Geophysical Research Letters

Supporting Information for

Continental Scale Assessment of Variation in Floodplain Roughness with Vegetation and Flow Characteristics

Gabriel Barinas[1, 2], Stephen P. Good[1, 2], and Desiree Tullos[1, 2]

[1]Water Resources Graduate Program, Oregon State University
[2]Department of Biological & Ecological Engineering, Oregon State University

Contents of this file

Additional Supporting Information (Files uploaded separately at Barinas et al., 2023)

Dataset file *fp_mannings.csv*
Script files *dataset.py* and *figures.py*
Required files: *NHDFlowline.csv*, *measurements.csv*, *gedi_L3L4.csv*, *ModisLC*, *geosfm.csv and kalyanapu.csv*.

Extended Methodology S1

1. Floodplain roughness definition

Manning's equation (Manning, 1891) is extensively applied in hydraulic modeling and is written as:

$$Q = \frac{k}{n} S^{1/2} R^{2/3} A, \quad \text{(S1)}$$

where Q is discharge [L3 t-1], S is the friction slope, defining the energy loss along a reach [L L-1], R the hydraulic radius [L], A is cross-sectional area [L2], k is a unit conversion factor, and n is Manning's roughness coefficient. The coefficient n is a representation of the roughness of the surface over which water is flowing and incorporates surface characteristics such as smoothness, grain size, vegetation and/or obstructions (Chow, 1959).

Here we conceptualize the floodplain as a wide, rectangular, cross-sectional area (see Figure S1) and apply Manning's equation explicitly to the floodplain alone, separate from the main river channel. The floodplain discharge, Q_{fp}, is isolated by taking the total measured discharge, Q_t, and subtracting the discharge within the main channel, Q_{mc}. The width of water in the floodplain, $w_{fp} = w_t - w_{mc}$, is assumed to be much greater than flow depth, z_{fp}, and therefore the hydraulic radius of flow in the floodplain is approximately equal to the floodplain flow depth (Reclamation, 2001). Rearranging Manning's equation (eq. S1) for the floodplain and solving for floodplain roughness, n_{fp}, yields the following relationship:

$$n_{fp} = \frac{k\, w_{fp}\, z_{fp}^{5/3}\, S^{1/2}}{Q_t - Q_{mc}} \quad \text{(S2)}$$

2. Bankfull width estimation

Parameters corresponding to the total flow (Q_t, w_t) are collected during overbank discharge measurements made by the United States Geological Survey (USGS) at gauging stations. These total parameters were used to derive floodplain specific parameters necessary to solve for the floodplain roughness. At bankfull depth, z_{bf}, the width of the main channel, w_{mc}, is not specified or measured, and must be estimated by determining the cross-sectional geometry of the main channel and floodplain. A piecewise function based on measurements of w and z was used to determine w_{mc} , with the main channel depth assumed a power function of the width, i.e. $z \propto w$ (Durand et al., 2016), and the floodplain as sloping linearly away from bankfull stage. The piecewise form of w as a function of z was then expressed as:

$$w = \frac{(z - z_0)^{1/s}}{u} \quad \text{when } z < z_{bf} \quad \text{(S3a)}$$

$$w = m(z - z_{bf}) + \frac{(z_{bf} - z_0)^{1/s}}{u}, \qquad \text{when } z \geq z_{bf} \tag{S3b}$$

where m is the cross-sectional up-slope of the floodplain, u and s are parameters that define the shape of the main channel curvature, and z_0 defines its starting point.

Discharge in the main channel above flood stage was represented as a rectangular cross-section, as such flow in the main channel Q_{mc} should assumed to be proportional to $(z)^{5/3}$, following Manning's function (eq S1). When the flow reaches bankfull condition i.e. $z = z_{bf}$, $Q_{mc} = Q_{bf}$, and thus the constant of proportionality is equal to $Q_{bf}/{z_{bf}}^{5/3}$, and Manning's function (eq. S1) for the main channel flow above bankfull was simplified as:

$$Q_{mc} = \left(\frac{Q_{bf}}{{z_{bf}}^{5/3}}\right) z^{5/3} \tag{S4}$$

where Q_{bf} is the flow at flood stage given that Q_{mc}= Q_t when $z = z_{bf}$.

3. Data sources and quality control

Most of the parameters (Q_t, w_t, z) required for the calculation of Manning's n with Eq. S2-S4, were obtained from the field measurements datasets available from the USGS WaterData platform (USGS, 2021a). The WaterData platform is part of the USGS efforts to monitor, assess, and deliver information about streamflow quality, use and availability. The platform provides access to field measurements at nearly 73,000 sites under USGS management. Consistent with Slater et al., (2015), the flood stage height (z_{bf}) was obtained from the WaterWatch platform (NWS, 2021). These values were determined by the National Weather Service by defining the flood stage as the lowest bank at which inundation of the surrounding area begins to cause damage. Friction slope estimates were obtained from the National Hydrography Dataset (USGS, 2021b), a database of features that includes a drainage network of US waterbodies.

Quality control measures on the floodplain data involved multiple steps. The rating curves at USGS sites are regularly adjusted to adapt the relationship to geometry changes associated with erosion or deposition at a gauging location. To account for this effect, only river discharge and geometry measurements where the measured values were within 10% of the respective rating curve value were included in the analysis. Additionally, sites with a low number of measurements over the flood stage (< 3) were also removed. As a way of avoiding the calculation of n with measurements with a high level of uncertainty in the width-depth relationship (eq. 3), as evidenced by regression curves with a high root mean squared error (RMSE), only samples with width measurements higher than 1.96 times the RMSE of the fit in eq. 3 were considered (Liu, 2011). Furthermore, a large percentage of sites from the resulting dataset (27%) had a slope set at a value of 0.0001 within the National Hydrography Dataset, representing a

minimum fixed value within the database. Due to the high uncertainty and potential error from including the fixed minimum slopes, these 298 sites were also excluded from analysis.

In addition, the results of this work are subject to limitations of the USGS gaging network and to uncertainties inherent in gaging stochastic and modified systems. The lack of representation of certain geographic areas within the USGS gaging network have been reported elsewhere (Kiang et al., 2013), as have some of the drivers of temporal noise and uncertainty in streamflow over time (Tu et al., 2023). Application of the results outside the geographic areas and LULC (Land Use Land Cover) conditions from which these data were derived may generate uncertainties that we were unable to quantify with this analysis.

4. Remote sensed vegetation datasets

Flow in floodplains is expected to be strongly influenced by the vegetation in the floodplain (Box et al., 2021) but vegetation characteristics (density, height, etc.) are not typically measured in the field during flood conditions. Here we used aboveground biomass density, B [M L-2], and vegetation canopy height, h_{veg}, as characterized by the NASA Global Ecosystem Dynamics Investigation (GEDI). GEDI utilizes a full waveform Light Detection and Ranging (LiDAR) system to make measurements of vegetation structure at 25m resolution (Potapov et al., 2021), which are then aggregated to a 1km spatial resolution grid. In this work, for each USGS site, we obtained the canopy height estimates from the L3B version 2 gridded product (Dubayah et al., 2021) and the aboveground biomass estimates from the L4B version 2 gridded product (Dubayah et al., 2022).

Within GEDI's L4B dataset, there is a Prediction Stratum (PS) classification, determined by plant functional types described as: Deciduous Broadleaf Trees, Evergreen Broadleaf Trees, Evergreen Needleleaf Trees, Deciduous Needleleaf Trees, and Grasses, Shrubs, and Woodlands grouped as one class. This classification was used to categorize our dataset based on the level of biomass and canopy height by extracting the GEDI data from the pixel where each gauge location fell within (See Table S1). It is important to note that Gridded GEDI datasets, while providing unique information about vegetation height and biomass, is limited by its 1km resolution, capable of measuring only vegetation above a certain height.

5. Theoretical modeling

Prior research suggests that *n* is proportional to the square-root of the vegetation inundation fraction, i.e. $n \propto (z_{fp}/h_{veg})^{1/2}$, and that it is also related to flow velocity and vegetation properties (Kouwen & Fathi-Moghadam, 2000). A mathematical model by

Kouwen was based on data from four tree species:

$$n = 0.228\left(\frac{V}{\sqrt{\frac{\xi E}{\rho}}}\right)^{-0.23}\left(\frac{y_n}{h}\right)^{0.5} \quad \text{(eq. 5)}$$

where V is flow velocity, ξE is a vegetation index, ρ is the density of the fluid, and yn/h is the depth of submergence (z_{fp}/h_{veg}), and 0.228 and -0.23 empirically fit. Based on this approach, we formulate an analogous expression incorporating GEDI derived vegetation properties as $c = a_1 V^{a_2} B^{-a_3}$, where a_1, a_2 and a_3 are model parameters, B is aboveground biomass, V is flow velocity, and c is Manning's n normalized by the square root of vegetation inundation fraction, i.e. $c = n/(z_{fp}/h_{veg})^{1/2}$. To ensure positive c (and n) values, the linearized equation:

$$\ln(c) = a_1 + a_2 \ln(V) + a_3 \ln(B) \quad \text{(eq. 6)}$$

was fit to values of c, V, and B at USGS sites in our dataset to determine a_1, a_2 and a_3.

To limit the uncertainty caused by outliers in the dataset during the development of the model, the range of c values was restricted with the use of the interquartile range (IQR) (Vinutha et al., 2018). The minimum c value included in the analysis was the first quartile minus 1.5 times the IQR and the maximum value was the third quartile plus 1.5 times the IQR, where the IQR is equal to the difference between the third and the first quartile.

The developed function underwent cross-validation by splitting the USGS dataset, after quality control, into five randomized equal subsets, with each subset serving as validation during separate simulations. Subsequently, we combined all five validation subsets to create a comprehensive validation dataset that includes all of the original USGS gauge locations. This approach enables us to thoroughly assess the applicability and representativeness of our empirical function across the entire set of gauge locations. This new validation set was then used to compare the performance of our model against other works.

Figures S1 and S2

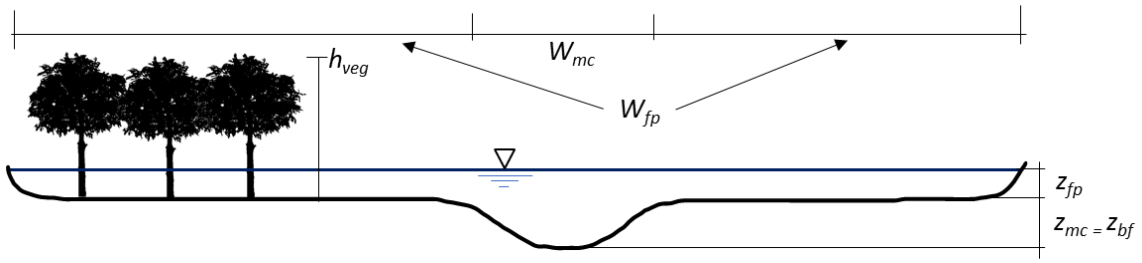

Figure S1 – Cross section diagram showing the variables used in the analysis. W_{mc} is the width of the main channel, W_{fp} is the width of the floodplain, Z_{mc} and Z_{bf} are the depth of the main channel during bankfull conditions, and h_{veg} is the height of the vegetation.

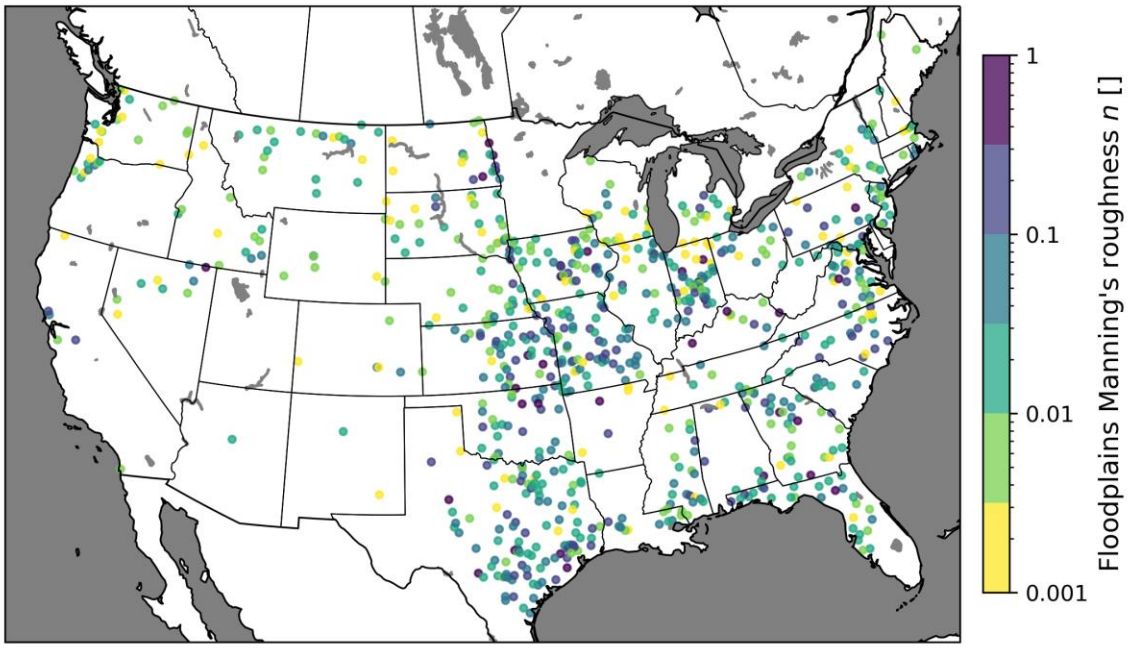

Figure S2 - Floodplain roughness (Manning's n) estimates for USGS sites.

Table S1

Mean floodplain Manning's n and Aboveground Biomass (B) [Mg/Ha] classified by tree structure and ranges of vegetation height. Values given are the median ± one median absolute deviation and the (samples count). The median vegetation height for these sites was 10m, while the 33rd and 66th percentiles were 7.4m and 13.5m. Values of 7.5m and 14m for vegetation height were selected for the ranges in order to have roughly the same number of total samples in each range.

Land Cover	**Floodplain Vegetation Biomass, *B* [Mg/Ha], floodplain *n,* and (sample count)**							
	***h*$_{Veg}$ <7.5m**		***h*$_{Veg}$ 7.5-14m**		***h*$_{Veg}$ >14m**		**All heights**	
	B	***n***	***B***	***n***	***B***	***n***	***B***	***n***
Deciduous Broadleaf Trees	25 ±6	0.023 ±0.022 (340)	63 ±20	0.026 ±0.022 (518)	130 ±29	0.026 ±0.020 (652)	77 ±45	0.025 ±0.022 (1514)
Evergreen Broadleaf Trees	37 ±0	0.005 ±0.004 (5)	50 ±0	0.025 ±0.022 (15)	98 ±28	0.032 ±0.022 (150)	95 ±25	0.030 ±0.022 (170)
Evergreen Needleleaf Trees	62 ±32	0.007 ±0.006 (10)	44 ±5	0.011 ±0.009 (37)	108 ±2	0.011 ±0.009 (79)	106 ±19	0.010 ±0.008 (126)
Grasses, Shrubs and Woodlands	12 ±8	0.012 ±0.010 (976)	28 ±16	0.038 ±0.034 (540)	51 ±2	0.022 ±0.013 (196)	18 ±12	0.017 ±0.014 (1737)
Unclassified	6 ±2	0.021 ±0.017 (141)	44 ±15	0.025 ±0.021 (264)	106 ±36	0.020 ±0.016 (279)	44 ±36	0.023 ±0.018 (787)
All GEDI land cover classes	15 ±10	0.014 ±0.012 (1472)	44 ±20	0.028 ±0.024 (1374)	110 ±38	0.023 ±0.017 (1356)	38 ±27	0.021 ±0.018 (4927)

Dataset

The dataset included as part of the Supplementary Information document is the result of the analysis that took place during this study. The dataset file named '*fp_mannings.csv*', consists of 4,927 calculations of Manning's n at each of the 804 USGS gauge sites that remained after quality control. The file also includes all variables collected and derived from USGS field measurements: discharge, width, depth (total, main channel, and floodplain), channel slope, site ID, coordinates, and number of estimates on that site.

Scripts

Included as supplementary information there are two scripts: *dataset.py* and *figures.py*.

The *dataset.py* script automates the process of calculating the Manning's *n* *fp_mannings.csv* file. It is divided into 3 sections: a SETUP section for module and file imports, a RUN section for defining the main function and running it for each state, and a MERGE section that puts together the results of each state into a single file. This script requires the *NHDFlowline.csv* and *measurements.csv* files which are included in Barinas et al., (2023). Other necessary files are downloaded automatically from USGS websites for each site: https://nwis.waterdata.usgs.gov/ for site coordinates; https://waterdata.usgs.gov/ for rating curves; and https://waterwatch.usgs.gov/ for flood stages.

The *figures.py* script creates the figures included in the main paper and in this document. This script is divided into 4 sections: IMPORTS loads the necessary modules and files required; MAP corresponds to Figure S2 in this document; MODEL corresponds to Figure 2 in the main manuscript; HEATMAP corresponds to figure 1 in the main manuscript, and VALIDATION corresponds to figure 3 in the main manuscript. This script requires remote sensed data by the Ecosystem Dynamics Investigation Mission GEDI. All GEDI data included in the *gedi_L3L4.csv* file refers to the pixel value where all USGS sites in the *fp_mannings.csv* file fell within and the data collected corresponds to the L3 and L4B version 2 gridded products (Dubayah et al., 2022). Finally, for the validation section of the script, the file *ModisLC.csv,* which contains the MODIS land cover classification corresponding to each gauge location, and the files *geosfm.csv* and *kalyanapu.csv*, which contains the values of *n* for each land cover type as presented in the original papers (Asante et al., 2008; Kalyanapu et al., 2009).