Predicting Reservoir Sedimentation and Capacity Loss Across the United States

Abigail Eckland, Intern, U.S. Bureau of Reclamation, Denver, Colorado, aeckland@usbr.gov
Melissa Foster, Geomorphologist, U.S. Bureau of Reclamation, Denver, Colorado, mfoster@usbr.gov
Aaron Hurst, Geomorphologist, U.S. Bureau of Reclamation, Denver, Colorado, ahurst@usbr.gov
Irina Overeem, Associate Professor, University of Colorado Boulder, Boulder, Colorado, irina.overeem@colorado.edu
Mussie Beyene, Hydrologic Engineer, U.S. Bureau of Reclamation, Denver, Colorado, mbeyene@usbr.gov

Extended Abstract

Dammed reservoirs across the United States are infilling with sediment, which reduces their water storage capacity and can hinder dam operations. However, at the majority of 90,000 reservoirs across the US, there are no bathymetric surveys to constrain the volume of sediment infill and remaining capacity. Our study compiles data relevant to surveyed reservoir sites and their upstream drainage basins to detect environmental and anthropogenic controls on sedimentation rates, to later predict infill of unsurveyed reservoir sites. We use repeat bathymetry data (first and last survey) at 535 sites to compute the reservoir sedimentation rate (SR) (Figure 1). SR (m^3 yr^{-1}) was calculated as:

$$SR = \frac{SV}{T} \tag{1}$$

where SV is the measured sediment volume deposited in the reservoir between first and last surveys (m^3) and T is the time between first and last surveys (years).

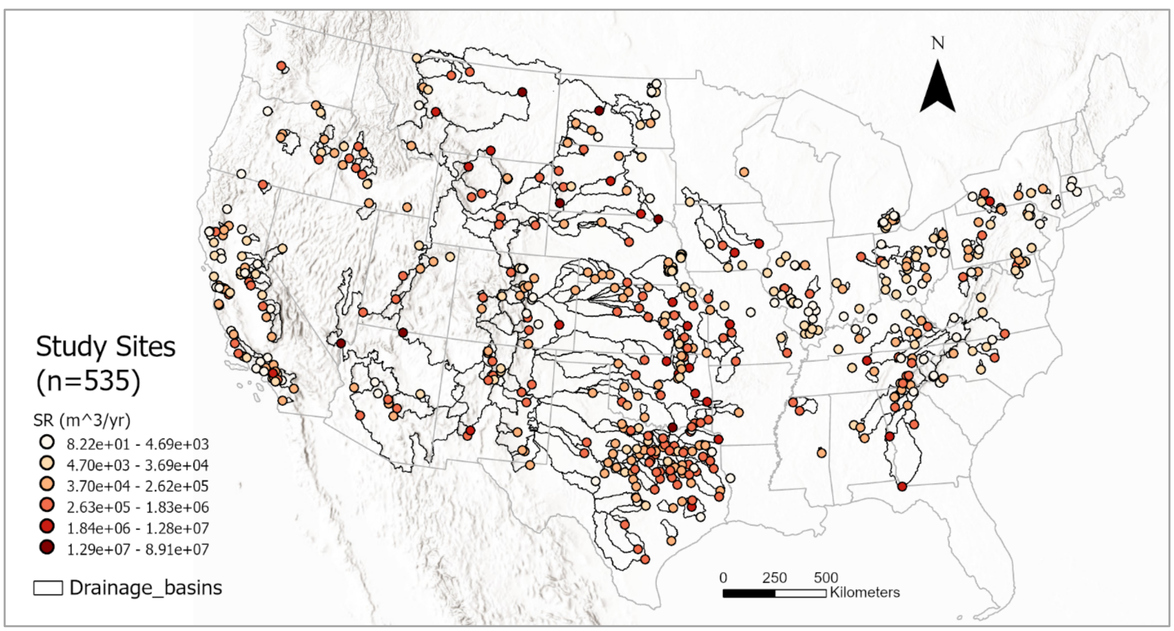

Figure 1. Map of 535 sites and their drainage basins, US. Circle colors show reservoir sedimentation rate (SR).

We are building a multiple linear regression (MLR) model in Python to predict SR at 535 surveyed reservoirs (Figure 1). Our reservoir sites and study basins are collocated with those described in Foster et al. (2023). We investigated environmental parameters that impact basin-sediment production and transport, such as: drainage area, inflow, relief, climate, and vegetation, as well as human impacts though proxies like population (Table 1). At this time, we have not reduced this parameter space based on sediment transport process assumptions alone, but rather to isolate the parameters that give the most statistical predictability.

We compiled upstream basin environmental and anthropogenic controls for 50+ first-order parameters in ArcGIS Pro. We then identified correlations between individual parameters and SR. Several parameters strongly correlated with SR (e.g., sediment-contributing drainage area, DA_C), while others correlated weakly or not at all (e.g., mean basin temperature and precipitation). Figure 2 shows a correlation matrix with 18 of the 50+ predictor variables and their R^2 correlation with SR and other parameters. Parameter definitions and specifications for the selected variables are given in Table 1.

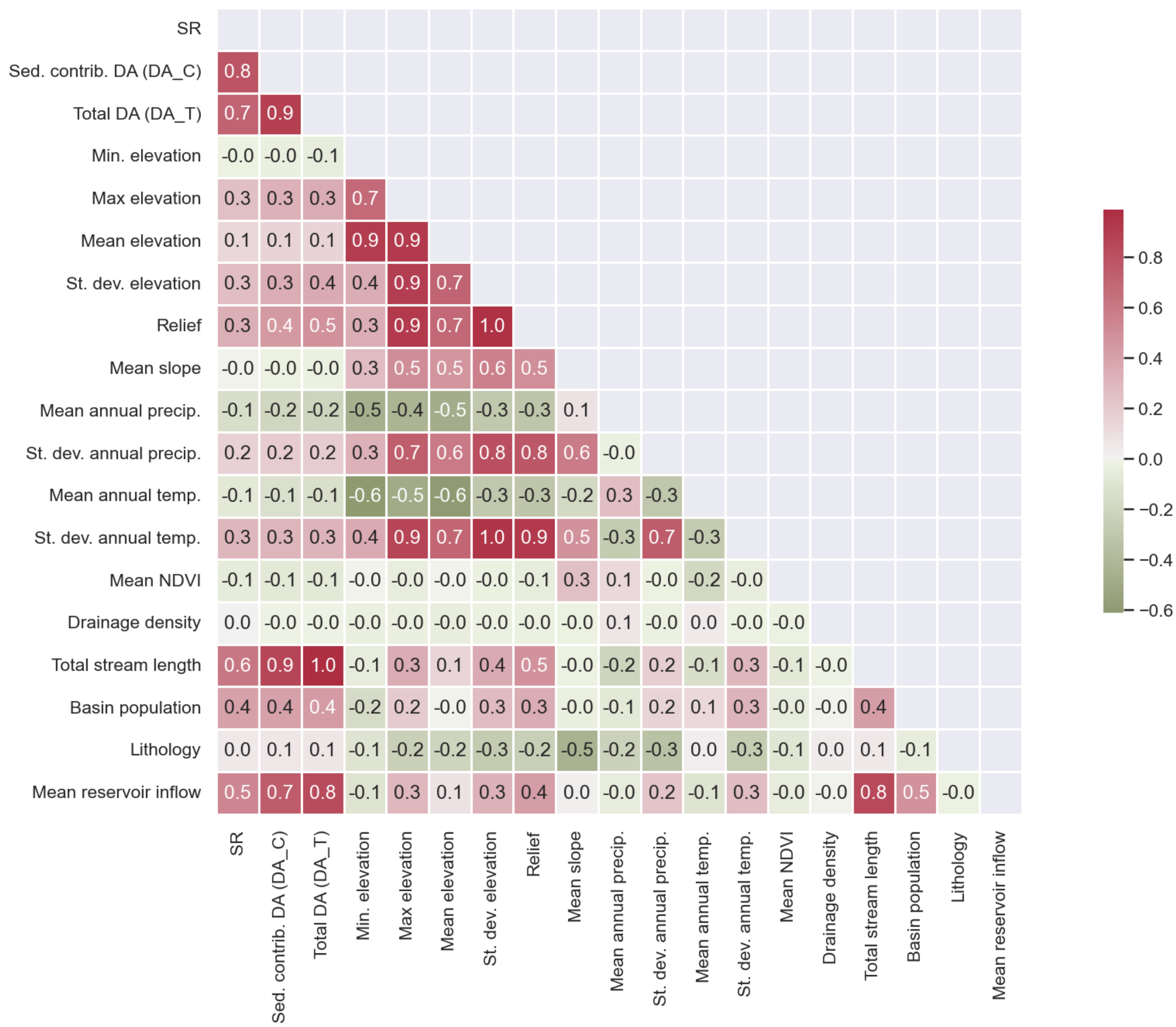

Figure 2. Correlation matrix displaying the coefficient of determination (R^2) for sedimentation rate (SR) and selected predictor variables for all 535 sites. Color bar represents R^2 values. Parameters are defined in Table 1.

Once we refine potential predictor variables for SR (Figure 2), we will determine which parameters to include in our MLR model through statistical analyses. First, to avoid overfitting the model, we will remove collinearity between variables. Collinearity is defined as two independent variables that correlate via an R^2 greater than 0.8. If two or more parameters of a similar data type are colinear, we may combine them into a single term using **principal component analysis** (PCA). PCA is a statistical tool used to reduce the complexity of a multidimensional dataset by transforming it into fewer dimensions while preserving the variation in the data. For example, since many of the elevation parameters are colinear (i.e., min, mean, and max elevation, Figure 2), we will explore the appropriateness of PCA to combine them into a single term.

Next, to untangle the best combination of parameters to use in our MLR model, including the derived principal components and raw parameter data, we will use the **regsubsets** function within the **leaps** package in R. This tool performs an exhaustive search over all parameters to find a combination that provides the best goodness of fit while minimizing the prediction error.

Finally, we will develop a MLR model using the **sklearn** package in Python, using 90% and 10% of the data to train and test the model, respectively. We will ensure that each parameter's relationship with SR is significant, with p-values below 0.05.

We have identified several parameters that predict SR well in a preliminary MLR model, such as DA_C (Figure 2). DA_C accounts for 63% of the proportion of variation (R^2) when predicting SR. Other parameters that explain a small proportion of the variation include mean reservoir inflow, total stream length, and basin population (Figure 2); however, these variables are largely controlled by basin size. For example, total stream length is a strong predictor of SR, but drainage density (Table 1) is not (Figure 2). We are still investigating how to address the influence of drainage area on our parameter data while maintaining the predictability of our model. Once we have finalized the parameters, we will sequentially add them to the model through **stepwise linear regression**, using regsubsets to inform the order in which to add the parameters.

We aim to develop our model to effectively predict SR in unsurveyed reservoirs across the US, beginning with unsurveyed reservoirs managed by the U.S. Bureau of Reclamation. The 535 drainage basins used to generate our statistical model range from $3.7 - 2.8 \times 10^6$ km^2 and span 6 IECC climate zones (ICC 2021), covering a large variety in basin size, topography, and rainfall regimes. Many of the identified controlling parameters are also identified in predictive models of fluvial suspended sediment flux to the coastal zone (Syvitski and Milliman 2007). Reservoir sedimentation may be impacted more by bedload transport, and we will explore notable differences between the respective predictive models. Future work will further examine basin runoff, wildfire history, land cover, and regulatory activities as potential additional controlling parameters in our MLR model. Our presentation will cover the preliminary results, report parameter updates, and provide an estimate of reservoir capacity loss for the nation.

Table 1. Selected environmental and anthropogenic parameters used to predict reservoir sedimentation rate (SR), along with data source, resolution, and approximate dates represented.

Parameter	Definition	Data source	Resolution	Approx. dates
SV	Sediment volume measured in reservoir	Foster et al. (2023)	-	1840-2017

T	Years between first and last survey	Foster et al. (2023)	-	-
DA_C	Time-weighted sediment-contributing drainage area above site	Foster et al. (2023)	-	-
DA_T	Total drainage area above site used to compute basin zonal statistics in ArcPro	Foster et al. (2023)	-	-
Min., Max, Mean., & St. dev. elevation; Relief, & Mean slope	Elevation statistics for drainage basin, calculated via the zonal statistics tool in ArcPro	USGS 3D Elevation Program (USGS 2019)	30m	2013
Mean annual precip., St. dev. annual precip., Mean annual temp., & St. dev. annual temp.	30-year normal climate statistics for drainage basin, calculated via the zonal statistics tool in ArcPro	PRISM (2014)	800m	1991-2020
Mean NDVI	Mean of the normalized difference vegetation index (NDVI) in drainage basin, calculated via the zonal statistics tool in ArcPro	USDA (2014)	1m	2010-2020
Drainage density	Sum of the length of NHDPlus High Resolution flowlines in drainage basin divided by DA_T	USGS (2021)	-	-
Total stream length	Sum of the length of NHDPlus High Resolution flowlines in drainage basin	USGS (2021)	-	-
Basin population	Total population in drainage basin (P)	EPA EnviroAtlas (Pickard et al. 2015)	30m	2010
Lithology	Lithology (L) value, computed as a weighted average over all surficial rock types in drainage basin	Syvitski and Milliman (2007)	-	-
Mean reservoir inflow	Mean reservoir inflow ($m^3\ s^{-1}$), computed from: mean daily[1], mean monthly[2], mean annual[3], and modeled mean annual reservoir inflow[4] data	HydroMet (2020)[1], ResOpsUS (Steyaert 2023)[2], RESIS-II (Ackerman et al. 2009)[3], StreamStats (USGS 2016)[3], USDA (2022)[4]	-	-

References

Ackerman, K.V., Mixon, D.M., Sundquist, E.T., Stallard, R.F., Schwartz, G.E., and Stewart, D.W. 2009. "RESIS-II: An Updated Version of the Original Reservoir Sedimentation Survey Information System (RESIS) Database," U.S. Geological Survey. Data accessed March 2023.

Foster, M., Hurst, A., and Eckland, A. 2023. Developing an Automated Method to Estimate Reservoir Sedimentation at ~30,000 Reservoirs across the United States. SEDHYD 2023, St Louis, Missouri.

HydroMet Tools Desktop Application. 2020. U.S. Bureau of Reclamation. https://github.com/usbr/HydrometTools. Data accessed March 2023.

International Code Council (ICC). 2021. Climate Zone Map from International Energy Conservation Code (IECC).

Parameter-elevation Regressions on Independent Slopes Model (PRISM) Climate Group, Oregon State University. 2014. https://prism.oregonstate.edu. Data accessed October 2022.
Pickard, B.R., Daniel, J., Mehaffey, M., Jackson, L.E., and Neale, A. 2015. "EnviroAtlas: A New Geospatial Tool to Foster Ecosystem Services Science and Resource Management," Ecosystem Services, 14:45-55. Data accessed May 2022.
Steyaert, J.C., Condon, L.E., Turner, W.D., and Voisin, N. 2022. "ResOpsUS, a Dataset of Historical Reservoir Operations in the Contiguous United States," Scientific Data, 9. Data accessed March 2023.
Syvitski, J. and Milliman, J. 2007. "Geology, Geography, and Humans Battle for Dominance over the Delivery of Fluvial Sediment to the Coastal Ocean," Journal of Geology, 115:1-19.
U.S. Department of Agriculture (USDA) Farm Services Agency. 2014. USA NAIP Imagery: Normalized Difference Vegetation Index (NDVI). Data accessed October 2022.
USDA Forest Service. 2022. Flow Metrics for the Contiguous United States (Historical), online at https://data.fs.usda.gov/geodata/edw/datasets.php?xmlKeyword=hydro%20flow%20metrics. Data accessed March 2023.
U.S. Geological Survey (USGS). 2016. The StreamStats program, online at http://streamstats.usgs.gov. Data accessed March 2023.
USGS. 2019. 3D Elevation Program 1-Meter Resolution Digital Elevation Model. Data accessed October 2022.
USGS. 2021. National Hydrography Dataset (NHD) Plus High Resolution Version 2.0, online at https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data. Beta version accessed in 2022.