepted Article

7 ACKNOWLEDGEMENTS

This work was primarily funded by a public agency, Yuba Water Agency (Marysville, California, USA; (Awards #201016094 and #10446). It was also supported by the USDA

National Institute of Food and Agriculture, Hatch project number CA-D-LAW-7034-H. We thank Sunny Bommasani, Jessica Aguilar, Sam Bohart, and Arthur Koehl for field assistance. We thank Recology Yuba-Sutter, Sierra Foothill Research & Extension Center, Eddie Powell, and Perry Sims for permission to access the river through their land.

8 CONFLICT OF INTEREST STATEMENT*

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

9 REFERENCES

Abell, D. L. (1989). Proceedings of the California riparian systems conference: protection, management and restoration for the 1990's. September 22-24, 1988, Davis, California. *General Technical Report - Pacific Southwest Forest and Range Experiment Station, USDA Forest Service*, (PSW-110), x-pp.

Adler, L. L. (1980). Adjustment of the Yuba River, California, to the Influx of Hydraulic Mining Debris, 1849-1979. Master of Arts Thesis, University of California, Los Angeles.

Amlin, N. A., & Rood, S. (2001). Inundation tolerances of riparian willows and cottonwoods. *Journal of the American Water Resources Association*, *37*(6), 1709–1720. https://doi.org/10.1111/j.1752-1688.2001.tb03671.x

Amlin, N. M., & Rood, S. B. (2002). Comparative tolerances of riparian willows and cottonwoods to water-table decline. *Wetlands*, *22*(2), 338–346. https://doi.org/10.1672/0277-5212(2002)022[0338:CTORWA]2.0.CO;2

Auchincloss, L. C., Richards, J. H., Young, C. A., & Tansey, M. K. (2012). Inundation depth, duration, and temperature influence Fremont cottonwood (Populus fremontii) seedling growth and survival. *Western North American Naturalist*, *72*(3), 323–333. https://doi.org/10.3398/064.072.0306

Benjankar, R., Burke, M., Yager, E., Tonina, D., Egger, G., Rood, S. B., & Merz, N. (2014). Development of a spatially-distributed hydroecological model to simulate cottonwood seedling recruitment along rivers. *Journal of Environmental Management*, *145*, 277–288. https://doi.org/10.1016/j.jenvman.2014.06.027

Benjankar, R., Tranmer, A. W., Vidergar, D., & Tonina, D. (2020). Riparian vegetation model to predict seedling recruitment and restoration alternatives. *Journal of Environmental Management*, *276*, 111339. https://doi.org/10.1016/j.jenvman.2020.111339

Bertoldi, W., Siviglia, A., Tettamanti, S., Toffolon, M., Vetsch, D., & Francalanci, S. (2014). Modeling vegetation controls on fluvial morphological trajectories. *Geophysical Research Letters*, *41*(20), 7167–7175. https://doi.org/10.1002/2014GL061666

Braatne, J. H., Rood, S. B., & Heilman, P. E. (1996). *Life history, ecology, and conservation of riparian cottonwoods in North America*. *Biology of Populus and its Implications for Management and Conservation*. Retrieved from https://www.researchgate.net/publication/284033185_Life_history_ecology_and_conservation_of_riparian_cottonwoods_in_North_America

Braatne, J. H., Jamieson, R., Gill, M., & Rood, S. B. (2007). Instream flows and the decline of riparian cottonwoods along the Yakima River, Washington, USA. *River Research and Applications*, *23*, 247–267. https://doi.org/10.1002/rra.978

Breiman, L. (2001). Random Forests by Leo Breiman. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021*, 542–545.

https://doi.org/10.1109/ICCECE51280.2021.9342376

Briggs, M. K., Roundy, B. A., & Shaw, W. W. (1994). Trial and error: Assessing the effectiveness of riparian revegetation in Arizona. *Ecological Restoration*, *12*(2), 160–167.

Camporeale, C., & Ridolfi, L. (2006). Riparian vegetation distribution induced by river flow variability: A stochastic approach. *Water Resources Research*, *42*(10), 1–13. https://doi.org/10.1029/2006WR004933

Camporeale, C., Perucca, E., Ridolfi, L., & Gurnell, A. M. (2013). Modeling the interactions between river morphodynamics and riparian vegetation. Reviews of Geophysics, 51(3), 379-414. https://doi.org/10.1002/rog.20014

Clipperton, G. K., Koning, C. W., Locke, A. G. H., Mahoney, J. M., & Quazi, B. (2003). *Instream Flow Needs Determinations for the South Saskatchewan River Basin, Alberta, Canada*. Retrieved from http://ssrb.environment.alberta.ca/pubs/IFN_Main_Report.pdf

Cooper, D. J., Merritt, D. M., Andersen, D. C., & Chimner, R. A. (1999). Factors controlling the establishment of Fremont cottonwood seedlings on the upper Green River, USA. *River Research and Applications*, *15*(5), 419–440. https://doi.org/10.1002/(sici)1099-1646(199909/10)15:5<419::aid-rrr555>3.0.co;2-y

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P. and Van Meirvenne, M. (2013). Application of the topographic position index to heterogeneous landscapes. *Geomorphology*. 186:39-49. doi: 10.1016/j.geomorph. 2012.12.015.

Accepted, Corrected

Dettinger, M., Udall, B., & Georgakakos, A. (2015). Western water and climate change. *Ecological Applications*, *25*(8), 2069–2093. https://doi.org/10.1890/15-0938.1

Diaz-Gomez, R., Pasternack, G. B., & Guillon, H. (2025). Guiding riparian vegetation planting with machine learning. *Journal of Ecohydraulics*. https://doi.org/10.1080/24705357.2025.2481033

Díaz Gómez, R., Pasternack, G. B., Guillon, H., Byrne, C. F., Schwindt, S., Larrieu, K. G., & Solis, S. S. (2022). Mapping subaerial sand-gravel-cobble fluvial sediment facies using airborne lidar and machine learning. *Geomorphology*, *401*, 108106. https://doi.org/10.1016/j.geomorph.2021.108106

Dixon, M. D., Johnson, W. C., Scott, M. L., Bowen, D. E., & Rabbe, L. A. (2012). Dynamics of plains cottonwood (Populus deltoides) forests and historical landscape change along unchannelized segments of the Missouri River, USA. *Environmental Management*, *49*(5), 990–1008. https://doi.org/10.1007/s00267-012-9842-5

Edmaier, K., Burlando, P., & Perona, P. (2011). Mechanisms of vegetation uprooting by flow in alluvial non-cohesive sediment. *Hydrology and Earth System Sciences*, *15*(5), 1615–1627. https://doi.org/10.5194/hess-15-1615-2011

Ellison, A.M., Bank, M.S., Clinton, B.D., Colburn, E.A., Elliott, K., Ford, C.R., Foster, D.R., Kloeppel, B.D., Knoepp, J.D., Lovett, G.M. and Mohan, J. (2005). Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment*, 3(9), 479-486. https://doi.org/10.1890/1540-9295(2005)003[0479:LOFSCF]2.0.CO;2

Escobar-Arias, M. I., & Pasternack, G. B. (2011). Differences in river ecological functions due to rapid channel alteration processes in two California rivers using the functional flows model,

Accepted, Corrected

Part 2 – Model Applications. *River Research and Applications*, *27*(January), 1–122. https://doi.org/10.1002/rra

Evans, J. S., & Murphy, M. A. (2023). spatialEco. *R Package Version 2.0-1, Https://Github.Com/Jeffreyevans/SpatialEco*.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fenner, P., Brady, W. W., & Patton, D. R. (1984). Observations on seeds and seedlings of Fremont cottonwood. *Journal of Desert Plants*, *6*, 55–58. Retrieved from http://completo

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://www.jstor.org/stable/2699986

Friedman, J. M., & Auble, G. T. (1999). Mortality of riparian box elder from sediment mobilization and extended inundation. *Regulated Rivers: Research and Management*, *15*(5), 463–476. https://doi.org/10.1002/(SICI)1099-1646(199909/10)15:5<463::AID-RRR559>3.0.CO;2-Z

Friedman, J. M., Scott, M. L., & Lewis, W. M. (1995). Restoration of riparian forest using irrigation, artificial disturbance, and natural seed fall. *Environmental Management*, *19*(4), 547–557. https://doi.org/10.1007/BF02471966

Gervasi, A. A., Pasternack, G. B., East, A. E. (2021). Flooding duration and volume more important than peak discharge in explaining 18 years of gravel-cobble river change. *Earth Surface Processes and Landforms* 46 (15), 3194-3212. DOI: 10.1002/esp.5230

Gilbert, G. K. (1917). Hydraulic-mining debris in the Sierra Nevada. In: Professional Paper, vol. 105. United States Geological Survey. Washington D.C.

González, E., Martínez-Fernández, V., Shafroth, P. B., Sher, A. A., Henry, A. L., Garófano-

Accepted, corrected

Gómez, V., & Corenblit, D. (2018). Regeneration of Salicaceae riparian forests in the Northern Hemisphere: A new framework and management tool. *Journal of Environmental Management*, *218*, 374–387. https://doi.org/10.1016/j.jenvman.2018.04.069

Gregory, S. V., Meleason, M. A., & Sobota, D. J. (2003). Modeling the dynamics of wood in streams and rivers. *American Fisheries Society Symposium*, *37*, 315–335.

Gregory, S.V., Swanson, F.J., McKee, W.A., and Cummins, K.W. (1991). An Ecosystem Perspective of Riparian Zones: Focus on links between land and water. *BioScience* 41(8), 540–551. https://doi.org/10.2307/1311607

Groves, J. H., Williams, D. G., Caley, P., Norris, R. H., & Caitcheon, G. (2009). Modelling of floating seed dispersal in a fluvial environment. *River Research and Applications*, *25*, 582–592. https://doi.org/10.1002/rra

Guisan, A., Weiss, S. B., & Weiss, A. D. (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, *143*(1), 107–122. Retrieved from https://link.springer.com/article/10.1023/a:1009841519580

Haga, H., Kumagai, T., Otsuki, K., & Ogawa, S. (2002). Transport and retention of coarse woody debris in mountain streams: An in situ field experiment of log transport and a field survey of coarse woody debris distribution. *Water Resources Research*, *38*(8), 1-1-1–16. https://doi.org/10.1029/2001wr001123

Hauer, F. R., & Lorang, M. S. (2004). River regulation, decline of ecological resources, and potential for restoration in a semi-arid lands river in the western USA. *Aquatic Sciences*, *66*(4), 388–401. https://doi.org/10.1007/s00027-004-0724-7

Hijmans, R. J. (2023). terra: Spatial Data Analysis. *R Package Version 1.7-29, <https://CRAN.R-Project.Org/Package=terra>*.

Accepted, Corrected

Jajarmizadeh, M., Harun, S., & Salarpour, M. (2012). A review on theoretical consideration and types of models in hydrology. *Journal of Environmental Science and Technology* 5(5): 249-261. DOI: 10.3923/jest.2012.249.261

James, L. A. (2005). Sediment from hydraulic mining detained by Englebright and small dams in the Yuba basin. *Geomorphology*, *71*(1–2), 202–226. https://doi.org/10.1016/j.geomorph.2004.02.016

Järvelä, J. (2004). Determination of flow resistance caused by non-submerged woody vegetation. *International Journal of River Basin Management*, *2*(1), 61–70. https://doi.org/10.1080/15715124.2004.9635222

Johnson, W.C. (1994). Woodland expansions in the Platte River, Nebraska: patterns and causes. Ecological monographs, 64(1), 45-84. https://doi.org/10.2307/2937055.

Johnson, D. H. (1980) The comparison of use and availability measurements for evaluating resource preference. *Ecology* 71, 61–65. https://doi.org/10.2307/1937156

Kammel, L. E., Pasternack, G. B., Massa, D. A., & Bratovich, P. M. (2016). Near-census ecohydraulics bioverification of Oncorhynchus mykiss spawning microhabitat preferences. *Journal of Ecohydraulics*, *1*(1–2), 62–78. https://doi.org/10.1080/24705357.2016.1237264

Karrenberg, S., Edwards, P.J. and Kollman, J. (2002). The life history of Salicaceae living in the active zone of floodplains. *Freshwater Biology*, 47, 733-748. https://doi.org/10.1046/j.1365-2427.2002.00894.x

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Springer. https://doi.org/10.1007/978-1-4614-6849-3

Accepted, Corrected

Latella, M., Notti, D., Baldo, M., Giordan, D. & Camporeale, C. (2024) Short-term biogeomorphology of a gravel-bed river: Integrating remote sensing with hydraulic modelling and field analysis. Earth Surface Processes and Landforms, 49(3), 1156–1178. https://doi.org/10.1002/esp.5760

Leclerc, M., Boudreault, A., Bechara, T. A., & Corfa, G. (1995). Two-Dimensional Hydrodynamic Modeling: A Neglected Tool in the Instream Flow Incremental Methodology. *Transactions of the American Fisheries Society*, *124*(5), 645–662. https://doi.org/10.1577/1548-8659(1995)124<0645:tdhman>2.3.co;2

Lopez, F., & Garcia, M. (1998). Open-channel flow through simulated vegetation: Suspended sediment transport modeling. *Water Resources Research*, *34*(9), 2341–2352.

Luhar, M., & Nepf, H. M. (2013). From the blade scale to the reach scale: A characterization of aquatic vegetative drag. *Advances in Water Resources*, *51*, 305–316. https://doi.org/10.1016/j.advwatres.2012.02.002

Mahoney, J. M., & Rood, S. B. (1991). A device for studying the influence of declining water table on poplar growth and survival. *Tree Physiology*, *8*(3), 305–314. https://doi.org/10.1093/TREEPHYS/8.3.305

Mahoney, J. M., & Rood, S. B. (1998). Streamflow requirements for cottonwood seedling recruitment-An integrative model. *Wetlands*, *18*(4), 634–645. https://doi.org/10.1007/BF03161678

Manly, B. F. J. & Alberto, J. A. (2014) *Introduction to ecological sampling*. Boca Raton, FL: Taylor and Francis Group LLC.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote*

Accepted, Corrected

Sensing, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Merritti, D. M., & Wohl, E. E. (2016). Processes governing hydrochory along rivers: hydraulics, hydrology, and dispersal phenology. *Ecological Applications 12*(4), 1071–1087. http://www.jstor.org/stable/3061037

Moggridge, H.L. and Gurnell, A.M. (2010), Hydrological controls on the transport and deposition of plant propagules within riparian zones. *River Research and Applications*, 26: 512-527. https://doi.org/10.1002/rra.1273

Moniz, P. J., Pasternack, G. B., Massa, D. A., Stearman, L. W., & Bratovich, P. M. (2020). Do rearing salmonids predictably occupy physical microhabitat? *Journal of Ecohydraulics*, *5*(2), 132–150. https://doi.org/10.1080/24705357.2019.1696717

Nagler, P. L., Hinojosa-Huerta, O., Glenn, E. P., Garcia-Hernandez, J., Romo, R., Curtis, C., Huete, A. R., & Nelson, S. G. (2005). Regeneration of native trees in the presence of invasive salt cedar in the Colorado River Delta, Mexico. *Conservation Biology*, *19*(6), 1842–1852. https://doi.org/10.1111/j.1523-1739.2005.00234.x

Naiman, R. J., Decamps, H., & Pollock, M. (1993). The Role of Riparian Corridors in Maintaining Regional Biodiversity. *Ecological Applications*, *3*(2), 209–212. http://www.jstor.org/stable/1941822 .

Pasternack, G. B. (2023). Autumn 2017 Lower Yuba River TUFLOW HPC 2D Model Description, Validation, and Exploratory Simulations. Prepared for Yuba County Water Agency. *University of California, Davis, CA*.

Pasternack, G. B., Tu, D., & Wyrick, J. R. (2014). Chinook Adult Salmon Spawning Physical Habitat of the Lower Yuba River. Prepared for the Yuba Accord River Management Team. *University of California, Davis, CA*.

Accepted, Corrected

Patten, D. T. (1998). Riparian ecosystems of semi-arid North America: Diversity and human impacts. *Wetlands*, *18*(4), 498–512. https://doi.org/10.1007/BF03161668

Phillips, S. J., & Pasternack, G. B. (2022). Development And Application Of Cottonwood Seedling Recruitment Potential Model At The Upper Long Bar Site On The Lower Yuba River. Prepared for Yuba Water Agency. *University of California, Davis, CA*.

Phillips, S. J., Pasternack, G. B., and Larrieu, K. 2025. Development and testing of a mechanistic potential niche model of riparian tree seedling recruitment. *Ecological Modeling*, 501 (2025), 110986. https://doi.org/10.1016/j.ecolmodel.2024.110986.

Phipps, R. L. (1979). Simulation of wetlands forest vegetation dynamics. *Ecological Modelling*, *7*(4), 257–288. https://doi.org/10.1016/0304-3800(79)90038-3

Pike, R. J., & Wilson, S. E. (1971). Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Geological Society of America Bulletin*, *82*(4), 1079-1084. https://doi.org/10.1130/0016-7606(1971)82[1079:ERHIAG]2.0.CO;2

Poff, N. L., Allan, J. D., Palmer, M. A., Hart, D. D., & Richter, B. D. (2003). River Flows and Water Wars: Emerging Science for Environmental Decision Making. *Biological Sciences Faculty Publications*.

Politti, E., Bertoldi, W., Gurnell, A., & Henshaw, A. (2018). Feedbacks between the riparian Salicaceae and hydrogeomorphic processes: A quantitative review. *Earth-Science Reviews*, *176*(August 2017), 147–165. https://doi.org/10.1016/j.earscirev.2017.07.018

Probst, P., Wright, M.N. and Boulesteix, A.L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: data mining and knowledge discovery, 9(3), p.e1301. https://doi.org/10.1002/widm.1301

Renard, K. G., Foster, G. R., Weesies, G. A., Mccool, D. K., & Yoder, D. C. (1997). *Predicting*

Accepted, Corrected

Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE) (703rd ed.). United States Department of Agriculture, Washington, DC.

Rew, L. J., Maxwell, B. D., & Aspinall, R. (2005). Predicting the occurrence of nonindigenous species using environmental and remotely sensed data. *Weed Science*, *53*(2), 236–241. https://doi.org/10.1614/ws-04-097r

Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). Index that quantifies topographic heterogeneity. intermountain Journal of sciences, 5(1-4), 23-27.

Schwindt, S., Larrieu, K., Pasternack, G. B., & Rabone, G. (2020). River Architect. *SoftwareX*, *11*. https://doi.org/10.1016/j.softx.2020.100438

Scott, M. L., Auble, G. T., & Friedman, J. M. (1997). Flood dependency of cottonwood establishment along the Missouri River, Montana, USA. *Ecological Applications*, *7*(2), 677–690. https://doi.org/10.1890/1051-0761(1997)007[0677:FDOCEA]2.0.CO;2

Scott, M. L., Shafroth, P. B., & Auble, G. T. (1999). Responses of Riparian Cottonwoods to Alluvial Water Table Declines. *Environmental Management*, *23*(3), 347–358. https://doi.org/10.1007/s002679900191

Shoutis, L., Patten, D. T., & McGlynn, B. (2010). Terrain-based predictive modeling of riparian vegetation in a northern rocky mountain watershed. *Wetlands*, *30*(3), 621–633. https://doi.org/10.1007/s13157-010-0047-5

Silva, P. V. and Pasternack, G. B. (2018). 2017 Lower Yuba River Topographic Mapping Report. Prepared for Yuba Water Agency. *University of California, Davis, CA*.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437.

Accepted, Corrected

https://doi.org/10.1016/j.ipm.2009.03.002

Solari, L., Van Oorschot, M. V., Belletti, B., Hendriks, D., Rinaldi, M., & Vargas-Luna, A. (2016). Advances on modelling riparian vegetation - hydromorphology interactions. *River Research and Applications*, *32*, 164–178. https://doi.org/10.1002/rra

Stella, J. C. (2005). A field-calibrated model of pioneer riparian tree recruitment for the San Joaquin Basin, CA. *Environmental Science, Policy and Management*, *Ph.D.*(January 2005), 215.

Stella, J. C., Battles, J. J., McBride, J. R., & Orr, B. K. (2010). Riparian Seedling Mortality from Simulated Water Table Recession, and the Design of Sustainable Flow Regimes on Regulated Rivers. *Restoration Ecology*, *18*(SUPPL. 2), 284–294. https://doi.org/10.1111/j.1526-100X.2010.00651.x

Stella, J.C., Battles, J.J., Orr, B.K. and McBride, J.R., 2006. Synchrony of seed dispersal, hydrology and local climate in a semi-arid river reach in California. Ecosystems, 9, pp.1200-1214. https://doi.org/10.1007/s10021-005-0138-y

Stella, J. C., Hayden, M. K., Battles, J. J., Piégay, H., Dufour, S., & Fremier, A. K. (2011). The Role of Abandoned Channels as Refugia for Sustaining Pioneer Riparian Forest Ecosystems. *Ecosystems*, *14*(5), 776–790. https://doi.org/10.1007/s10021-011-9446-6

Stromberg, J. (1998). Dynamics of Fremont cottonwood (Populus fremontii) and saltcedar (Tamarix chinensis) populations along the San Pedro River, Arizona. *Journal of Arid Environments*, *40*(2), 133–155. https://doi.org/10.1006/jare.1998.0438

Stromberg, J. C. (1993). Arizona-Nevada Academy of Science Riparian Mesquite Forests: A Review of Their Ecology, Threats, and Recovery Potential. *Source: Journal of the Arizona-Nevada Academy of Science*, *27*(1), 111–124.

Accepted, Corrected

Stromberg, J. C. (2001). Restoration of riparian vegetation in the south-western United States: Importance of flow regimes and fluvial dynamism. *Journal of Arid Environments*, *49*(1), 17–34. https://doi.org/10.1006/jare.2001.0833

Tranmer, A.W., Benjankar, R., Vidergar, D. and Tonina, D. (2023). Identifying failure mechanisms of native riparian forest regeneration in a variable-width floodplain using a spatially-distributed riparian forest recruitment model. *Ecological Engineering*, 187, 106865. https://doi.org/10.1016/j.ecoleng.2022.106865

Villanueva, V. R., Castellet, E. B., Díez-Herrero, A., Bodoque, J. M., & Sánchez-Juny, M. (2014). Two-dimensional modelling of large wood transport during flash floods. *Earth Surface Processes and Landforms*, *39*(4), 438–449. https://doi.org/10.1002/esp.3456

Vogiatzakis, I. N., & Griffiths, G. H. (2006). A GIS-based empirical model for vegetation prediction in Lefka Ori, Crete. *Plant Ecology*, *184*(2), 311–323. https://doi.org/10.1007/s11258-005-9075-2

Weber, M. D., & Pasternack, G. B. (2017). Valley-scale morphology drives differences in fluvial sediment budgets and incision rates during contrasting flow regimes. *Geomorphology*, *288*, 39–51. https://doi.org/10.1016/j.geomorph.2017.03.018

Whitham, T.G., Bailey, J.K., Schweitzer, J.A., Shuster, S.M., Bangert, R.K., LeRoy, C.J., Lonsdorf, E.V., Allan, G.J., DiFazio, S.P., Potts, B.M. and Fischer, D.G. (2006). A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics*, 7(7), 510-523. https://doi.org/10.1038/nrg1877

(YCWA), Y. C. W. A. (2013). Technical Memorandum 2-1 Hydrologic Alteration Yuba River Development Project, (2246).

You, X., Liu, J., & Zhang, L. (2015). Ecological modeling of riparian vegetation under disturbances: A review. *Ecological Modelling*, *318*, 293–300. https://doi.org/10.1016/j.ecolmodel.2015.07.002

Zamora-Arroyo, F., Nagler, P. L., Briggs, M., Radtke, D., Rodriquez, H., Garcia, J., et al. (2001). Regeneration of native trees in response to flood releases from the United States into the delta of the Colorado River, Mexico. *Journal of Arid Environments*, *49*(1), 49–64. https://doi.org/10.1006/jare.2001.0835

Zhang, M., Shi, W., Xu, Z. (2020). Systematic comparison of five machine-learning models in classification and interpolation of soil particle size fractions using different transformed data. *Hydrology and Earth System Sciences*. 24, 2505–2526. https://doi.org/10.5194/hess-24-2505-2020

Zong, L., & Nepf, H. (2011). Spatial distribution of deposition within a patch of vegetation. *Water Resources Research*, *47*(3), 1–12. https://doi.org/10.1029/2010WR009516

Accepted, Corrected

Title: Predictability of cottonwood recruitment along a dynamic, regulated river

Authors: Rachel E. Wright[1], Sierra J. Phillips[1], Romina Diaz-Gomez[1], Yufang Jin[1], Mary L. Cadenasso[2], Gregory B. Pasternack[1*]

[1]Department of Land, Air, and Water Resources, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

[2]Department of Plant Sciences, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

Supplementary Materials

Content in this file is organized with section headings that match those in the main manuscript.

1 INTRODUCTION

Identifying ideal environmental conditions from a physical process viewpoint is complex, as the characteristics of a river system are a product of dynamic hydrogeomorphic processes. River bank and floodplain heterogeneity provide the physical template for the spatial pattern and development of varying riparian vegetation communities (Gregory et al., 1991). Topographic gradients impact the amount of energy and force of river flows, resulting in areas of erosion or deposition (Swanson et al., 1982). These sediment transport processes may impact vegetation through uprooting, burial, and erosion (Politti et al., 2018), and affect the substrate's ability to retain the moisture needed for root growth by influencing sediment composition and grain size (Camporeale et al., 2013). Microtopography, or the localized topographic variability in soil surface elevation and roughness, also impacts the immediate hydrologic conditions experienced by a seed or plant (Moser et al., 2007; Pollock et al., 1998). Environmental characteristics resulting from varying microtopographic patterns may influence the distribution of plants through the creation of differing habitats (Titus, 2016). These variations in soil conditions and topography result in a high diversity of riparian plant species that can coexist (Naiman et al., 1993).

Successful natural recruitment of cottonwoods and other pioneer species is also crucial for the continuation of riparian forests. The ecophysiological requirements for cottonwood recruitment and survival are intricately linked to fluvial hydrologic and geomorphic processes. Recruitment of these pioneer species is largely dependent on large, infrequent flows, which provide the necessary physical disturbance to create open space for colonization, dispersal of seeds, and substrate moisture for early root growth and consequent seedling recruitment (Benjankar et al., 2020; Rood et al., 2003; Stella et al., 2010). After recruitment, seedling survival is then reliant on environmental factors such as access to sunlight and a root growth rate comparable to recession rate declines of the water table in order to maintain access to moisture (Amlin & Rood, 2002; Benjankar et al., 2020; Stella et al., 2010).

1.1 Cottonwood Seedling Recruitment

In non-cohesive soils, uprooting vegetation may occur through Type I or Type II uprooting mechanisms, which respectively reference early germinated or mature vegetation (Edmaier et al., 2011). Type I uprooting occurs when the drag force exceeds the root resistance of the plant, while Type II uprooting is when the scouring near the base of the plant exposes the root system and decreases the root anchoring resistance until turning into Type I (Edmaier et al., 2011).

1.2 Riparian Vegetation Modeling

Like statistical-empirical models in general, AI/ML models are most useful and can be highly accurate for professional practice when tuned on and applied to a local setting, staying within the range of conditions for which tuning was done. Relationships between species and the environment are often complex and nonlinear, allowing classification ML procedures to provide more meaningful analysis of ecological data then traditional statistical methods may be able to (Cutler et al., 2007; De'Ath & Fabricius, 2000).

3 METHODOLOGY

3.1 Experimental Design

Most spatial analyses were performed in ArcGIS Pro (ESRI, Redlands, CA), while machine learning modeling was performed in R. Most of the data in this study was collected in American customary units and then converted to SI units, leading to non-integer values being reported in some instances where one might expect the selection of integer values, such as raster cell size.

3.2.1 Hydrophysical variables

The preparation of the bed before seed dispersal and the scouring flows after dispersal are both analyzed through the dimensionless bed shear stress ($\tau *$) with the equation from Schwindt et al., (2019):

$$\tau *= \frac{1}{D_{84} g(s-1)} \left[\frac{u}{5.75 log_{10}(12.2h/2D_{84})}\right]^2 \tag{1}$$

where D_{84} is the grain diameter approximated as D_{84}=2.2D_{50} (Rickenmann & Recking, 2011), g is the gravitational acceleration (9.81 m/s^2), s is sediment grain and water density ratio (2.68 g/cm^3), u is the water velocity (m/s), and h is the water depth (m). The dimensionless bed shear stress is calculated for each provided discharge and compared against the thresholds for partially and fully mobilized sediment transport in every cell. The flows during the 2 years prior to seed dispersal are analyzed to determine the bed preparation, while the flows after germination during the seed dispersal period through the following May are used for the scouring survival analysis.

Inverse distance weighted (IDW) interpolation is used in River Architect to spatially extrapolate the WSE raster for the river's wetted area at a given discharge to estimate the water level elevation (WLE) beyond the wetted area (Larrieu et al., 2021). When WLE < DEM, then WLE is the groundwater level. When WLE > DEM, then it is the WSE of disconnected ponds, swales, and floodplain channels. Inundation duration is tracked for every cell when the WLE has a greater value than the DEM. Consecutive days of inundation are counted throughout the inundation survival period during the seed dispersal period after germination through the following May.

Desiccation stress may occur if seedling roots cannot maintain contact with the soil moisture as WLE recedes. A recession rate of 1 cm/day was considered stressful and 2.5 cm/day was considered lethal (Amlin & Rood, 2002; Mahoney & Rood, 1998; Phillips & Pasternack, 2022; Stella et al., 2010). The mortality coefficient is used to quantify the recession rate and is calculated using a 3-day moving average for each cell, as this accounts for a time lag associated with the capillary fringe and the rate at which seedlings can grow roots (Braatne et al., 2007; Burke et al., 2009). The desiccation survival period begins during the seed dispersal period when germination begins and ends when baseflow starts.

3.2.2 Model inputs

Silva and Pasternack, (2018) detailed the 2017 meter-resolution topo-bathymetric surveys and DEM production. Details about spatial coverage, resolution, and accuracy for the digital elevation model (DEM) used in this study are in **Table S1.**

Table S1. 2017 LYR topographic data summary from Pasternack, (2023).

Attribute	Description
Aerial extent	Entire river from Englebright Dam to the confluence with the Feather River including the Narrows Reach.
Years of data collection	Airborne LiDAR surveys were done on September 18 and 19, 2017. Multibeam echosounder surveys were done on August 16-18 and August 21, 2017.
Bathymetric resolution	10.28 points per m^2.
Topographic resolution	Ground and bathymetric bottom classified density of LiDAR data was 8.45 points/m^2.
Bathymetric uncertainty	Comparison of 76 bathymetric (submerged or along the water's edge) check points yielded a root mean square difference of 9.7 cm.
Topographic uncertainty	Comparison of 21 ground check points yielded a non-vegetated vertical accuracy of 4.2 cm.

The 2017 LYR 2D hydrodynamic models were made using TUFLOW GPU (Huxley & Syme, 2016; WBM Pty Ltd, 2016). **Table S2** provides a summary of model parameters and **Table S3** includes performance metrics for 2D model validation.

Table S2. TUFLOW GPU model parameters specifications from Pasternack, (2023).

Model Parameter	Specification
Mesh Resolution	0.9144 m (3 feet) for flows < 849.5 m^3/s 3.04 m (10 feet) for flows > 849.5m^3/s
Cell Wet/Dry Depth*	0.09144 m
Timestep	Adaptive timestepping
Discharge range of model	50 discharges spanning 300 to 198,885 cfs (unable to model 155,977 and 198,885 for DPDMRY and MRYFR due to flooding into the Goldfields).
Downstream WSE data/model source	Stage-discharge rating curves (DPD with a weir equation, MRY with USGS rating table, FR with stacked polynomial approach).
Eddy Viscosity Formulation	Smagorinsky
Eddy Viscosity Coefficient	0.5
Eddy Viscosity Constant	0 m^2/s for validation and 0.03716 m^2/s for all other runs
Manning's n < 28.32 m^3/s	One set with a global value followed by one set with a spatially distributed topographic surface roughness based on lidar data per method of Abu-Aly et al., (2014).

Manning's n ≥ 28.32 m^3/s	One set with a global value, one set with a spatially distributed topographic surface roughness, and one set with a combination of spatially distributed topographic surface roughness and spatially distributed vegetation roughness per method of Abu-Aly et al., (2014).
LiDAR test of WSE prediction accuracy	Below DPD, the root mean square difference in observed versus predicted WSE was 0.11-0.16 ft with approximately 90-99% of points were within ± 0.25 ft. Above DPD, the root mean square difference in observed versus predicted WSE was 0.22-.23 ft with approximately 72-79% of points were within ± 0.25 ft.
Wading depth prediction accuracy	A total of 86 field observations of depth were made on September 26, 2017 in the Timbuctoo Bend Reach when discharge was 1,021 cfs, which yielded an absolute median error of 4% with a coefficient of 0.90 and a regression slope of 0.99. A total of 126 field observations of depth were made on October 10, 2017 in the Parks Bar Reach when discharge was 1,082 cfs, which yielded an absolute median error of 5% with a coefficient of 0.87 and a regression slope 0.89. A total of 54 field observations of depth were made on October 31, 2017 below Daguerre Point Dam when discharge was 548 cfs, which yielded an absolute median error of 3% with a coefficient of 0.99 and a regression slope of 1.00.
Wading velocity magnitude prediction	A total of 86 field observations of velocity were made on September 26, 2017 in the Timbuctoo Bend Reach when discharge was 1,021 cfs yielded an absolute median error of 11% with a coefficient of 0.92 and a regression slope of 0.92. A total of 126 field observations of velocity were made on October 10, 2017 in the Parks Bar Reach when discharge was 1,082 cfs, which yielded an absolute median error of 14% with a coefficient of 0.81 and a regression slope of 1.14. A total of 54 field observations of velocity were made on October 31, 2017 below Daguerre Point Dam when discharge was 548 cfs, which yielded an absolute median error of 10% with a coefficient of 0.90 and a regression slope of 0.66.
Kayak-based Lagrangian velocity magnitude prediction accuracy	A total of 2,686 observation points of velocity were made below Highway 20 on December 5, 2017 yielding an absolute median error of 11% with a coefficient of determination of 0.85 and a regression slope of 0.86. A total of 3,702 observation points of velocity were made below Daguerre Point Dam on December 21, 2017 yielding an absolute median error of 19% with a coefficient of determination of 0.82 and a regression slope of 0.89. A total of 2,899 observation points of velocity were made below the Marysville gaging station on December 21, 2017 yielding an absolute median error of 19% with a coefficient of determination of 0.72 and a regression slope of 0.88.

Kayak-based Lagrangian velocity direction prediction accuracy	

*This does not dry out cells if they have depths < 0.3 ft; it is just an internal parameter.

Table S3. Summary of key metrics for all performance indicator variables.

			Wading**		Kayak**		
Domain	Mann-ings n	WSE*	V	D	V	Further tweaks possible?	Performance
EDH20	0.040	-0.05	6%	-1%	N/A	Could increase Manning's n slightly	Excellent
H20DPD	0.035	0.12	12%	-1.2%	4.3%	Cannot be improved	Excellent
DPDMRY	0.034	0.07	.8%	3%	12.5	Could decrease Manning's n slightly	Excellent
MRYFR	0.030	-0.04	N/A	N/A	6.6	Could increase Manning's n slightly	Excellent

*signed median deviation

**signed median error %

Table S4. Hydrologic data sources for each modeling domain.

Domain	Domain Description	Flow Data
1) MRYFR	Reach extends from just upstream of the confluence with the Feather River to 3.4-km downstream of the Marysville gaging station.	USGS Marysville gage (11421000)
2) DPDMRY	Reach extends from 3.4-km below the Marysville gaging station to just downstream of Daguerre Point Dam.	USGS Marysville gage (11421000)
3A) EDDPD (downstream)	Reach extends from to just upstream of Daguerre Point Dam to confluence with Dry Creek	The Yuba Water Agency has projected flows through 2017 for both above and below the Dry Creek tributary. A linear regression comparing the flows above and below Dry Creek was used

		to find the slope for flows <1,000 cfs, 1,000-10,000 cfs, and > 10,000 cfs. USGS Deer Creek gage (11418500) and Englebright near Smartsville gage (11418000) were added together. Flows were respectively multiplied by the slope for each category above.
3B) EDDPD (upstream)	Extends from just upstream of Dry Creek to Englebright Dam	USGS Deer Creek gage (11418500) and Englebright near Smartsville gage (11418000) were added together

3.2.5 LYR Field Data Collection

A hand-held Trimble GeoXH mapping-grade GPS was used to navigate to and mark the boundaries of each site to be surveyed. If a slope, boundary, or other obstacle prevented close contact to the base of a tree, the GPS point was collected at the closest possible location along with the distance and compass direction to the tree; coordinates were adjusted later in ArcGIS Pro.

Observation methods differed by plant height class. For cottonwoods < 2-m tall, a tape measure and caliper were used to measure height and stem diameters, respectively (Error! Reference source not found.). Diameters were measured above the root collar and at 50% of the height. If the tree was > 2-m tall, diameter at breast height (DBH) was recorded using a diameter tape, while height was measured using Pythagorean relationships between a set distance to the base of the tree and the angle to both the top of the canopy and the base of the tree. The angle was collected using a clinometer, while a measured distance from the base of the tree was collected using a long tape measure.

3.2.6 RSRM Bioverification

To compute FR values, the percent utilization and percent availability were needed. The percent utilization for each class was determined by dividing the total number of juvenile cottonwoods across the sampled sites in each recruitment potential class by the total number of juvenile cottonwoods found in the whole dataset. The total area for each riparian recruitment potential class was calculated by summing the area of all sampled sites for a given class, using the 2017-2021 mode recruitment potential class values from field site selection. Total class areas were then summed to compute total model-prediction area. The percent area for each riparian recruitment potential class was then calculated by dividing each individual class area by the overall total area of predicted classes.

Statistical bootstrapping is a method used for determining a measure of accuracy for sample estimates, with random sets of the same sample size created and used with the test metrics to quantify the statistical confidence limits and evaluate whether the observations behave like a random variable or not. This study did not require the additional steps with statistical bootstrapping, because the results were so extreme that they could not be random, given the sample sizes.

3.3 Random Forest Algorithm

RF's use an ensemble of decision trees that are repeatedly aggregated using different combinations of explanatory variables to make a more accurate classification decision (Breiman, 2001), and have been observed to have a high classification accuracy when compared to other classification methods (Cutler et al., 2007). RF's allow for the exploration of prediction patterns and processes through the use of both discrete and continuous explanatory variables, variable importance measures, and graphical representations.

3.3.2 Predictors

A correlation matrix was generated for the 16 topographic variables to identify any similar variables that provide the same information **Figure S1**. The Pearson correlation method was used, which produces values ranging from -1 to 1 and can only be used with continuous variables. A value of -1 indicates a total negative correlation, a value of 0 is no correlation, and 1 is a total positive correlation. There was a negative correlation between curvature, profile, and TPI. There was also a negative correlation between roughness, pixel-scale topographic variability, slope, and TRI.

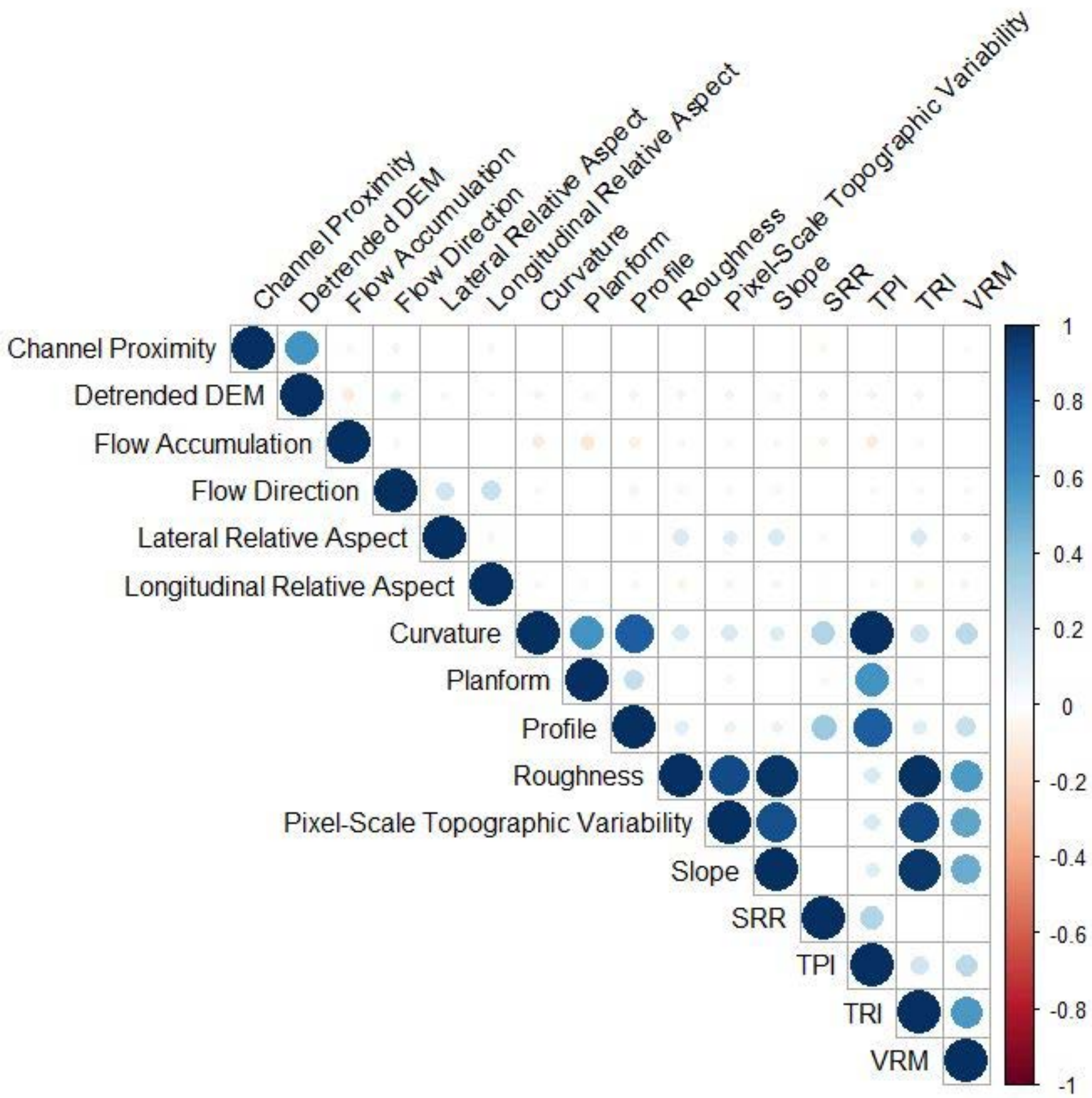

Figure S1. Correlogram using the Pearson coefficient for the 16 topographic predictor variables.

3.3.3 Implementation

The repeated k-fold cross validation effectively captures the generalization performance of the RF model, by ensuring that the predictive model's skill report does not depend on the way that training and testing data are chosen, which in this case is the difference between the model estimated and true values (Kuhn & Johnson, 2013).

3.3.1 RF model performance

A confusion matrix was used to visualize the accuracy and relative error among presence and absence classes using the hold out testing data, for each of the 10-folds from the 10 repetitions (**Table 5**). Overall accuracy was calculated as the number of correct predictions to the overall number of predictions, relaying the effectiveness of the model. The producer's accuracy portrays the sensitivity, the ratio of correctly classified presence points, and the specificity, the ratio of correctly classified absence points. The best sensitivity or specificity is 1.0, meaning all the predictions were correct, while 0.0 would be the worst. Omission and commission errors respectively represent the reference or classified cells omitted from the correct class, with a balance between these errors as the ideal. These metrics provide a deeper understanding of the performance of the model beyond the accuracy, portraying how well it classified both presence and absence, which may not result in the same ratio value.

The ROC plots the proportion of true positives (i.e. proportion of presence points correctly identified as presence) on the y-axis against the proportion of false positives on the x-axis (i.e. proportion of absence points classified as presence) (Fawcett, 2006). The ROC space is conceptually simple, with the point (0, 0) representing no positive classifications, the point (0, 1) representing unconditional positive classifications, and the point (1, 0) representing a perfect classification. The AUC indicates the area between the ROC curve and the diagonal from (0, 0) to (1, 1) and represents the probability that a randomly chosen positive instance will rank higher than a randomly chosen negative instance (Fawcett, 2006). It ranges from 0 to 1, with 0.5 indicating random predictions, equaling the diagonal line the area is calculated between, and 1.0 indicating a perfect classification, with no realistic classification model having an AUC<0.5 (Fawcett, 2006).

A strength of the RF algorithm is the ability to generate a ranking of variable importance. Permutation-based feature importance was used, which is measured by the decrease in the model's prediction accuracy when a variable is permuted (Breiman, 2001). Partial dependence plots were also used to further examine the marginal effect a variable has on the predicted outcome of the model when all other explanatory variables are held constant at their mean values. In this study, it would be the effect a predictor has on the probability of the RF model predicting cottonwood presence, allowing visual examination of how the probability of predicting presence increases or decreases as the variable value changes.

To be considered an accurate and useful model, important variables were not only considered statistically important for prediction but also biophysically realistic and meaningful to understand the natural phenomenon of cottonwood recruitment. In evaluating the directionality of a predictor variable given its median and quartile values, if the model were to predict that presence points were at a detrended elevation corresponding to the bottom of pools in the river, then there might be high statistical predictability in the model but it is biophysically wrong, as cottonwoods cannot recruit at that type of permanently inundated location.

4 RESULTS

4.2 Question 1: Does the RSRM accurately predict cottonwood seedling presence?

A summary of the calculation steps used to determine the FR is provided in **Table S5** and **Table S6**, using the mode and maximum recruitment potential class values, respectively. **Figure S2** portrays a similar result to **Figure 9**, where bioverification criteria one is met but fails criteria two.

Table S5. FR results for bioverification using the 2017-2021 mode recruitment potential class values.

Recruitment Class		% Predicted Area	# Cottonwoods	% Cottonwoods	FR
Lethal	0	88.3	1467	49.6	0.6
Tolerable	0.125	0.8	63	2.1	2.7
Stressful	0.25	6.1	5	0.2	0.0
Favorable	0.5	4.0	14	0.5	0.1
Optimal	1	0.9	1	0.0	0.0
Outside Modeled Cells			1408	47.6	

Table S6. FR results for bioverification using the 2017-2021 maximum recruitment potential class values.

Recruitment Class		% Predicted Area	# Cottonwoods	% Cottonwoods	FR
Lethal	0	87.3	1485	48.9	0.6
Likely Lethal	0.0625	0.0	0	0.0	0.0
Tolerable	0.125	0.6	63	2.1	3.2
Stressful	0.25	4.9	9	0.3	0.1
Favorable	0.5	5.6	25	0.8	0.1
Optimal	1	1.6	1	0.0	0.0

Outside Modeled Cells	1451	47.8

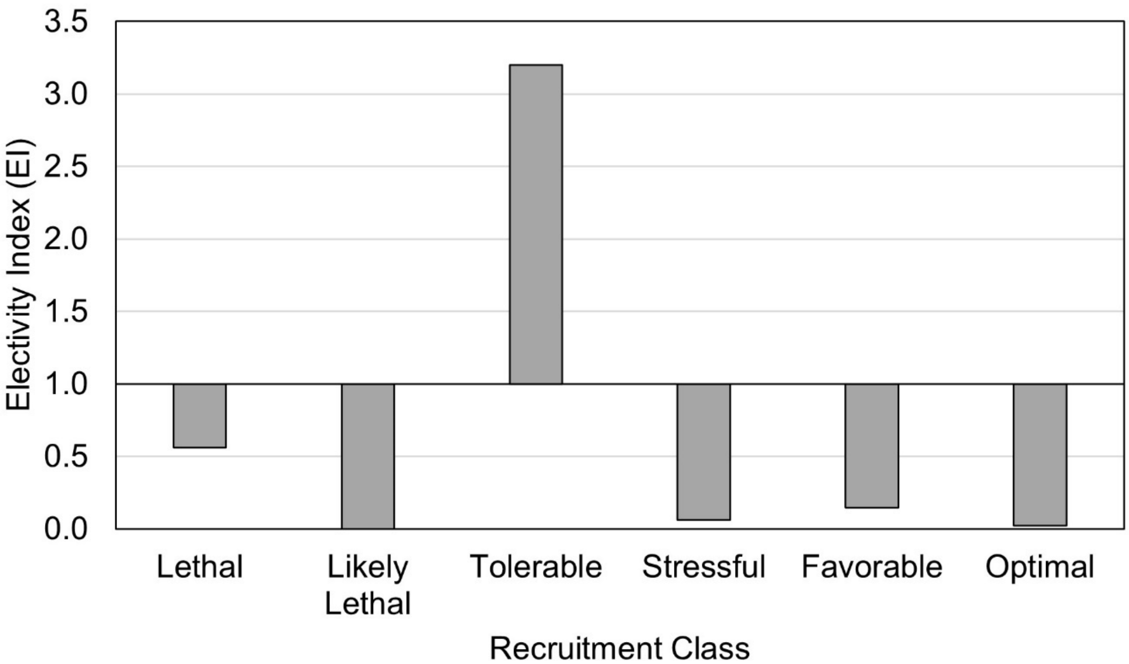

Figure S2. Forage ratio test results using the 2017-2021 maximum recruitment class values for the expert-parameterized implementation of the RSRM.

4.3 Question 2: Does the RF Model accurately predict presence and absence?

For cottonwood presence the producer's accuracy (sensitivity) was 89% (omission error of 11%) and the user's accuracy was 86% (commission error of 14%). For cottonwood absence the producer's accuracy (specificity) was 85% (omission error of 15%) and the user's accuracy was 88% (commission error of 12%). In other words, 89% of cottonwood presence cells were predicted to be presence cells, while 14% of absence cells were predicted to be presence cells.

On the other side, 85% of cottonwood absence cells were predicted as absence, while 12% of presence cells were predicted to be absence. An AUC-ROC value of 94% was reached (**Figure S3**).

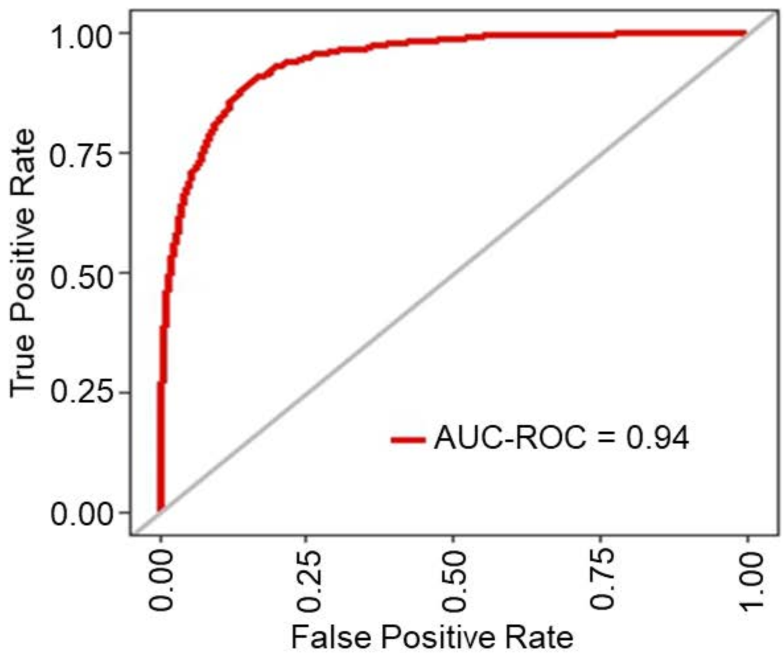

Figure S3. AUC-ROC value by number of randomly selected predictors.

The RF was run multiple times with varying categories of the predictor variables to compare performance metrics (**Table S7**). Run 1 included only the hydrophysical variables from the RSRM and run 2 included only the topographic variables. Run 4 includes an additional 10 topographic and hydraulic predictor variables. A topographic change raster from 2014 to 2017 was used to identify areas of scour or deposition. An incremental wetted area raster simulated the wetted area of the river at different discharges to explore where juvenile cottonwoods were occurring relative to hydrologic regime and channel dimensions. Eight flow convergence routing landform (FCRL) rasters at varying discharges were also used. Flow convergence routing is a morphodynamic mechanism and classifies locations in the river corridor based on hydraulics and sediment dynamics (MacWilliams et al., 2006).

Table S7. Comparison of model performance metrics between the RF runs.

Run	Predictors Used	# Predictors	Accuracy	Sensitivity	Specificity	p-value	AUC-ROC
1	Hydrophysical	4	76%	71%	81%	< 2e-16	78%
2	Topographic	16	85%	87%	82%	< 2e-16	93%
3	Hydrophysical & Topographic	20	87%	89%	85%	< 2e-16	94%
4	Run 3 + Additional Rasters	30	91%	93%	89%	< 2.2e-16	97%

4.4 Question 3: Drivers for cottonwood presence and absence?

The detrended elevation, which was found to be the most important predictor variable, was a representation of the land surface when the down valley slope is removed while still preserving local topographic variations. It was assigned a relative importance of 100%, with all the following variables' relative importance compared against it. The probability of cottonwood presence increased from detrended elevations of 0 to 2 m before a sharp peak, then decreased as the elevation increased further (**Figure 12**), which is consistent with the logic of cottonwood recruitment. Channel proximity was the 2nd ranked variable with relative importance of 71% (**Figure 11**). The predicted outcome of cottonwood presence decreased as the distance to the channel increased (**Figure 12**), which is also realistic. Inundation survival was 3rd at 43% in terms of predictive power, however, PDP showed a negative relationship between presence probability and more favorable inundation metrics (i.e., the probability of presence predictions decreased as inundation become more favorable), which is not initially biophysically sensible (**Figure 12**). Lateral relative aspect was 4th at 21% (**Figure 11**). The presence probability initially decreased as lateral relative aspect increased from -1.0 (facing towards the river) to -0.75, before stabilizing until the probability decreased again at around 0.75 (facing away from the river) (**Figure 12**), which is realistic. Vector ruggedness measure (VRM) was close behind the lateral relative aspect, with a relative importance of 20% (**Figure 11**). The probability of cottonwood presence being predicted increased rapidly following a VRM of 0.0, before decreasing after a sharp peak and then stabilizing after a value of 0.02 (**Figure 12**).

The median, lower quartile, and upper quantile values for the top five explanatory variables was further examined to interpret biophysical realism. For detrended elevation, presence cells were found at a lower elevation than the absence points. The upper quantile value for presence was equivalent to the lower quantile value for the absence cells. Presence cells also occurred closer to the wetted baseflow channel than the absence points. For the inundation survival, presence cells occurred at a lethal inundation and the absence cells at a favorable inundation. The median lateral relative aspect for presence was negative, indicating pixels facing towards the river while the absence points faced away. The VRM was similarly very small for both presence and absence, with an average presence VRM of 0.0018 and absence VRM of 0.0016.

Figure S4 compares the ranking of predictor importance for each RF run. The detrended DEM and channel proximity were the most important predictors for all of the RF runs except run 1, which only had the hydrophysical predictors. For run 4, the 84,400 cfs FCRL and the incremental wetted area were the 3rd and 4th most important, respectively. For run 3, the third and fourth most important predictors were inundation survival and lateral relative aspect, while for run 2 they were lateral relative aspect and VRM. Inundation survival was the most important in run 1, with bed preparation the second most important at 3% and scour survival and recession rate close to 0%. When comparing runs 2 and 3, the general ranking of the topographic variables stayed relatively the same, with only a few small variations in the order in the less important variables. The relative importance shifted only a little when the hydrophysical variables were included.

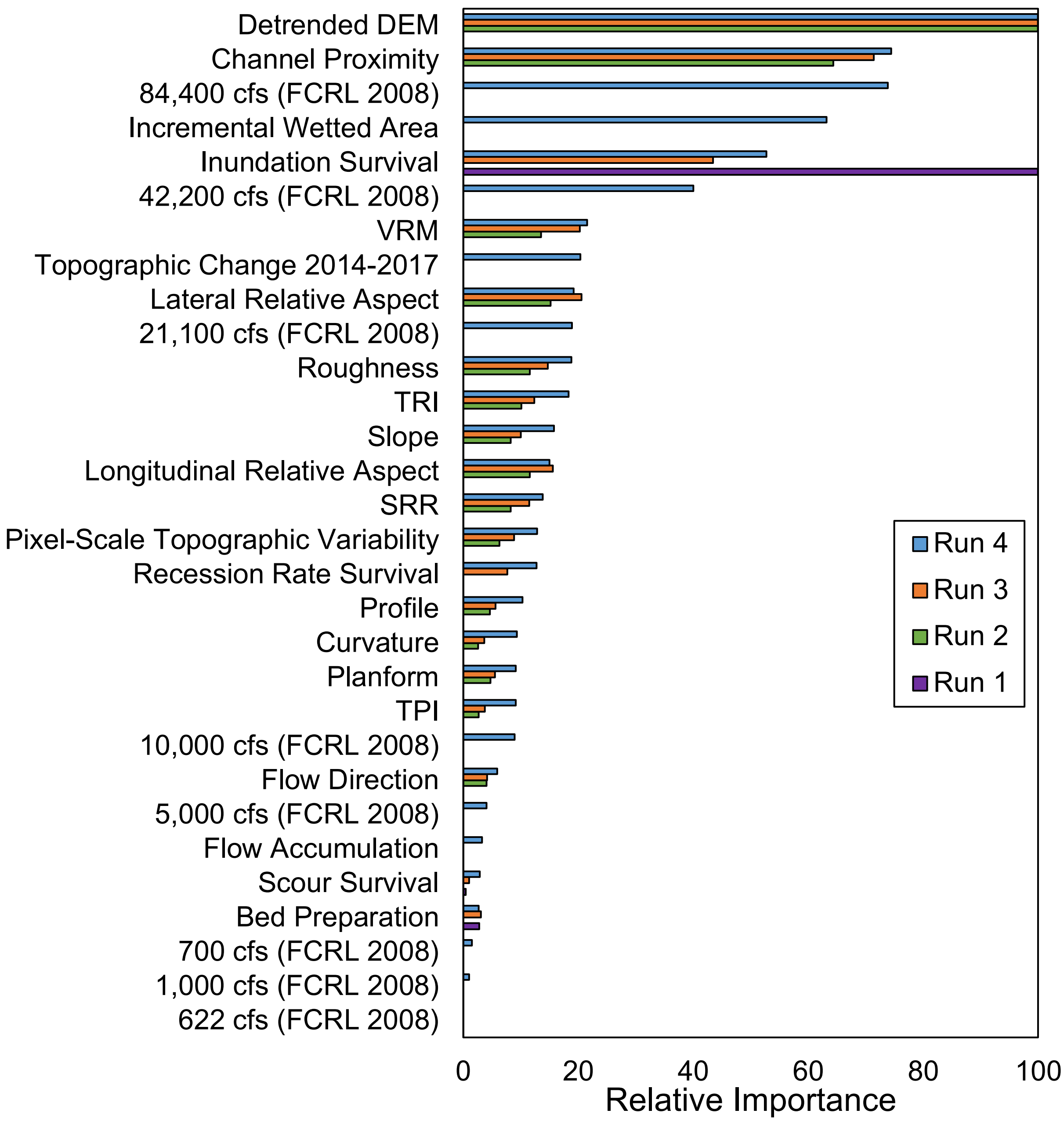

Figure S4. The relative importance of the predictor variables for the RF runs 1-4.

When the inundation survival was removed from the RF model, detrended elevation, channel proximity, vector ruggedness measure (VRM), and lateral relative aspect remained in the top four variables **(Figure S5**). The lateral relative aspect and VRM actually had a reduced importance from the original RF model (**Figure 11**), going from about 20% to 14% relative importance. The RF model without inundation survival had the same overall accuracy of 87%. This highlights the overall high importance detrended elvevation and channel proximity have for predicting cottonwood presence and absence.

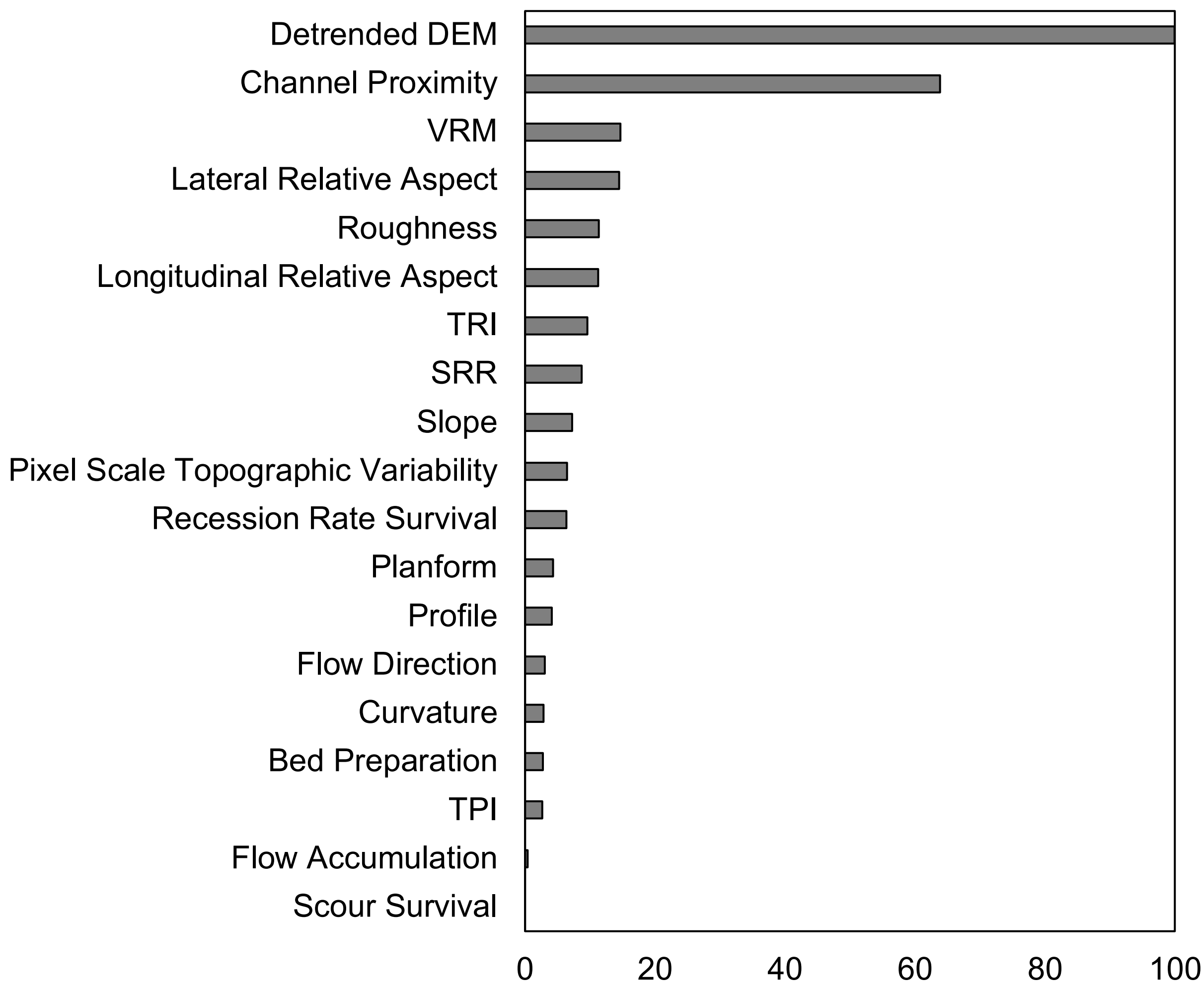

Figure S5. Ranked variables of importance when inundation survival was removed.

9 REFERENCES

Abu-Aly, T. R., Pasternack, G. B., Wyrick, J. R., Barker, R., Massa, D., & Johnson, T. (2014).

Effects of LiDAR-derived, spatially distributed vegetation roughness on two-dimensional hydraulics in a gravel-cobble river at flows of 0.2 to 20 times bankfull. *Geomorphology*, *206*, 468–482. https://doi.org/10.1016/j.geomorph.2013.10.017

Amlin, N. M., & Rood, S. B. (2002). Comparative tolerances of riparian willows and cottonwoods to water-table decline. *Wetlands*, *22*(2), 338–346. https://doi.org/10.1672/0277-5212(2002)022[0338:CTORWA]2.0.CO;2

Benjankar, R., Tranmer, A. W., Vidergar, D., & Tonina, D. (2020). Riparian vegetation model to predict seedling recruitment and restoration alternatives. *Journal of Environmental Management*, *276*, 111339. https://doi.org/10.1016/j.jenvman.2020.111339

Braatne, J. H., Rood, S. B., & Heilman, P. E. (1996). *Life history, ecology, and conservation of riparian cottonwoods in North America*. *Biology of Populus and its Implications for Management and Conservation*. Retrieved from https://www.researchgate.net/publication/284033185_Life_history_ecology_and_conservation_of_riparian_cottonwoods_in_North_America

Braatne, J. H., Jamieson, R., Gill, M., & Rood, S. B. (2007). Instream flows and the decline of riparian cottonwoods along the Yakima River, Washington, USA. *River Research and Applications*, *23*, 247–267. https://doi.org/10.1002/rra.978

Breiman, L. (2001). Random Forests by Leo Breiman. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021*, 542–545. https://doi.org/10.1109/ICCECE51280.2021.9342376

Burke, M., Jorde, K., & Buffington, J. M. (2009). Application of a hierarchical framework for assessing environmental impacts of dam operation: Changes in streamflow, bed mobility and recruitment of riparian trees in a western North American river. *Journal of Environmental Management*, *90*(SUPPL. 3), S224--S236. https://doi.org/10.1016/j.jenvman.2008.07.022

Burman, S., & Pasternack, G. B. (2017). Riparian Canopy Abundance, Distribution, and Height on the Lower Yuba River in 2008. Retrieved from http://www.yubaaccordrmt.com/Studies Reports/Riparian/UCDR36_LYR_UCDriparianreport_20170423.pdf

Camporeale, C., Perucca, E., Ridolfi, L., & Gurnell, A. M. (2013). Modeling the interactions between river morphodynamics and riparian vegetation. *Reviews of Geophysics*, *51*(3), 379–414. https://doi.org/10.1002/rog.20014

Carley, J. K., Pasternack, G. B., Wyrick, J. R., Barker, J. R., Bratovich, P. M., Massa, D. A., et al. (2012). Significant decadal channel change 58-67 years post-dam accounting for uncertainty in topographic change detection between contour maps and point cloud models. *Geomorphology*, *179*, 71–88. https://doi.org/10.1016/j.geomorph.2012.08.001

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

De'Ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, *81*(11), 3178–3192.

https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Dixon, M. D., Johnson, W. C., Scott, M. L., Bowen, D. E., & Rabbe, L. A. (2012). Dynamics of plains cottonwood (Populus deltoides) forests and historical landscape change along unchannelized segments of the Missouri River, USA. *Environmental Management*, *49*(5), 990–1008. https://doi.org/10.1007/s00267-012-9842-5

Edmaier, K., Burlando, P., & Perona, P. (2011). Mechanisms of vegetation uprooting by flow in alluvial non-cohesive sediment. *Hydrology and Earth System Sciences*, *15*(5), 1615–1627. https://doi.org/10.5194/hess-15-1615-2011

Escobar-Arias, M. I., & Pasternack, G. B. (2011). Differences in river ecological functions due to rapid channel alteration processes in two California rivers using the functional flows model, part 2 - model applications. *River Research and Applications*, *27*(January), 1–122. https://doi.org/10.1002/rra

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Garvelmann, J., Pohl, S., & Weiler, M. (2015). Spatio-temporal controls of snowmelt and runoff generation during rain-on-snow events in a mid-latitude mountain catchment. *Hydrological Processes*, *29*(17), 3649–3664. https://doi.org/10.1002/hyp.10460

Gervasi, A. A., Pasternack, G. B., & East, A. E. (2021). Flooding duration and volume more important than peak discharge in explaining 18 years of gravel–cobble river change. *Earth Surface Processes and Landforms*, *46*(15), 3194–3212. https://doi.org/10.1002/esp.5230

Gordon, E., & Meentemeyer, R. K. (2006). Effects of dam operation and land use on stream channel morphology and riparian vegetation. *Geomorphology*, *82*(3–4), 412–429. https://doi.org/10.1016/j.geomorph.2006.06.001

Gregory, S. V., Swanson, F. J., McKee, W. A., & Cummins, K. W. (1991). An Ecosystem Perspective of Riparian Zones. *BioScience*, *41*(8), 540–551. https://doi.org/10.2307/1311607

Guinn, J. M. (1890). Exceptional years: a history of California floods and drought. *Historical Society of Southern California, Los Angelos*, *1*(5), 33–39.

Huxley, C., & Syme, B. (2016). TUFLOW GPU-Best Practice Advice for Hydrologic and Hydraulic Model Simulations. *Hydrology and Water Resources Symposium*, *Queenstown*, 9.

James, L. A. (2005). Sediment from hydraulic mining detained by Englebright and small dams in the Yuba basin. *Geomorphology*, *71*(1–2), 202–226. https://doi.org/10.1016/j.geomorph.2004.02.016

Kammel, L. E., Pasternack, G. B., Massa, D. A., & Bratovich, P. M. (2016). Near-census ecohydraulics bioverification of Oncorhynchus mykiss spawning microhabitat preferences . *Journal of Ecohydraulics*, *1*(1–2), 62–78. https://doi.org/10.1080/24705357.2016.1237264

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. *Applied Predictive Modeling*. https://doi.org/10.1007/978-1-4614-6849-3

Larrieu, K. G., Pasternack, G. B., & Schwindt, S. (2021). *Automated analysis of lateral river* *connectivity and fish stranding risks—Part 1: Review, theory and algorithm*. *Ecohydrology* (Vol. 14). https://doi.org/10.1002/eco.2268

Legendre, P., & Legendre, L. (1998). *Numerical Ecology* (2nd ed.). Amsterdam, NL: Elsevier.

Ltd, W. P. (2016). *TUFLOW Users Manual-build 2016-03-AE*. WBM Oceanics Australia.

MacWilliams, M. L., Wheaton, J. M., Pasternack, G. B., Street, R. L., & Kitanidis, P. K. (2006). Flow convergence routing hypothesis for pool-riffle maintenance in alluvial rivers. *Water Resources Research*, *42*(10), 1–21. https://doi.org/10.1029/2005WR004391

Mahoney, J. M., & Rood, S. B. (1998). Streamflow requirements for cottonwood seedling recruitment-An integrative model. *Wetlands*, *18*(4), 634–645. https://doi.org/10.1007/BF03161678

Moniz, P. J., & Pasternack, G. B. (2021). Chinook salmon rearing habitat–discharge relationships change as a result of morphodynamic processes. *River Research and Applications*, *37*(10), 1386–1399. https://doi.org/10.1002/rra.3855

Moser, K., Ahn, C., & Noe, G. (2007). Characterization of microtopography and its influence on vegetation patterns in created wetlands. *Wetlands*, *27*(4), 1081–1097. https://doi.org/10.1672/0277-5212(2007)27[1081:COMAII]2.0.CO;2

Naiman, R. J., Decamps, H., & Pollock, M. (1993). The Role of Riparian Corridors in Maintaining Regional Biodiversity Author (s): Robert J . Naiman , Henri Decamps and Michael Pollock Published by : Ecological Society of America Stable URL :

http://www.jstor.org/stable/1941822 . THE ROLE OF RIPARIAN CO. *Ecological Applications*, *3*(2), 209–212.

Pasternack, G. B. (2023). Autumn 2017 Lower Yuba River TUFLOW HPC 2D Model Description, Validation, and Exploratory Simulations. *Prepared for Yuba County Water Agency*.

Pasternack, G. B., Tu, D., & Wyrick, J. R. (2014). Chinook Adult Salmon Spawning Physical Habitat of the Lower Yuba River, (2000), 1–30.

Patten, D. T. (1998). Riparian ecosystems of semi-arid North America: Diversity and human impacts. *Wetlands*, *18*(4), 498–512. https://doi.org/10.1007/BF03161668

Phillips, S. J., & Pasternack, G. B. (2022). Development And Application Of Cottonwood Seedling Recruitment Potential Model At The Upper Long Bar Site On The Lower Yuba River November Prepared for :, (November).

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., et al. (1997a). The Natural Flow Regime. *BioScience*, *47*(11), 769–784. https://doi.org/10.2307/1313099

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., et al. (1997b). The Natural Flow Regime. *BioScience*, *47*(11), 769–784. https://doi.org/10.2307/1313099

Politti, E., Bertoldi, W., Gurnell, A., & Henshaw, A. (2018). Feedbacks between the riparian

Salicaceae and hydrogeomorphic processes: A quantitative review. *Earth-Science Reviews*, *176*(August 2017), 147–165. https://doi.org/10.1016/j.earscirev.2017.07.018

Pollock, M. M., Naiman, R. J., & Hanley, T. A. (1998). Plant Species Richness in Riparian Wetlands--A Test of Biodiversity Theory. *Ecology*, *79*(1), 94. https://doi.org/10.2307/176867

Rickenmann, D., & Recking, A. (2011). Evaluation of flow resistance in gravel-bed rivers through a large field data set. *Water Resources Research*, *47*(7). https://doi.org/10.1029/2010WR009793

Rood, S. B., Braatne, J. H., & Hughes, F. M. R. R. (2003). Ecophysiology of riparian cottonwoods: Stream flow dependency, water relations and restoration. *Tree Physiology*, *23*(16), 1113–1124. https://doi.org/10.1093/treephys/23.16.1113

Schwindt, S., Pasternack, G. B., Bratovich, P. M., Rabone, G., & Simodynes, D. (2019). Lifespan map creation enhances stream restoration design. *MethodsX*, *6*, 756–759. https://doi.org/10.1016/j.mex.2019.04.004

Senter, A., Pasternack, G., Piégay, H., & Vaughan, M. (2017). Wood export prediction at the watershed scale. *Earth Surface Processes and Landforms*, *42*(14), 2377–2392. https://doi.org/10.1002/esp.4190

Silva, P. V. and Pasternack, G. B. (2018). 2017 Lower Yuba River Topographic Mapping Report Nov. Prepared for:Yuba Water Agency., (November), 48.

Southall, N., Sawyer, A., Diaz, S., Hammersmark, C., & Bowles, C. (2022). Golden Opportunity. *Civil Engineering Magazine Archive*, *92*(2), 50–59.

Stella, J. C., Battles, J. J., McBride, J. R., & Orr, B. K. (2010). Riparian Seedling Mortality from Simulated Water Table Recession, and the Design of Sustainable Flow Regimes on Regulated Rivers. *Restoration Ecology*, *18*(SUPPL. 2), 284–294. https://doi.org/10.1111/j.1526-100X.2010.00651.x

Strom, M. A., Pasternack, G. B., & Wyrick, J. R. (2016). Reenvisioning velocity reversal as a diversity of hydraulic patch behaviours. *Hydrological Processes*, *30*(13), 2348–2365. https://doi.org/10.1002/hyp.10797

Stromberg, J. C. (1993). Arizona-Nevada Academy of Science Riparian Mesquite Forests: A Review of Their Ecology, Threats, and Recovery Potential. *Source: Journal of the Arizona-Nevada Academy of Science*, *27*(1), 111–124.

Swanson, F. J., Gregory, S. V., Sedell, J. R., & Campbell, A. G. (1982). Land-water interactions: the riparian zone. *Analysis of Coniferous Forest Ecosystems in the Western United States*, 267–291.

Titus, J. H. (2016). Microtopography and Woody Plant Regeneration in a Hardwood Floodplain Swamp in Florida Author (s): Jonathan Harold Titus Source : Bulletin of the Torrey Botanical Club , Vol . 117 , No . 4 (Oct . - Dec ., 1990), pp . 429- Published by : Torrey Botanic, *117*(4), 429–437.

Vaughan, M. C. (2013). Large Streamwood Storage Does Not Decrease Downstream through a

Watershed. *University of California, Davis, CA*, *MS thesis*, 61.

Wyrick, J., & Pasternack, G. B. (2012). Landforms of the Lower Yuba River. Prepared for the Yuba Accord River Management Team., (April). https://doi.org/10.13140/RG.2.2.27716.63365

Wyrick, J. R., & Pasternack, G. B. (2014). Geospatial organization of fluvial landforms in a gravel-cobble river: Beyond the riffle-pool couplet. *Geomorphology*, *213*, 48–65. https://doi.org/10.1016/j.geomorph.2013.12.040

(YCWA) Yuba County Water Agency. (2013). TECHNICAL MEMORANDUM 7-10 Instream Flow Downstream of Englebright Dam Yuba River Development Project FERC Project No . 2246, (2246), 194.

Zhang, G., wang, J., Xue, Y., Zhang, C., Xu, B., Cheng, Y., & Ren, Y. (2020). Comparison of sampling effort allocation strategies in a stratified random survey with multiple objectives. *Aquaculture and Fisheries*, *5*(3), 113–121. https://doi.org/10.1016/j.aaf.2020.02.002

View publication stats