Title

Mapping subaerial sand-gravel-cobble fluvial sediment facies using airborne lidar and machine learning

Authors

Gómez, Romina Díaz
Pasternack, Gregory B
Guillon, Hervé
et al.

Publication Date

2022-03-01

DOI

10.1016/j.geomorph.2021.108106

Mapping subaerial sand-gravel-cobble fluvial sediment facies using airborne lidar and machine learning

Romina Díaz Gómez [a*], Gregory B. Pasternack[a], Hervé Guillon[a], Colin F. Byrne[b], Sebastian Schwindt[c], Kenneth G. Larrieu [d], Samuel Sandoval Solis[a]

[a] Department of Land, Air, and Water Resources, University of California at Davis, Davis, CA

[b] Technical Service Center, United States Bureau of Reclamation, Denver, CO, USA

[c] Institute for Modeling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Germany[d] Department of Civil and Environmental Engineering, University of California at Davis, Davis, CA

* Corresponding author: rdiazgomez@ucdavis.edu, + 1 530-302-5658

Abstract

Substrate facies monitoring is critical for the understanding of fluvial geomorphologic and ecohydraulic patterns and processes. However, direct substrate measurement is time-consuming and subjected to data sparsity because of small sample, size, and limited data collections within an area of interest, which make it difficult to capture facies patterns. Most new experimental studies focus on mapping substrate based on median grain size of a specific grain size class using automatic or semiautomatic photosieving techniques. This study aimed to develop and apply a method to accurately predict size-mixture facies patterns on exposed riverbeds with minimal ground truth plots (100) using airborne lidar and machine learning. The selected testbed river was a 37.5-km stretch of the regulated lower Yuba River in California, USA, mapped at sub-meter resolution in 2017. First, we designed a grid-by-point grain size sampling method and binned grain sizes into representative mixtures, such as fine or large gravel, to assign subaerial facies labels. Second, we classified facies based on a multivariate cluster analysis. Third, we generated 15 lidar-derived topographic and spectral predictors. Six distinct size-mixture facies were identified from field data and a seventh, pure sand facies, from UAV data. A random forest predictive model with an 86% 10-fold cross-validation accuracy was applied to produce a facies map at the 1.54 m pixel scale. The detrended elevation was identified as the most important variable for predicting facies spatial patterning, followed by baseflow, wetted area proximity, and green lidar intensity. We conclude that machine learning combined with intensity lidar data is highly effective

for distinguishing mixed classes of substrates. Ultimately, the new substrate mixture-binning approach also provides novel insights into the arrangement of river sediment facies patterns.

Keywords: grain size, lidar, substrate, gravelometry, machine learning, remote sensing, fluvial geomorphology

Highlights

- Cluster analysis of field-plot grain-size data yields mixed-size sediment facies to train machine learning.
- A supervised random forest machine learning model accurately predicts subaerial facies using only airbone lidar.
- The top lidar predictor is detrended elevation.

1. Introduction

1.1. Need for rapid, detailed substrate mapping

Substrate facies are broadly defined as areas of sediment that are recognizably different from adjacent sediment patches because every patch formed in a different depositional context. Visible substrate facies constitute the top layer of the active riverbed with variable thickness and potentially high lateral spatial variability (Brierly, 1991). The characteristics of substrate facies can also be attributed to hydrodynamics (Jackson et al., 2013; Hou et al., 2019) and morphodynamics, which drive differential grain size mobility (Nelson et al., 2010, 2009; Venditti et al., 2017). Thus, facies differences can be identified through grain size distributions (GSDs), spatial sorting of different sizes, degree of weathering, and patterning of grain stacking and orientation (Buffington and Montgomery, 1999; Heritage and Milan, 2009; Wittenberg, 2002). In addition, a spatio-temporal reciprocal link exists between facies, landform units, and hydraulic response (Abu-Aly et al., 2014; An et al., 2017; Chanson, 2004; Nelson et al., 2015). It is therefore no surprise that facies are ecologically important for: (i) maintaining

sustainable habitats in rivers (Fryirs and Brierley, 2012; Hauer et al., 2018; Moniz et al., 2019; Pasternack, 2019); (ii) efficient restoration design (Hauer et al., 2018); and (iii) fluvial morphodynamics that renew physical habitats (Escobar-Arias and Pasternack, 2010).

Common methods for obtaining spatially explicit facies patterning include visual facies mapping, GSD measurement methods coupled with a spatial sampling scheme, photosieving techniques, and aerial methods. Visual facies mapping is fast, but potentially inaccurate and unable to yield desired quantitative metrics for ecohydraulic or geomorphic applications (Jackson et al., 2013). Field GSD sampling methods typically measure coarse bed sediment directly in the field (e.g., using Wolman pebble counts coupled with a spatial sampling pattern; Wolman, 1954), which is labor intensive (Kondolf and Li, 1992; Marcus et al., 1995; Whitaker and Potts, 2007). Photosieving techniques assess individual grains and calculate percentiles using a graphical algorithm or semiautomatic segmentation software (e.g., BASEGRAIN). They are relatively low-cost but only cover the small viewing area captured by a digital camera (Barnard et al., 2007; Bertin and Friedrich, 2016; Detert and Weitbrecht , 2012; Di Francesco et al., 2016; Graham et al., 2010; Groom et al., 2018; Weitbrecht et al., 2013). Aerial methods combine airborne survey data with digital image texture analysis, such as aerial photo-sieving. The accuracy of aerial methods depends on the range of grain size variability and imagery quality. Yet, aerial methods can require extensive and time-consuming calibration (Carbonneau et al., 2005a, 2004; Dugdale et al., 2010; Verdú et al., 2005). Generally, manual post-processing is required to delineate the real shape of grains for reducing misidentification. On a riverbed with overlapping grains and

a fine matrix, segmentation methods are prone to errors when detecting the grain size because of weak edge clarity of the grains or shadows, which leads to over-segmentation (Purinton and Bookhagen, 2019).

At a time when topography, water surface elevation, water depth, and vegetation are being widely mapped at high resolution, substrate facies pattern remain the most important variable yet to benefit from technological revolution (Piégay et al., 2020; Tomsett and Leyland, 2019). One of the main reasons for this discrepancy is that mapping river substrate facies is labor-intensive, time-consuming, expensive, and difficult to deploy over large areas at high resolution. Further, fluvial dynamism necessitates that all these problems be addressed on a repeat basis. Consequently, many significant spatial details of substrate heterogeneity are missed and unavailable for use in understanding and managing rivers. Yet, remote sensing data processed with machine learning (ML) algorithms could offer a viable solution.

1.2. Recent advances in remote sensing and machine learning

In recent years, researchers have employed lidar remote-sensing technologies to estimate the GSDs on exposed riverbeds using multiple sensors and methodologies. In a fluvial river system, Škarpich et al. (2019) analyzed variations in bar material GSDs and the hydraulic conditions of managed and re-naturalized gravel reaches. The study used lidar to obtain a digital elevation model (DEM) (grid size 5 m x 5 m) with a photosieving software for estimating gravel size metrics. Chardon et al. (2020) presented the first riverbed evidence that lidar intensity standard deviations provide a GSD proxy for non-submerged areas. In addition, grain size measurement, delineation,

and sorting on a gravel riverbed was successfully conducted with terrestrial laser scanning (Heritage and Milan, 2009; Hodge et al., 2009; Neverman et al., 2019; Wu et al., 2018). From this data a GSD map could be produced, where the median particle size ranged from less than 32 mm to 181 mm (Brasington et al., 2012). However, this is a site-specific method with limited areal extent.

The use of airborne green lidar enables submerged substrate mapping (Carrivick and Smith, 2019) and is well-suited for long (> 5 km) and wide river corridors (Chardon et al., 2020). However, there is no standard workflow for processing lidar cloud point data (Kashani et al., 2015), especially to estimate GSDs, rather than just qualitative substrate facies labels. Another challenge is that the need for frequently repeated surveys and high temporal resolution for river analyses results in high survey costs (Chardon et al., 2020; Tomsett and Leyland, 2019). Thus far, no studies of continuous substrate mapping in rivers have used lidar.

To facilitate airborne surveying of rivers, Unmanned Aerial Vehicles (UAVs) have been used to decrease costs, ease manipulation, increase spatio-temporal resolution, and obtain high-quality orthophoto mosaics (Buscombe, 2020; Carbonneau et al., 2018; Hemmelder et al., 2018; Langhammer et al., 2017; Langhammer and Vacková, 2018; Rusnák et al., 2018; Vázquez-Tarrío et al., 2017; Woodget et al., 2018a, 2017). The structure-from-motion (SfM) survey technique provides high-quality dense 3D point clouds, high spatial density (points/m^2), and high point accuracy (Carrivick and Smith, 2019; Leduc et al., 2019). GSDs maps have been generated using UAS photogrammetry combining semi-automatic grain size estimation methods, imagery

texture processing, and point cloud roughness at the reach scale distribution at a 2-cm pixel size, for a 2.5-km reach extent (Carrivick and Smith, 2019; Langhammer et al., 2017; Pearson et al., 2017; Vázquez-Tarrío et al., 2017; Woodget et al., 2018b, 2017). Fluvial substrates based on dominant grain size were also mapped using object-based image analysis (OBIA) with an RGB orthomosaic acquired from a UAV (Arif et al., 2017). Nevertheless, this method currently is highly problematic, complicated, and labor intensive when used over wide areas.

Recently, ML frameworks have advanced the development of GSD assessment tools that improve edge detection between grains and their shadowed interstitial space (Buscombe, 2020; Lang et al., 2020; Purinton and Bookhagen, 2019). Combined with emerging remote sensing techniques, GSDs can today be measured and mapped by imaging the ground from elevated platforms, such as poles, balloons, UAVs, airplanes, helicopters, and satellites (Carbonneau et al., 2020, 2018, 2005b, 2004,; Dietrich, 2016; Dugdale et al., 2010; Pearson et al., 2017; Piégay et al., 2020). Thus, remote sensing data and ML is used extensively in environmental science to resolve classification mapping challenges (Lary et al., 2016; Maxwell et al., 2018; Sofia, 2020) such as mapping soil types, glacier facies, and coastal sediments (Liu et al., 2019; Misiuk et al., 2019; Zhang et al., 2020). In beach environments, ML algorithms, such as support vector machine (SVM) and maximum likelihood, have been successfully used to classify cobble coverage (Matsumoto and Young, 2018). In fluvial environmental research, substrate classification can be enhanced using the spectral and geomorphometry information with ML algorithms that accept a variety of predictors and produce higher accuracy (Lary et al., 2016; Legleiter et al., 2016; Sofia, 2020).

1.3. Study purpose

This study aims to develop state-of-the-art substrate mapping methods that are less limited with respect to the spatial extent of images (mostly mapped from UAV), one characteristic grain size (e.g., the median grain size), and topographic roughness proxies that are typically used to infer grain size ranges. These methods primarily yield qualitative facies labels, but they do not meet scientific and management needs because there is no procedure for quantitatively mapping subaerial substrate facies in high (~1-3 m) resolution over long (~20-50 km) river stretches with remote sensing data.

This study hypothesizes that identifying substrate facies based on GSD data can provide quantitative metrics to efficiently support a variety of geomorphic and ecohydraulic applications. Given the evidence that grain size is linked with spectral-topographic signals, mapping broad areas might capitalize upon the substrate's information and the relation between facies and topographic variability derived from a DEM. To investigate this possibility, this study asks the fundamental question: can substrate facies patterns and GSDs be accurately predicted using lidar datasets and low-cost fieldwork with ML?

This study addresses the challenge of developing an accurate ML facies predictive model based on airborne lidar, including quantitative substrate metrics as labels at the lower Yuba River (LYR) in California, USA, as a testbed. Specifically, it determines subaerial substrate river facies based on GSD abundance, predicts facies spatial patterns, and quantifies subaerial surficial riverbed sediment facies with minimal ground truth using an ML framework. The prediction performance is statistically assessed and

the most useful predictors are identified to better understand physical processes controlling substrate facies patterning. Based on a review of ML articles making geoscience related predictions, 83% concluded that an ML model was successful when it achieved an accuracy of > 70%. Therefore, this threshold was adopted as the performance indicator to determine whether the test in this study resulted in a successful outcome.

2. Study area: Low Yuba River

The Yuba River is a tributary of the Feather River in north-central California that drains ~3480 km^2 of the western Sierra Nevada (Fig. 1). The ~37.5 km (by valley centerline) segment between the 1941 Englebright Dam (ED) and the confluence of the Feather River is referred to as the lower Yuba River (LYR). The LYR is a regulated river with (~20 emergent bars/islands at bankfull), low sinuosity, high width-to-depth ratio, and slight to no entrenchment. The geomorphically determined bankfull discharge was estimated as 141.6 $m^3 \cdot s^{-1}$, with ~82% annual exceedance probability (Wyrick and Pasternack, 2014).

Numerous sedimentary processes have influenced LYR substrate facies. The watershed experienced a strong anthropogenic impact with historic hydraulic gold mining activities and management strategies. The consequences were severe channel aggradation with deposition of millions of tons of mining sediment during the mid to late-nineteenth century (James et al., 2009). Since its construction, Englebright Dam is a sediment barrier promoting downstream geomorphic recovery (Carley et al., 2012; Gilbert, 1917). Daguerre Point Dam (DPD) splits the LYR into an upstream and

downstream section and it was constructed in 1910. Levees, bank stabilization, and channelization were constructed for flood control and reduce the complexity of in-channel landform pattern (Wyrick et al., 2014; Wyrick and Pasternack, 2012; Wyrick and Pasternack, 2014).

The LYR substrate lithology corresponds to auriferous, rounded and subangular-to-subrounded clasts of coarse granite, rhyolite, and metavolcanic rocks, which dominate the lithology of the quaternary gravels in a matrix of sand and silt (Adler, 1980). The LYR bed material source is a relatively homogenized mixture of sand, gravel, and cobble that is being exhumed from the river corridor by the flows coming out of and over ED. A little downstream, 11.9 million m^3 of sediment was stored in a horse-shoe bend, where it forms alluvial terraces on the hillsides (Pasternack, 2008). Boulders are found primarily along the banks where bedrock has broken off the hillside and tumbled into the channel. Downstream from ED, the LYR flows into a 12-km stretch dominated by extensive dredge spoils of the Yuba Goldfields, piled in gravel ridges 7-20 m high on both sides of the channel, providing additional material to the river (Jackson et al., 2013). The LYR's cobble-gravel bed has well-rounded imbricated coarse substrate (0.30 m) at ED and much finer substrate (0.04 m) close to its confluence with the Feather River. In the bankfull wetted area downstream of ED, cobble is the dominant size class, but a range of silt to bedrock can be observed in the channel. The mean particle size for the entire bankfull segment is 97.4 mm. In-channel landforms are primarily composed of small or medium cobble (i.e., 32-90 mm and 90-128 mm, respectively; Jackson et al., 2013). The longitudinal substrate distribution shows that 10-

50% of riverbed sediment serves as rearing cover for Chinook salmon (i.e., large cobble and boulders; Moniz et al., 2019; Pasternack et al., 2013).

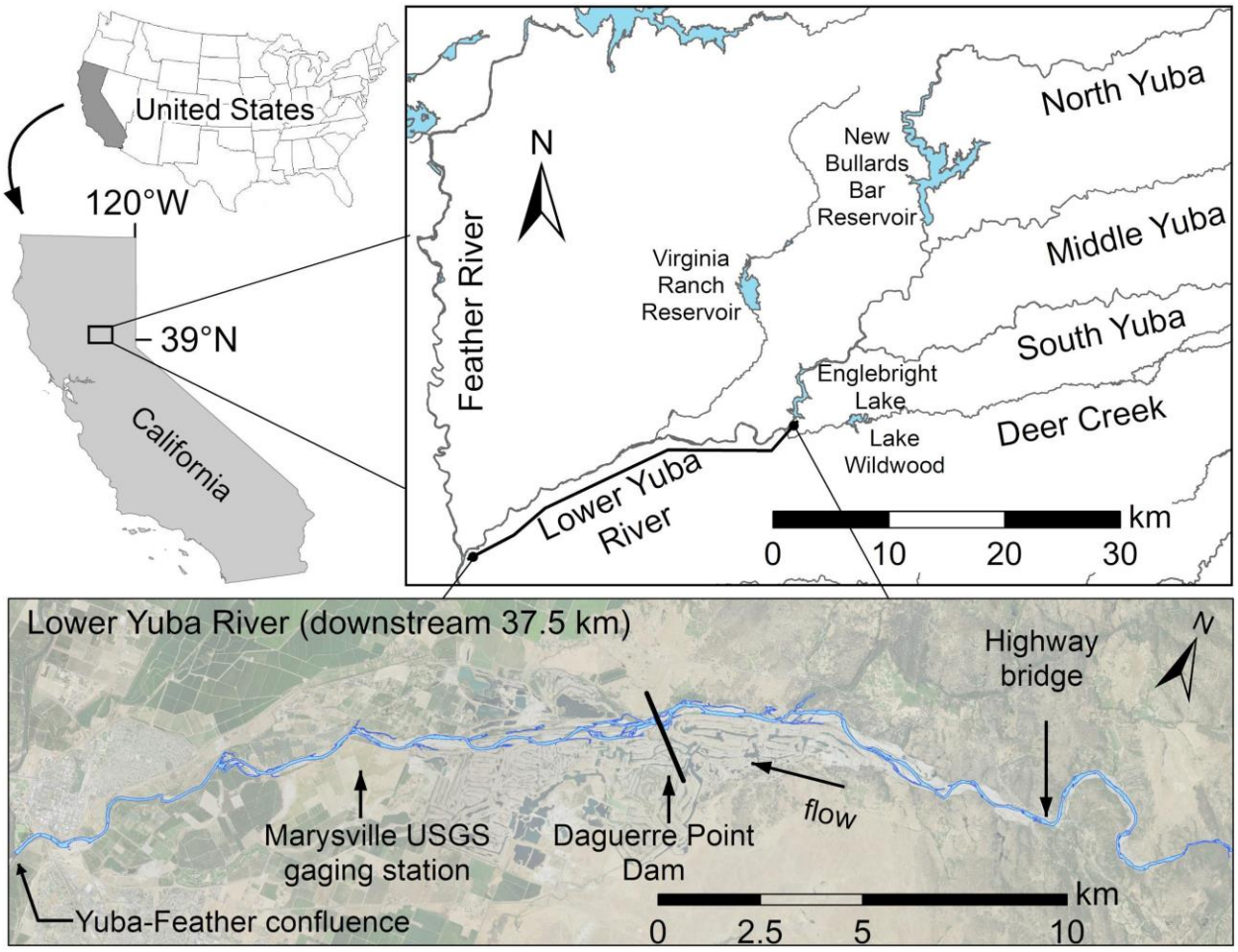

Fig. 1. Maps of the lower Yuba River (California, USA), which serves as a testbed in this study.

3. Material and methods

3.1. Substrate mapping workflow

A novel substrate mapping workflow was designed and applied to the LYR (Fig. 2). A pre-processing phase comprised two activities. First, preparing lidar data by identifying and producing potential substrate facies predictors. All predictors were stacked into a set and clipped to each field observation (5 X 5 m plot to generate training datasets for a random forest (RF) algorithm. Second, identifying substrate facies clusters based on GSDs from field data using hierarchical cluster analysis. Facies clusters were assigned categorical names, referred to as “labels” in ML training, which include GSD information and are consequently more than just nominal labels.

A processing phase included ML implementation with a supervised classification learning task. In this study, ML algorithms did not predict the sizes of individual grains, but instead predicted which substrate facies each raster pixel belonged to. Because each substrate facies has a measured representative GSD, every pixel’s GSD is inferred from its ML-predicted label. Predicting substrate facies is easier and more suitable than trying to predict individual grains because airborne lidar cannot resolve individual sand, gravel, and cobble grains. Finally, an accuracy assessment using performance measurement was applied.

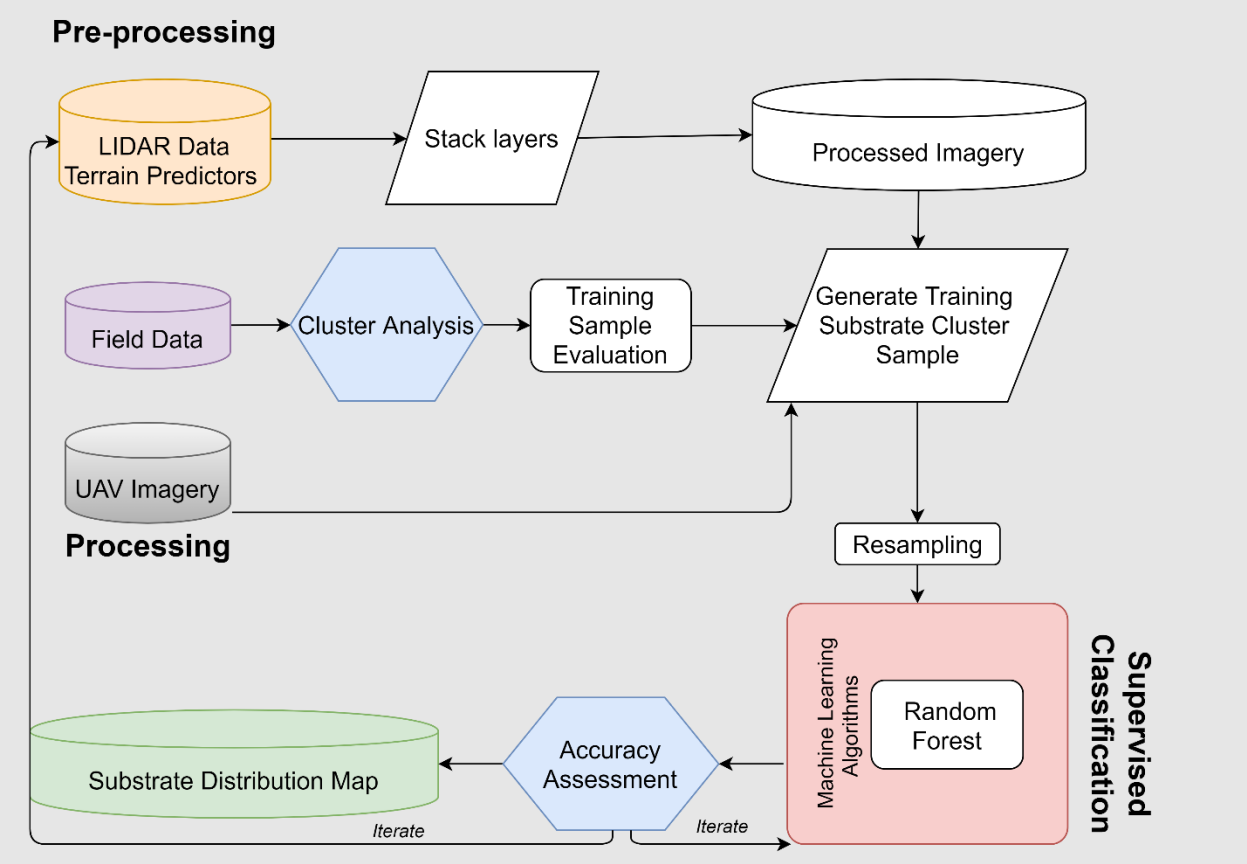

Fig. 2. Substrate mapping workflow with pre-processing (lidar-derived predictors development and substrate facies label generation based on field data and UAV imagery) and ML processing (supervised learning classification, substrate predictive model performance assessment, and model application for substrate facies mapping).

3.2. Data collection

This study focused on an exposed river corridor of the LYR, outside the wetted area at a baseflow discharge of 28.43 $m^3 s^{-1}$ (~20% of bankfull discharge). Further, two-dimensional, validated numerical hydrodynamic models of steady baseflow were

available in high resolution (1.54 m) for 2017 (Barker et al., 2018), and were used to clip out the wetted area.

3.2.1. Airborne lidar

Airborne lidar data were collected in September 2017 at a survey altitude of 500 m with a beam size of 50 cm and a scan angle reduced to a maximum of ±20^{o} from nadir (Quantum Spatial, 2017). Discharge during flight days was between 29.45 m^3 s^{-1} and 22.5 m^3 s^{-1} upstream and downstream of DPD, respectively. Near-infrared (NIR, 1064 nm) and green (532 nm) lidar laser first-return (33.78 points/m^2) rasters were generated at 0.45 m (1.5 ft) pixel resolution. The lidar data contains the occurrence of a return from every location on the ground, and the intensity (strength) of the backscattered laser signal on a scale of 0-255.

3.2.2. UAV imagery

UAV imagery was acquired across the LYR from 2018 to 2020 (recorded with a Matrix 600 platform with an FC550 camera model), at a pixel resolution of 4608 X 3456 and a focal length of 12 mm. RGB imagery was collected across four reaches (Sierra Overhead Analitycs, 2018). More than 3000 images were captured for every reach, with 75-80% overlapping at 44-52 m above ground level. Between 40 and 80 ground control points were recorded on the riverbanks (with a Trimble R10 device) prior to the UAV surveys. The flight mission was planned using the Pix4d app for data collection and operated with an automatic pilot. An orthophoto mosaic was derived with a pixel size of $1.5 \cdot 10^{-2}$ m.

3.3. Ground truth field data to create training variables

3.3.1. Pebble counts

A GSD dataset of large (>1000 m^2) alluvial bars above the highest 2018 flood stage (to ensure no changes in the substrate since the 2017 aerial mapping) was generated with Wolman pebble counts (Wolman, 1954). One hundred bar polygons were identified for the LYR in 2018 based on Google Earth. Bar subsets were selected randomly. Then, for every bar, sample locations were chosen based on a stratified random approach to yield an equal effort sampling for every stratum (facies cluster) to be representative for the patch-pattern (Chawla et al., 2002; Lane et al., 2017).

Two site characterization approaches were used. First, 35 out of 100 sample locations were collected with a transversal line approach, which consisted of collecting grains along a transversal line crossing the bar. Second, the remaining 65 samples were observed with the grid-by-number Wolman pebble count method (Wolman, 1954). We used a 5 X 5 m plot, discretized into 100 nodes spaced 0.5 m apart from each other (around twice the largest grain size) (Fig. 3). The grain size of each particle was measured using a metal gravelometer, providing higher accuracy than ruler measurements and reducing the variability between different operators (Bunte and Abt, 2001). The individual grains located at every node in a plot were handpicked, with pinky finger extended and eyes closed, to pick up the first grain touched (Leopold, 1970), yielding 100 samples measured per plot. In total, 100 ground truth sample plots were selected, each with 100 grains measured, yielding a total of 10,000 grains measured from December 2018 to June 2019. A GPS position was recorded at the center of every

plot using a Trimble GeoExplorer 2008 polygon file that was created to overlay the sample locations in the later geospatial analyses. Collection time was ~15 min per plot with two operators working simultaneously. Field data sheets consisted of 16 size measurement bins. Counts for every size bin were normalized by the total count to obtain the percentage of each size class per plot.

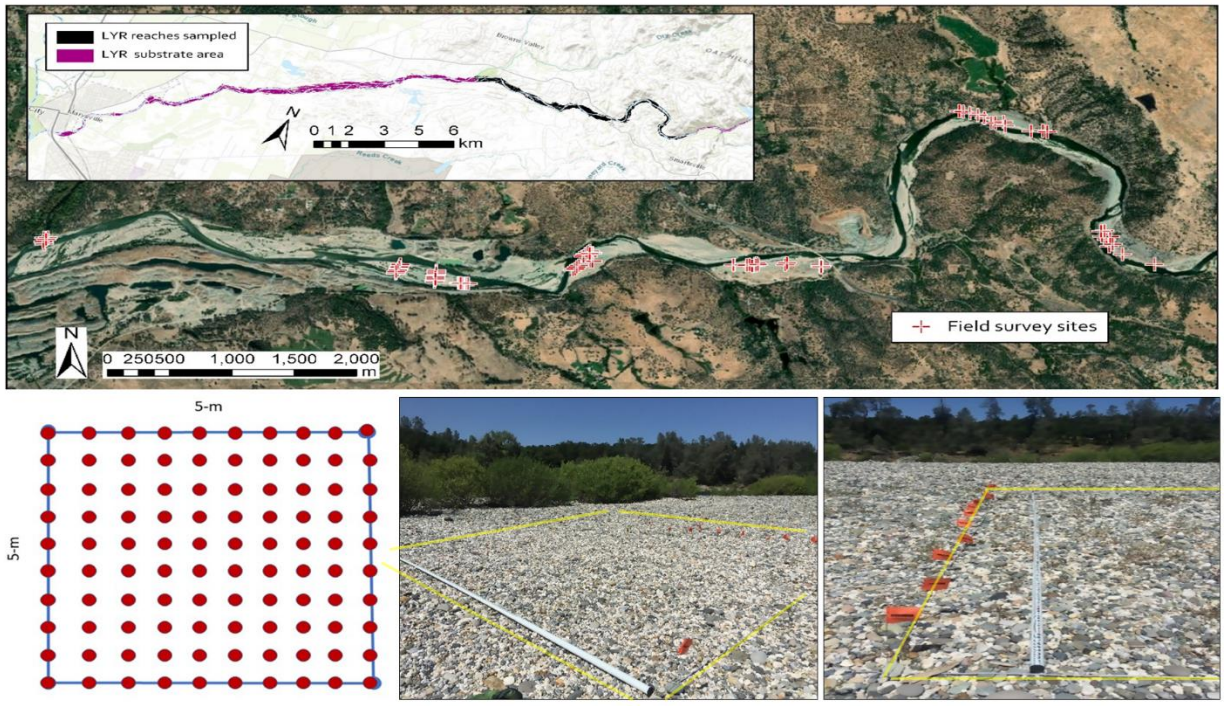

Fig. 3. Example of substrate sample plots. Observation points were spaced 0.5 m apart and the total grid covered 25 m^2.

3.3.2. Manual sand facies clustering using UAV imagery

Sand facies are easy to identify on cm-resolution UAV imagery (Bae et al., 2019). Sixty training samples of different sizes and shapes of the sand facies with highly homogeneous fine substrate (grain size < 2 mm) were collected from the 2018 UAV

orthomosaic (Figs. 4 and 5). This approach was used instead of collecting and sieving bulk samples as the particles were too small for a gravelometer. All sampled locations were above the 2018 and 2019 flood extents.

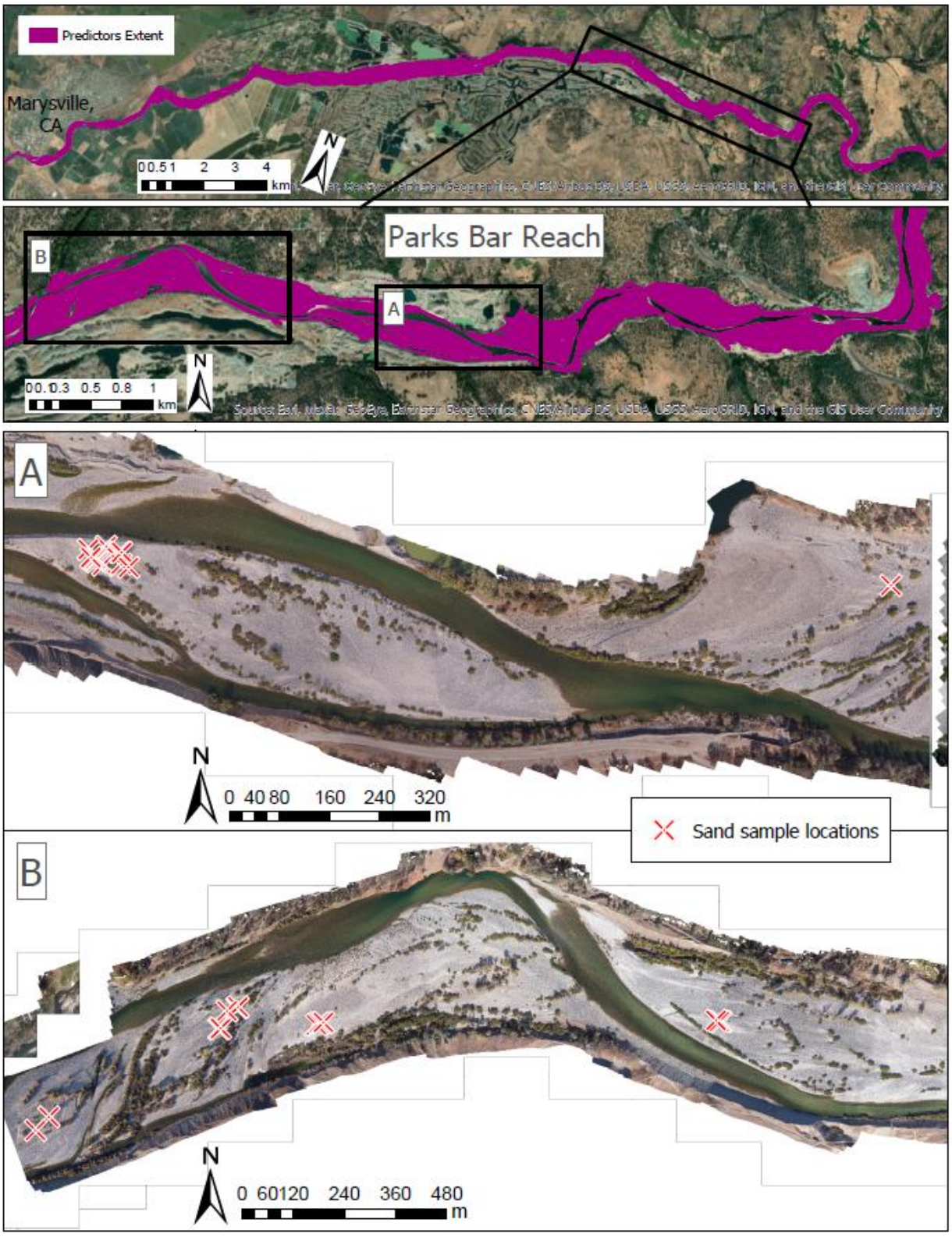

Fig. 4. Sand samples site locations.

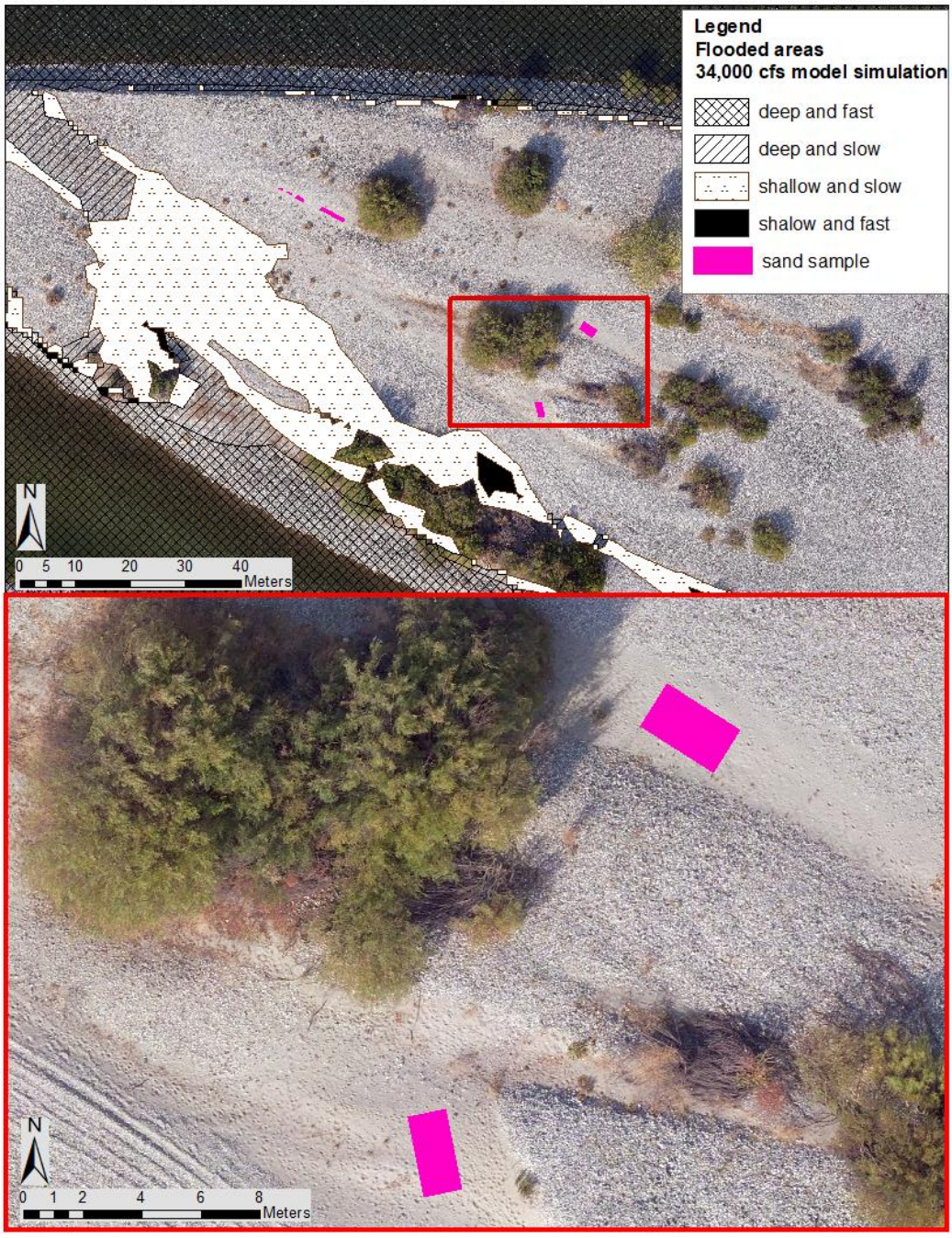

Fig. 5. Locating the sand facies clusters above flooded areas using ~1.5 cm resolution UAV imagery (Wolman pebble counts are not intended for sand and silt).

3.3.3. Extracting predictors from airborne lidar

Possible substrate predictors were generated from airborne lidar in raster format at 1.54 m (5 ft) resolution. Topographic predictors were derived from a high quality 2017 DEM generated from the lidar point cloud with extensive quality control measures (Silva and Pasternack, 2018). Fifteen predictor layers were used in this model (Table 1). The raster for every predictor was processed to clip out inundated terrain, upland areas outside the river corridor, and ground covered by vegetation (details in supplemental material).

Table 1. Variables used in the ML algorithm to predict substrate facies.

Predictor	Method of
Detrended Elevation	Pasternack et al., 2018
Green Lidar Return Intensity	Lidar survey
NIR Lidar Return Intensity	Lidar survey
Std Lidar Intensity	Matsumoto and Young, 2018
Mean Lidar Intensity	Matsumoto and Young, 2018
Baseflow Wetted Area Proximity	This study

Longitudinal relative Aspect	This study
Lateral relative Aspect	This study
Topographic Slope	Hijmans et al., 2018
Terrain Ruggedness Index (TRI)	Hijmans et al., 2018
Topographic Roughness	Hijmans et al., 2018
Topographic Position Index (TPI)	Hijmans et al., 2018
Topographic Flow Direction	Hijmans et al., 2018
Topographic Curvature	Evans, 2020
Pixel-scale Topographic Variability	Weber and Pasternack, 2017

3.3.4. Classifying GSDs for cluster analysis

A combination of statistical methods was applied to the field data to identify the river's differentiable substrate facies (Fig. 6). The rational classification method used here is almost identical to that used in Lane et al. (2017) and Byrne et al. (2020), but there is only one source variable to differentiate substrate facies: observed GSDs. The strategy to turn GSDs into a set of variables involves binning grain size data into discrete ranges and computing the abundance of particles in each bin as a variable. The key challenge is to mindfully decide how many bins are needed in support of cluster analysis. Too

detailed, excessive size-class binning is not optimal for cluster analysis, because it yields clustering variables with extreme highs and lows that produce many individual clusters with only a few field plots in each. The goal is to have broader clusters with many field samples. To create a parsimonious number of clustering variables, the initial 16 GSD size bins were aggregated to four. Every aggregated size class was given a name reflecting the size of the particles in that class: sand and fine gravel (< 5.6 mm), gravel (5.6 – 22.6 mm), large gravel (22.6 – 64 mm) and cobble (64-256 mm). The percent abundance of every bin was computed for every plot. Then, data in every bin were normalized to a 0-1 scale and a hierarchical cluster analysis (Murtagh and Legendre, 2014; Ward, 1963) was performed to reveal substrate facies.

The choice to use four bins was not only made considering conventional sedimentology and the impact of bin number on statistical analysis performance, but also recognizing the eco-geomorphic significance of different size classes. Habitat protection requires a 64 mm grain size class threshold to differentiate larger grains that function as cover that protects smaller organisms from predators and direct sunlight (Fischer, 2000). Also, 63 mm is the differentiable threshold of gravel from cobble (ISO, 2017; Wentworth, 1922). Chinook spawning requires a large gravel grain size range (22.6 – 64mm) in the LYR (Kondolf, 2000). Spawning steelhead trout prefer the gravel substrate bin range (5.6-22.6 mm) (Kammel et al., 2016; Kondolf and Wolman, 1993). Grains smaller than 5.6 were lumped into one class of sand and fine gravel.

3.3.5. Identifying LYR substrate facies

Cluster analysis required an expert decision regarding the number of size classes to use as input variables, which was chosen as four classes. Cluster analysis then analyzed all of the field locations to reveal groups of locations with similar mixtures of those four sizes. Theoretically, cluster analysis can yield any combination of the four variables, based on what is actually present at the observed locations. For example, it is plausible to end up with a sandy cobble mixture as a cluster, if that is what is found on the riverbed. Similarly, there could be a cluster composed of an equal mixture of all size classes. One situation in which each data-driven cluster could theoretically yield groupings that match the four input size classes would be if the river exhibited near-perfect hydraulic sorting to yield highly homogenous patches in just the same size range as the input size classes consist of. Therefore, regardless of the number of final substrate facies clusters that an expert chooses based on the hierarchical clustering diagram, it is extremely unlikely that the clusters will coincide with a single bin range of sizes exactly matching the input size classes. After all, the whole point of the cluster analysis is to identify the mixtures of sediment among the four size classes that are revealed in the field training samples.

In this study, a single threshold distance value was used to delineate substrate facies in the hierarchical clustering diagram. The choice of this value can be difficult to make and justify in some applications, but in this study it turned out to be very simple, as will be presented in the results. To make that decision, the key insight comes from interpreting

what mixture of the four grain size class variables each grouping of sites represents at different hierarchical levels.

Therefore, a suite of analyses was performed to evaluate differences in the resulting facies and to decide at what level in the hierarchy to formalize facies delineations. A one-way analysis of variance (ANOVA) was done to compare the mean values of the aggregated grain size classes among facies clusters. A post-hoc Tukey's honestly significant differences (HSD) test was applied at the 95% confidence level to indicate the most representative grain sizes of every substrate facies cluster (Lane et al., 2017). A stability analysis was conducted by comparing the results of clustering analysis when every aggregated grain size class column was removed one at a time. Four stability metrics were used to aid in the selection of the optimal number of clusters: (1) the average proportion of non-overlap, (2) the average of distance, (3) the average distance between means, and (4) the figure of merit (Datta and Datta, 2003).

Finally, facies clusters names were assigned on an expert basis by interpreting facies GSD attributes and considering 1.5 cm UAV imagery of individual facies clusters plots. Considering traditional classification system naming conventions, facies clusters names were selected based on the dominant grain size class, when such was present. Some facies clusters ended up with mixtures of different aggregated size classes, so they were given mixed names.

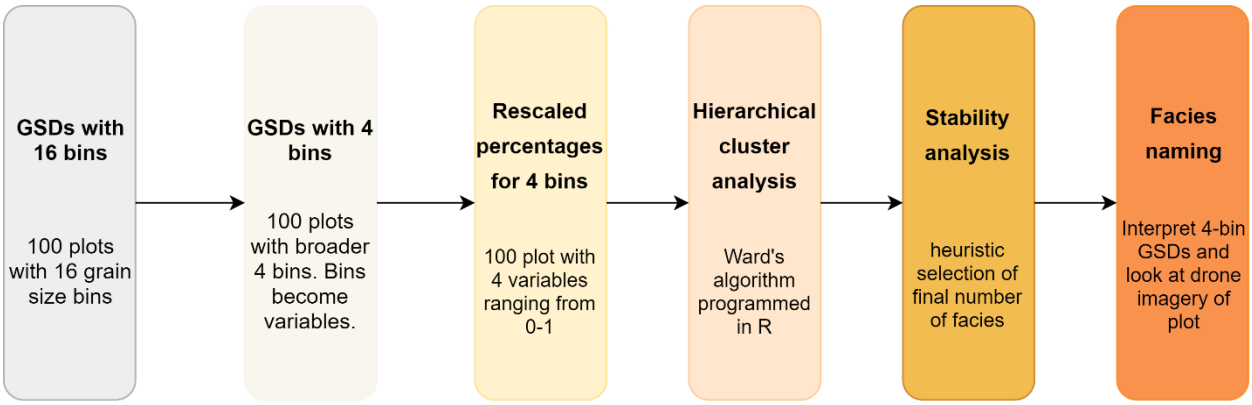

Fig. 6. Identifying LYR substrate facies clusters workflow.

3.4. Machine learning implementation

Given 100 field plots with substrate facies clusters labels identified from statistical clustering serving as training data, the next step involved evaluating an ML algorithm and hyperparameters regarding their ability to predict those labels from a set of 15 predictors (Table 1, Fig. 2). The hyperparameters are parameters related to the model's architecture and have effects on the resulting model performance. The process of tuning consisted of finding the hyperparameters that yielded the best model performance (Kuhn and Johnson, 2013).

A supervised classification task with a (RF) algorithm was performed. A statistical resampling method was applied to segregate observation data into training and validation subsets. A suite of performance metrics was analyzed. The best-performing model was applied to the entire study area.

3.4.1. Supervised classification with ML

The ML supervised learning task applied the known information (ground truth), in this case labels, and was trained to recognize similar patterns distinguishing label boundaries in multidimensional predictor spaces. The RF algorithm by Breiman (2001) is an ensemble of tree decisions and depends on the value of an independently sampled random vector. RF has shown high performance and accuracy compared to other ML models in mapping with remote sensing derived variables and small training samples (Liu et al., 2019; Zhang et al., 2020), as well as with high dimensional and non-normally distributed data (Guillon et al., 2020; Maxwell et al., 2018; Millard and Richardson, 2015). The number of tree (ntree) selected was 500 as conservative default value (Maxwell et al., 2018). The number of predictor variables randomly sampled as candidates at every split (mtry) was defined by a grid-search method with a grid resolution of 1, and mtry values between 1 and the number of predictors (15) (Probst et al., 2018; Zhang et al., 2020). Implementation was performed using the caret package in R (Kuhn et al., 2008).

Resampling is a technique to evaluate the model's efficacy by splitting the data into a sample subset to fit and a subset to test the model. Selecting a robust resampling method in spatial classification problems is a key step in the ML framework (Guillon et al., 2020; Ramezan et al., 2019). The process of doing multiple cross-validations is called "repeated" *k*-fold cross-validation, where *k* indicates the number of groups that a given dataset will be split into. The cross-validation method increases the statistical significance and reduces model overfitting by using all observations to train and validate

the model (Burman, 1989; Hawkins et al., 2003). Twenty repeats of the 10-fold cross-validation resampling method were used to calibrate and validate the RF model. Training data were split randomly (10-fold), where one was used for model assessment and the rest for training. A pre-processed facies dataset with 1113 training pixels each having 15 predictor variables was used to build the model.

Despite efforts to achieve equal sampling across facies clusters, there were fewer sand facies training samples compared to large gravel facies samples. To address the imbalance, three methods were considered. The up-sampling method consisted of simulating additional minority classes (synthetic observations) to improve the balance across classes, while the down-sampling method reduced randomly chosen cases among the majority classes to decrease their effect on the classifier. The synthetic minority over-sampling technique (SMOTE) used both up- and down-sampling (Chawla et al., 2002; Kuhn and Johnson, 2013). However, the application of the imbalance methods did not improve the performance of the model and a model without these methods was generated as the most representative.

3.4.2. Measuring performance

The ML facies predictive model's statistical performance was assessed using median cross-validation overall accuracy as a result of the repeated cross-validation resampling method. Computing the ratioof correct predicted facies to the total number of facies clusters, the model with the highest median cross-validation was selected (Hawkins et al., 2003). To evaluate the relative error among facies, a confusion matrix and sensitivity metric were computed. The sensitivity metric represents the ratio between correctly and

incorrectly predicted observations (Deng et al., 2016; Kuhn and Johnson, 2013; Ramezan et al., 2019). To identify the main facies predictors, the so-called variable of importance was computed (Breiman, 2001). In addition, a visual inspection was done to qualitatively compare the 1.5 cm pixel size UAV orthomosaic versus the final substrate predictive map. To understand the relationship between predictors and facies, histograms between predicted facies and most important predictors were generated for the total overbank area in the river and a representative bar.

4. Results

4.1. Facies training data binning results

Grain size distributions of 100 field plots show significant substrate variability between plots but a general abundance of large gravel (Fig. 7). Histograms of the same data aggregated from 16 size classes to four show that this amount of simplification in GSD numbers yields meaningful variability across all four classes (i.e., a wide range of %-abundance) with few empty abundance bins in every size class, supporting the suitability of a four-variable scheme in cluster analysis (Fig. 8). As an example of how to read the plot, the top left results for the “sand and fine gravel class” show that 76% of field plots have 0 -10% abundance of this class. Thus, there was little sand and fine gravel in the field plots. Because sand and fine gravel are highly skewed, these grain size classes yield a single grouping. The gravel class result shows that 30% of the sites have 10-20% of abundance. Less than 10% of field plots have more than 40% abundance of gravel. The result for large gravel shows that 22% of the field plots have

most of the gravel abundance between 40 and 50%. For large cobble, ~20% of field plots show 10-80% abundance of cobble. Large cobble abundance exhibits more uniform distributions, helping to explain their dominant roles in structuring the six facies cluster.

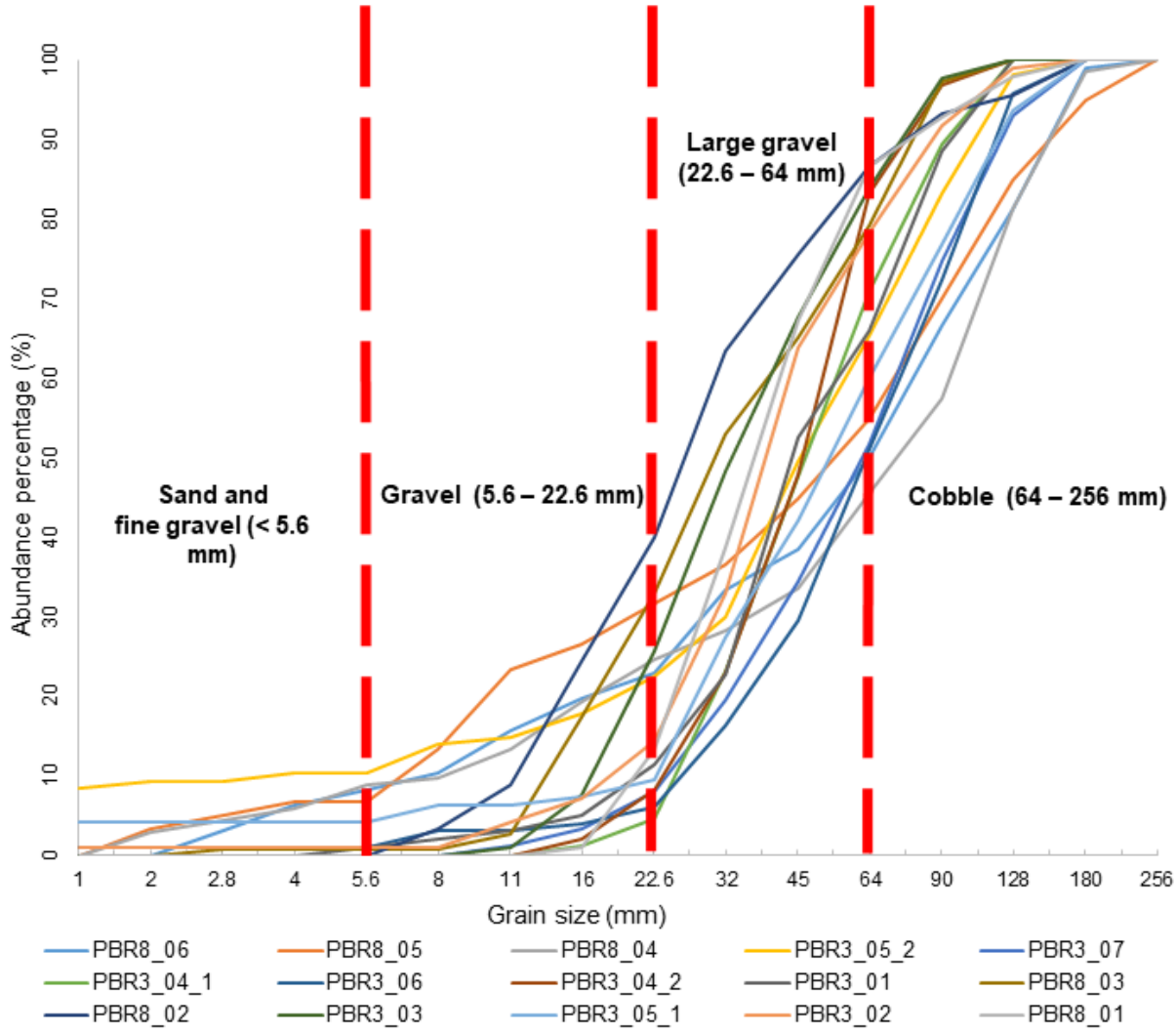

Fig. 7. Raw GSDs for 15 sites on one large alluvial bar (called PBR). Vertical red lines break the GSDs into four aggregated grain size classes.

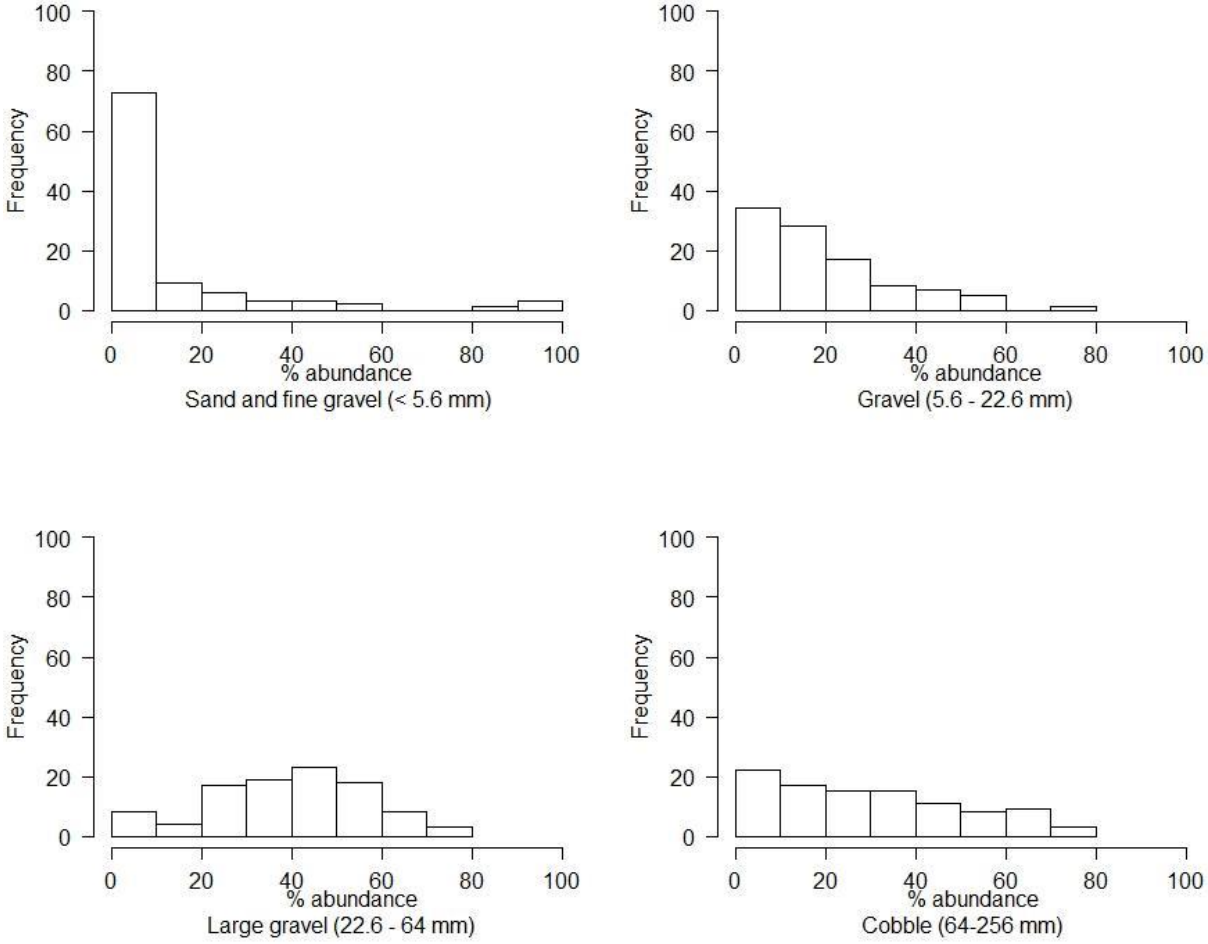

Fig. 8. Histograms of grain size across 100 ground plot samples. The x-axis shows bins of percent abundance of every size class. The y-axis shows the number (and percentage) of samples with abundance percentage bin (because there are exactly 100 samples and frequency equals percentage).

4.2. Facies clusters in LYR

Hierarchical clustering using Ward's methods (Murtagh and Legendre, 2014; Ward, 1963) showed the clustering structure of the 100 ground truth substrate field plots (Fig.

9). Evaluation of the classification performance metrics for scenarios with different numbers of clusters revealed that six substrate facies clusters is the most representative number of groups to understand LYR substrate diversity. To clarify why this is the case, consider the consequences of choosing a four-facies system based on a higher threshold distance of ~ 10 in Fig. 9. A consolidation from six down to four clusters would result in the mixed gravel and equal mix clusters merging into one cluster, while cobble/gravel and large gravel would merge into another cluster. This would be a poor outcome. First, the equal mix cluster has ~8.5 times more cobble and ~2.4 times less gravel than the mixed gravel cluster. Results found that the two clusters have roughly similar, low relative abundances on the LYR (5.2 and 4.0%, respectively). Therefore, merging those units would have significant and negative consequences for ecological analysis, as it would obscure the relatively rare locations of good cobble cover available in areas of actual equal mix and over-represent the possibility of cover in the rare areas of actual mixed gravel. Second, similar ecological reasoning applies when comparing the consequences of merging cobble/gravel and large gravel clusters, but this time the result could be a significant overestimation of available cover habitat, because cobble/ gravel is the most abundant class (42.3% of study area), whereas large gravel is only present at 12.7% of the study area. Merging the two could increase estimated cover habitat by a significant 57.5 ha. Therefore, maintaining the differentiation of substrate facies with six compared to four clusters is not only sedimentologically justified given the wide statistical differences among six clusters, but is ecologically important and helpful for river management. Note that decreasing the distance threshold to obtain more than six clusters is difficult to support on the basis of

available mixed gravel and equal mix sites and their relative infrequency on the LYR (Fig. 9), as those two clusters would be the first to split. Therefore, six clusters is the best result.

Names for the six facies clusters reflect the nature of each mixture and the cluster input variable (i.e., one of four grain size classes) with the largest abundance in every facies clusters, as per the average abundance of every size class among the field plots grouped into every facies (Table 2). The first dendrogram split, which represents the maximum level of dissimilarity between facies, distinguishes three substrate facies clusters with smaller sizes (left side of dendrogram) from three predominantly large substrate-size facies (right side of the dendrogram). Three fine grain substrate facies clusters splits can be identified and named based on the proportion of gravel: sand/fine gravel, mixed gravel, and equal mix. Three substrate facies with larger grain sizes are named as large cobble, cobble-gravel, and large gravel. The only exception is the equal mix facies, which has an equal composition (~25%) of fine gravel, gravel, large gravel and cobble (Table 2).

Individual one-way ANOVA results indicate that grain size class abundance varied significantly among the six substrate facies clusters ($p < 0.05$). Box-and-whisker plots (Fig. 10) show the relative differences in the percentage of grain size classes across the six identified substrate facies. These plots indicate that substrate facies cluster 4 (sand/fine gravel) have a higher percentage of fine gravel than the other facies clusters. Multiple comparisons of the mean of the grain size class percentages using Tukey's HSD post hoc at the 95% confidence interval suggest that substrate facies clusters

exhibit significant statistical differences. For example, the cobble grain size is significantly higher for substrate facies cluster 6 (large cobble) than for all the other substrate clusters. Conversely, the fine gravel grain size differs significantly between substrate facies clusters 1 and 4, but there is no significant difference among facies clusters 1, 2 and 6 (Fig. 10).

Table 2. Substrate facies clusters names. Average abundance of every size class among field plots for every facies cluster.

		Percent abundance of every grain size class			
Facies Cluster Name	Facies Cluster Number	Fine gravel (<5.6 mm) (%)	Gravel (5.6 – 22.6 mm) (%)	Large Gravel (22.6 – 64 mm) (%)	Cobble (64 – 256 mm) (%)
Large Gravel	1	2.22	36.6	50.6	10.6
Cobble/Gravel	2	2.67	12.1	53.5	31.7
Equal Mix	3	24.2	21	25.6	29.2
Sand/Fine Gravel	4	87.2	6.97	3.39	2.4

Mixed Gravel	5	26.1	50.3	20.2	3.41
Large Cobble	6	3.12	7.37	29.5	60

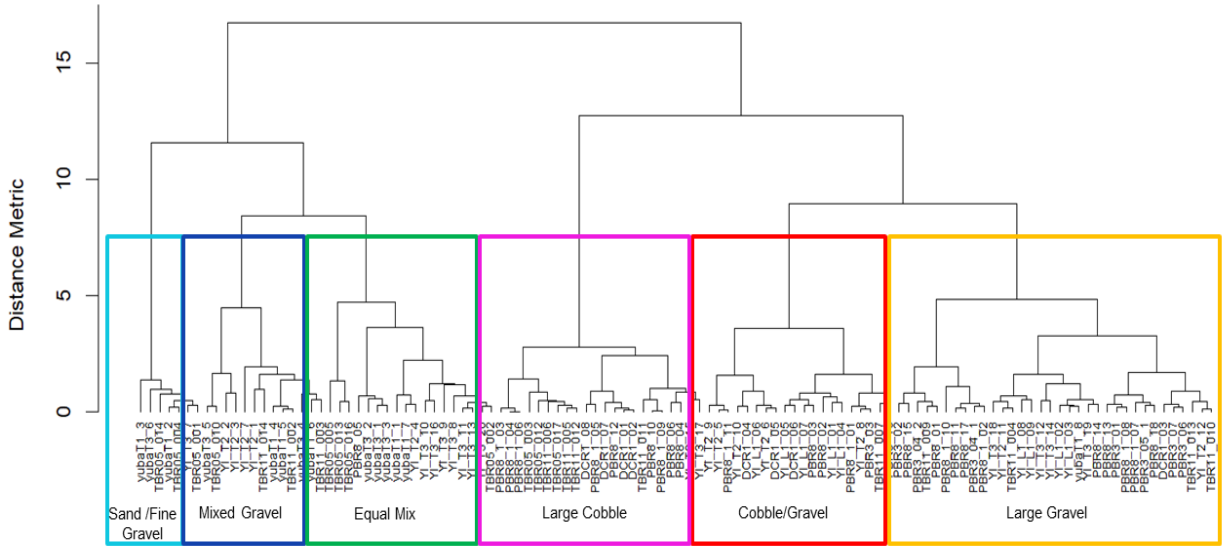

Fig. 9. Hierarchical clustering using Ward's method of 100 ground truth substrate sample plots across the LYR, revealing six distinct substrate facies clusters. Every line represents a substrate site. Colors highlight identified facies clusters.

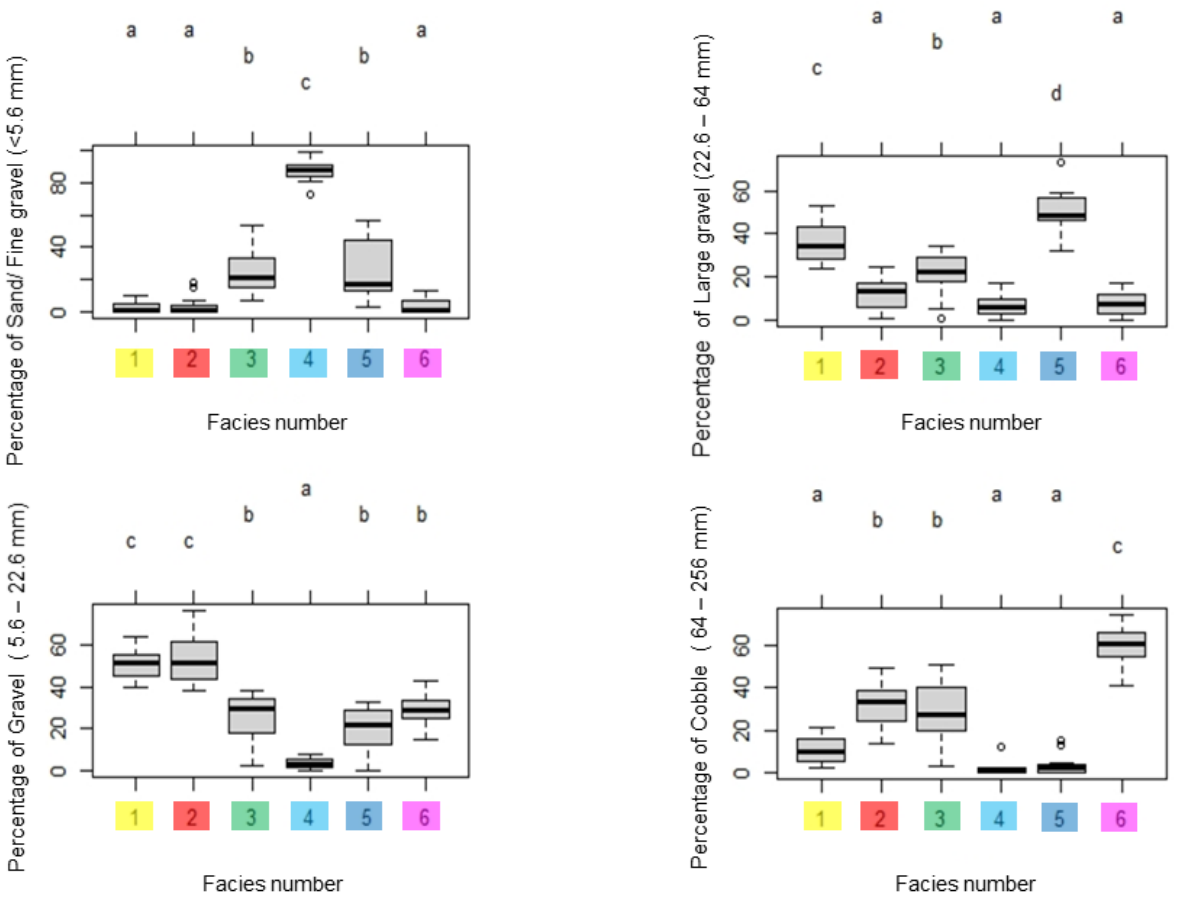

Fig. 10. Box-and-whisker plots and Tukey's honestly significant differences (HSD) test indicate differences in grain size variability across the six identified facies clusters. Facies cluster number 1. Cobble/Gravel, 2. Large gravel, 3. Cobble/Gravel, 4. Mixed, 5. Equal mix, 6. Sand/Fine Gravel. Letters on the top indicate the significance of differences. Same letters indicate no significant difference between facies clusters.

Stability measures reveal the minimum variation among the possible number of clusters. The average proportion of non-overlap is relatively low for all possible numbers of facies, which means that the clustering is stable. The figure-of-merit measure

represents the average intra-facies variance of the removed variable and the lowest value is at 10 clusters (0.47). The average of distance measures the average distance between observations for every site in the same facies cluster and the average distance between means measure the average distance between facies cluster centers. Smaller values are preferred. Both metrics showed the minimum value at ten clusters. However, ten facies clusters were deemed too many because of the resulting small number of sites in certain clusters. Six clusters were selected to represent LYR substrate facies because that number was deemed optimal to describe physical differences in real-world substrate characteristics and the stability measures show a decreasing degree of stability improvement as the number of clusters increases from six to ten (Fig. 11).

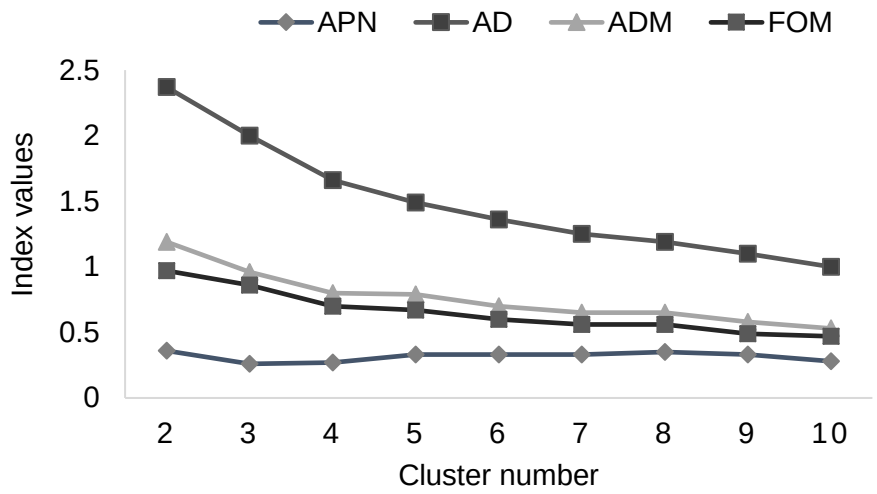

Fig. 11. Stability measure by facies cluster number. The average proportion of non-overlap (APN), the average of distance (AD), the average distance between means (ADM), and the figure of merit (FOM).

4.3. Substrate facies predictive model

The RF model predicted different abundances of the six clusters, as touched on earlier. Cobble/gravel was present for 191.7 ha (43.2% of study area), while sand/fine gravel only covered 7.8 ha (1.7% of area). Large cobble was second in abundance (28.2%). Sand, equal mix, and mixed gravels were in the 4-6% of area range. The LYR stands out for its widespread occurrence of cobbles on open dry terrain in the river corridor, and hence supports less riparian vegetation than rivers elsewhere with abundant mud and sand.

The RF model predicted facies with a cross-validation accuracy of 86% with 15 variables available for splitting at every tree node (Fig. 12). Predictive success varied by facies. The confusion matrix (Table 3) suggests that all seven labels corresponded to the predicted value with a sensitivity higher than 0.73. Sand/fine gravel had the highest sensitivity value of 0.95, which is remarkable. That was followed by sand, with a sensitivity of 0.91 corresponding to 7487 correctly identified sand pixels. Nevertheless, cobble/gravel facies were confused in a higher proportion by large gravel facies, and vice versa. Large gravel's sensitivity of 0.88 is the highest of the gravel-abundant facies. In addition, large cobble facies, with a sensitivity of 0.81, was confused in a higher proportion with cobble/gravel facies observations.

Mixture facies had the lowest sensitivity. Mixtures have relatively low abundance on the LYR, which means they can be difficult to adequately sample in the field and characterize using lidar. The equal mix's sensitivity was 0.73 and highly confused with large cobble facies. This can be understood by the field observation that large cobbles protrude well above sand, fine gravel, and gravel, making the equal mixture appear more like the large cobble facies. Because they are mixed, it is unsurprising that the mixture facies are the hardest to distinguish. However, mixed gravel's sensitivity was 0.87, confused mainly by sand facies. Interestingly, every mixture was not confused much with another mixture.

The detrended DEM was the most important variable in the model compared to other topographic and laser intensity predictors (Fig. 13). The next influential facies predictors were baseflow wetted area proximity (77% importance), and green lidar return intensity (32% importance).

A visual inspection was done to qualitatively compare the 1.5 cm pixel size UAV orthomosaic and the final substrate prediction map. In the UAV orthomosaic, centered on the so-called Long Bar, large cobble patches have clear boundaries, but in the predictor map, large cobble shows a speckled pattern with some large gravel pixels. Sand and mixed gravel facies show random pixels distributed in the facies map that are not presented in the UAV orthomosaic. However, visually narrow sand patches can be observed behind shrub vegetation as well in the prediction map. Also, homogenous large cobble patches in front of shrubs are observed as dark substrate in UAV imagery and in the predicted map. To this end, the model correctly distinguishes coarse facies

from fine facies (Fig. 14) and visual validation indicates that the model is highly accurate.

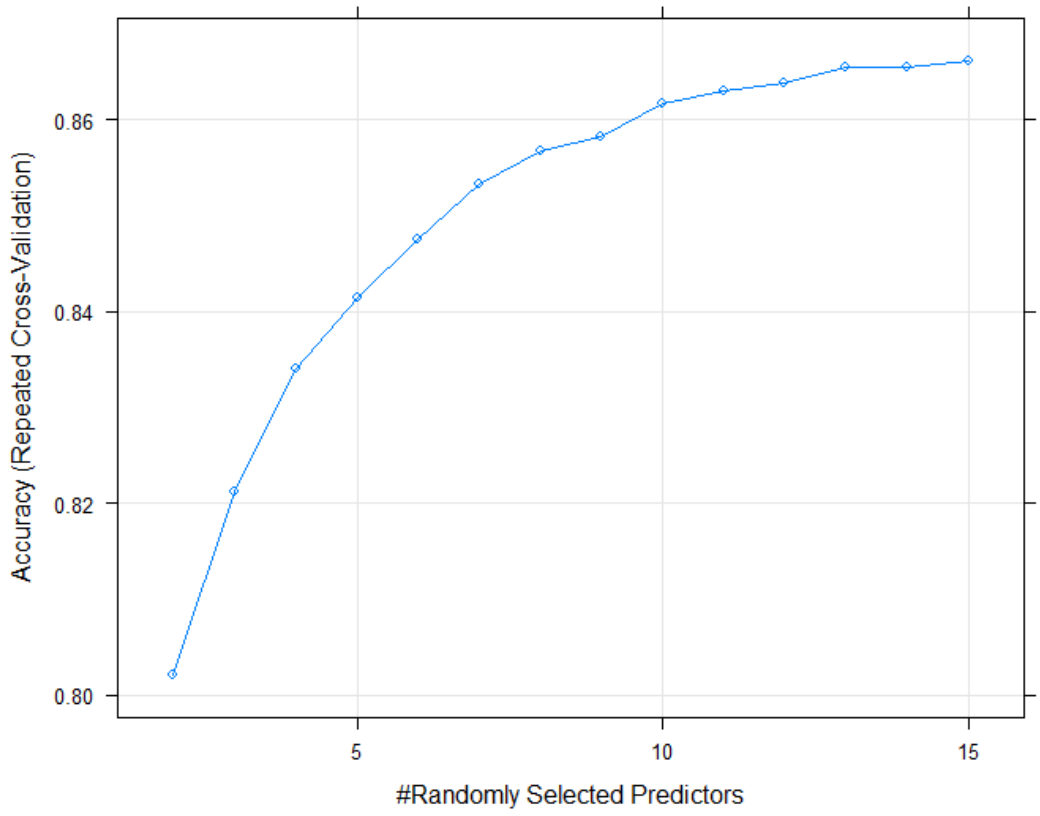

Fig. 12. Overall accuracy for the facies random forest predictive model by number of predictor variables

Table 3. Confusion matrix; the major diagonal represents the facies cluster observations that were correctly predicted.

Reference observations

Predicted observations	Large Gravel	Cobble/ Gravel	Equal Mix	Sand/ Fine Gravel	Mixed Gravel	Large Cobble	Sand
Large Gravel	10179	1123	65	0	88	255	132
Cobble/Gravel	1404	16918	633	61	162	1322	290
Equal Mix	21	112	5857	13	13	241	36
Sand/Fine Gravel	32	0	1	1604	0	0	4
Mixed Gravel	163	282	67	0	3672	200	125
Large Cobble	642	877	1013	0	12	10426	86
Sand	99	188	284	2	253	336	7487
Total	12540	19500	7920	1680	4200	12780	8160
Sensitivity	0.88	0.86	0.73	0.95	0.87	0.81	0.91

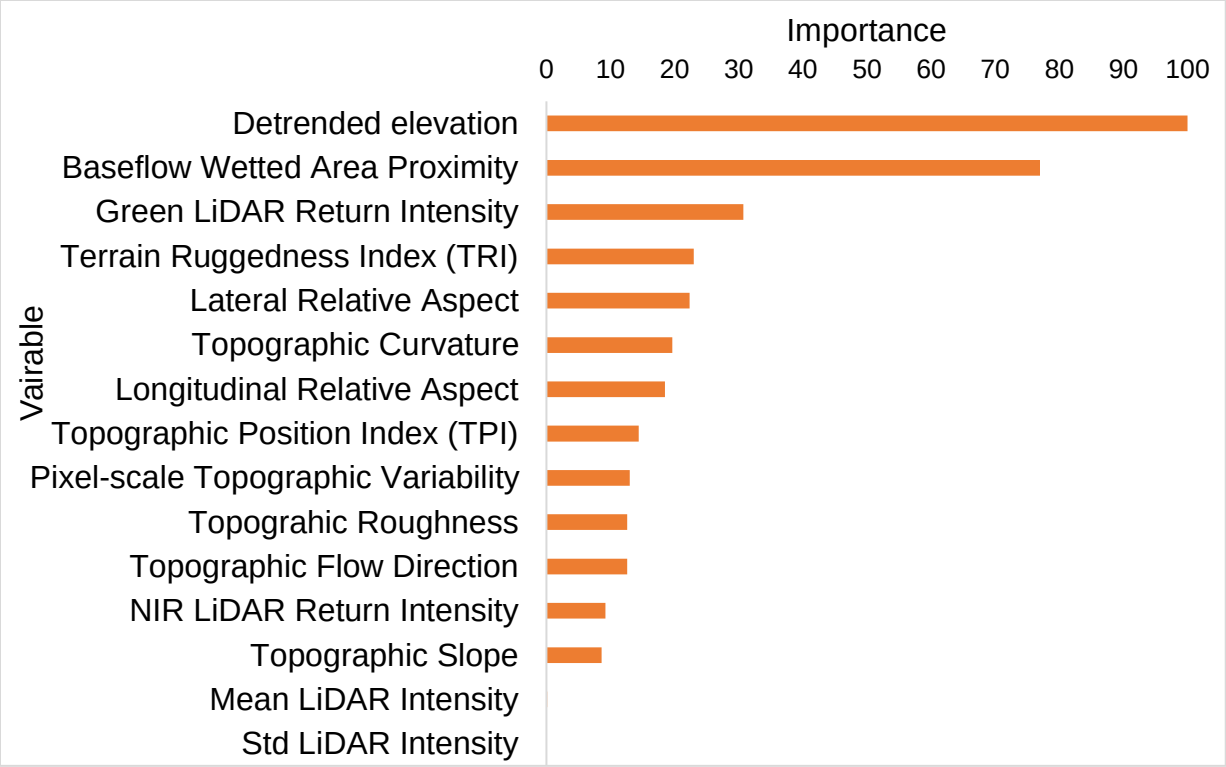

Fig. 13. Variable importance for the random forest facies predictive model.

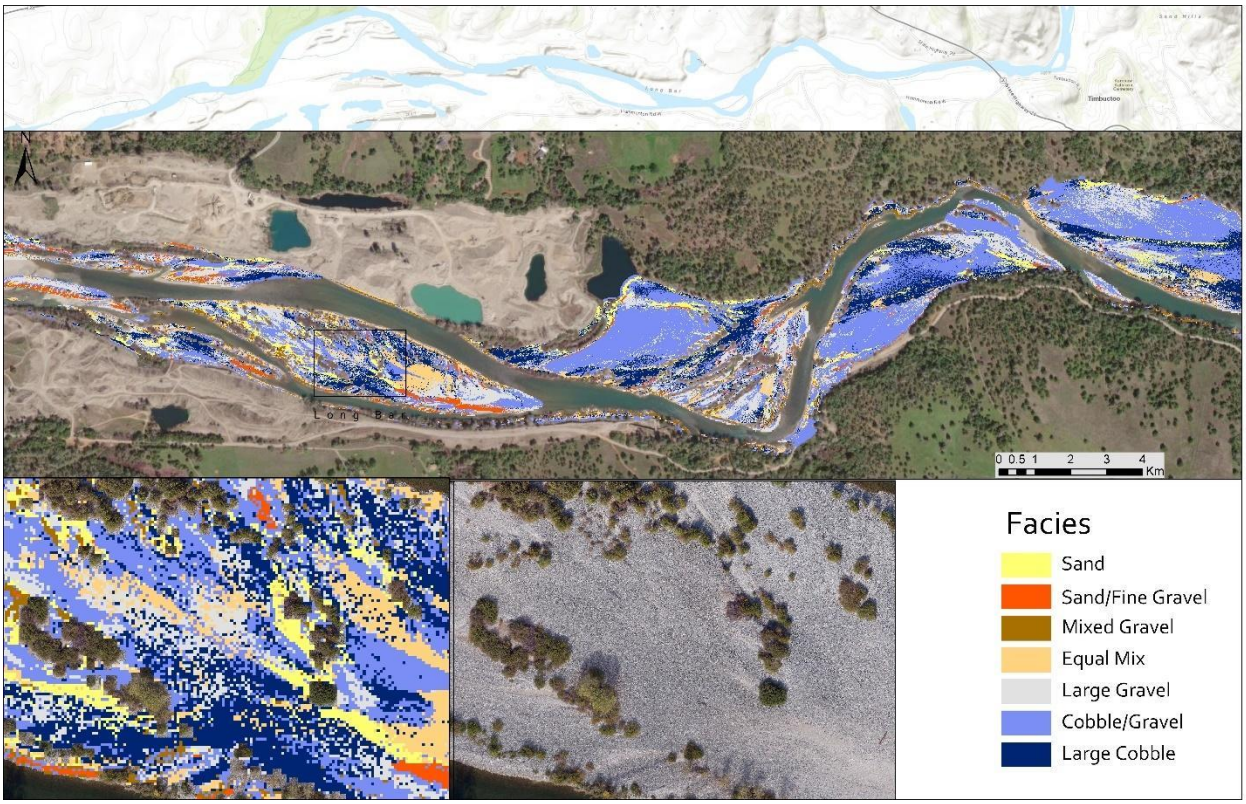

Fig. 14. Substrate facies distribution derived from the random forest predictive model (top). Visual inspection (bottom) shows correspondence between facies map (1.54 m pixel size) and UAV orthomosaic (1.5 cm pixel size).

As a second example, facies spatial patterns correlation with predictor variability on a bar in the so-called Timbuctoo Bend are featured in Fig. 15. The statistical predictive performance was analyzed by comparing the distribution of predictor values and facies for the whole study area and for just the bar shown in Fig. 16. Results reveal that distance from baseflow wetted area was slightly higher for sand/finer facies compared to coarser facies in the total overbank areas (Fig. 16). For example, mean distance from the baseflow wetted area is 77.7 m for the sand facies at the bar and the total overbank. Meanwhile, detrended elevation mean values were distributed differently. Finer facies

with sand abundance were at higher elevations compared to coarser facies. Green intensity mean values were higher for finer facies. Large cobble showed lowest mean green lidar return intensity values. Green lidar return intensity values gradually increased from coarse to fine facies at the total overbank scale.

The lateral relative aspect distributions reveal a strong difference between sand and sand/fine gravel facies compared to coarser facies in the overbank areas. For the total overbank areas, finer facies face away (-1) from the channel. However, this distribution is not clear at the Timbuctoo Bend bar site. Meanwhile, longitudinal relative aspect mean values showed similar distribution among facies, failing to distinguish between them.

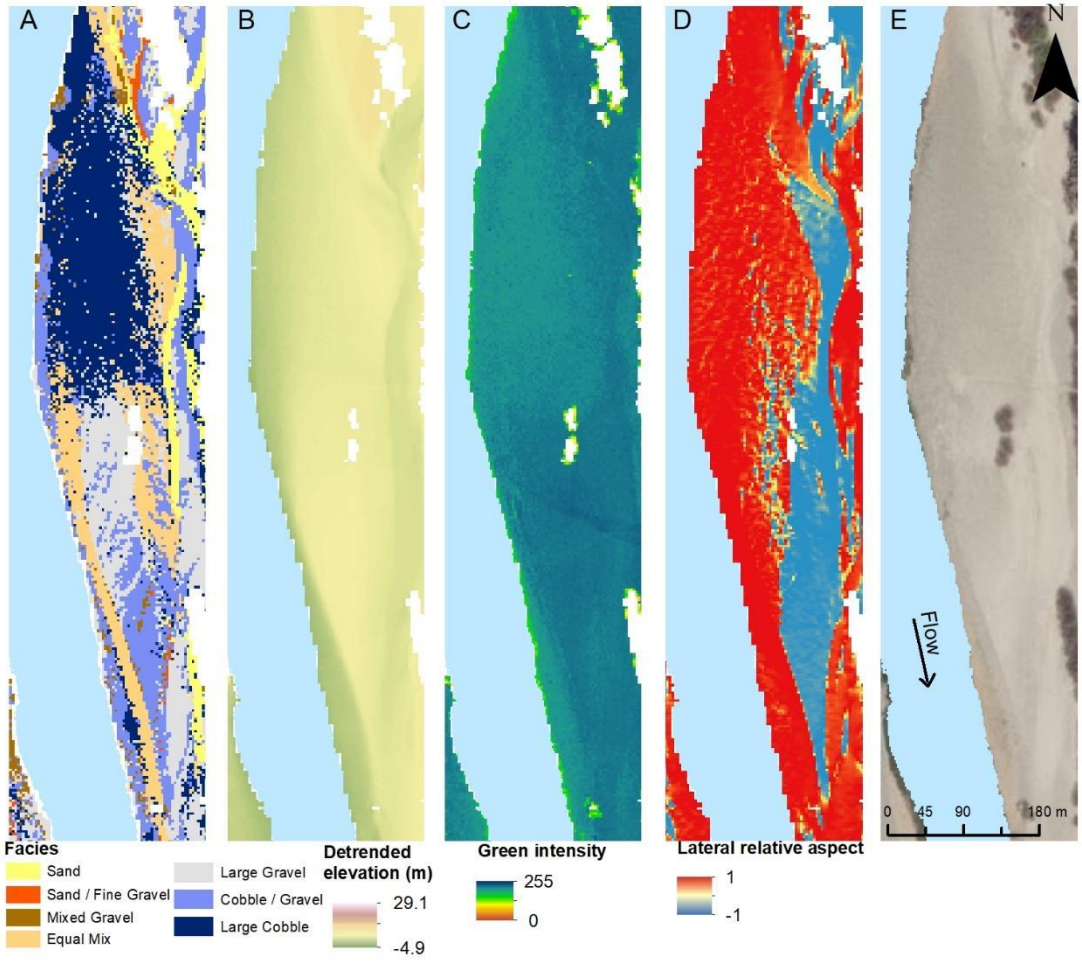

Fig. 15. Facies map comparison of the bar in the so-called Timbuctoo Bend with main facies predictors (lidar-derived). Coarser facies correlate with low green intensity values. A) facies map B) detrended elevation, C) green intensity, D) lateral relative aspect, E) aerial imagery.

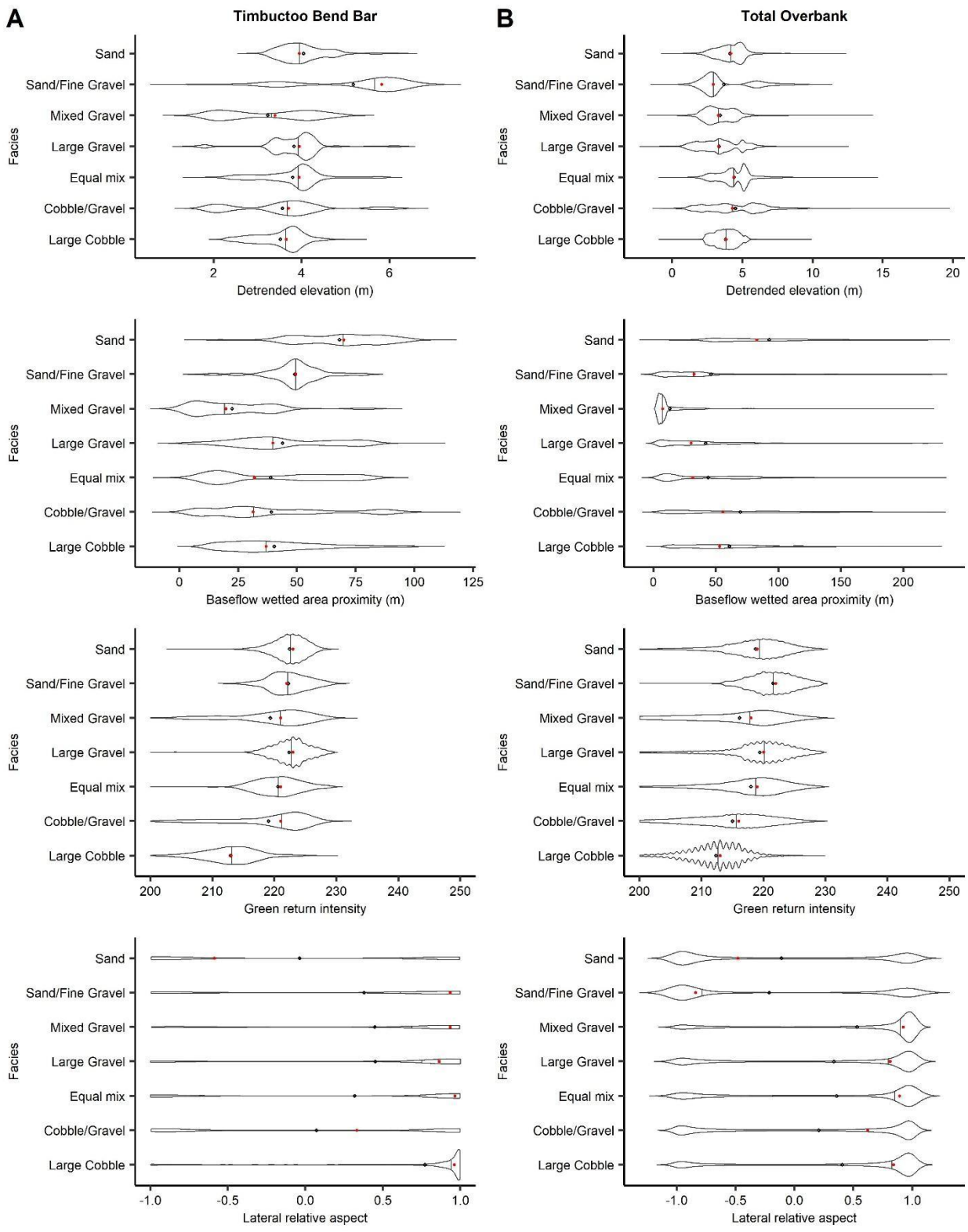

Fig. 16. Statistical distribution between facies clusters and the most important

predictors at A) the bar in Timbuctoo Bend and B) total overbank area in LYR. Red dots indicate mean values.

5. Discussion

In this study, we set out to determine if one could accurately predict the roughly meter-resolution spatial pattern of subaerial substrate facies in a large gravel-cobble river knowing only what is obtained from an airborne lidar campaign and field training data. The source lidar point cloud did not resolve individual sand, gravel, and cobble grains. Nevertheless, our results suggest that, yes, airborne lidar can detect and map substrate facies with sufficient accuracy for many real-world and scientific applications. The extensive model validation and analysis of facies patterns with respect to important predictors is discussed. Then, substrate mapping workflow challenges are described. Finally, application opportunities are highlighted.

5.1. Gravel-cobble river substrate facies model

A subaerial facies map was developed along 37.5 km of the LYR at a spatial resolution of 1.54 m, thereby distinguishing seven facies clusters using airborne lidar predictors and ML, achieving 86% cross-validation accuracy. RF delivered strong performance, in line with similar multi-classification learning tasks showing that RF is a powerful ML method (Maxwell et al., 2018; Zhang et al., 2020). Currently, no similar models exist for fluvial systems for mixed-size facies.

Large gravel, cobble-gravel, and large cobble with an abundance of gravel-cobble were accurately predicted by the model (Table 3). Sand/fine gravel and sand showed the highest accuracy, which might be because facies with sand abundance were under-sampled because of a limited presence in the river, thus overestimating the accuracy level. However, based on a visual assessment, finer facies are coherently predicted (e.g., examples in Fig. 14). The predictive model demonstrated low sensitivity for mixed facies, both equal mix and mixed gravel, which might be caused by unclear boundaries and the limited amount of training data for mixed gravel facies, thereby limiting the algorithm's learning from the predictors. Overall, the developed model succeeds in identifying and mapping facies.

In environmental sciences, clustering analysis has been broadly used to group objects (Jain et al., 1999). In river substrate studies, cluster analysis has been applied to delineate riverbed patchiness in flume experiments (Nelson et al., 2014). In this study, cluster analysis was used to identify six facies classes for dry terrain on the basis of GSDs by grid sampling methods. Incorporating a mixture approach that includes four bins of grain size class as the structure variables and considering fish habitat requirement through an eco-based approach maximizes the usefulness in distinguishing meaningful facies classes. The presented approach is beyond traditional ternary classification, where mean grain size quantifications use only the abundance of a specific grain size (Buffington and Montgomery, 1999). Consequently, eco-facies could be determined, including bin thresholds that can be adjusted by particular requirements. The main advantage of the proposed method is that it can provide insight in quantitative facies spatial patterns analysis, diversity, and connectivity – critical in habitat

restoration. While this study investigated the utility of aerial near-infrared lidar observing dry terrain with a sampling density of ~ 32 pts/m^2, the same framework should also currently work for submerged riverbed data collected with multibeam echosounders (a technology with far higher point density) where they can be used. When green lidar (currently sampling ~ 8 pts/m^2) achieves a similar point density as current near-infrared lidar, then the framework should apply to that data as well, though limited to relatively shallow and clear water that green light can penetrate.

5.2. Laser intensity and topographic predictors influencing facies patterns

The variable importance in analysis of the RF model indicated that detrended elevation, proximity to the baseflow wetted area, and green lidar return intensity are the most useful facies predictors. Moreover, topographic roughness and lateral and longitudinal relative aspect predictors explained facies distribution. It is important to understand the mechanistic drivers of these statistical outcomes. Detrended elevation was identified as the main driver in predicting the spatial distribution of facies classes. The facies patterns analyzed showed that detrended elevation varies among facies, with sand facies found 0.60 – 1.2 m (2-4 ft) higher compared to the other facies at the overbank and bar scales in the LYR. These patterns are consistent with the literature that low elevations with more dynamic and energetic fluid mechanics have coarser facies, where sediment transport processes are different from those at higher elevation (Rice and Church, 1998). Facies patterns and sorting are linked to elevation variability and are associated with hydraulic variables and topography (Powell, 1998; Thonon et al., 2007; Wittenberg, 2002). Simultaneously, facies are influenced by vegetation at different spatial scales

(McKenney et al., 1995). This interaction was coherently observed on the substrate map, where shrubs dissipate energy and generate the accumulation of sand behind them.

Proximity to the baseflow wetted area was revealed as the second-most important predictor to determine facies distribution. The facies spatial pattern was such that coarser facies are closer to the wetted area than finer facies for both the whole river and intra-bar variability. This accords with Asselman (1999) and He and Walling (1998). These studies observed that distance to the river controls the grain size patterns on overbank floodplain deposits and depend on flow frequency. Because flow velocity decreases over the floodplain, finer sediment deposits there, while velocity is high near to the channel and can transport coarser sediment. Similarly, Groom et al. (2018) found the same distance signal consistent with lateral accretion at the bar scale in gravel rivers using photogrammetry. Consequently, the developed substrate predictor model coherently captured the facies spatial patterns signals, enabled substrate mapping along the 37.5 km, and provided insight into floodplain sediment transport processes over a large area.

Green lidar return intensity was the third useful predictor of facies distribution. These values are higher on finer facies. Recently, Near-Infrared has also been linked to grain size distribution at the bar scale (Chardon et al. 2020). However, this study found that Near-Infrared variability was not as important in predicting facies distribution at a large scale. This suggests that there could be important benefits to multi- and hyper-spectral lidar technology for understanding substrate facies and fluvial geomorphology as a

whole (Kaasalainen et al., 2009; Shao et al., 2019), rather than resting with Near-Infrared alone. The developed model is highly sensitive in predicting large cobble facies with an abundance of large cobble as well as facies with an abundance of fine gravel and sand. This may be explained by the fact that green lidar return intensity captures the reflectance variability of the grain size range in every facies. The level of detection of the grain size range in this model ranges from 2 to 256 mm. Including green and Near-Infrared intensity provided insight in distinguishing facies in overbank areas. Also, return intensity can be affected by petrographic characteristics of the grains (Burton et al., 2011; Chardon et al., 2020; Kashani et al., 2015). Likely, facies with a marked difference in lithological composition could be mapped with greater accuracy.

Previous studies have found topographic roughness to be a direct grain size proxy, although at high spatial resolution with data collected from a UAV, in flume experiments, or with the point sampling method in the field (Bertin et al., 2017; Gomez, 1993; Heritage and Milan, 2009; Pearson et al., 2017; Smart et al., 2004; Vázquez-Tarrío et al., 2019). Our results reveal that DEM-derived topographic variability predictors have high predictive power to distinguish facies distribution in the LYR. Specifically, the topographic variability, the amount of elevation difference in the terrain, was captured by the Terrain Ruggedness Index. For example, a facies cluster with a wide range of grain sizes and shapes generates micro-topographic altitude differences at the pixel scale that can be captured by topographic variability at the same scale. The longitudinal relative aspect predictor indicating if the pixel is facing upstream or downstream showed slightly different patterns among the facies at the bar and overbank scales. However, the substrate model visually indicates that coarser facies are distributed against the flow

direction on the head of the bar and there is a reduction of the grain size in accordance with other field studies (Rice and Church, 2010).

The lateral relative aspect emerged as the fourth most important variable. The lateral relative aspect predicted power is linked with the effect of hydraulic processes and morphodynamic pattern described above. The facies spatial distributions revealed that the model correctly maps coarser facies facing towards the channel, supporting the hydraulic influence, while sand/fine gravel and sand facies face away from the channel. The lateral variability pattern is captured at the overbank and bar scales. This finding might support the concept that the mobility of larger grains is negatively related to the distance from the thalweg because of high flow velocity and flood-driven elevation change (Thonon et al., 2007). This provides insight to the hydrodynamic influence on sediment transport and it is evident that hydraulic characteristics, such as flow velocity and water depth, are strong predictors. However, our model excluded potential hydraulic predictors, avoiding the need for hydraulic models. The proposed method allowed the generation of a 37.5 km continuous facies map, exceeding the extension of existing substrate maps and mapping methods. Understanding the connection between predictors and facies could lead to new knowledge of facies patterns and connectivity with complex hydraulic, morphodynamic, and morphological processes in fluvial environments.

5.3. Key challenges and opportunities

We found multiple challenges in the implementation of the substrate mapping workflow. Generating facies labels for training was a key challenge that emerged during the data

collection and was related to maintaining representative and balanced training data. This step is where the most expert knowledge is required. Determining the plot sample size and plot spacing needs can be adjusted based on the total area to be mapped, by considering different river scales (reach, segment, morphological unit), and by measuring the maximum grain size that can be found in the river. The grain size data may be obtained with sieves and/or semiautomatic methods, but the mixture abundance approach is challenging. Moreover, UAV data can be used to generate facies from pure sand as training data. Yet, the main novel finding in distinguishing facies is the evidence that the mixture and eco-based approach enhances the cluster analysis method to distinguish representative facies clusters.

Hydrodynamic variables for specific discharges (e.g., velocity and depth) are potentially strong predictors. However, the model application might be limited to areas with multi-dimensional numerical simulations available. Predictors derived from multispectral or hyperspectral sensors may increase model accuracy while reducing the interpretability level. Developing interpretable models with predictors that can provide insight into facies distribution is key.

Temporal challenges emerged from the need to map broad areas to conduct field data collection at the same time, or at least close to the time when the data acquisition was carried out by the remote sensor. Many factors can change a river's substrate, such as floods and human activities between the field collection dates and the predictors' acquisition date used in the model. Training data and predictors need to be as close in time as possible.

The developed substrate mapping workflow constitutes an important step forward for river science, because substrate data is currently one of the most limited types of information derived from remote sensing. Notably, the LYR is only lightly vegetated and the climate (including its disturbance flow regime) supports a large area of subaerial sediment. Nevertheless, if this method were applied to similar exposed rivers globally, it would likely yield novel and transformative geomorphic understanding.

The approach presented here can improve change detection studies as the frequency of remote sensing missions increases. Having a continuous substrate map before and after flood events would provide insight in sediment transport processes using a quantitative abundance metric for every facies. Furthermore, quantitative facies metrics can identify pattern transitions and enhance ecohydraulics analysis needed for restoration projects. This novel approach can enhance the speed of substrate monitoring needed in flood event analysis, ecohydraulic river design, and river management.

6. Conclusions

Combining airborne lidar with ML is an efficient strategy for large scale predictions of substrate facies distribution. A continuous substrate map spatial of seven facies clusters was developed at the 1.54 m pixel scale with high accuracy (86%). This study found that the mixture approach captured the substrate facies variability, while considering the sediment's mixed nature enabled a better understanding of the facies spatial pattern over 37.5 km of the Lower Yuba River. Six substrate facies clusters were identified using the novel eco-mixture approach, namely three fine grain substrate facies that

were categorized based on the abundance proportion of gravel (fine gravel, mixed gravel, and equal mix) and three substrate facies with larger grain sizes (large cobble, cobble-gravel, and large gravel). Field data collection combined with UAV data are thus shown to be reliable methods to identify facies across large areas providing opportunities for fluvial morphology research and environmental management.

Acknowledgement

Primary support for this project was provided by Yuba Water Agency (Award #201016094). This project was also supported by the USDA National Institute of Food and Agriculture, Hatch project number #CA-D-LAW-7034-H. We thank Duane Massa and the PSMFC field crew for their efforts in helping to collect and process substrate data. We also thank Owen Ransom and Sierra Overhead Analytics for the UAV area imagery as well as those from our lab group who helped with fieldwork: John Dean, Arielle Gervassi, and Peter Moniz.

1. Declaration of competing interests

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

2. References

Adler, L. L., 1980. *Adjustment of the Yuba River, California, to the influx of hydraulic mining debris, 1849-1979.* Master Thesis. Univeristy of California, Los Angeles.

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

An, C., Fu, X., Wang, G., Parker, G., 2017. Effect of grain sorting on gravel bed river evolution subject to cycled hydrographs: Bed load sheets and breakdown of the hydrograph boundary layer. J. Geophys. Res. Earth Surf. 122, 1513–1533. https://doi.org/10.1002/2016JF003994

Arif, M.S.M., Gülch, E., Tuhtan, J.A., Thumser, P., Haas, C., 2017. An investigation of image processing techniques for substrate classification based on dominant grain size using RGB images from UAV. Int. J. Remote Sens. 38, 2639–2661. https://doi.org/10.1080/01431161.2016.1249309

Asselman, N.E.M., 1999. Grain-size trends used to assess the effective discharge for floodplain sedimentation, River Waal, the Netherlands. J. Sediment. Res. 69, 51–61. https://doi.org/10.2110/jsr.69.51

Bae, S., Yu, J., Wang, L., Jeong, Y., Kim, J., Yang, D.Y., 2019. Experimental analysis of sand grain size mapping using UAV remote sensing. Remote Sens. Lett. 10, 893–902. https://doi.org/10.1080/2150704X.2019.1629713

Barker, J.R., Pasternack, G.B., Bratovich, P.M., Massa, D.A., Wyrick, J.R., Johnson, T.R., 2018. Kayak drifter surface velocity observation for 2D hydraulic model validation. River Res. Appl. 34, 124–134. https://doi.org/10.1002/rra.3238

Barnard, P.L., Rubin, D.M., Harney, J., Mustain, N., 2007. Field test comparison of an autocorrelation technique for determining grain size using a digital 'beachball' camera versus traditional methods. Sediment. Geol. 201, 180–195. https://doi.org/10.1016/j.sedgeo.2007.05.016

Bertin, S., Friedrich, H., 2016. Field application of close-range digital photogrammetry (CRDP) for grain-scale fluvial morphology studies. Earth Surf. Process. Landforms 41, 1358–1369. https://doi.org/10.1002/esp.3906

Bertin, S., Groom, J., Friedrich, H., 2017. Isolating roughness scales of gravel-bed patches. Water Resour. Res. 53, 6841–6856. https://doi.org/10.1002/2016WR020205

Brasington, J., Vericat, D., Rychkov, I., 2012. Modeling river bed morphology, roughness, and surface sedimentology using high resolution terrestrial laser scanning. Water Resour. Res. 48, 1–18. https://doi.org/10.1029/2012WR012223

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brierly, G.J., 1991. Floodplain sedimentology of the Squamish River, British Columbia: relevance of element analysis. Sedimentology 38, 735–750.

https://doi.org/10.1111/j.1365-3091.1991.tb01017.x

Buffington, J.M., Montgomery, D.R., 1999. A Procedure for classifying textural facies in gravel-bed rivers. Water Resour. Res. 35, 1903–1914. https://doi.org/10.1029/1999WR900041

Bunte, K., Abt, S.R., 2001. Sampling Surface and Subsurface Particle-Size Distributions in Wadable Gravel-and Cobble-Bed Streams for Analyses in Sediment Transport, Hydraulics, and Streambed Monitoring.

Burman, P., 1989. A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods. Biometrika 76, 503. https://doi.org/10.2307/2336116

Burton, D., Dunlap, D.B., Wood, L.J., Flaig, P.P., 2011. Lidar Intensity as a Remote Sensor of Rock Properties. J. Sediment. Res. 81, 339–347. https://doi.org/10.2110/jsr.2011.31

Buscombe, D., 2020. SediNet: a configurable deep learning model for mixed qualitative and quantitative optical granulometry. Earth Surf. Process. Landforms 45, 638–651. https://doi.org/10.1002/esp.4760

Byrne, C.F., Pasternack, G.B., Guillon, H., Lane, B.A., Sandoval-Solis, S., 2020. Reach-scale bankfull channel types can exist independently of catchment hydrology. Earth Surf. Process. Landforms 45, 2179–2200. https://doi.org/10.1002/esp.4874

Carbonneau, P.E., Belletti, B., Micotti, M., Lastoria, B., Casaioli, M., Mariani, S., Marchetti, G., Bizzi, S., 2020. UAV-based training for fully fuzzy classification of Sentinel-2 fluvial scenes, Earth Surface Processes and Landforms. https://doi.org/10.1002/esp.4955

Carbonneau, P.E., Bergeron, N., Lane, S.N., 2005a. Automated grain size measurements from airborne remote sensing for long profile measurements of fluvial grain sizes. Water Resour. Res. 41, 1–9. https://doi.org/10.1029/2005WR003994

Carbonneau, P.E., Bergeron, N., Lane, S.N., 2005b. Automated grain size measurements from airborne remote sensing for long profile measurements of fluvial grain sizes. Water Resour. Res. 41, 1–9. https://doi.org/10.1029/2005WR003994

Carbonneau, P.E., Bizzi, S., Marchetti, G., 2018. Robotic photosieving from low-cost multirotor sUAS: a proof-of-concept. Earth Surf. Process. Landforms 43, 1160–1166. https://doi.org/10.1002/esp.4298

Carbonneau, P.E., Lane, S.N., Bergeron, N.E., 2004. Catchment-scale mapping of surface grain size in gravel bed rivers using airborne digital imagery. Water Resour. Res. 40. https://doi.org/10.1029/2003WR002759

Carley, J.K., Pasternack, G.B., Wyrick, J.R., Barker, J.R., Bratovich, P.M., Massa, D.A., Reedy, G.D., Johnson, T.R., 2012. Significant decadal channel change 58–67years 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

Carrivick, J.L., Smith, M.W., 2019. Fluvial and aquatic applications of Structure from Motion photogrammetry and unmanned aerial vehicle/drone technology. Wiley Interdiscip. Rev. Water 6, e1328. https://doi.org/10.1002/wat2.1328

Chanson, H., 2004. Hydraulics of open channel flow. Elsevier.

Chardon, V., Schmitt, L., Piégay, H., Lague, D., 2020. Use of terrestrial photosieving and airborne topographic LiDAR to assess bed grain size in large rivers: a study on the Rhine River. Earth Surf. Process. Landforms 45, 2314–2330. https://doi.org/10.1002/esp.4882

Chawla, N. V, Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Datta, S. and Datta, S.: Comparisons and validation of statistical clustering techniques for microarray gene expression data, Bioinformatics, 19, 459–466, 2003. https://doi.org/10.1093/bioinformatics/btg025

Deng, X., Liu, Q., Deng, Y., Mahadevan, S., 2016. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. Inf. Sci. (Ny). 340–341, 250–261. https://doi.org/10.1016/j.ins.2016.01.033

Detert, M., Weitbrecht, V., others, 2012. Automatic object detection to analyze the geometry of gravel grains--a free stand-alone tool, River flow.

Di Francesco, S., Biscarini, C., Manciola, P., 2016. Characterization of a Flood Event through a Sediment Analysis: The Tescio River Case Study. Water 8, 308. https://doi.org/10.3390/w8070308

Dietrich, J.T., 2016. Riverscape mapping with helicopter-based Structure-from-Motion photogrammetry. Geomorphology 252, 144–157. https://doi.org/10.1016/j.geomorph.2015.05.008

Dugdale, S.J., Carbonneau, P.E., Campbell, D., 2010. Aerial photosieving of exposed gravel bars for the rapid calibration of airborne grain size maps. Earth Surf. Process. Landforms 35, n/a-n/a. https://doi.org/10.1002/esp.1936

Escobar-Arias, M.I., Pasternack, G.B., 2010. A hydrogeomorphic dynamics approach to assess in-stream ecological functionality using the functional flows model, part 1-model characteristics. River Res. Appl. 26, 1103–1128. https://doi.org/10.1002/rra.1316

Evans, J.S., 2020. Package ‘ spatialEco .’

Fischer, P., 2000. An experimental test of metabolic and behavioural responses of benthic fish species to different types of substrate. Can. J. Fish. Aquat. Sci. 57, 2336–2344. https://doi.org/10.1139/f00-211

Fryirs, K.A., Brierley, G.J., 2012. Geomorphic analysis of river systems: an approach to reading the landscape. John Wiley & Sons.

Gilbert, G.K., 1917. Hydraulic-Mining Debris in the Sierra Nevada, U.S. Geological Survey Professional Paper 105. https://doi.org/10.3133/pp105

Gomez, B., 1993. Roughness of stable, armored gravel beds. Water Resour. Res. 29, 3631–3642. https://doi.org/10.1029/93WR01490

Graham, D.J., Rollet, A.J., Piégay, H., Rice, S.P., 2010. Maximizing the accuracy of image-based surface sediment sampling techniques. Water Resour. Res. 46, 1–15. https://doi.org/10.1029/2008WR006940

Groom, J., Bertin, S., Friedrich, H., 2018. Assessing Intra-Bar Variations in Grain Roughness Using Close-Range Photogrammetry. J. Sediment. Res. 88, 555–567. https://doi.org/10.2110/jsr.2018.30

Guillon, H., Byrne, C.F., Lane, B.A., Sandoval Solis, S., Pasternack, G.B., 2020. Machine Learning Predicts Reach-Scale Channel Types From Coarse-Scale Geospatial Data in a Large River Basin. Water Resour. Res. 56, 1–22. https://doi.org/10.1029/2019WR026691

Hauer, C., Leitner, P., Unfer, G., Pulg, U., Habersack, H., Graf, W., 2018. The role of sediment and sediment dynamics in the aquatic environment, in: Riverine Ecosystem Management. Springer, Cham, pp. 151–169.

Hawkins, D.M., Basak, S.C., Mills, D., 2003. Assessing Model Fit by Cross-Validation. J. Chem. Inf. Comput. Sci. 43, 579–586. https://doi.org/10.1021/ci025626i

He, Q., Walling, D.E., 1998. An investigation of the spatial variability of the grain size composition of floodplain sediments. Hydrol. Process. 12, 1079–1094. https://doi.org/10.1002/(SICI)1099-1085(19980615)12:7<1079::AID-HYP642>3.0.CO;2-E

Hemmelder, S., Marra, W., Markies, H., De Jong, S.M., 2018. Monitoring river morphology & bank erosion using UAV imagery – A case study of the river Buëch, Hautes-Alpes, France. Int. J. Appl. Earth Obs. Geoinf. https://doi.org/10.1016/j.jag.2018.07.016

Heritage, G.L., Milan, D.J., 2009. Terrestrial Laser Scanning of grain roughness in a gravel-bed river. Geomorphology 113, 4–11. https://doi.org/10.1016/j.geomorph.2009.03.021

Hijmans,, R. J., Van Etten, J., Gheng, J., Greenberg, J. A., Lamigueiro, O.P., Bevan, A., 2018. Package ‘raster’. version 2.6-7. Package raster.

Hodge, R., Brasington, J., Richards, K., 2009. Analysing laser-scanned digital terrain models of gravel bed surfaces: Linking morphology to sediment transport processes and hydraulics. Sedimentology 56, 2024–2043. https://doi.org/10.1111/j.1365-3091.2009.01068.x

Hou, Z., Scheibe, T.D., Murray, C.J., Perkins, W.A., Arntzen, E. V., Ren, H., Mackley, R.D., Richmond, M.C., 2019. Identification and mapping of riverbed sediment facies in the Columbia River through integration of field observations and numerical simulations. Hydrol. Process. 33, 1245–1259. https://doi.org/10.1002/hyp.13396

ISO, 2017. ISO 14688-1:2017 Geotechnical investigation and testing — Identification and classification of soil — Part 1: Identification and description, 2nd ed. ISO,pp.1-23.Jackson, J. R., Pasternack, G. B., Wyrick, J.R., 2013. Substrate of the Lower Yuba River. Prepared for the Yuba Accord River Management Team. Davis, CA, 61pp.

Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering. ACM Comput. Surv. 31, 264–323. https://doi.org/10.1145/331499.331504

James, L.A., Singer, M.B., Ghoshal, S., Megison, M., 2009. Historical channel changes in the lower Yuba and Feather Rivers, California: Long-term effects of contrasting river-management strategies, in: Management and Restoration of Fluvial Systems with Broad Historical Changes and Human Impacts. Geological Society of America Boulder, USA, pp. 57–81.

Kaasalainen, S., Hyyppä, H., Kukko, A., Litkey, P., Ahokas, E., Hyyppä, J., Lehner, H., Jaakkola, A., Suomalainen, J., Akujärvi, A., Kaasalainen, M., Pyysalo, U., 2009. Radiometric calibration of LIDAR intensity with commercially available reference targets. IEEE Trans. Geosci. Remote Sens. 47, 588–598. https://doi.org/10.1109/TGRS.2008.2003351

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

Kashani, A., Olsen, M., Parrish, C., Wilson, N., 2015. A Review of LIDAR Radiometric Processing: From Ad Hoc Intensity Correction to Rigorous Radiometric Calibration. Sensors 15, 28099–28128. https://doi.org/10.3390/s151128099

Kondolf, G.M., 2000. Assessing Salmonid Spawning Gravel Quality. Trans. Am. Fish. Soc. 129, 262–281. https://doi.org/10.1577/1548-8659(2000)129<0262:ASSGQ>2.0.CO;2

Kondolf, G.M., Wolman, M.G., 1993. The sizes of salmonid spawning gravels. Water Resour. Res. 29, 2275–2285. https://doi.org/10.1029/93WR00402

Kondolf, M., Li, S., 1992. The Pebble Count Technique for Quantifying Surface Bed Material Size in Instream Flow Studies Integrated analysis of Mediterranean climate dam-river-coast systems View project Apalachicola River Mussel Habitat and Geomorphic Assessment View project The Pe 3, 80–87.

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

Kuhn, M., others, 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.

Lane, B.A., Pasternack, G.B., Dahlke, H.E., Sandoval-Solis, S., 2017. The role of topographic variability in river channel classification. Prog. Phys. Geogr. Earth Environ. 41, 570–600. https://doi.org/10.1177/0309133317718133

Lang, M.W., Kim, V., McCarty, G.W., Li, X., Yeo, I.Y., Huang, C., Du, L., 2020. Improved detection of inundation below the forest canopy using normalized lidar intensity data. Remote Sens. 12. https://doi.org/10.3390/rs12040707

Langhammer, J., Lendzioch, T., Miřijovský, J., Hartvich, F., 2017. UAV-Based Optical Granulometry as Tool for Detecting Changes in Structure of Flood Depositions. Remote Sens. 9, 240. https://doi.org/10.3390/rs9030240

Langhammer, J., Vacková, T., 2018. Detection and Mapping of the Geomorphic Effects of Flooding Using UAV Photogrammetry. Pure Appl. Geophys. 175, 3223–3245. https://doi.org/10.1007/s00024-018-1874-1

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7, 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

Leduc, P., Peirce, S., Ashmore, P., 2019. Short communication: Challenges and applications of structure-from-motion photogrammetry in a physical model of a braided river. Earth Surf. Dyn. 7, 97–106. https://doi.org/10.5194/esurf-7-97-2019

Legleiter, C.J., Stegman, T.K., Overstreet, B.T., 2016. Spectrally based mapping of riverbed composition. Geomorphology 264, 61–79. https://doi.org/10.1016/j.geomorph.2016.04.006

Leopold, L.B., 1970. An Improved Method for Size Distribution of Stream Bed Gravel. Water Resour. Res. 6, 1357–1366. https://doi.org/10.1029/WR006i005p01357

Liu, H., Xu, K., Li, B., Han, Y., Li, G., 2019. Sediment Identification Using Machine Learning Classifiers in a Mixed-Texture Dredge Pit of Louisiana Shelf for Coastal Restoration. Water 11, 1257. https://doi.org/10.3390/w11061257

Marcus, W.A., Ladd, S.C., Stoughton, J.A., Stock, J.W., 1995. Pebble counts and the role of user-dependent bias in documenting sediment size distributions. Water Resour. Res. 31, 2625–2631.

Matsumoto, H., Young, A., 2018. Automated Cobble Mapping of a Mixed Sand-Cobble Beach Using a Mobile LiDAR System. Remote Sens. 10, 1253. https://doi.org/10.3390/rs10081253

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. Int. J. Remote Sens. 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

McKenney, R., Jacobson, R.B., Wertheimer, R.C., 1995. Woody vegetation and channel morphogenesis in low-gradient, gravel-bed streams in the Ozark Plateaus, Missouri and Arkansas. Geomorphology 13, 175–198. https://doi.org/10.1016/0169-555X(95)00034-3

Millard, K., Richardson, M., 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. Remote Sens. 7, 8489–8515. https://doi.org/10.3390/rs70708489

Misiuk, B., Diesing, M., Aitken, A., Brown, C.J., Edinger, E.N., Bell, T., 2019. A spatially explicit comparison of quantitative and categorical modelling approaches for mapping seabed sediments using random forest. Geosci. 9, 1–34. https://doi.org/10.3390/geosciences9060254

Moniz, P.J., Pasternack, G.B., Massa, D.A., Stearman, L.W., Bratovich, P.M., 2019. Do rearing salmonids predictably occupy physical microhabitat? J. Ecohydraulics 0, 1–19. https://doi.org/10.1080/24705357.2019.1696717

Murtagh, F., Legendre, P., 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J. Classif. 31, 274–295. https://doi.org/10.1007/s00357-014-9161-z

Nelson, P.A., Bellugi, D., Dietrich, W.E., 2014. Delineation of river bed-surface patches by clustering high-resolution spatial grain size data. Geomorphology 205, 102–119. https://doi.org/10.1016/j.geomorph.2012.06.008

Nelson, P.A., Dietrich, W.E., Venditti, J.G., 2010. Bed topography and the development of forced bed surface patches. J. Geophys. Res. Earth Surf. 115, 1–19. https://doi.org/10.1029/2010JF001747

Nelson, P.A., McDonald, R.R., Nelson, J.M., Dietrich, W.E., 2015. Coevolution of bed surface patchiness and channel morphology: 1. Mechanisms of forced patch formation. J. Geophys. Res. F Earth Surf. 120, 1687–1707. https://doi.org/10.1002/2014JF003428

Nelson, P.A., Venditti, J.G., Dietrich, W.E., Kirchner, J.W., Ikeda, H., Iseya, F., Sklar,

L.S., 2009. Response of bed surface patchiness to reductions in sediment supply. J. Geophys. Res. Earth Surf. 114, 1–18. https://doi.org/10.1029/2008JF001144

Neverman, A.J., Fuller, I.C., Procter, J.N., Death, R.G., 2019. Terrestrial laser scanning and structure-from-motion photogrammetry concordance analysis for describing the surface layer of gravel beds. Prog. Phys. Geogr. Earth Environ. 43, 260–281. https://doi.org/10.1177/0309133318822966

Pasternack, G.B., 2019. Natural Fluvial Ecohydraulics, Environmental Science. New York: Oxford University Press, New York. https://doi.org/10.1093/obo/9780199363445-0111

Pasternack, G.B., 2008. SHIRA-Based River analysis and field-based manipulative sediment transport experiments to balance habitat and geomorphic goals on the lower Yuba River. Cooperative Ecosystems Studies Unit (CESU) 81332 6 J002 Final Report.

Pasternack, G.B., Baig, D., Weber, M.D., Brown, R.A., 2018. Hierarchically nested river landform sequences. Part 1: Theory. Earth Surf. Process. Landforms 43, 2510–2518. https://doi.org/10.1002/esp.4411

Pasternack, G.B., Tu, D., Wyrick, J.R., 2013. Chinook adult salmon spawning physical habitat of the lower Yuba River. Prep. Low. Yuba River Accord River Manag. Team, Univ. California, Davis, CA (Available www. yubaaccordrmt. com).

Pearson, E., Smith, M.W., Klaar, M.J., Brown, L.E., 2017. Can high resolution 3D topographic surveys provide reliable grain size estimates in gravel bed rivers? Geomorphology 293, 143–155. https://doi.org/10.1016/j.geomorph.2017.05.015

Piégay, H., Arnaud, F., Belletti, B., Bertrand, M., Bizzi, S., Carbonneau, P., Dufour, S., Liébault, F., Ruiz-Villanueva, V., Slater, L., 2020. Remotely sensed rivers in the Anthropocene: state of the art and prospects. Earth Surf. Process. Landforms 45, 157–188. https://doi.org/10.1002/esp.4787

Powell, D.M., 1998. Patterns and processes of sediment sorting in gravel-bed rivers. Prog. Phys. Geogr. Earth Environ. 22, 1–32. https://doi.org/10.1177/030913339802200101

Probst, P., Wright, M.N., Boulesteix, A.L., 2018. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. https://doi.org/10.1002/widm.1301

Purinton, B., Bookhagen, B., 2019. Introducing PebbleCounts: A grain-sizing tool for photo surveys of dynamic gravel-bed rivers. Earth Surf. Dyn. Discuss. 1–33. https://doi.org/10.5194/esurf-2019-20

Quantum Spatial, 2017. Yuba River, California Topobathymetric LiDAR Technical Data Report. Corvalis, OR.

Ramezan, C.A., Warner, T.A., Maxwell, A.E., 2019. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. Remote Sens. 11. https://doi.org/10.3390/rs11020185

Rice, S., Church, M., 1998. Grain size along two gravel-bed rivers: statistical variation, spatial pattern and sedimentary links. Earth Surf. Process. Landforms 23, 345–363. https://doi.org/10.1002/(SICI)1096-9837(199804)23:4<345::AID-ESP850>3.0.CO;2-B

Rice, S.P., Church, M., 2010. Grain-size sorting within river bars in relation to downstream fining along a wandering channel. Sedimentology 57, 232–251. https://doi.org/10.1111/j.1365-3091.2009.01108.x

Rusnák, M., Sládek, J., Kidová, A., Lehotský, M., 2018. Template for high-resolution river landscape mapping using UAV technology. Meas. J. Int. Meas. Confed. 115, 139–151. https://doi.org/10.1016/j.measurement.2017.10.023

Shao, H., Chen, Y., Yang, Z., Jiang, C., Li, W., Wu, H., Wang, S., Yang, F., Chen, J., Puttonen, E., Hyyppä, J., 2019. Feasibility Study on Hyperspectral LiDAR for Ancient Huizhou-Style Architecture Preservation. Remote Sens. 12, 88. https://doi.org/10.3390/rs12010088

Sierra Overhead Analitycs, 2018. Low Yuba River, UAV proccesing report. Sonora.

Silva, P. V. R. M. and Pasternack, G.B., 2018. 2017 Lower Yuba River Topographic Mapping Report Nov. Prepared for:Yuba Water Agency. Davis, CA.

Škarpich, V., Galia, T., Ruman, S., Máčka, Z., 2019. Variations in bar material grain-size and hydraulic conditions of managed and re-naturalized reaches of the gravel-bed Bečva River (Czech Republic). Sci. Total Environ. 649, 672–685. https://doi.org/10.1016/j.scitotenv.2018.08.329

Smart, G., Aberle, J., Duncan, M., Walsh, J., 2004. Measurement and analysis of alluvial bed roughness. J. Hydraul. Res. 42, 227–237. https://doi.org/10.1080/00221686.2004.9728388

Sofia, G., 2020. Combining geomorphometry, feature extraction techniques and Earth-surface processes research: The way forward. Geomorphology 355, 107055. https://doi.org/10.1016/j.geomorph.2020.107055

Thonon, I., Middelkoop, H., van der Perk, M., 2007. The influence of floodplain morphology and river works on spatial patterns of overbank deposition. Netherlands J. Geosci. - Geol. en Mijnb. 86, 63–75. https://doi.org/10.1017/S0016774600021326

Tomsett, C., Leyland, J., 2019. Remote sensing of river corridors: A review of current trends and future directions. River Res. Appl. 35, 779–803. https://doi.org/10.1002/rra.3479

Vázquez-Tarrío, D., Borgniet, L., Liébault, F., Recking, A., 2017. Using UAS optical imagery and SfM photogrammetry to characterize the surface grain size of gravel bars in a braided river (Vénéon River, French Alps). Geomorphology 285, 94–105. https://doi.org/10.1016/j.geomorph.2017.01.039

Vázquez-Tarrío, D., Recking, A., Liébault, F., Tal, M., Menéndez-Duarte, R., 2019. Particle transport in gravel-bed rivers: Revisiting passive tracer data. Earth Surf. Process. Landforms 44, 112–128. https://doi.org/10.1002/esp.4484

Venditti, J.G., Nelson, P.A., Bradley, R.W., Haught, D., Gitto, A.B., 2017. Bedforms, structures, patches, and sediment supply in gravel-bed rivers. Gravel-Bed Rivers Process Disasters 439–466. https://doi.org/10.1002/9781118971437.ch16

Verdú, J.M., Batalla, R.J., Martínez-Casasnovas, J.A., 2005. High-resolution grain-size characterisation of gravel bars using imagery analysis and geo-statistics. Geomorphology 72, 73–93. https://doi.org/10.1016/j.geomorph.2005.04.015

Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc. 58, 236–244. https://doi.org/10.1080/01621459.1963.10500845

Weitbrecht, Volker, Zurich, E., Detert, M., Weitbrecht, V, 2013. User guide to gravelometric image analysis by BASEGRAIN.

Wenthworth, C. K., 1922. A Scale of Grade and Class Terms for Clastic Sediments. J. Geol. 30 (5), 377–399. doi:https://doi.org/10.1086/622910.

Whitaker, A.C., Potts, D.F., 2007. Coarse bed load transport in an alluvial gravel bed stream, Dupuyer Creek, Montana. Earth Surf. Process. Landforms 32, 1984–2004. https://doi.org/10.1002/esp.1512

Wittenberg, L., 2002. Structural patterns in coarse gravelriver beds: typology, survey and assessment of the roles of grain size and river regime. Geogr. Ann. Ser. A, Phys. Geogr. 84, 25–37. https://doi.org/10.1111/j.0435-3676.2002.00159.x

Wolman, M.G., 1954. A method of sampling coarse river-bed material. EOS, Trans. Am. Geophys. Union 35, 951–956.

Woodget, A.S., Austrums, R., Maddock, I.P., Habit, E., 2017. Drones and digital photogrammetry: from classifications to continuums for monitoring river habitat and hydromorphology. Wiley Interdiscip. Rev. Water. https://doi.org/10.1002/wat2.1222

Woodget, A.S., Fyffe, C., Carbonneau, P.E., 2018a. From manned to unmanned aircraft: Adapting airborne particle size mapping methodologies to the characteristics of sUAS and SfM. Earth Surf. Process. Landforms 43, 857–870. https://doi.org/10.1002/esp.4285

Woodget, A.S., Fyffe, C., Carbonneau, P.E., 2018b. From manned to unmanned aircraft: Adapting airborne particle size mapping methodologies to the

characteristics of sUAS and SfM. Earth Surf. Process. Landforms 43, 857–870. https://doi.org/10.1002/esp.4285

Wu, F.-C., Wang, C.-K., Huang, G.-H., 2018. Delineation of gravel-bed clusters via factorial kriging. Geomorphology 308, 161–174. https://doi.org/10.1016/j.geomorph.2018.02.013

Wyrick, J. R. and Pasternack, G. B. 2012. Landforms of the Lower Yuba River. Prepared for the Yuba Accord River Management Team. University of California at Davis, Davis, CA, 91pp. DOI: 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

Wyrick, J.R., Senter, A.E., Pasternack, G.B., 2014. Revealing the natural complexity of fluvial morphology through 2D hydrodynamic delineation of river landforms. Geomorphology 210, 14–22. doi:10.1016/j.geomorph.2013.12.013.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. Hydrol. Earth Syst. Sci. 24, 2505–2526. https://doi.org/10.5194/hess-24-2505-2020

Supplementary Materials for

Mapping subaerial sand-gravel-cobble fluvial sediment facies using airborne lidar and machine learning

Correspondence to: rdiazgomez@ucdavis.edu

Methods

Developing Facies Predictors lidar DEM-derived

Although the factors that control and correlate with substrate facies can be conjectured, there is no a priori knowledge and certainty as to what the best substrate predictors would be for the LYR. As a result, the typical strategy in ML application is to extract all the potential predictors that can be surmised and computed as raster. this study possible substrate predictors that were generated included topographic and spectral lidar predictors in raster format, at 1.5 m pixel size resolution. Fifteen predictor layers were used in this model.

Topographic predictors

Topographic variables were selected as possible substrate predictors. A previous analysis of LYR grain size distributions (GSDs) revealed strong correlations between slope and mean grain size at the reach scale (Jackson et al., 2013). A reach-scale slope raster, derived from 2017 DEM (from lidar), represented the rate of change of elevation between one pixel and the eight pixels surrounding it. Inclination values were calculated in percent using the “slope” tool in ArcGIS.

Terrain variability can be analyzed based on the topographic index, leading to improved terrain characterization. The topographic roughness variable was calculated based on 2017 DEM, which is the difference between the maximum and minimum values of a pixel and its eight surrounding pixels. The terrain ruggedness index (TRI) was defined as the mean of the absolute differences between the value of a pixel and the values of its eight neighbors (the amount of elevation difference between adjacent pixels of a DEM). TRI provided a relative measure of the elevation changes. TRI values are lower in flatter areas and higher in steep or rugged areas. The topographic position index (TPI) is the difference between the value of a pixel and the mean value of its eight surrounding pixels. Positive TPI values represent locations that are higher than the average of their surroundings, as defined by the neighborhood (ridges). Negative TPI values represent locations that are lower than their surroundings (valleys). TPI values near zero are either flat areas (where the slope is near zero) or areas of constant slope. The topographic position is an inherent scale-dependent phenomenon. The topographic

terrain variables were generated according to Wilson et al. (2007) using the terrain function in the raster package in R (Hijmans et al., 2018). Also, a topographic flow direction raster was generated, representing the 'flow direction' (of water), i.e. the direction of the highest drop in elevation (or the smallest rise if all neighbors were higher).

Furthermore, the curvature topographic (concavity/convexity) index was also used, topographic curvature was computed according to NcNab (1993). This index was based on features that confine the view from the center of a 3x3 pixel window. Negative values in the profile curvature indicate that the surface is upwardly convex, whereas positive values indicate that the surface is upwardly concave. It was computed using the Spatial ECO package in R (Evans, 2020).

Topographic roughness at the pixel scale (derived from the point cloud) has a high correlation with grain size (Vázquez-Tarrío et al., 2017; Woodget and Austrums, 2017). For this reason, a pixel-scale topographic variability raster defined as the standard deviation of lidar points within one pixel, was used as a possible facies model predictor. It was computing the Standard Deviation for the 25 points that fall inside every 5x5 pixel window in the output raster using the Point Statistics tool from ArcGIS.

Although a river's slope is too low to be discernable at the scale of an individual pixel, it is large enough that the same local topographic feature (e.g. pool bottom, bar top, bank top) will have significantly different elevations down a river. Ideally, if the overall slope of the river could be removed, a computation called "detrending", then it might be possible to find that every type of feature has a common elevation. Because individual landscape features tend to be composed of the same material with the same GSD, knowing the relative elevation of a feature type could be a good predictor of substrate facies. As a result, a detrended raster DEM of the LYR was created via linear regression of the river channel's longitudinal profile using an open-source Python code developed for this purpose by the Pasternack Lab group and available online at: https://github.com/klarrieu/GCS_Scripts .The algorithm uses the theory and conceptual methods explained in Pasternack et al.(2018).

Another topographic variable that was generated for testing was longitudinal relative aspect, defined as the compass direction that a slope faces. However, these landform aspects are not too meaningful when provided as absolute compass directions, because rivers meander. The knowledge that a pixel faces due north does not inform about whether that landform is a riverbank, scroll bar top, or anything else. Meandering has to be taken into account. This is analogous to the need to detrend elevation to characterize the relative elevations of landforms along a river. Thus, the aim here is to obtain a landform-centric aspect that is not relative to due north like a compass but instead relative to whatever direction the river's centerline points in adjacent to every pixel of interest.

Riverbanks face the river, so the bank on every side of the river has an opposing aspect. Meanwhile, other landforms have characteristic relative aspects to the river. Simplifying aspect down to the four cardinal direction relative to the river's centerline a landform's aspect can be facing the centerline, facing away from it, or parallel to the

centerline direction facing upstream or downstream. As with elevation, it is possible that substrate facies vary with aspect because it is a feature of landform type, and landform type is related to substrate facies.

Adjusting aspect to a centerline requires a decision about what centerline to use to characterize a river's meandering and how to compute the relative aspect. In this study, the centerline used was the river's thalweg at 530 cfs. To avoid wildly changing angles arising at locations where the thalweg is highly tortuous due to local noise that is not real meandering, the thalweg pathway was smoothed using the ArcGIS "Smooth Line" function, applying the PAEK algorithm. The smoothing distance was 20'. Then, the thalweg was stationed at 5' intervals and the local thalweg angle was computed. The thalweg was stationed at 5' increments using the ETGeoWizards "Create Station Lines" function, which automatically creates an attribute in the XS lines with the angle of the thalweg. Thalweg intersection points were then created from the XS lines and the "thalweg angle" attribute was interpolated over the DEM area using the nearest neighbor method. ArcGIS's nearest neighbor algorithm was used to assign a thalweg angle to every raster pixel away from the thalweg. Lastly, the raster calculator was used in ArcGIS to calculate the relative aspect map using the dot product of the aspect and the nearest thalweg angle vectors as sin (thalweg angle) * sin(aspect) + cos(thalweg angle) * cos(aspect). The longitudinal relative aspect raster ranges between -1 and 1, with -1 indicating that the aspect is aligned parallel to the nearest thalweg point facing downstream, 0 indicating that the aspect is perpendicular to the thalweg, and 1 indicating that the aspect is parallel to the thalweg facing upstream.

Additionally, a lateral relative aspect was computed to distinguish if a perpendicular pixel is facing toward or away from the river. It was computed using the thalweg line to split a polygon of the DEM area in two and then applied the raster values for every side of the river separately. The lateral relative aspect raster ranges between -1 and 1, with -1 indicating that aspect is facing toward the river and 1 indicating that the aspect facing away from the river.

Distance from the river channel was correlated with grain size variability in gravel-bed floodplains. A baseflow wetted area proximity raster was generated based on the baseflow wetted area polygon and Euclidean distance using proximity tool of Arcgis. Every pixel indicates the distance from the channel in LYR.

Spectral predictors

For the LYR there were both green and NIR wavelengths, meaning there are two intensity spectral variables available (Figure S 2). How low or high the value of intensity is depends on the nature of the surface the light interacts with. For example, when NIR light hits water, most of it gets absorbed, returning a low-intensity value unless the surface of the water is rough (i.e., "rippled"), enabling light to reflect off it. That would yield a high-intensity value. Similarly, the thicker the vegetation is, the lower the intensity of light return will be for green lidar.

Green and NIR lidar intensity spectral datasets were available for the LYR as rasters. In this study, four such variables were extracted: NIR intensity, green intensity, mean and standard deviation of Lidar intensity. Std Lidar intensity and mean lidar intensity were calculated by combining the green and NIR intensity rasters using ArcGIS's raster calculator tool.

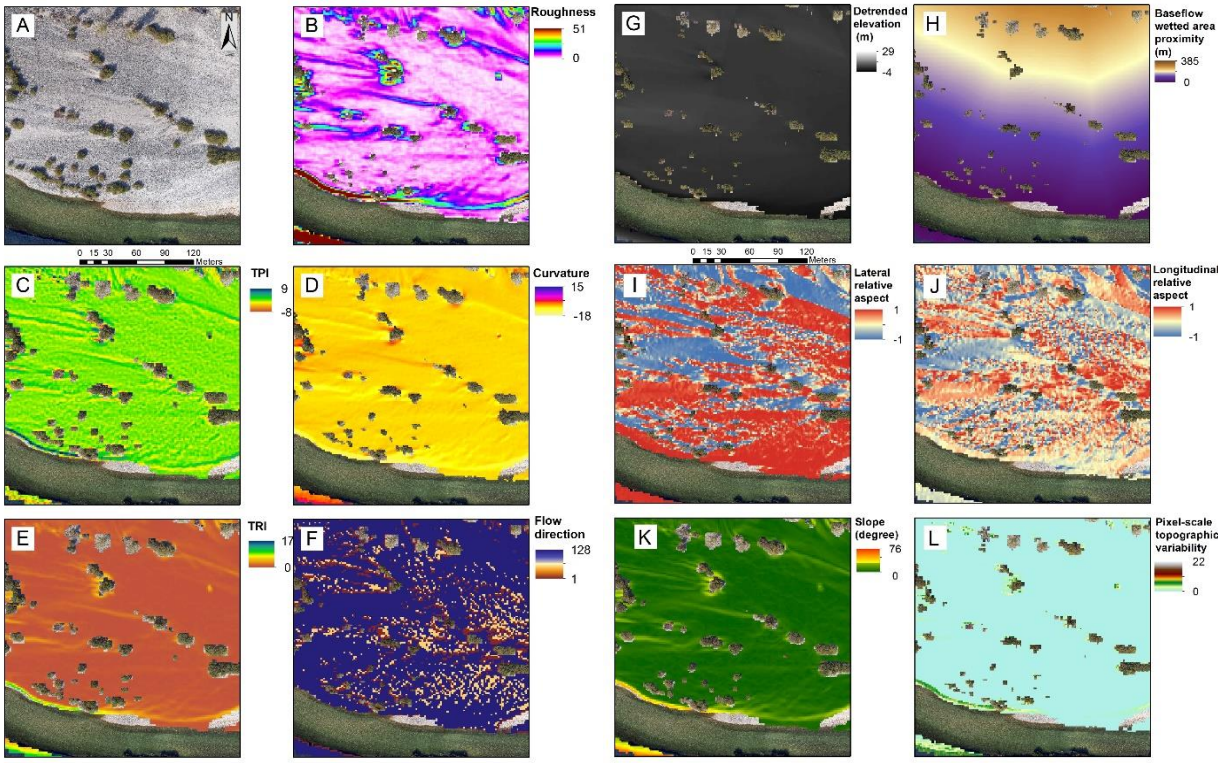

Figure S 1. Topographic lidar DEM derived predictors.

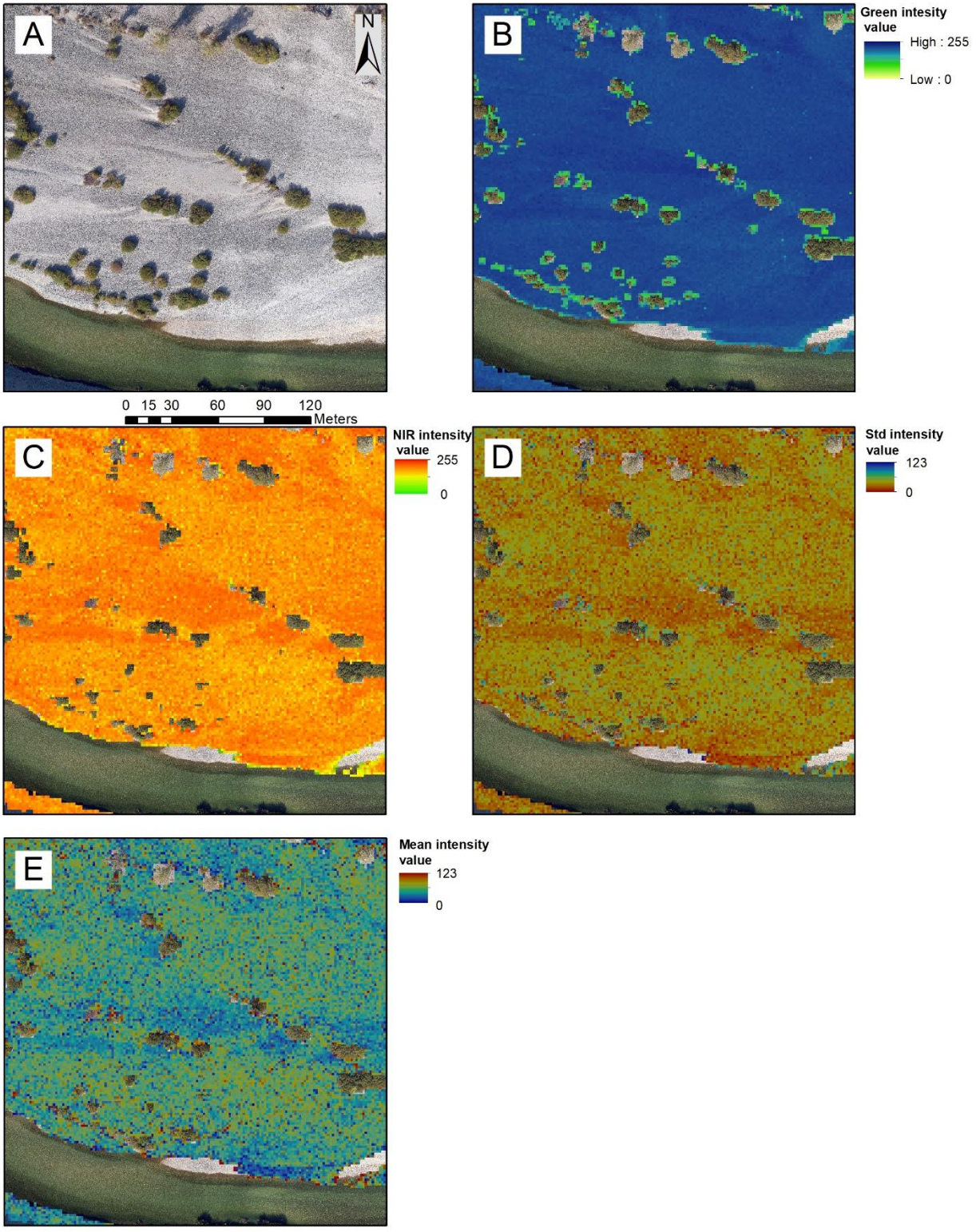

Figure S 2. Spectral predictors lidar-derived.