5-2024

Exploring the Role of Near Channel Geospatial Attributes to Predict Suspended Sediment Concentration Patterns Across the CONUS Region

Aaron J. Sigman
Utah State University, aaron.sigman@usu.edu

EXPLORING THE ROLE OF NEAR CHANNEL GEOSPATIAL ATTRIBUTES TO PREDICT SUSPENDED SEDIMENT CONCENTRATION PATTERNS ACROSS THE CONUS REGION

by

Aaron J. Sigman

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

Approved:

Colin B. Phillips, Ph.D. Major Professor	Belize Lane, Ph.D. Committee Member
Jack C. Schmidt, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2023

ABSTRACT

Exploring the role of near channel geospatial attributes to predict Suspended Sediment Concentration patterns across the CONUS region

by

Aaron J. Sigman, Master of Science

Utah State University, 2023

Major Professor: Dr. Colin B. Phillips
Department: Civil and Environmental Engineering

High concentrations of suspended sediment (SSC) in a river can represent a critical water quality concern, reduce the storage capacity of reservoirs, and impact aquatic habitat. The flux of suspended sediment is a complex function of river hydraulics controlling the available shear stress and the concentration of fine sediment supplied by the watershed. Local hydraulic controls can be reliably measured at the reach scale, however untangling SSC requires determining which factors may impact sediment supply. To understand the role of geospatial watershed processes on sediment supply, we utilized SSC data from over 1,000 US Geological Survey stations. We find that SSC at a site is generally well described by a lognormal distribution with median concentrations spanning over five orders of magnitude across the continental United States. Similar SSC values are clustered throughout the continental US indicating a potential dependence on regional watershed properties. Here we utilize readily available geospatial watershed and network scale attributes (topography, soil, vegetation, land use, and climate) to explore

how geospatial attributes extracted at the near channel network and the watershed scale impact average SSC patterns throughout the country. Through principal component analysis and multiple non-linear regression, we reduce a wide variety of geospatial attributes down to seven key predictors. Combined these predictors provide a reasonable explanation of the mean SSC pattern across the CONUS region. The combination of a common probability distribution and geospatial estimation of the mean and standard deviation sets the basis for probabilistic predictions and forecasts of SSC within the CONUS region.

(60 pages)

PUBLIC ABSTRACT

Exploring the role of near channel geospatial attributes to predict Suspended Sediment Concentration patterns across the CONUS region

Aaron J. Sigman

High concentrations of suspended sediment (SSC) in a river can represent a critical water quality concern, reduce the storage capacity of reservoirs, and impact aquatic habitat. The total amount of sediment is calculated from a combination of river properties, including the amount of available sediment and the flow of water carrying the sediment. Water flow properties can be found using local information about the channel, however understanding the concentration of sediment in the river requires understanding the supply of sediment from the watershed. To understand where sediment is coming from, we examined over 1000 United States Geological Survey sites with SSC data. Across the country there is an extraordinary range in the measured values for SSC, however, the median value of SSC for a site generally describes the regional concentration. Similar concentrations are grouped within certain regions of the continental United States, showing areas of higher or lower concentrations highlighting the importance of local watershed properties. For this research, we use maps of elevation, soil properties, vegetation, land use, and climate to explore how the geospatial information alongside and upstream of the river affects SSC. With multiple types of data processing, the most important mapping factors can be extracted and used to predict SSC. Combined these datasets provide a reasonable explanation of the regional SSC patterns across the

continental United States. Understanding and reliably estimating SSC is an important first step for predicting and managing physical water quality.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Phillips for all his knowledge and experience, continuously providing advice, support, and patience throughout this project. My committee members, Dr. Lane and Dr. Schmidt, provided new ideas and points of view on how to explore and expand this research, that proved incredibly helpful in key points. Many thanks to the professors from the Utah Water Research Laboratory at Utah State University for their quality teaching and genuine interest in their students, renewing my interests in continuing research. Lastly, I would like to thank all my family and friends for their encouragement and support through my studies.

Aaron J. Sigman

CONTENTS

Page

LIST OF FIGURES

Page

INTRODUCTION

The suspended sediment concentration (SSC) within a river is a function of the sediment particles, turbulent stresses within the flow, and sediment availability within the channel and watershed (Garcia, 2008; Garcia & Parker, 1991). Sediment transport is a natural process and plays an important role in the geomorphic, biological, and ecological functioning of fluvial systems. Excessive amounts of suspended sediment can negatively impact water quality, harm stream ecosystems, and alter water chemistry and temperature (Bilotta & Brazier, 2008; Boano et al., 2014; Drummond et al., 2022). Chronically high concentrations of sediment can cause siltation, affecting the abundance of fish food organisms and salmonid spawning beds (Cederholm, 1979; Schwartz et al., 2011). Suspended sediment loads with higher concentrations of silt or clay can represent environmentally significant carriers of nutrients and contaminants due to physical and chemical properties that allow them to bind or encapsulate other particles (Gottselig et al., 2014; Milligan & Loring, 1997; Phillips et al., 2019). Suspended sediment represents one of the primary mode through which rivers transport eroded sediment out of watersheds (Merritt et al., 2003; Shen & Julien, 1992) and a major source of reservoir infilling reducing their overall storage capacity and operational lifetimes (Dutta, 2016; Einsele & Hinderer, 1997; Wohl & Cenderelli, 2000). Identifying non-point source mechanisms contributing to SSC remains an outstanding challenge as watersheds possess a large spatial range in erosional and climatic factors as well as temporal changes to land use and frequent disturbances (e.g., floods, landslides, and wildfire). Understanding how watershed processes affect the sources of sediment and the transport mechanisms of

sediment within a river system is essential in predicting how changes to and within watersheds will impact water quality and management decisions.

A variety of natural factors are hypothesized to influence the amount and size of sediment supplied to watersheds that may show up in suspended sediment records. Local characteristics of the watershed, such as lithology, precipitation, slope and relief, and land use and land cover are all reported to impact catchment erosion and the supply of sediment within a river network (Czuba & Foufoula-Georgiou, 2014; Langbein & Schumm, 1958; Merritts et al., 1994). Short term high-resolution measurements of suspended sediment highlight significant variability in both concentration and flux (Jung et al., 2020; Lana-Renault et al., 2007; Topping et al., 2003), whereas longer-term averaging of suspended sediment flux measurements reveal a relatively steady erosional process and can even dilute the signal of large disturbances such as landslides, wildfires, or flooding from large storm events (Coombs & Melack, 2013; Lin et al., 2008; Sadler & Jerolmack, 2015). For such a highly variable process, watershed-scale characteristics provide good explanation of suspended sediment yield on annual to decadal timescales with models such as SPARROW, BQART, WBMsed all utilizing regional and watershed attributes to estimate sediment flux (Anderson & Macdonald, 1998; Cohen et al., 2013, 2022; Schwarz et al., 2006; Syvitski & Milliman, 2007). However, near-channel sediment sources and local characteristics show more importance for understanding regional patterns in suspended sediment concentrations (Belmont et al., 2011; Stout et al., 2014; Vaughan et al., 2017). Regional properties, such as precipitation, soil erodibility, elevation, and land use change (e.g. timber harvests), have been shown to affect sediment rating curves (Fisher et al., 2021; Karwan et al., 2007; Vaughan et al., 2017;

Zabaleta et al., 2007). Significant progress has been made on the prediction of suspended sediment flux; however, this may primarily be due to our ability to predict discharge at large scales. Discharge scales nearly linearly with catchment area (Burgers et al., 2014; Frasson et al., 2019; Galster, 2007) once gradients in rainfall are accounted for. However, suspended sediment concentration remains remarkably challenging to accurately predict. Despite the challenge in predicting SSC, we find that there is a geographically clustered spatial pattern evident across the United States indicating that regional watershed or climatic factors may provide a first order predictor on the mean SSC.

To understand SSC, it is useful to understand both how the data are collected and the available mechanics-based models that describe SSC locally. Suspended sediment measurements from natural rivers within the US Geological Survey's (USGS) National Water Information System (NWIS) represent depth integrated water samples composed of a collection of particle sizes whose size composition depends on both the flow strength of the river and the availability of the particle. To further explore and separate the impact of flow strength from particle availability or sourcing we first consider the Rouse profile as it represents a useful mechanistic framework for understanding suspended sediment transport.

The concentration profile of suspended sediment with depth tends to follow an 'S' shape or Rouse profile with higher concentrations and larger particle sizes near the bed of the river (Garcia & Parker, 1991; Lamb et al., 2020; Rouse, 1937). For the theoretical Rouse profile, sediment in suspension is sourced from the riverbed without supply limitations (Garcia & Parker, 1991). For such a profile to exist within a reach of a river,

there must be ample suspended sediment available to transport from sources near the sampling location, which would indicate significant storage of sand and silt sized particles (the coarse fraction of suspended sediment) within the riverbed or near the channel. The finest particles are commonly attributed to nonlocal sources via the concept of 'washload,' particles that are typically characterized to be less than 62 microns in diameter whose concentrations within the bed are insufficient to explain their concentrations within the water column (Einstein & Chien, 1953) or alternatively particles whose accumulation do not change the morphology and structure of the riverbed (Hill et al., 2017). For washload, we might expect that SSC within rivers is controlled primarily by watershed factors, whereas for larger particles SSC may be a function of local storage and reach scale hydraulics. However, due to flocculation even the finest particles aggregate to larger sizes (Lamb et al., 2020) such that their transport and sample concentration may depend on reach-scale hydraulics in the short term (hydrograph to sub-annual timescales). As such understanding short term SSC dynamics may remain the role of calibrated physics-based models and frequent sampling (Rubin et al., 2020; Rubin & Topping, 2001), whereas exploring the role of geospatial watershed characteristics may be better suited to disentangling cumulative loads (Brakebill et al., 2010) which integrate over short term fluctuations and annual flux models (Cohen et al., 2013).

The USGS water quality database represents the largest collection of sites with SSC measurements to date. The USGS has established numerous sites across the United States where measurements of suspended sediment are or were collected within the past 50 years (Fig. 1). When observed in aggregate, high and low values of the geometric

mean SSC are spatially clustered within specific geographic regions within the USA (Fig. 1), which suggests that spatial attributes within these regions may be able to predict the observed pattern. However, despite the spatial pattern there are not yet an agreed upon set of watershed attributes necessary to predict the pattern across and within regions. Are there a collection of geospatial attributes watershed or near-channel attributes that can explain the observed pattern? Here we seek to utilize publicly available geospatial data to determine if the time-averaged SSC pattern can be determined from watershed characteristics near and upstream the sample location. An understanding of how watershed attributes combine to predict the geometric mean for SSC represents a critical path forward for understanding how landscape and spatial climate attributes impact physical water quality. Additionally, understanding which watershed factors and at what scale they impact SSC can help determine how historical and future change scenarios (variability or change in space and time in land use and climate) within a watershed have or may impact SSC and therefore impair water quality. Further, this information could be used to inform policy, management and restoration of stream and river networks on issues including infrastructure, water security, reservoir sedimentation and hydropower supply, and aquatic biodiversity (Dutta, 2016; Huettel et al., 1998; Palmer et al., 2000; Schleiss et al., 2016; Vörösmarty et al., 2003).

The overarching goal of this research is to determine if local watershed properties can be used to determine the time averaged concentration of suspended sediment within a watershed. We sought to answer the following questions: (1) how do watershed properties relate to suspended sediment concentrations and ultimately impact a river's sediment water quality? and (2) at what scale (point, local, or basin) do these watershed

properties best contribute to suspended sediment? To address these questions, a method to extract river corridor properties at different scales was created to collect statistics for each watershed attribute. This geospatial extraction allows for comparison between the watershed attribute descriptive statistics and each stream gage's historically averaged suspended sediment concentration.

METHODS

All discharge and SSC data for each stream gage was collected from the USGS NWIS. From the NWIS database we utilize 1272 sites, each with discretely sampled partial time series for discharge and suspended sediment concentration. Water quality data within the NWIS system are collected for specific purposes and therefore do not all have the same sampling frequency or quantity. The data quality at all sites is not equivalent as some sites contain too few measurements or highly biased temporal sampling to reliably estimate the descriptive statistics. We have conservatively chosen to limit our analysis to sites with greater than 100 unique samples to increase the likelihood that descriptive statistics (i.e. the average) will describe the underlying data. At a small number of sites (124) the time series may represent sporadic sampling, only low or high flows, and highly discontinuous sampling (periods of brief intense sampling spread across multiple decades). We further exclude sites with main channel lengths under 5 km as these areas are too small to effectively differentiate between the point and basin scale for many of the geospatial variables. We also focused on the continental United States, removing sites from Alaska, Hawaii, Puerto Rico, and other US territories. For the methods, USGS web data retrieval was used, so all sites were tested to check for valid and working web links. These filtering steps result in the exclusion of 124, 51, 18, and 37 sites for improperly sampled time series, extracontinental origin, broken web links, and small basin sizes, respectively. Following the aforementioned filtering we are left with 1042 sites (Fig. 1).

Of the 1042 remaining sites, 76 were selected to run an initial training analysis. These sites were selected from seven basins across the contiguous United States highlighting an array of geospatial and climatic attributes. In the selection of these basins, the goal was to identify sites with a diverse range of climatic values, suspended sediment concentrations, and multiple land uses. This initial step was implemented to reduce the processing time required in developing tools to batch process the extraction of geospatial and climatic attributes within all basins upstream of the broader set of NWIS gages used in this study. The reduced set of sites used in this preprocessing step produced stronger statistical trends between the average SSC and geospatial attributes than when considering the full collection of sites. This pre-analysis highlights an important cautionary note in that sample site selection can strongly bias your resulting trends and inferences. From each selected NWIS site, we compute descriptive statistics for the suspended sediment concentration, flux and water discharge.

All datasets for land use, hydrology, topography, precipitation, air temperature, and evapotranspiration represent nationally continuous data products for the continental United States (CONUS) and were accessed from publicly available sources. Rainfall and temperature data were collected from the PRISM Climate Group ((Daly et al., 2008; PRISM Climate Group, 2014). This data provides 30-year normal values (1990-2020) for precipitation in millimeters and average minimum, mean, and maximum temperatures in degrees Celsius at an 800x800m grid scale for the CONUS region. Elevation data was accessed from the USGS's 3D elevation program (U.S. Geological Survey, 2019). This dataset covers the entire CONUS region at scales from approximately 30 meters to 1 meter in resolution. The data extraction method for elevation accesses the highest

resolution elevation data available for the study area. The land cover map was downloaded at a national scale from the Multi-Resolution Land Characteristics Consortium (Dewitz & U.S. Geological Survey, 2021). This is a group of federal agencies who coordinate and generate consistent and relevant land cover information at the national scale for a wide variety of environmental, land management, and modeling applications (Dewitz & U.S. Geological Survey, 2021). We utilized the most recent dataset available at the time of access which was the 2019 National Land Cover Dataset (NLCD). The NLCD is available at a scale of 30×30 meters and separates each pixel into one of multiple possible categories, including but not limited to: water, developed (urban), forest, agriculture, and shrublands. The 2019 NLCD map represents land attributes at a significantly later date than the majority of SSC timeseries, however, as we seek to understand the general CONUS wide patterns the mismatch in dates is unlikely to significantly alter the trends as NLCD datasets typically remain regionally consistent while undergoing changes at a more local scale. Lastly, a worldwide evapotranspiration dataset was collected from the Global Aridity Index and Potential Evapotranspiration Climate Database (Trabucco & Zomer, 2018). This dataset provides a measure of evapotranspiration worldwide at a 30-arc-second, or approximately 1km, scale updated in 2022. Aridity was derived by dividing the spatial precipitation data by the estimated spatial evapotranspiration following Trabucco & Zomer (2018).

To perform the spatial extraction of watershed attributes we utilize the HyRiver software libraries within Python (Chegini et al., 2021). The HyRiver package allows for the extraction of upstream river basin and network attributes utilizing National Hydrography Dataset plus (NHD+) data through the USGS Hydro Network-Linked Data

Index and WaterData web services (Moore et al., 2019). For each USGS stream site used, the upstream river basin and the main channel of the river at a specified distance can be extracted and buffered by a set distance to create a spatial polygon. This buffer can then be used to extract the values from each of the aforementioned geospatial data products at a predefined distance from the mainstem channel (Fig. 2). For numerical quantities, statistics of the spatial layer are extracted including minimum, maximum, average, median, standard deviation, range, and inner quartile range. Spatial raster layers consisting of classified categorical data were summarized to record the percentage of each category within the buffered stream network. For the following analysis, we utilize a 10 km buffer distance to define the stream corridor (5 km on both sides) and extract geospatial attributes at the point, 20%, 40%, 60%, 80% and full basin scale (Fig.2). Here the percentage of the mainstem attributes extracted is determined as the length of the polygon upstream along the streamline divided by the total length of the mainstem within the basin. We recognize that rivers are more than their mainstem channel and that a fuller analysis of geospatial attributes should likely consider a fraction of the full stream network rather than a fraction of the mainstem, however, as a first step we stick to the main channel and incorporate tributaries within the buffer distance. We note that in many of the sites we examined, the geospatial attributes do not necessarily change dramatically throughout the basin laterally.

Following geospatial data extraction, we compiled a database where each gage and its associated NWIS attributes (statistics for SSC and discharge) are further associated with a host of geospatial attributes for varying buffer distances from the main channel and fractions of the mainstem stream network from the point to basin scale. We

explore the potential dependence of SSC metrics on geospatial attributes through multivariate regression and principal component analysis (PCA), which allow for substantial explanatory variable reduction through iterative application of the method. This variable reduction stage is essential to limit the number of total variables necessary to understand the spatial pattern of average SSC.

RESULTS

We utilize 1042 sites (Fig. 1), each with discretely sampled time series for discharge and suspended sediment concentration. Exploratory analysis of these records reveals that the distribution of SSC at each site is non-normally distributed. We find that a natural log transformation of the data provides a more symmetric distribution which is generally well described by a lognormal distribution. Such a transformation and the nonnormality of the data is not unexpected because water discharge time series are typically lognormally distributed as well (Helsel et al., 2020). For the majority of sites, the USGS provides a suite of whole watershed scale attributes including: number of samples, drainage area, 30-year normal average precipitation, soil characteristics (particle size, R and K factor), land use (restricted to forest, agriculture, and developed), and statistics on upstream dams (number of large dams and total storage capacity). Examinations of individual sites within the database reveal that that the majority (approximately 90%) of the probability density functions are unimodal skewed distributions. Following a natural log transformation, the distributions become nearly symmetric indicating that the appropriate statistical descriptors for the central tendency and dispersion are the geometric mean and geometric standard deviation, respectively (Helsel et al., 2020). Throughout the following results and discussion references to the mean SSC refer to the geometric mean. Across the CONUS region at the selected sites the values for mean SSC range by five orders of magnitude, while the standard deviation ranges by one order of magnitude. Regions of high concentration are clustered into the arid Southwest and the Midwestern plains, specifically the western plains.

We begin our analysis by exploring relations between basin characteristics and mean SSC. Comparing drainage area to average values of SSC and suspended sediment flux (Q_s), we can see a positive trend in loglog space for both SSC and flux (Fig. 3). These data cover over six orders of magnitude in drainage area. The overall trend makes sense for flux as discharge grows with drainage area (Galster, 2007; Leopold & Maddock, 1953) and flux is the product of discharge and SSC. However, SSC and flux differ in the degree to which they scale with drainage area. In comparing the regression trends between mean SSC ($SSC = 17.8DA^{0.15}$, R^2=0.058, Pval < 0.001) and mean flux ($Qs = 3.7e\text{-}7DA^{0.83}$, R^2=0.53, Pval < 0.001) it is evident that SSC only weakly depends on drainage area. Both trends are statistically significant, however the SSC-drainage relation is far from predictive as the variability in SSC for a given drainage area (roughly three orders of magnitude) is larger than the observed dependence on drainage area (one order of magnitude, Fig. 3a). Variation in Qs for a given drainage area is also approximately three orders of magnitude, however the trend covers over seven orders of magnitude in flux. Variation about the Qs drainage area trend correlates strongly with the SSC values, indicating that understanding the controls on SSC will lead to improved flux predictions.

The spatial pattern (Fig. 1) across the CONUS region highlights the potential importance of a climatic gradient and indeed we observe that sediment concentration declines with basin scale precipitation (Fig. 3b). All things being equal within a catchment, that concentration declines with increased precipitation is not unexpected. The median trend between the average basin precipitation (30 year normal) is remarkably well described by a power function (Fig. 3 b & c), however, this function has limited

predictability beyond the median concentration due to the high degree of variability within the SSC data. We don't believe this trend should be discounted as it certainly explains a first order continental scale relation, however, it also indicates that other variables are likely equally or potentially more important.

We utilized principal component analysis (PCA) to explore a larger collection of basin scale averaged attributes including: drainage area, number of large dams upstream, percentage of the basin with forest, urban or agriculture land use types, 30-year normal rainfall, and the percentage of sand, silt and clay within the basin soils. We combine these variables with the natural log transformed SSC and the number of SSC samples per site. Drainage area and rainfall were log transformed prior to analysis to account for their non-normal distributional shape. The first four principal components describe 65.5% of the variance within the dataset with ~44% accounted for by the first two principal components. The PCA (Fig. 4a) highlights how SSC varies across a gradient of land use, soil composition, and rainfall. Specifically, higher concentrations occur for sites with higher drainage areas, low rainfall, more clay rich soils, and agricultural land use; while, lower concentrations occur for sites with higher rainfall, more forest and sandy soils (Fig. 4a). It is important to note here that PCA does not provide a physical relation or suggest that these variables are the only important predictors. Many of these variables are naturally co-varying in that smaller catchments are likely to be more forested and agricultural land use occurs within favorable soils for farming. However, the basin scale results do not reveal how the geospatial attributes are arrayed throughout the catchment or even if a particular land use variable is located near the river. The basin scale PCA reveals that with these particular variables the SSC pattern is not randomly arrayed across

these sites, rather it forms a gradient of lower to higher concentrations. Here we leave the basin scale geospatial results to bring in a wider array of geospatial attributes extracted from near and upstream of the river channel.

For the near-channel analysis, we limit our geospatial extraction to 20% of the mainstem channel length upstream of the sample site with a buffer distance of 5 km adjacent to the channel on either side. Within the near-channel analysis we can extract topographic features like local elevation, relief and channel gradient, and soil properties adjacent to the channel. For each site, many statistical outputs of each dataset were collected. These included major quantiles, percentiles, maximums, minimums, and ranges. Near-channel properties have been identified as reasonable predictors of SSC within regional settings (Stout et al., 2014; Vaughan et al., 2017). The number of variables expanded at the local scale as we computed a range of descriptive statistics within each buffer distance for each variable including the: mean, median, minimum, maximum, range, interquartile range, and standard deviation. For each variable (see Appendix for a complete list of variables) we explored correlations and bivariate plots to understand if and how the variables were interrelated and relate to SSC. The majority contained either no relation or power relations with considerable scatter. No single variable emerged as a particularly strong predictor. Iterative PCA at the near-channel scale allows for the reduction in total dataset dimensionality. We performed an initial PCA with all attributes, and iteratively removed attributes with the lowest correlation coefficients between either of the first two principal components. Our final PCA resulted in the inclusion of 16 variables (see Fig. 4b for list) and mean SSC, with the first four principal components explaining 63.1% of the total variance with 41.8% accounted for by

the first two components. The overall pattern indicates that higher SSC values are correlated with sites with larger temperature and temperature ranges, high elevations, higher evapotranspiration rates, agricultural land use, and higher frequencies of potential flooding. Lower SSC values correlate with sites with wetter climates with increased vegetation. We note though that the addition of more geospatial datasets in the near-channel scale PCA doesn't increase the overall explained variance compared to the basin variables (63.1 to 65.5%) but did reveal relations between variables and highlighted climatic and topographic variables over land use as larger drivers of variance within the data.

The switch from a basin scale analysis to the near-channel analysis highlights a few fundamental differences. Near-channel relief highlights bluffs and canyons, while basin scale relief would reveal the presence or absence of high relief (e.g., mountains), and potentially, the proximity to mountains. Some basin scale parameters are ambiguous, in that a site within the mountains would have lower relief than a site downstream on the same river, while near-channel relief can distinguish between the two. We highlight the change from basin to near-channel through the use of the three land use categories (Fig. 5). Similar to the PCA, higher SSC values tend to occur in regions with larger percentages of agriculture (Fig. 5). At the basin scale, land use is primarily a gradient between agriculture and forest, while at the near-channel scale there is a greater mix across land use types including a growth in the percent of the urban land use type (Fig. 5 a & b). In general, moving from basin descriptions of land use to the near-channel values increases the percentages of the urban land use type or eliminates it. For the near-channel scale description of land use we combined multiple NLCD classification categories to

limit the analysis to the three urban, agriculture, and forest. Urban consists of the four developed categories, open space, low intensity, medium intensity, and high intensity. Agriculture consists of barren land, shrub/scrub, pasture/hay, and cultivated crops. Forest consists of deciduous, evergreen, and mixed forests, as well as herbaceous, and the wetland categories or woody and herbaceous wetlands. Considering each land use type individually (Fig. 6 a-c), SSC possesses a statistically significant though small positive trend with agriculture, a negative trend with forest coverage near the channel, and no trend with urban land cover. The median trend for both forest and agriculture land covers is well fit by an exponential function with a constant variation about the trend, however the overall variability remains considerable. Ultimately, these results highlight that land use at both the basin and near-channel scale is minimally predictive of the mean site SSC.

We explore individual relations between geospatial attributes linked to or believed to be linked to erosive processes. While we explored a large variety of potential geospatial and topographic variables, we focus on the following: precipitation, evapotranspiration, annual temperature and temperature variability, elevation, flood frequency (flood factor) as determined by topography, river slope, time integrated NDVI (TIN), soil erodibility (K-factor), soil tolerance (T-factor), and hydrologic group. The variability across all factors remains quite high, however some notable trends emerge. Increases in precipitation result in decreases in concentration, while increasing evapotranspiration rate within the near-channel region results in high concentrations (Fig. 7 d & e). As the annual average temperature increases, we observe an increase in the mean SSC. Precipitation and evapotranspiration can be combined into an aridity index (Trabucco & Zomer, 2018), however the resulting pattern with SSC is nearly a replica of

the observed pattern with precipitation. While not captured by the median trend line, these three climatic variables do reveal that the highest SSC concentrations occur at sites with low rainfall, high ET, and high annual temperatures (e.g., deserts). Slope, elevation, and the flood frequency factor all represent different aspects of the topography near the SSC sampling site (Fig. 7 a-c). The elevation term represents the median elevation within the near-channel buffer. The correlation between elevation and mean SSC is weak, however, we note that there is a strong pattern of increasing variability in the data as elevation increases. The flood frequency factor represents the frequency at which the near overbank area floods. In a sense this term may reveal the frequency at which a river deposits and erodes overbank and floodplain material. A positive trend is observed indicating that as the near channel area floods at a greater rate there is a higher mean SSC. Here the slope represents the linear regression of the upstream elevation along the mainstem channel. This slope value is not the hydraulic slope of the river reach except in small basins but represents a general steepness of the channel upstream. No trend is observed for this along channel slope with SSC (Fig. 7a).

To further understand how the many geospatial factors may interact to affect the mean SSC pattern we performed an iterative multiple nonlinear regression (MNLR). Initially, we incorporated over 60 geospatial attributes extracted at the near-channel scale (see list of variables in Appendix A). We iteratively performed MNLR on the natural log transformed variables and used the significance level of each variable to select which variables to keep (a p-value less than 0.05 was utilized as the cutoff for inclusion in the next iteration). Seven resulting variables were used to run a final regression to determine the leading coefficient and exponents for each variable (Eq 1).

(1) $$SSC = 7.36E^{0.39} P^{-0.85} T^{1.12} Kf^{0.67} Tf^{1.63} Ff^{1.6} HG^{0.93}$$

where the seven variables are: median elevation (E), mean annual precipitation (P), mean annual temperature (T), average K-factor (K_f), soil-tolerance factor (T_f), flood factor (F_f), and hydrologic group (HG). Predicted values compare favorably with the observed mean SSC values with over 50% of the observed data within a factor of two of the predicted values (Fig. 8). The multiple nonlinear regression captures the behavior of mean SSC for the majority of sites with mean SSC values less than 10^3 mg/L, however, we find that no regression model predicts the highest SSC values.

In returning to the spatial motivation for this project, we explore the accuracy of the model spatially across the CONUS region (Fig 9). This map highlights regions of overprediction (blue), where predicted values for mean SSC are greater than the observed values, underprediction (red), where predicted values for mean SSC are less than the observed values, and 'well' predicted (gray), where predicted values are within a factor of two from the observed values. About 39 percent of the sites fall within a factor of two, and approximately 91 percent of the sites fall within a factor of ten. This map captures the regions of the model's accuracies and inaccuracies.

DISCUSSION

Accurately estimating the mean SSC at basin-wide scales remains challenging, we explored if near-channel geospatial attributes could determine reliable estimates of the mean and standard deviation of the site SSC distribution. We focused on SSC over suspended sediment flux because SSC is an important water quality indicator. Predictions for suspended sediment flux are already reasonably good for larger rivers and capture the primary trend of increasing flux with water discharge. As flux is the product of SSC and water discharge, it is expected that flux increases with basin area, however any improvement in our ability to predict SSC would transfer to an improved prediction of flux as well, because the variation about the flux drainage area trend scales with SSC. We anticipated that with larger watersheds the basin wide statistics would obscure important patterns through the incorporation of significant area away from the channel and that point scale attributes would exclude important features upstream of the sample locations. For a reduced set of test basins (see green shaded basins in Fig. 1) we explored how changing the extraction distance along the mainstem from the point to the basin in increments of 20% of the basin length and the orthogonal distance from the channel from 1 to 10 km. However, moving from the basin to the point scale results in a smaller than expected shift in the geospatial attributes overall (Fig 2), while some changes are large in practice (a reduction in up to half a meter in 30 year normal rainfall) the trends of the data were the same. One potential confounding factor may be that we chose climatic variables that represent well average quantities and that many of the topographic variables which exhibit more change with distance upstream had only a slight correlation with SSC values.

Land use variables do exhibit a larger degree of change as we consider different extraction distances along and from the channel (Fig. 5), however, our analysis indicates that land use is not primarily responsible for the SSC pattern at the CONUS scale. Land use was observed to be an important component of the basin scale PCA, however, at the near-channel scale with the inclusion of more specific soil data products the importance of land use decreased. This likely occurred due to forest and agricultural land use types largely reflecting climatic and soil properties, while urban areas represent a smaller fraction of the total area within a catchment or near-channel region for most gages. Shifting from basin to the near-channel scale results in numerous sites gaining in the urban land use type as indicated by the increased spread of sites across the ternary diagram (Fig. 5 b). The largest changes in land use were observed as tradeoffs between forest and agriculture, however these do not appreciably change the observed patterns in the ternary diagrams. These methods could still be utilized within more regionalized studies where land use may represent a more definitive feature, as drastic changes in land use such as deforestation and urbanization are known to result in substantial increases in SSC (Lewis et al., 2001; Wolman, 1967). The near-channel extraction methods could also be useful when considering basins with point sources or hot spots of erosion. At the CONUS scale, these behaviors are likely transitory or obscured within the larger variability within the full dataset. Additionally, there may be some extended utility at the more regional scale to break up land use categories as not all agriculture or forest types behave similarly in terms of erosion. For this study 20 land use categories were reduced into four categories (urban, forest, agricultural, and water), which may oversimplify what is seen in each region, possibly resulting in an increased degree of variability in the

dataset. If this project were to be repeated, the entire NLCD range could be included to identify how this simplification impacts the results. A potential extension of this method could explore extraction of geospatial attributes based on a percentage of the full tributary network, as the current method is restricted to the main channel.

A challenge with many of the existing geospatial attributes is that they inherently are covarying or redundant. The variable reduction steps within both the PCA and the multiple nonlinear regression reinforce the idea of the covariation as numerous variables provided minimal explanatory power to the methods. Land use in particular covaries with many climatological and soil properties and is therefore potentially redundant within this analysis, however, land use may be a more valuable explanatory variable for understanding changes in SSC for studies exploring pre and post land use change or utilizing paired watersheds with similar climatic and soil properties.

A limiting aspect of the analysis presented here are the widely available geospatial products. Datasets that uniformly provide coverage across the CONUS region are often simplified in many regions, however the CONUS wide coverage and more standardized methods are important for any data comparison across sites. Datasets of higher resolution such as rainfall intensity, are often hyperlocal and missing large swaths of the CONUS region rendering them less useful for intercomparison at this point. Climate variables (precipitation, ET, temperature) were collected as 30-year normals to identify the historical trend for these climates, however, CONUS wide rainfall data beyond annual means remains a key area needed for exploration as erosion scales with rainfall intensity (Berger et al., 2010; Cammeraat, 2004; Römkens et al., 2002). A

limitation of the 30-year normal approach to climate in this particular study is that the SSC data for a site often do not cover the entire or for some sites even the same 30-year range. However, we are not convinced that a more closely aligned 30-year normal precipitation would reveal stronger trends as the difference between 30-year normal periods does not appreciably change the CONUS wide rainfall pattern. A potential alternative approach could be to create rainfall statistics based strictly on the basin or near-channel region for the periods when sediment was actively sampled. This approach should be verified within high-resolution data where a clear signal could be extracted and used as a guide when interpreting CONUS wide results.

The most significant data limitations for this research are related to the national soil and lithology data products. Data for soils and land use were collected from SSURGO and the NLCD and lithology was examined from USGS sources, however, each dataset necessarily represents a statewide compilation of data. Different states may have different collection methods for these datasets, possibly resulting in increased variability in watersheds that span over multiple states, particularly in the case of lithology where rock types can inexplicably change at state boundaries. A further limiting aspect of the soil and lithology data is the inherent categorical nature of the data and the absence of a reliable method to convert each into relevant erosion metrics such as fractions of grain size present and erodibility. Existing metrics for soils data utilized here were important factors representing four of seven statistically significant components within the multiple nonlinear regression. This highlights how more accurate soils data may be key in predicting SSC and unraveling spatial water quality patterns more broadly.

The SSC data themselves represent an additional potentially limiting factor, while we are fairly confident in the distributional shapes for each site a more stringent test should be employed to remove sites that do not accurately capture the true mean and standard deviation. A stricter filtering of sites based on their temporal coverage over their sampling record could reduce the overall variability within many of the bivariate regressions, however, we do not believe that this would necessarily result in different trends as the computed statistics for sites used here are independent of the sample size. We performed a quality control analysis where we restricted our correlation analysis to only sites where the discharge values sampled at the same time as the SSC samples matched the distributions of the full hydrograph for that period. This filtering resulted in reduced scatter for lower mean SSC values but did not appreciably change the overall patterns or trends. That the trends generally remain the same may not be surprising given the variability across site mean SSC values. Values for the mean SSC range across over five orders of magnitude naturally, the estimated mean SSC values would have to be off by more than an order of magnitude or be systematically biased for the resulting observed trends to change. Nonetheless the limited temporal sampling of SSC across the CONUS region remains a challenge for making broad management or operational decisions. Variability in space across the CONUS region represents another potential pitfall of these data. While there are a large number of sample sites, especially for sediment transport, they are not evenly spread across the CONUS region. We address this spatial sampling issue by resampling the data (Fig. 2 & 4-6) when making correlations between climatic or geospatial variables. For example, this resampling by number of sites for a given climatic value is necessary as the uneven spatial sampling does not guarantee that the predictor (x-

axis variable) is representatively sampled across the CONUS region. These binned medians generally represent robust trends through the data though with limited range in terms of prediction of the highest and lowest values. Because we have resampled the data, the regression equations fit to the data contain unrealistic explanatory statistics but do represent a more realistic trend within the data.

Many of the quantiles and percentages were well fit around the median and mean values, and often the range values had little to no explanatory information. Since most of the datasets fit best with log-scaling, we found that the median value for most datasets best described their value and correlation with mean SSC. The only dataset with a variable outside of median or mean, was temperature range. This value of mean annual maximum minus mean annual minimum seemed to show a prominence among the rest of the variable statistics (Appendix 4, Fig 10).

Within the CONUS wide SSC dataset there are many reasons to be cautious with the data. There are generally well described median patterns, but the variability within mean SSC strongly limits their utility. However, the multiple nonlinear regression analysis provides good estimation of the mean SSC, and highlights the combined role of topography, climate and soil properties in describing the spatial pattern of mean SSC across the CONUS region. That the pattern is reasonably well described by seven variables (median elevation, flood frequency, precipitation, temperature, and soil parameters K-factor, T-factor, and soil hydrologic group) all of which are geospatial in nature provides an interesting description on sediment supply and potential water quality within these basins. Median elevation and flood frequency describe the topographical

regions near each site. Increases in SSC with median elevation are slight (exponent of 0.39 within the MNLR) and indicates a tendency for higher elevation sites like mountains and the arid plateaus of the southwest to have higher concentrations. Flood frequency, as defined by SSURGO is the likelihood for a region to be flooded during the year and can be related to the amount of lowland and or adjacent flood plains surrounding the channel. Areas with high flood frequency tend to be regions with high sinuosity or large floodplains with high potential for sediment storage and remobilization when river levels rise. Median precipitation and temperature range describe the climatological factors used in this analysis. Decreasing SSC with increased mean annual precipitation potentially results in a direct dilution of the concentration, however, increasing rainfall also results in increased overall erosion (Ferrier et al., 2013; Langbein & Schumm, 1958). The inverse relation with precipitation could be modulated by vegetation as high rainfall correlates with more dense vegetation and increased forest land use percentages, however, median NDVI was excluded from the final multiple nonlinear regression due to low explanatory power. Nevertheless, it remains reasonable to suggest that high rainfall regions are likely to have more vegetation, limiting hillslope erosion. A positive dependence on temperature range may highlight either the role of variability in temperature dependent erosion processes and/or indicate a more arid location. The soil parameters describe soil erodibility (K-factor), availability (T-factor), and infiltration potential (hydrologic group), respectively. The first two directly relate to the potential quantities of fine particles that can be mobilized, while hydrologic group is a categorical scale which progressively defines soils with finer particle sizes and lower infiltration capacity (i.e., higher runoff potential). The final MNLR model predictions fall along a one-to-one line with 50 and

80% of the predictions within a factor of two and ten, respectively. Maximum and minimum variability of the dataset still cover a large range of values, and MNLR model does not predict the highest mean SSC values. Model predictions do not exceed mean values of 10^3 mg/L whereas the observed data extends up to 10^5 mg/L within the Colorado River and several arroyos and washes within the arid southwest. Overall, the MNLR and variable exploration here demonstrates that mean SSC is predictable across the CONUS region based on geospatial attributes alone. However, the vast variability within mean SSC across the CONUS region remains a considerable challenge.

Regionally, this MNLR model has good predictive power, but there are also regions of overprediction and underprediction. Regions along the Mississippi and Ohio rivers are generally well fit. Some disconnects in prediction accuracy are found within state lines, with Oklahoma and Nebraska generally predicting values higher than the observed values and Idaho and Virginia generally predicting values lower than the observed values, while their neighboring states have a more well fit spread in their site's accuracies. Many of the geological inputs to this regression model are regionally or state-based datasets, and discontinuities between survey borders may change the results for the model's predictions. Sites of major differences may also be sites at the extreme ends of variability within the values for mean SSC, that the model struggles to predict. Arizona is a good example of this, as their values for mean SSC fell within the highest and lowest values, resulting in the largest differences between predictions and observed values. This highlights the importance of the soils data and its state-to-state differences, as well as a need to explore the variability of mean SSC and the input datasets.

CONCLUSION

We created a model to predict total suspended sediment concentration using publicly available near channel river properties. The properties for each USGS water quality gage were gathered at 20% of the length of the channel upstream and buffered five kilometers on either side of the river. The final variables used in the model include median elevation, median 30-year normal precipitation and temperature, erodibility factor, soil tolerance, flood frequency, and hydrologic group. Using these variables allows for a reasonable (within factor two) estimation of the mean SSC for 50% and within a factor of 10 for 80% of the SSC sampling sites across the CONUS region.

More work is needed on this method to fully understand its impacts and improve its predictions. Further research into the relationships between the near-channel distances and river types could draw out more information on sources of variability within these results. With a method like this, more datasets (rainfall intensity) and sites could be added to continue explore SSC values and their variability. Especially exploration of high concentration sites as these are persistently underfit and represent important management concerns. By adding more depth and robustness to this method, measurements of suspended sediment concentration and flux could be predicted and forecasted with geospatial data and potentially track the mean SSC through time as geospatial and climatic variables change. This would have implications for sites with no water quality gaging equipment or sites experiencing changing climate and anthropogenic alterations to the near channel properties.

REFERENCES

Anderson, D. M., & Macdonald, L. H. (1998). Modelling road surface sediment production using a vector geographic information system. *Earth Surface Processes and Landforms*, *23*(2), 95–107. https://doi.org/10.1002/(SICI)1096-9837(199802)23:2<95::AID-ESP849>3.0.CO;2-1

Belmont, P., Gran, K. B., Schottler, S. P., Wilcock, P. R., Day, S. S., Jennings, C., Lauer, J. W., Viparelli, E., Willenbring, J. K., & Engstrom, D. R. (2011). Large shift in source of fine sediment in the Upper Mississippi River. *Environmental Science & Technology*, *45*(20), 8804–8810.

Berger, C., Schulze, M., Rieke-Zapp, D., & Schlunegger, F. (2010). Rill development and soil erosion: A laboratory study of slope and rainfall intensity. *Earth Surface Processes and Landforms*, *35*(12), 1456–1467. https://doi.org/10.1002/esp.1989

Bilotta, G. S., & Brazier, R. E. (2008). Understanding the influence of suspended solids on water quality and aquatic biota. *Water Research*, *42*(12), 2849–2861. https://doi.org/10.1016/j.watres.2008.03.018

Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, *52*(4), 603–679. https://doi.org/10.1002/2012RG000417

Brakebill, J. W., Ator, S. W., & Schwarz, G. E. (2010). Sources of Suspended-Sediment Flux in Streams of the Chesapeake Bay Watershed: A Regional Application of the SPARROW Model1. *JAWRA Journal of the American Water Resources Association*, *46*(4), 757–776. https://doi.org/10.1111/j.1752-1688.2010.00450.x

Burgers, H. E. (Robin), Schipper, A. M., & Jan Hendriks, A. (2014). Size relationships of water discharge in rivers: Scaling of discharge with catchment area, main-stem length and precipitation. *Hydrological Processes*, *28*(23), 5769–5775. https://doi.org/10.1002/hyp.10087

Cammeraat, E. L. H. (2004). Scale dependent thresholds in hydrological and erosion response of a semi-arid catchment in southeast Spain. *Agriculture, Ecosystems & Environment*, *104*(2), 317–332. https://doi.org/10.1016/j.agee.2004.01.032

Cederholm, C. J. (1979). *The effects of logging road landslide sitlation on the salmon and trout spawning gravels of Stequaleho Creek and the Clearwater River Basin, Jefferson County, Washington, 1972-1978. 1*.

Chegini, T., Li, H.-Y., & Leung, L. R. (2021). HyRiver: Hydroclimate Data Retriever. *Journal of Open Source Software*, *6*(66), 3175. https://doi.org/10.21105/joss.03175

Cohen, S., Kettner, A. J., Syvitski, J. P. M., & Fekete, B. M. (2013). WBMsed, a distributed global-scale riverine sediment flux model: Model description and validation. *Computers & Geosciences*, *53*, 80–93. https://doi.org/10.1016/j.cageo.2011.08.011

Cohen, S., Syvitski, J., Ashley, T., Lammers, R., Fekete, B., & Li, H.-Y. (2022). Spatial Trends and Drivers of Bedload and Suspended Sediment Fluxes in Global Rivers. *Water Resources Research*, *58*(6), e2021WR031583. https://doi.org/10.1029/2021WR031583

Coombs, J. S., & Melack, J. M. (2013). Initial impacts of a wildfire on hydrology and suspended sediment and nutrient export in California chaparral watersheds. *Hydrological Processes*, *27*(26), 3842–3851. https://doi.org/10.1002/hyp.9508

Czuba, J. A., & Foufoula-Georgiou, E. (2014). A network-based framework for identifying potential synchronizations and amplifications of sediment delivery in river basins. *Water Resources Research*, *50*(5), 3826–3851. https://doi.org/10.1002/2013WR014227

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Dewitz, J. & U.S. Geological Survey. (2021). *National Land Cover Database (NLCD) 2019 Products* [dataset]. U.S. Geological Survey. https://doi.org/10.5066/P9KZCM54

Drummond, J. D., Schneidewind, U., Li, A., Hoellein, T. J., Krause, S., & Packman, A. I. (2022). Microplastic accumulation in riverbed sediment via hyporheic exchange from headwaters to mainstems. *Science Advances*, *8*(2), eabi9305. https://doi.org/10.1126/sciadv.abi9305

Dutta, S. (2016). Soil erosion, sediment yield and sedimentation of reservoir: A review. *Modeling Earth Systems and Environment*, *2*(3), 123. https://doi.org/10.1007/s40808-016-0182-y

Einsele, G., & Hinderer, M. (1997). Terrestrial sediment yield and the lifetimes of reservoirs, lakes, and larger basins. *Geologische Rundschau*, *86*(2), 288–310. https://doi.org/10.1007/s005310050141

Einstein, H. A., & Chien, N. (1953). Can the rate of wash load be predicted from the bed-load function? *Eos, Transactions American Geophysical Union*, *34*(6), 876–882. https://doi.org/10.1029/TR034i006p00876

Ferrier, K. L., Perron, J. T., Mukhopadhyay, S., Rosener, M., Stock, J. D., Huppert, K. L., & Slosberg, M. (2013). Covariation of climate and long-term erosion rates across a steep rainfall gradient on the Hawaiian island of Kaua‘i. *GSA Bulletin*, *125*(7–8), 1146–1163. https://doi.org/10.1130/B30726.1

Fisher, A., Belmont, P., Murphy, B. P., MacDonald, L., Ferrier, K. L., & Hu, K. (2021). Natural and anthropogenic controls on sediment rating curves in northern California coastal watersheds. *Earth Surface Processes and Landforms*, *46*(8), 1610–1628. https://doi.org/10.1002/esp.5137

Frasson, R. P. de M., Pavelsky, T. M., Fonstad, M. A., Durand, M. T., Allen, G. H., Schumann, G., Lion, C., Beighley, R. E., & Yang, X. (2019). Global Relationships Between River Width, Slope, Catchment Area, Meander Wavelength, Sinuosity, and Discharge. *Geophysical Research Letters*, *46*(6), 3252–3262. https://doi.org/10.1029/2019GL082027

Galster, J. C. (2007). Natural and anthropogenic influences on the scaling of discharge with drainage area for multiple watersheds. *Geosphere*, *3*(4), 260–271. https://doi.org/10.1130/GES00065.1

Garcia, M. (Ed.). (2008). *Sedimentation Engineering: Processes, Measurements, Modeling, and Practice* (110th ed.). American Society of Civil Engineers. https://doi.org/10.1061/9780784408148

Garcia, M., & Parker, G. (1991). Entrainment of bed sediment into suspension. *Journal of Hydraulic Engineering*, *117*(4), 414–435.

Gottselig, N., Bol, R., Nischwitz, V., Vereecken, H., Amelung, W., & Klumpp, E. (2014). Distribution of Phosphorus-Containing Fine Colloids and Nanoparticles in Stream Water of a Forest Catchment. *Vadose Zone Journal*, *13*(7), vzj2014.01.0005. https://doi.org/10.2136/vzj2014.01.0005

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., & Gilroy, E. J. (2020). Statistical methods in water resources. In *Techniques and Methods* (4-A3). U.S. Geological Survey. https://doi.org/10.3133/tm4A3

Hill, K. M., Gaffney, J., Baumgardner, S., Wilcock, P., & Paola, C. (2017). Experimental study of the effect of grain sizes in a bimodal mixture on bed slope, bed texture, and the transition to washload. *Water Resources Research*, *53*(1), 923–941.

Huettel, M., Ziebis, W., Forster, S., & Luther, G. W. (1998). Advective Transport Affecting Metal and Nutrient Distributions and Interfacial Fluxes in Permeable Sediments. *Geochimica et Cosmochimica Acta*, *62*(4), 613–631. https://doi.org/10.1016/S0016-7037(97)00371-2

Jung, B. M., Fernandes, E. H., Möller, O. O., & García-Rodríguez, F. (2020). Estimating Suspended Sediment Concentrations from River Discharge Data for Reconstructing Gaps of Information of Long-Term Variability Studies. *Water*, *12*(9), Article 9. https://doi.org/10.3390/w12092382

Karwan, D. L., Gravelle, J. A., & Hubbart, J. A. (2007). Effects of Timber Harvest on Suspended Sediment Loads in Mica Creek, Idaho. *Forest Science*, *53*(2), 181–188. https://doi.org/10.1093/forestscience/53.2.181

Lamb, M. P., de Leeuw, J., Fischer, W. W., Moodie, A. J., Venditti, J. G., Nittrouer, J. A., Haught, D., & Parker, G. (2020). Mud in rivers transported as flocculated and suspended bed material. *Nature Geoscience*, *13*(8), Article 8. https://doi.org/10.1038/s41561-020-0602-5

Lana-Renault, N., Regüés, D., Martí-Bono, C., Beguería, S., Latron, J., Nadal, E., Serrano, P., & García-Ruiz, J. M. (2007). Temporal variability in the relationships between precipitation, discharge and suspended sediment concentration in a small Mediterranean mountain catchment. *Hydrology Research*, *38*(2), 139–150. https://doi.org/10.2166/nh.2007.003

Langbein, W. B., & Schumm, S. A. (1958). Yield of sediment in relation to mean annual precipitation. *Eos, Transactions American Geophysical Union*, *39*(6), 1076–1084. https://doi.org/10.1029/TR039i006p01076

Leopold, L. B., & Maddock, T. (1953). *The Hydraulic Geometry of Stream Channels and Some Physiographic Implications*. U.S. Government Printing Office.

Lewis, J., Mori, S., Keppeler, E., & Ziemer, R. (2001). Impacts of Logging on Storm Peak Flows, Flow Volumes and Suspended Sediment Loads in Caspar Creek, California. *Water Science and Application*, *2*, 85–125.

Lin, G.-W., Chen, H., Chen, Y.-H., & Horng, M.-J. (2008). Influence of typhoons and earthquakes on rainfall-induced landslides and suspended sediments discharge. *Engineering Geology*, *97*(1), 32–41. https://doi.org/10.1016/j.enggeo.2007.12.001

Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling & Software*, *18*(8), 761–799. https://doi.org/10.1016/S1364-8152(03)00078-1

Merritts, D. J., Vincent, K. R., & Wohl, E. E. (1994). Long river profiles, tectonism, and eustasy: A guide to interpreting fluvial terraces. *Journal of Geophysical Research: Solid Earth*, *99*(B7), 14031–14050. https://doi.org/10.1029/94JB00857

Milligan, T. G., & Loring, D. H. (1997). The Effect of Flocculation on the Size Distributions of Bottom Sediment in Coastal Inlets: Implications for Contaminant Transport. *Water, Air, and Soil Pollution*, *99*(1), 33–42. https://doi.org/10.1023/A:1018307710140

Moore, R. B., McKay, L. D., Rea, A. H., Bondelid, T. R., Price, C. V., Dewald, T. G., & Johnston, C. M. (2019). User's guide for the national hydrography dataset plus (NHDPlus) high resolution. In *Open-File Report* (2019–1096). U.S. Geological Survey. https://doi.org/10.3133/ofr20191096

Palmer, M. A., Covich, A. P., Lake, S., Biro, P., Brooks, J. J., Cole, J., Dahm, C., Gibert, J., Goedkoop, W., Martens, K., Verhoeven, J., & Van De Bund, W. J. (2000). Linkages between Aquatic Sediment Biota and Life Above Sediments as Potential Drivers of Biodiversity and Ecological Processes: A disruption or intensification of the direct and indirect chemical, physical, or biological interactions between aquatic sediment biota and biota living above the sediments may accelerate biodiversity loss and contribute to the degradation of aquatic and riparian habitats. *BioScience*, *50*(12), 1062–1075. https://doi.org/10.1641/0006-3568(2000)050[1062:LBASBA]2.0.CO;2

Phillips, C. B., Dallmann, J. D., Jerolmack, D. J., & Packman, A. I. (2019). Fine-particle deposition, retention, and resuspension within a sand-bedded stream are determined by streambed morphodynamics. *Water Resources Research*, *55*(12), 10303–10318.

PRISM Climate Group. (2014). *PRISM Gridded Climate Data* [dataset]. https://prism.oregonstate.edu

Römkens, M. J. M., Helming, K., & Prasad, S. N. (2002). Soil erosion under different rainfall intensities, surface roughness, and soil water regimes. *CATENA*, *46*(2), 103–123. https://doi.org/10.1016/S0341-8162(01)00161-8

Rouse, H. (1937). Modern Conceptions of the Mechanics of Fluid Turbulence. *Transactions of the American Society of Civil Engineers*, *102*(1), 463–505. https://doi.org/10.1061/TACEAT.0004872

Rubin, D. M., Buscombe, D., Wright, S. A., Topping, D. J., Grams, P. E., Schmidt, J. C., Hazel Jr., J. E., Kaplinski, M. A., & Tusso, R. (2020). Causes of Variability in Suspended-Sand Concentration Evaluated Using Measurements in the Colorado River in Grand Canyon. *Journal of Geophysical Research: Earth Surface*, *125*(9), e2019JF005226. https://doi.org/10.1029/2019JF005226

Rubin, D. M., & Topping, D. J. (2001). Quantifying the relative importance of flow regulation and grain size regulation of suspended sediment transport α and tracking changes in grain size of bed sediment β. *Water Resources Research*, *37*(1), 133–146. https://doi.org/10.1029/2000WR900250

Sadler, P. M., & Jerolmack, D. J. (2015). Scaling laws for aggradation, denudation and progradation rates: The case for time-scale invariance at sediment sources and

sinks. *Geological Society, London, Special Publications*, *404*(1), 69–88. https://doi.org/10.1144/SP404.7

Schleiss, A. J., Franca, M. J., Juez, C., & De Cesare, G. (2016). Reservoir sedimentation. *Journal of Hydraulic Research*, *54*(6), 595–614. https://doi.org/10.1080/00221686.2016.1225320

Schwartz, J. S., Simon, A., & Klimetz, L. (2011). Use of fish functional traits to associate in-stream suspended sediment transport metrics with biological impairment. *Environmental Monitoring and Assessment*, *179*(1), 347–369. https://doi.org/10.1007/s10661-010-1741-8

Schwarz, G., Hoos, A. B., Alexander, R. B., & Smith, R. A. (2006). Section 3. The SPARROW Surface Water-Quality Model—Theory, application and user documentation. In *Techniques and Methods* (6-B3). U.S. Geological Survey. https://doi.org/10.3133/tm6B3

Shen, H. W., & Julien, P. Y. (1992). Erosion and sediment transport. *Handbook of Hydrology.*, 12–61.

Stout, J. C., Belmont, P., Schottler, S. P., & Willenbring, J. K. (2014). Identifying Sediment Sources and Sinks in the Root River, Southeastern Minnesota. *Annals of the Association of American Geographers*, *104*(1), 20–39. https://doi.org/10.1080/00045608.2013.843434

Syvitski, J. P. M., & Milliman, J. D. (2007). Geology, Geography, and Humans Battle for Dominance over the Delivery of Fluvial Sediment to the Coastal Ocean. *The Journal of Geology*, *115*(1), 1–19. https://doi.org/10.1086/509246

Topping, D. J., Schmidt, J. C., & Vierra Jr, L. E. (2003). *Computation and analysis of the instantaneous-discharge record for the Colorado River at Lees Ferry, Arizona—May 8, 1921, through September 30, 2000*. US Geological Survey.

Trabucco, A., & Zomer, R. J. (2018). Global aridity index and potential evapotranspiration (ET0) climate database v2. *CGIAR Consort Spat Inf*, *10*, m9.

U.S. Geological Survey. (2019). *USGS 3D Elevation Program Digital Elevation Model* [dataset]. https://elevation.nationalmap.gov/arcgis/rest/services/3DEPElevation/ImageServer

Vaughan, A. A., Belmont, P., Hawkins, C. P., & Wilcock, P. (2017). Near-Channel Versus Watershed Controls on Sediment Rating Curves. *Journal of Geophysical Research: Earth Surface*, *122*(10), 1901–1923. https://doi.org/10.1002/2016JF004180

Vörösmarty, C. J., Meybeck, M., Fekete, B., Sharma, K., Green, P., & Syvitski, J. P. M. (2003). Anthropogenic sediment retention: Major global impact from registered river impoundments. *Global and Planetary Change*, *39*(1), 169–190. https://doi.org/10.1016/S0921-8181(03)00023-7

Wohl, E. E., & Cenderelli, D. A. (2000). Sediment deposition and transport patterns following a reservoir sediment release. *Water Resources Research*, *36*(1), 319–333. https://doi.org/10.1029/1999WR900272

Wolman, M. G. (1967). A Cycle of Sedimentation and Erosion in Urban River Channels. *Geografiska Annaler: Series A, Physical Geography*, *49*(2–4), 385–395. https://doi.org/10.1080/04353676.1967.11879766

Zabaleta, A., Martínez, M., Uriarte, J. A., & Antigüedad, I. (2007). Factors controlling suspended sediment yield during runoff events in small headwater catchments of the Basque Country. *CATENA*, *71*(1), 179–190. https://doi.org/10.1016/j.catena.2006.06.007

FIGURES

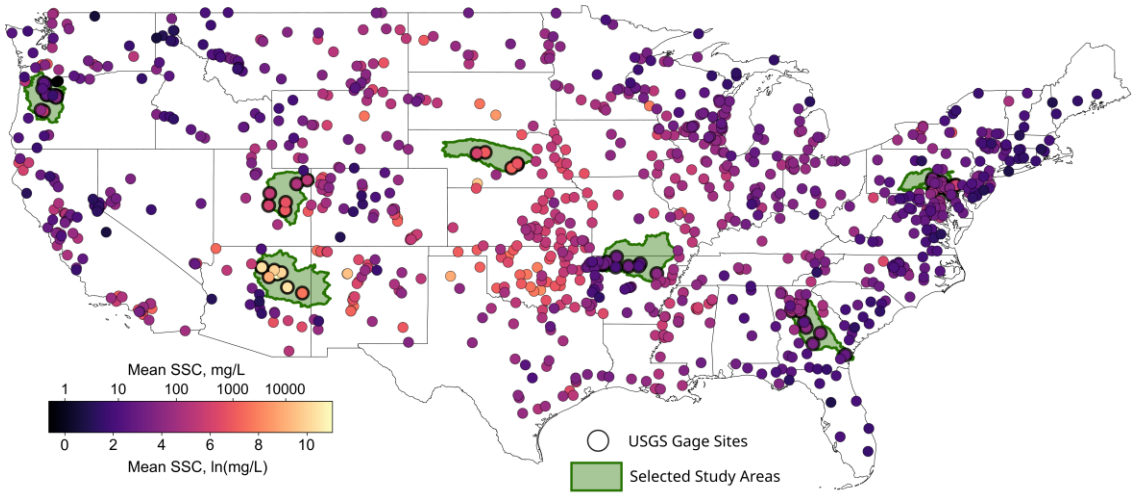

Figure 1. Map of 966 Suspended Sediment Concentration sites from the USGS National Water Information Service across the CONUS region.
The points are shaded by the mean of each site's natural log transformed concentration. A spatial pattern of high and low concentrations can be seen across the CONUS region with particularly high concentrations occurring in the plains and southwest. Green regions are the initially selected HUC6 basins for exploratory analysis and analysis algorithm development. These basins include: Willamette-OR, Little Colorado-AZ, Lower Green-UT, Loup-NE, Upper White-AR/MO, Altamaha-GA, and Lower Susquehanna-PA (west-to-east).

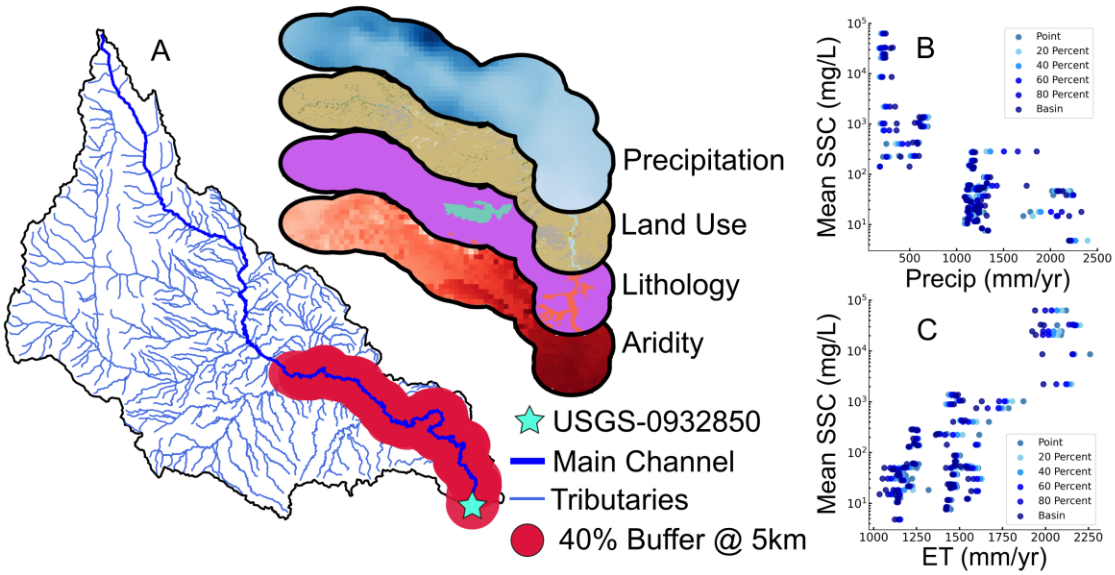

Figure 2. Example of methods for near-channel extraction of geospatial attributes for the basin upstream of the USGS gage 0932850 (San Rafael River near Green River, UT). The main channel of the river is the thicker blue line while the tributaries are thinner blue lines (A). In this example, the buffered region represents 40% of the total main channel length upstream and extends 5 kilometers on either side of the main channel. Geospatial attributes are extracted and quantified for datasets within the buffer, with examples shown for the mean annual 30-year normal precipitation, land use, lithology, and calculated aridity. Scatterplots of precipitation (B) and evapotranspiration (C) against mean SSC are shown for the sites within the selected basins (n=76). Points are colored by their extraction method, from a point scale, along increments of buffer distances (20%, 40%, 60%, 80%), and the whole basin.

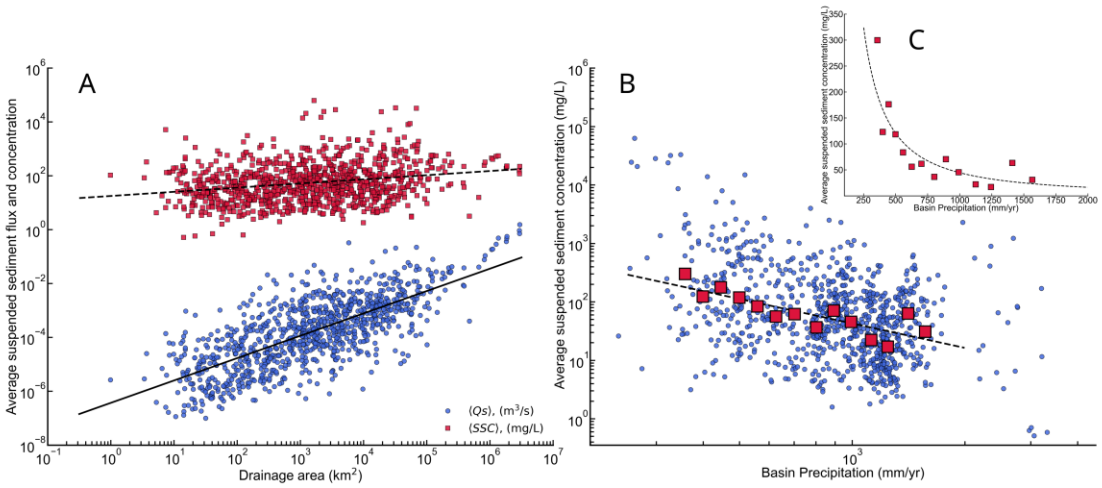

Figure 3. Site mean SSC (n=966) correlates weakly with drainage area while sediment flux displays a strong correlation across over eight orders of magnitude of drainage area. (A) The fitted trend lines are power functions of drainage area for both SSC (mg/L, red squares) and flux (Qs, m^3/s, blue circles). (B) SSC plotted against 30-year normal precipitation (mm/yr) averages across each basin. The trend line is fit to log spaced binned median values (red squares) for bins containing more than 10 points. (C) Binned median data and the fitted trend line for linear scale axes to highlight the strong curvature within the relation between precipitation and SSC.

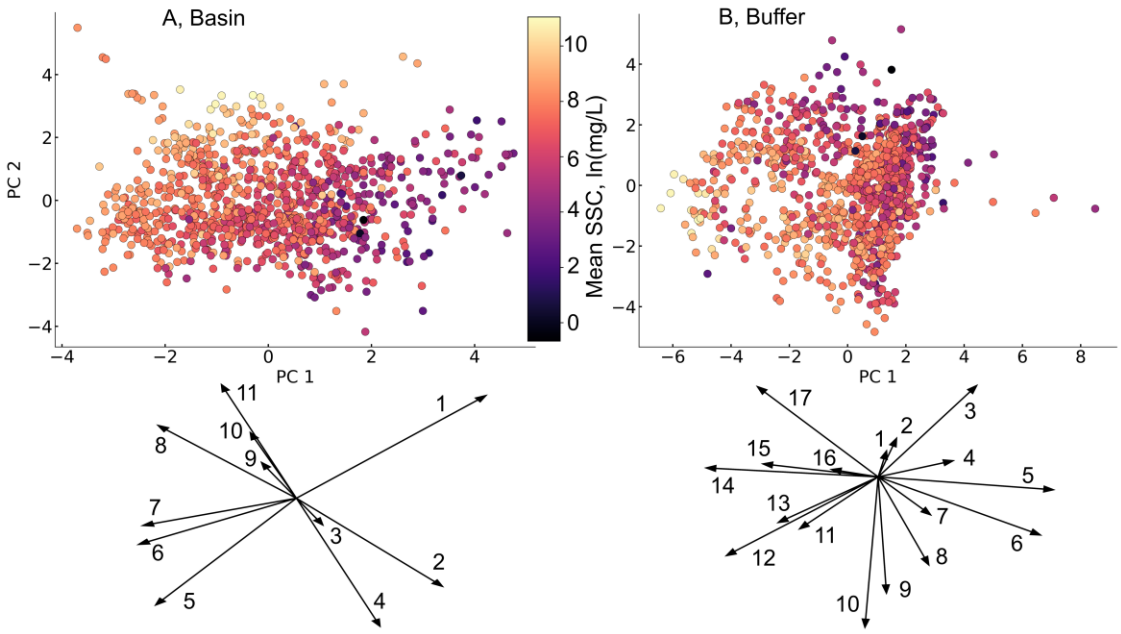

Figure 4. Principal component analysis (PCA) of basin (left) and near-channel (right datasets.

Each PCA is shaded by the mean of the natural log transformed suspended sediment concentration (n=966). Variables included in each PCA are listed with their corresponding vector components (vector diagrams are offset from the zero-zero point for clarity). The vector diagrams below each PCA denote the projected orientation of each variable within the data and the length of the vector denotes the overall importance of the variable for explaining the variance within the dataset. For the basin (left panels), datasets are: percent sand (1), percent forest (2), percent urban (3), annual precipitation (4), percent silt (5), percent clay (6), percent agriculture (7), mean SSC (8), number of SSC samples (9), number of upstream dams (10), and drainage area (11). The first four components of the basin data explain 65.5% of the total variance. For the near-channel (right panels), datasets are: slope (1), k-factor (2), NDVI TIN (3), percent forest (4), aridity (5), annual precipitation (6), percent urban (7), soil tolerance (8), flooding frequency (9), temperature (10), drainage area (11), evapotranspiration (12), SSC (13), temperature range (14), percent agriculture (15), hydrologic group (16), and median elevation (17). The first four components of the near-channel data explain 63% of the variance.

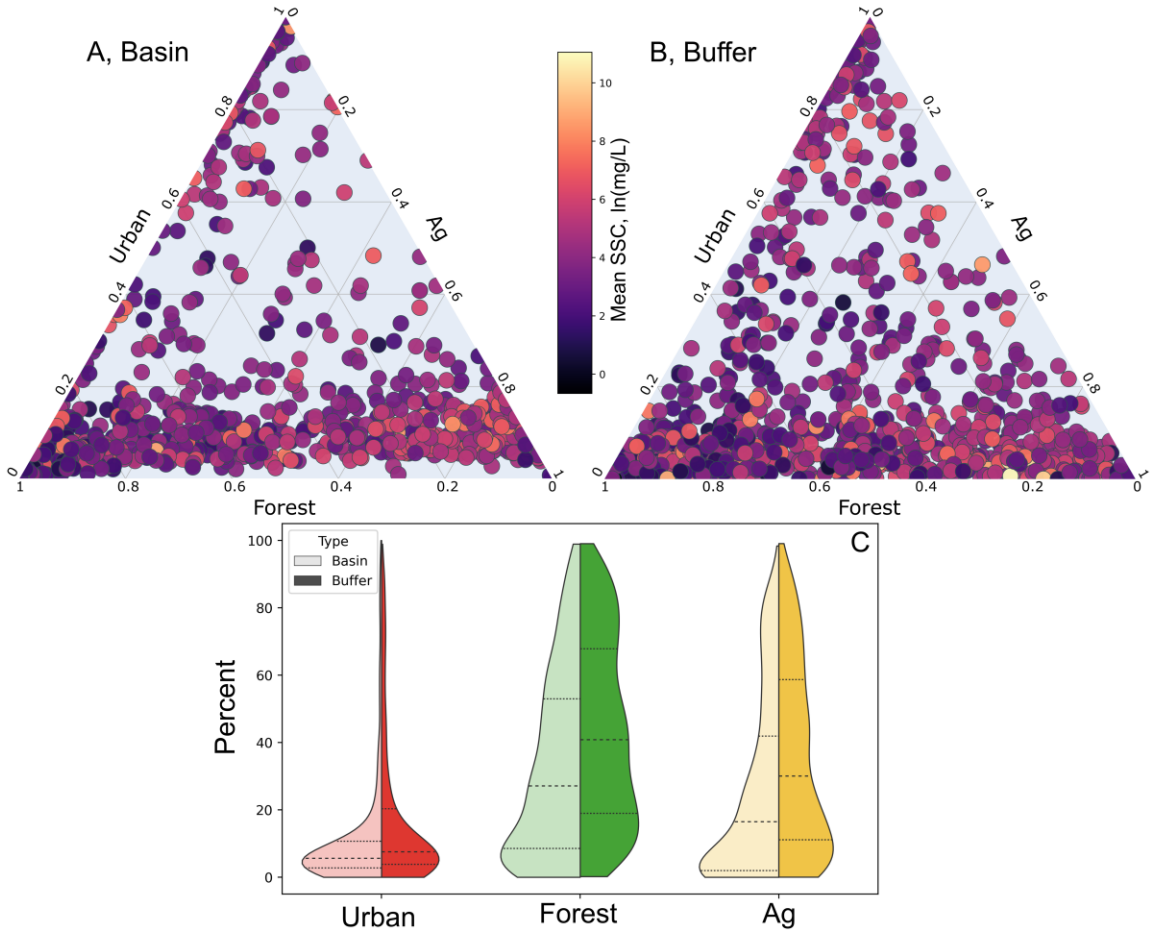

Figure 5. Relation between mean SSC across three land use categories for the basin (A) and near-channel (B) scales.

Ternary plots of land use show percentages of the three categories of land use: urban, forest and agricultural (Ag). Comparison between the two panels highlights how land use changes between the different spatial scales. The points are shaded by the mean of the natural log-transformed SSC (n=966). (C) Split violin plots showing the overall change within the distributions of land use between basin (lighter shade on left) and near-channel (darker shade on right).

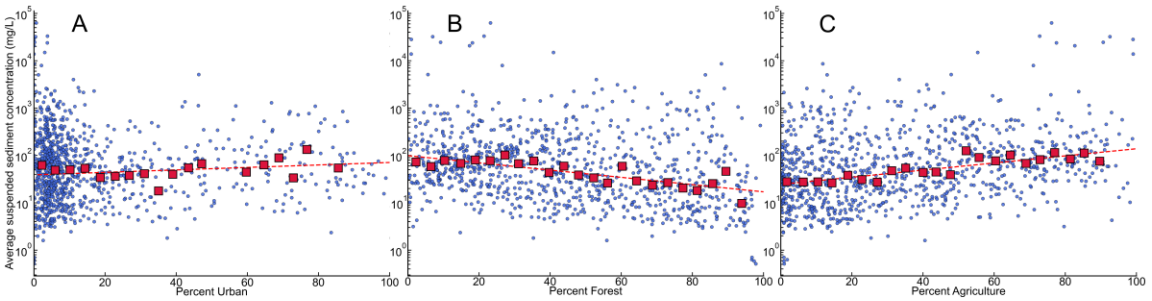

Figure 6. Individual relations between land use categories and mean SSC at the near-channel scale.

Scatterplots of all buffer-extracted points with their land use percentages, urban (A), forest (B), agriculture (C), against suspended sediment concentration (mg/L) (n=966). The red squares represent the medians of log-spaced binned data for all bins containing more than 10 points. An exponential function was fit to the binned medians (red dotted line).

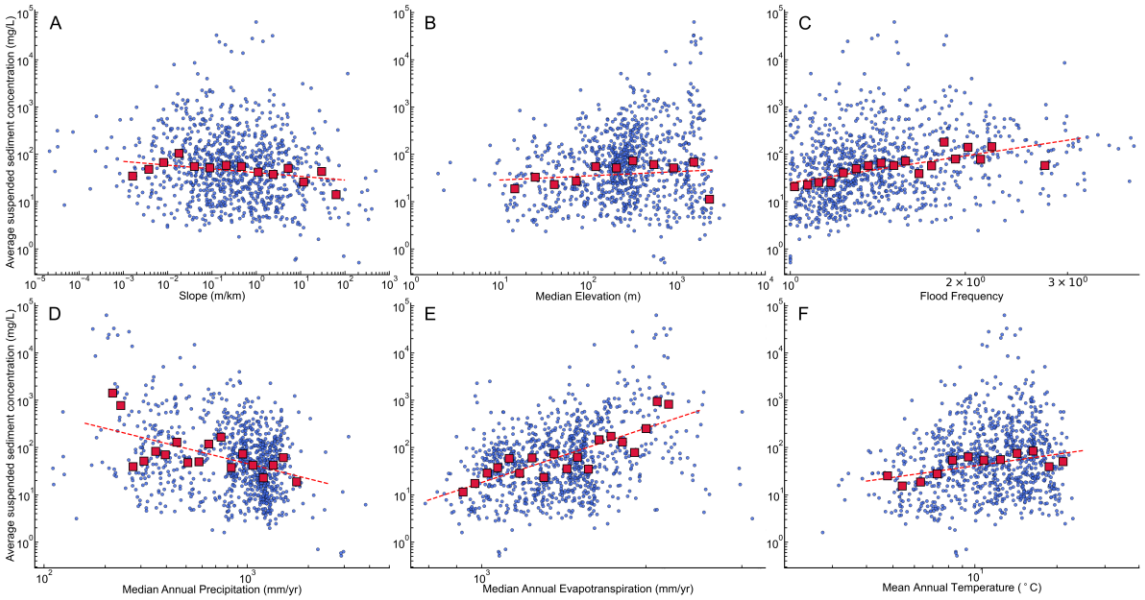

Figure 7. Relations between mean SSC and selected topographical and climatological variables.

Scatterplots of all buffered points plotted against suspended sediment concentration (mg/L) (n=966). The variables are: median slope (A), median elevation (B), flood frequency (C), median annual precipitation (D), median annual evapotranspiration (E), and annual temperature (F). The red squares represent the binned medians of log-spaced bins containing more than 10 points. Each regression is a power function fit to the binned medians (red dotted line).

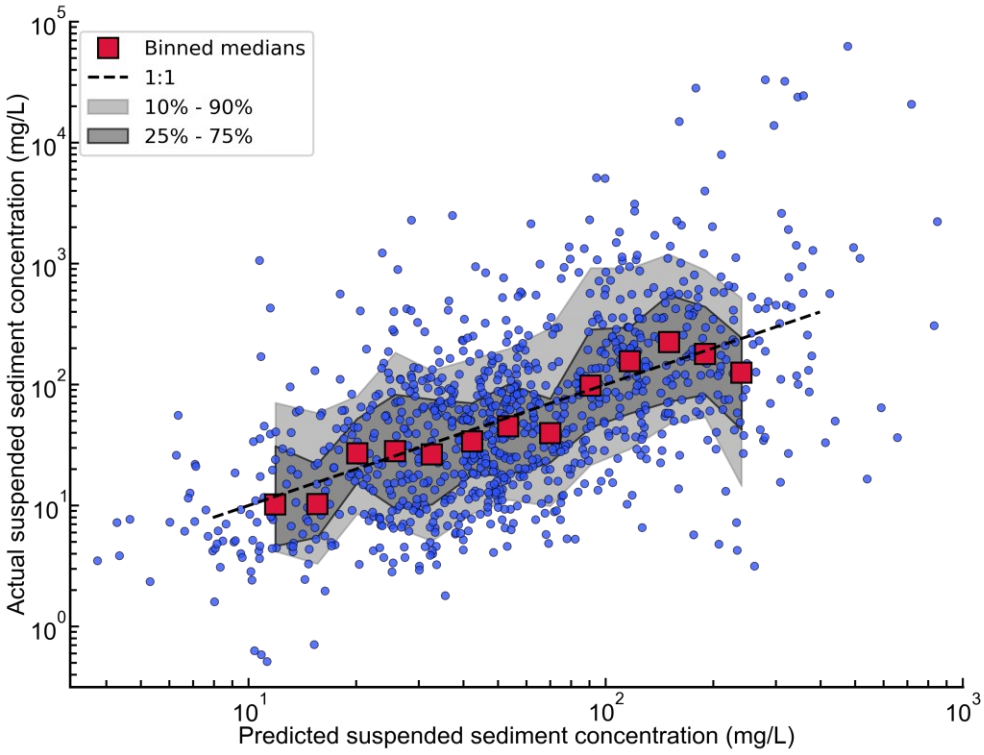

Figure 8. Multiple nonlinear regression model prediction against observed mean SSC. Black dashed line represents the 1:1 line. Log-spaced bins containing more than 25 points show median values (red squares) and contain 875/966 total sites. Data are shaded by the interquartile range (dark gray, contains 50% of total data within binned range) and 10^{th} to 90^{th} percentiles (light gray, contains 80% of total data within binned range) of the data along the binning region. Data are binned here to highlight how the center of the data accurately follows the 1:1 line.

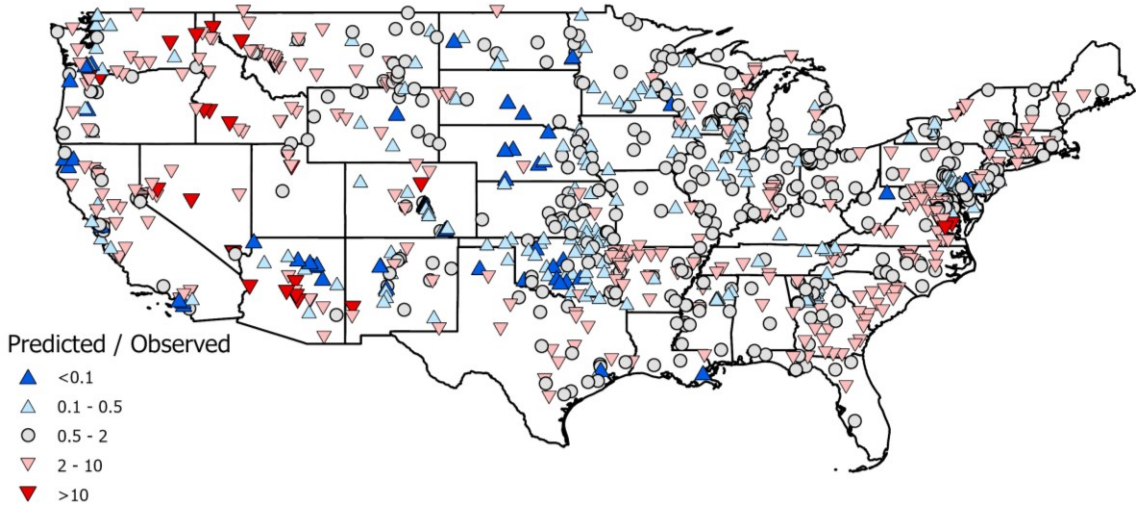

Figure 9. Results of multiple nonlinear regression model across the CONUS region. Triangles pointing up are where observed SSC values are a factor of 2 (light blue) or factor of 10 (blue) greater than those predicted in the model. Circles (gray) are within a factor of 2, and triangles pointing down are where observed SSC values are a factor of 2 (light pink) and factor of 10 (red) less than the predicted values.

APPENDIX

Appendix A. List of Variables Extracted (78)

Variables preceded by USGS were calculated by the USGS.

Latitude, longitude, number of water quality samples, drainage area, basin annual precipitation, USGS percent urban, USGS percent forest, USGS percent agricultural, USGS number of major dams upstream, USGS percent clay, USGS percent sand, USGS percent silt, main channel length.

The following were calculated based on the SSC data:geometric mean SSC, SSC geometric standard deviation, geometric mean sediment flux, sediment flux geometric standard deviation

Near channel elevation mean, median, maximum, minimum, 10^{th} percentile, 90^{th} percentile, standard deviation, range, inner quartile range

Near channel percent urban, forest, agriculture

Near channel annual precipitation mean, median, maximum, minimum, 10^{th} percentile, 90^{th} percentile, standard deviation, range, inner quartile range

Near channel annual evapotranspiration mean, median, maximum, minimum, 10^{th} percentile, 90^{th} percentile, standard deviation, range, inner quartile range

Near channel mean aridity

Near channel annual temperature mean, median, standard deviation

Near channel maximum annual temperature mean, median, standard deviation

Near channel minimum annual temperature mean, median, standard deviation

Near channel temperature range (maximum annual mean temperature minus minimum annual temperature mean)

Near channel slope

Near channel TIN (Time integrated NDVI) mean and standard deviation for years 2010, 2011, 2012, 2013, 2014, and all averaged

Near channel k-factor mean and standard deviation

Near channel soil tolerance mean and standard deviation

Near channel flooding frequency mean and standard deviation

Near channel hydrologic group

Appendix B. List of Variables After PCA (14)

Drainage Area

Near channel median elevation

Near channel percent urban, forest, agricultural

Near channel median annual precipitation

Near channel median annual evapotranspiration

Near channel mean aridity

Near channel median temperature

Near channel temperature range

Near channel slope

Near channel TIN

Near channel K-factor

Near channel soil tolerance

Near channel flooding frequency

Near channel hydrologic group

Appendix C. List of Final Variables (7) within Multiple nonlinear regression model

Near channel median elevation

Near channel median annual precipitation

Near channel median annual temperature

Near channel K-factor

Near channel soil tolerance

Near channel flooding frequency

Near channel hydrologic group

Appendix D. Supplemental Figures

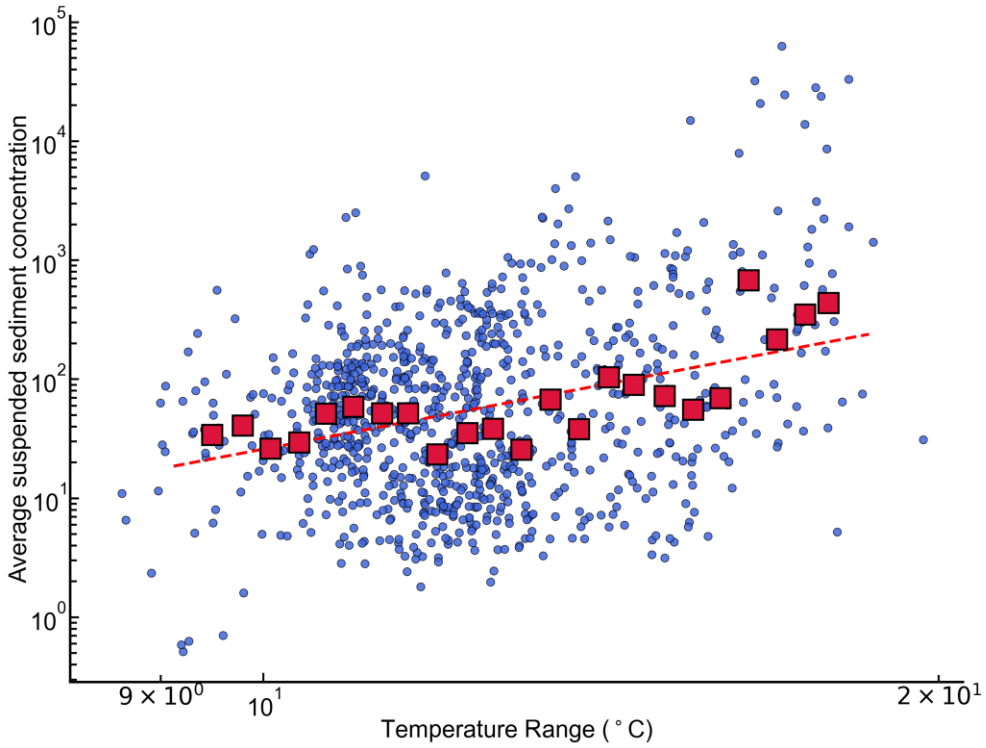

Figure 10. Relation between mean SSC and temperature range.
Scatterplot of all buffered points plotted against suspended sediment concentration (mg/L) (n=966). The red squares represent the binned medians of log-spaced bins containing more than 10 points. Each regression is a power function fit to the binned medians (red dotted line).