Advances in Water Resources 119 (2018) 188–196

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Stochastic modeling of suspended sediment load in alluvial rivers

Shahab Aldin Shojaeezadeh[a], Mohammad Reza Nikoo[a], James P. McNamara[b], Amir AghaKouchak[c,d], Mojtaba Sadegh[e,*]

[a] Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran
[b] Department of Geosciences, Boise State University, Boise, Idaho, USA
[c] Department of Civil and Environmental Engineering, University of California, Irvine, California, USA
[d] Department of Earth System Sciences, University of California, Irvine, California, USA
[e] Department of Civil Engineering, Boise State University, Boise, ID, USA

ARTICLE INFO

Keywords:
Sediment transport
Suspended sediment load
Stochastic modeling
Copula
Uncertainty analysis

ABSTRACT

Sediment is a major source of non-point pollution. Suspended sediment can transport nutrients, toxicants and pesticides, and can contribute to eutrophication of rivers and lakes. Modeling suspended sediment in rivers is of particular importance in the field of environmental science and engineering. However, understanding and quantifying nonlinear interactions between river discharge and sediment dynamics has always been a challenge. In this paper, we introduce a parsimonious probabilistic model to describe the relationship between Suspended Sediment Load (SSL) and discharge volume. This model, rooted in multivariate probability theory and Bayesian Network, infers conditional marginal distribution of SSL for a given discharge level. The proposed framework relaxes the need for detailed information about the physical characteristics of the watershed, climatic forcings, and the nature of rainfall-runoff transformation, by drawing samples from the probability distribution functions (PDFs) of the underlying process (here, discharge and SSL data). Discharge and SSL PDFs can be simplified into a joint distribution that describes the relationship between SSL and discharge, in which the latter acts as a proxy for different predictors of SSL. The joint distribution is created based on historical discharge and SSL data, and stores information about the discharge-SSL relationship and sediment transport process of the watershed of interest. We test this framework for seven major rivers in the U.S., results of which show promising performance to predict SSL and its likelihood given different discharge levels.

1. Introduction

Understanding and quantifying nonlinear interactions between river discharge and sediment dynamics, and its short- and long- term impacts on the river morphology, and terrestrial and aquatic ecosystems is challenging, given the heterogeneity of watersheds, stochastic nature of turbulent flow, and characteristics of erosion processes (Garcia, 2008; Malmon et al., 2002). Sediment transport and deposition can induce major environmental problems such as deteriorating aquatic habitat and surface water quality, filling reservoirs, redirecting the course of streams, widening floodplains of rivers, and destroying buildings and properties (Best and Bristow, 1993; Morris and Fan, 1998; Parker and Ikeda, 1989). Hence, understanding and modeling sedimentation processes in streams and rivers is important in different fields of study, including environmental quality and resource management, river management and river transportation (Aytek and Kişi, 2008; Lovejoy et al., 1997; Schumm et al., 1984).

Sediment transport can occur in two general forms: 1. suspended load, and 2. bed load, which in great part depends on sediment particle size distribution (Garcia, 2008). For example, cohesive clay particles form consolidated particles, which can increase critical shear stress and subsequently fall velocity of the consolidated particles (Berlamont et al., 1993). This property decreases sediment transport rate and hence lessen suspended sediment concentration (Wiberg and Smith, 1987). On the other hand, silt particles have lower fall velocity compared to the consolidated clay particles. Silt is usually eroded from watershed surface and is dominantly transported as suspended load in rivers (Bagnold, 1962). Sand contents, on the contrary, mostly source from the river bed, and their suspension depends on discharge value and shape of the river (Garcia, 2008). Moreover, land use and cover can also affect sediment transport. Urban land cover, for example, can augment impervious area and increase flood peaks, which in turn decreases erosion at the source but enhances suspension of bed load (Wolman, 1967). Agricultural land use can also increase watershed's sediment yield as a result of removing natural vegetal cover and disturbing soil surface (Walling, 1999). On

* Corresponding author.
E-mail address: mojtabasadegh@boisestate.edu (M. Sadegh).

https://doi.org/10.1016/j.advwatres.2018.06.006
Received 2 March 2018; Received in revised form 12 June 2018; Accepted 28 June 2018
Available online 26 July 2018

the contrary, forest can protect land surface from erosion and diminish soil mobility (Leopold, 1968). These parameters along with flow shear stress govern sediment transport in rivers (Yang, 1996).

Sediment transport modeling is commonly interconnected to hydraulic engineering (Yang, 1996). Traditionally, physical models have been used to tackle this challenge (Van Rijn, 1984; Vanoni, 2006), but more recently computational and empirical models have gained popularity with the advancement of computational power (Garcia, 2008; Morgan et al., 1998; Warner et al., 2008). Also, black-box approaches such as artificial intelligence techniques and machine learning algorithms have recently been used in various fields of environmental modeling, including sediment transport (Cobaner et al., 2009; Kisi et al., 2012). Stochastic approaches for modeling sediment transport and yield, however, received less attention, and were only sparsely used to account for processes related to rivers' hydraulic properties, such as turbulence effect on incipient motion of bed load material, particle motion on a mobile bed, and particle trajectory estimation (Kleinhans and van Rijn, 2002; Lisle et al., 1998; Van Rijn et al., 1993).

In this study, we introduce a stochastic approach for modeling Suspended Sediment Load (SSL) in alluvial rivers. This event-based probabilistic approach draws insights from the multivariate probability theory and Bayesian Networks, and estimates the conditional marginal distribution of SSL in a high flow event given the discharge volume. SSL consists of wash load (from upstream erosion) and suspended bed load, which are both dependent on stochastic processes such as flow turbulence, and precipitation intensity, energy and angle, among others (Garcia, 2008). SSL is highly interrelated with high flow volume (Chalov et al., 2015). In fact, a majority of watersheds sediment yield is transported during flood events (Rankl, 2004). We model this correlation structure using copulas, and exploit this dependence structure to estimate SSL based on discharge volume. This approach has a sound probabilistic basis, and estimates likelihood of different SSL levels, which in turn provides insights for uncertainty quantification of predictions. The proposed approach characterizes the underlying data generating Probability Distribution Function (PDF), from which random samples are drawn to construct the conditional marginal distribution of SSL. The underlying data generating PDF is constructed from historical observed data and stores key components of the watersheds' sedimentation processes (such as slope, vegetation cover, and soil characteristics, among others). This framework not only relaxes the need for detailed information about watershed characteristics and climatic forcing, but also requires a significantly lower computational cost to model SSL compared to its computational and physical counterparts. Note that in this study, the joint distribution of SSL and discharge has been exploited given the availability of long-term observations of both variables. One may construct the joint distribution of SSL, discharge, velocity profile and pressure field, among others; and condition the marginal distribution of SSL on multiple drivers for a more confined uncertainty range. However, lack of long-term observation of such detailed information might be prohibitive.

2. Methodology

2.1. Stochastic modeling of environmental processes

Environmental models strive to capture response of a system to some drivers, which might not be necessarily unique (Sadegh et al., 2018a). If the relationship between observed inputs, $\mathbf{R} = \{r_1, r_2, \cdots, r_n\}$, states, $\mathbf{X}$, and system response, $\mathbf{Y}$, is stochastic, the modeling approach should also be stochastic. Indeed, most environmental modeling frameworks should adopt a stochastic nature, considering the random characteristics of the measurement errors. The underlying data generating process for an environmental system, then, can be described as a conditional PDF, $F(\mathbf{Y}|\mathbf{X},\mathbf{R})$, derived from the joint PDF $F(\mathbf{Y},\mathbf{X},\mathbf{R})$, constrained by the conservation principles and the physical limitations of the system (Bulygina and Gupta, 2009; 2010; Bulygina et al., 2009; 2011).

2.2. Bayesian network

We can employ Bayesian Networks (BNs) to derive the underlying data generating conditional PDF, $F(.|.)$, for environmental modeling. BNs are probabilistic graphical models that describe a set of random variables, $\mathbf{R}$, and their conditional dependencies, $F(r_i|r_j)$, through a directed acyclic graph (Thulasiraman and Swamy, 1992). The joint density function, $f(.)$, of random variables, $\mathbf{R}$, that collectively form a BN, given an ordering similar to that of the parent variables, can be described as Russel and Norvig (2003),

$$f(\mathbf{R}) = \prod_{\forall r_j \in \mathbf{R}} f(r_j(n)|r_j(n-1), \cdots, r_j(1)), \tag{1}$$

which using the chain rule in the probability theory can be reorganized to Madadgar and Moradkhani (2013),

$$f(r_j(n)|r_j(n-1), \cdots, r_j(1)) = \frac{f(r_j(n), r_j(n-1), \cdots, r_j(1))}{f(r_j(n-1), \cdots, r_j(1))}. \tag{2}$$

The joint density function, $f(.)$, in Eq. 2 can then be modeled using copulas.

2.3. Copulas

Copulas are mathematical functions that describe the dependence structure between two, or more, time-independent random variables, regardless of their marginal distributions (Joe, 2014). A bivariate copula, informally defined, maps from a space $\mathbf{I} \times \mathbf{I}$ ($\mathbf{I} \in [0, 1]$) to the space $\mathbf{I}$ (Nelsen, 2007). If $F_1(r_1) = u_1$ and $F_2(r_2) = u_2$ are marginal distributions of random variables r_1 and r_2, according to Sklar's theorem (Sklar, 1959) there is a copula that explains their multivariate cumulative distribution function,

$$F(r_1, r_2) = C(F_1(r_1), F_2(r_2)) = C(u_1, u_2), \tag{3}$$

and the copula function, C, is unique if the marginal distributions, F_1 and F_2, are continuous. The copula density function, $c(.)$, can then be simply derived as Sadegh et al. (2018b),

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}. \tag{4}$$

The joint density function can be decomposed as Madadgar and Moradkhani (2013),

$$f(r_1, r_2) = c(u_1, u_2) \prod_{i=1}^{2} f_i(r_i), \tag{5}$$

which reorganizes Eq. 2 to Madadgar et al. (2017); Mazdiyasni et al. (2017),

$$f(r_2|r_1) = \frac{c(u_1, u_2) f(r_2) f(r_1)}{f(r_1)} = c(u_1, u_2) f(r_2) \tag{6}$$

This framework can be readily extended to higher dimensions. To find the best copula model to describe the dependence structure between SSL and discharge, we employ the 26 copula families of the Multivariate Copula Analysis Toolbox (MvCAT, Sadegh et al. (2017)), namely: 1. Gaussian, 2. t-, 3. Clayton, 4. Frank, 5. Gumbel, 6. Independence, 7. Ali-Kikhail-Haq (AMH), 8. Joe, 9. Farlie-Gumbel-Morgenstern (FGM), 10. Gumbel-Barnett, 11. Plackett, 12. Cuadras-Auge, 13. Raftery, 14. Shih-Louis, 15. Linear-Spearman, 16. Cubic, 17. Burr, 18. Nelsen, 19. Galambos, 20. Marshall-Olkin, 21. Fischer-Hinzmann, 22. Roch-Alegre, 23. Fischer-Kock, 24. BB1, 25. BB5, and 26. Tawn copulas (refer to Table 1 of Sadegh et al. (2017) for more information). Also, to find the best marginal distribution of the observed data, r_1 and r_2, we use 17 different continuous marginal distribution functions, including: 1. Beta, 2. Birnbaum-Saunders, 3. Exponential, 4. Extreme value, 5. Gamma, 6. Generalized extreme value, 7. Generalized Pareto, 8. Inverse Gaussian, 9. Logistic, 10. Log-logistic, 11. Lognormal, 12. Nakagami, 13. Normal, 14. Rayleigh, 15. Rician, 16. t location-scale, and 17. Weibull distributions (Bowman and Azzalini, 1997; Johnson et al., 1993; 1994).

Table 1
Average annual SSL and discharge of the nine USGS stations of this study, and the fraction of annual SSL transported during the high flow ($>70^{th}$ percentile of discharge distribution) events. Last column presents Pearson correlation coefficient between high flow events' SSL and discharge volume.

River	Location	USGS station number	Average		Fraction of SSL in high flow events (%)	Pearson correlation coefficient: SSL and discharge
			SSL $\times 10^6$ (ton/year)	Discharge (km^3/year)		
Colorado (A)	near Grand Canyon, AZ	9402500	111.5	14.27	86.4	0.907
Colorado (B)	near Cisco, UT	9180500	9.15	6.09	87.8	0.954
Mississippi (C)	at Mcgregor, IA	5389500	1.6	37.37	68.3	0.851
Hudson (D)	at Stillwater, NY	1331095	0.078	5.87	89.0	0.889
Sacramento (E)	near Freeport, CA	11447650	1.95	20.24	86.0	0.908
Missouri (F)	at Hermann, MO	6934500	104.03	64.36	80.2	0.903
San Joaquin (G)	near Vernalis, CA	11303500	0.313	3.65	78.5	0.949
Arkansas (H)	at Arkansas City, KS	7146500	2.38	1.82	95.1	0.966
Mississippi (I)	at Thebes, IL	7022000	97.84	219.76	78.8	0.946

2.4. Modeling SSL given discharge volume

High flow events are major contributors (70-80%) of the annual sediment yield of most basins (Chalov et al., 2015). We therefore focused our attention on modeling the relationship between discharge volume and SSL during high flow events, as defined by the daily discharge levels exceeding x^{th} percentile of the long-term discharge series. We have considered different levels of $x = \{65\text{th}, 70\text{th}, \cdots, 90\text{th}\}$ percentiles of discharge series, and selected 70^{th} percentile as a good representative for the threshold of high flows. This threshold warrants that the high flow events are associated with a majority of the sediment load of the watersheds (Table 1). Indeed, ~70-90% (Table 1, column 6) of annual suspended sediment yield of our case studies are transported in such events (more in Section 3).

Our proposed framework is graphically presented in Fig. 1. We first identify high flow events (as defined by discharge above 70^{th} percentile of historical flow record, Fig. 1.1) and devise a set of discharge volumes and associated SSLs from these events (Fig. 1.2). We then fit marginal distributions to the SSL and discharge volume observations (Figs. 1.3.1 and 3.2), which are in turn used in constructing copula models (Fig. 1.4). We then, using fitted marginal and copula models, exploit Eq. 6 to derive the conditional marginal distribution of SSL given discharge volume (Fig. 1.5).

It is worth mentioning that we have carefully examined the high flow events for temporal independence. Each single event is at least two days apart from the following and preceding events. The SSL associated with each flood event is estimated as total sediment load during the extended period of the event, which includes one day before the start and one day after the finish date of the event. Discharge volume and SSL show significant inter-dependency (at 1% significance level) as documented by the Pearson correlation coefficient in Table 1 (last column). It is also noteworthy that we have tested the historical data for hysteresis effects (Buendia et al., 2016; Sherriff et al., 2016), and concluded that such effect is negligible given the watershed sizes and average annual discharge of the rivers (Table 1, column 5).

3. Case studies

We test the capability of the proposed model (Fig. 1) for prediction of SSL given discharge volumes over seven major rivers (nine United States Geological Survey, USGS, stations), geographically dispersed across the contiguous U.S. We carefully select stations, obtained from the sediment portal of the USGS (https://cida.usgs.gov/sediment/), with a historical record of at least 15 years of daily river discharge and SSL. These stations are listed in Table 1, and Fig. 2 displays their geographical locations in the U.S. as well as their long-term daily river discharge and SSL observations. It is noteworthy that USGS usually estimates SSL based on Suspended Sediment Concentration (SSC) measurements. SSC measurement and derivation of SSL as well as discharge measurement are prone to errors, which in turn introduce a level of uncertainty to the modeling approach.

4. Results and discussion

To evaluate the performance of the proposed framework, the derived event-based data are divided into calibration and evaluation periods, with the calibration section including 70% of the record (randomly selected) and the evaluation section comprising of the rest of the record. We then fit the proposed model (Eq. 6) to the calibration data, and evaluate the fitted model against the out-of-sample evaluation data. In doing so, the first step is to find the marginal distributions, $F_1(r_1)$ and $F_2(r_2)$, that best fit the observed data, r_1 and r_2. We fit the 17 aforementioned marginal distribution functions to the log-transformed discharge, r_1, and SSL, r_2, observations (Sadegh et al., 2018a) , and select the best model according to the Bayesian Information Criterion (BIC), as formulated by,

$$\mathrm{BIC} = n \log\left(\frac{\sum_{i=1}^{n}\left(r_j(i) - \hat{r}_j(i)\right)^2}{n}\right) + k\log(n), \tag{7}$$

in which, n denotes the length of record, k signifies the model's degrees of freedom, and $r_j(i)$ and $\hat{r}_j(i)$ are the i^{th} elements of the modeled and observed values of the j^{th} $(j = 1, 2)$ driver (i.e. discharge and SSL), respectively. Next step is to find the copula model, $C(r_1, r_2)$ in Eq. 3, from the aforementioned 26 copula models that best describes the dependence structure of the two drivers in this study, i.e. river discharge, r_1, and SSL, r_2. One approach to select the best bivariate model could be to find the copula function, $C(r_1, r_2)$, with the best performance metric (i.e. minimum BIC). An alternative approach could be selection of the copula model, $C(r_1, r_2)$, that collectively with the marginal distribution function of SSL, $F_2(r_2)$, as characterized by the Eq. 6, yield the best prediction of SSL compared to the observed data. Both approaches merge to the selection of a common model, mostly; but we have adopted the latter method, and selected a copula model, $C(r_1, r_2)$, for which the most likely prediction of SSL (mode in the conditional marginal distribution of Eq. 6) is closest to that of the observation, as documented by the Nash-Sutcliffe Efficiency,

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(r_2(i) - \hat{r}_2(i)\right)^2}{\sum_{i=1}^{n}\left(\hat{r}_2(i) - \hat{\bar{r}}_2\right)^2}, \tag{8}$$

in which $\hat{\bar{r}}_2$ represents average value of the second driver, i.e. SSL. BIC and NSE hold properties that fit the interests of this study. BIC takes into account the length of record and model complexity to select the best model. Usually models with higher degrees of freedom are more flexible, and hence tend to yield a closer fit to the observed data. This, however, might result in overfitting the calibration data, in which the model closely follows the observed data but fails to accurately describe the out-of-sample data record. BIC, hence, fairly compares the models

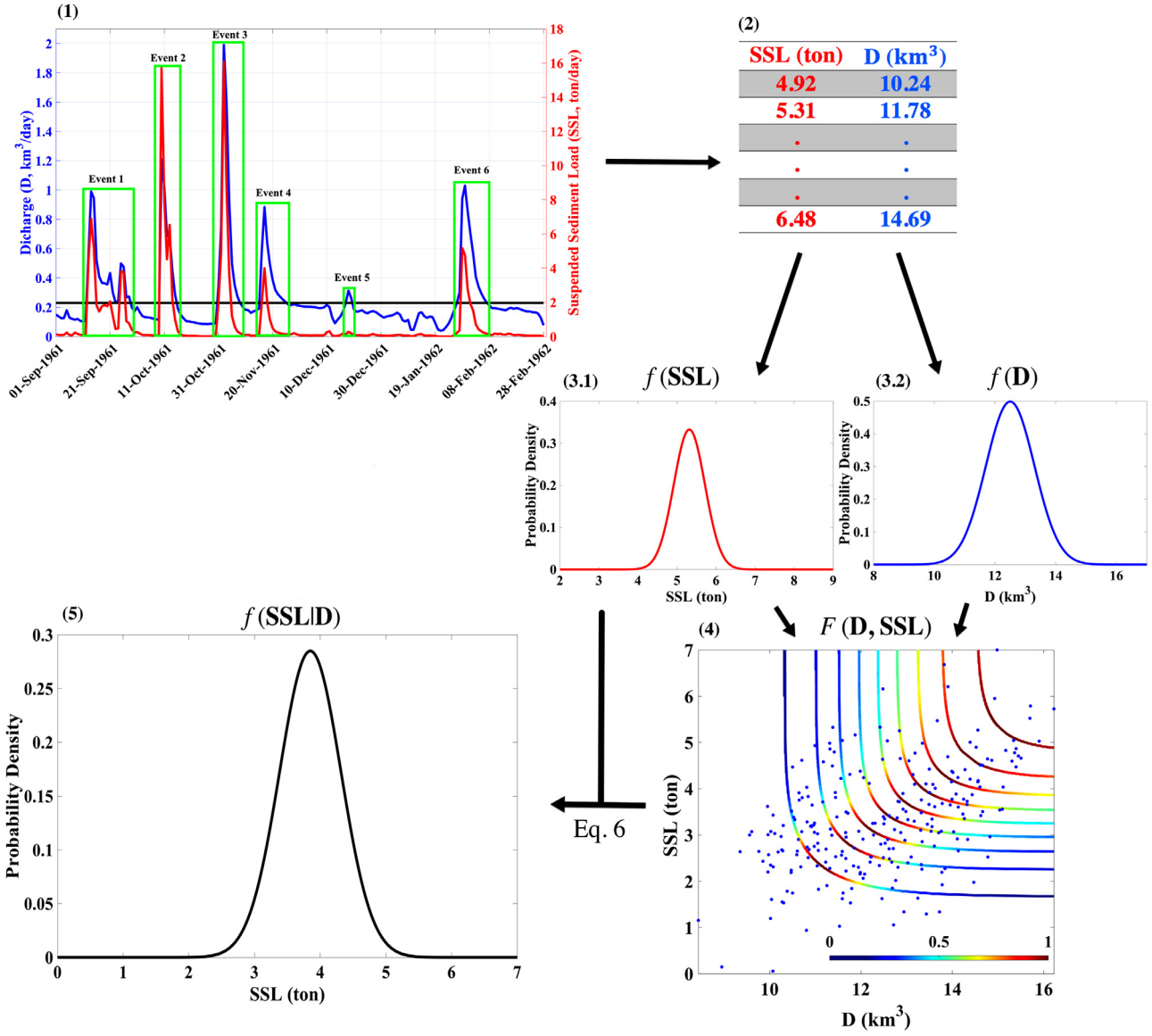

Fig. 1. The proposed framework for modeling Suspended Sediment Load (SSL) given river discharge volume.

with different degrees of freedom. NSE, on the other hand, scales the goodness of performance of the models into the interval of $(-\infty, 1]$, for which NSE = 1 is associated with a perfect fit. This helps us compare performance of the proposed model for different rivers.

Table 2 lists the marginal distributions that best fit the log-transformed discharge and SSL data for each of the nine USGS stations, as well as the selected copula model. Interestingly, the Generalized Pareto (GP) distribution is consistently selected to represent the discharge distribution, except for the Mississippi river (I). This selection is consistent with the literature that use the GP distribution to model hydroclimate variables above a certain threshold. In our study, 70th percentile of long-term flow time series is selected as the threshold. The SSL distribution, however, is represented by a variety of different distribution functions, namely Generalized Extreme Value (GEV), Generalized Pareto (GP), Gamma, and Inverse Gaussian distribution functions for different stations. As for the copula models, Placket seems to be more frequently selected, but Gumbel, Burr and Galambos copulas are also used for some stations. Table 2 also presents the statistics of model performance for prediction of observed SSL values in terms of NSE, as well as the coverage and spread of the uncertainty ranges. NSE is calculated based on the divergence of the highest likely estimation of SSL from the observed value. The results show good performance of the proposed model with NSE values in the range of ~0.7 to ~0.9 for all the stations (Table 2). Moreover, the model prediction performance is consistent over the calibration and evaluation periods (in terms of similar NSEs). The coverage of the observed SSL within the 68% and 95% predictive uncertainty ranges, along with the average width (spread, in terms of fraction of average annual SSL delivery) of the uncertainty bounds, are also presented in Table 2 (last eight columns). As expected, the coverage within the 68% and 95% uncertainty bounds fall in the vicinity of 68% and 95%. This confirms that our methodology is indeed able to capture the overall behavior of the underlying data generating PDF.

Fig. 3 visually demonstrates SSL model simulations given discharge volume for the calibration data for all the nine USGS stations of this study. High flow discharge volume (x-axis, km^3) and SSL (y-axis, ton) are both presented in log-scale. Each vertical line represents the conditional marginal distribution of SSL for a given discharge volume (as presented on the x-axis), color coded to display the density levels. Red sections represent higher density levels (more likely), whereas blue sections portray lower densities (less likely). As expected, greater discharge volumes (bigger high flow events) are associated with larger SSL levels. However, the relationship between them is nonlinear (for example see

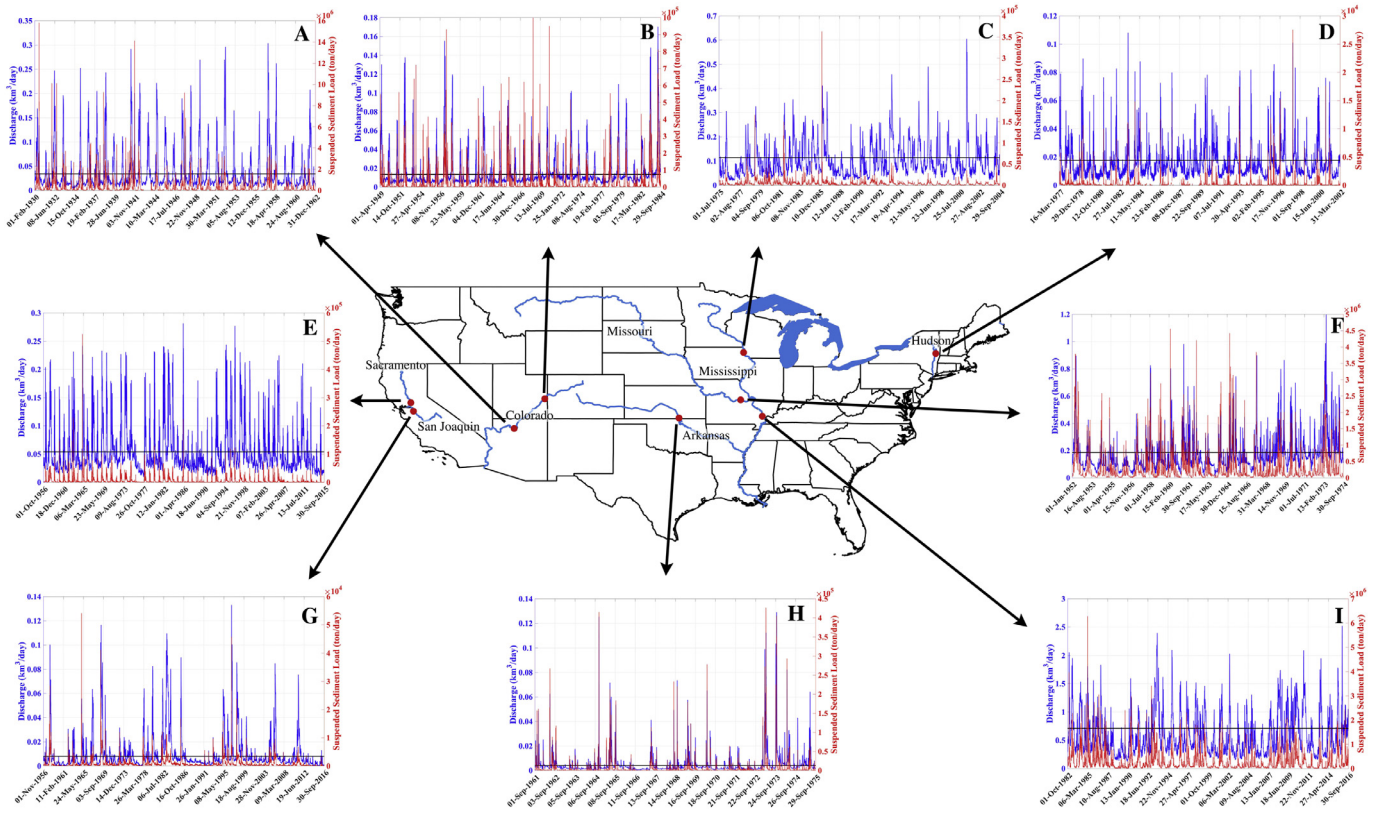

Fig. 2. Daily historical time-series of discharge (blue line) and SSL (red line) for the nine USGS stations of this study. The horizontal black line in each figure shows the 70th percentile of the discharge distribution, which is selected as the threshold for high flows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Selected marginal distributions and copula functions for each station, as well as the NSE for the best prediction of SSL, and the statistics of 68% and 95% predictive uncertainty ranges in the calibration and evaluation periods. Coverage describes the fraction of observation points encapsulated in the uncertainty bounds, and the average uncertainty spread is presented in terms of fraction of the average annual SSL delivery of each watershed (river section).

River name	Selected model			Best prediction		Statistics of 68% uncertainty range				Statistics of 95% uncertainty range			
	Marginal distribution		Copula	NSE		Coverage (%)		Spread		Coverage (%)		Spread	
	Discharge	SSL		Calib.	Eval.	Calib.	Eval.	Calib.	Eval.	Calib.	Eval.	Calib.	Eval.
Colorado (A)	GP*	GEV**	Placket	0.692	0.747	75.29	54.05	0.349	0.370	100	97.29	1.654	1.925
Colorado (B)	GP	Gamma	Gumbel	0.900	0.673	72.52	61.53	0.278	0.631	98.90	92.30	1.938	5.153
Mississippi (C)	GP	Gamma	Burr	0.773	0.700	67.47	65.71	0.541	0.384	93.98	97.14	1.472	0.971
Hudson (D)	GP	GEV	Placket	0.661	0.770	74.74	69.04	0.200	0.245	99.0	97.61	1.546	1.734
Sacramento (E)	GP	GP	Galambus	0.775	0.794	69.11	77.58	1.005	0.789	94.11	94.82	2.804	2.638
Missouri (F)	GP	GEV	Placket	0.671	0.659	64.95	54.83	0.121	0.152	98.29	98.33	0.534	0.669
San Joaquin (G)	GP	GEV	Placket	0.866	0.901	56.31	59.09	0.690	0.803	97.08	100	9.976	9.142
Arkansas (H)	GP	Inverse Gaussian	Placket	0.662	0.731	61.11	64.51	0.118	0.180	95.83	96.77	0.895	1.296
Mississippi (I)	Rayleigh	GEV	Placket	0.837	0.845	70.58	69.76	0.448	0.412	99.01	100	1.644	1.843

*GP: Generalized Pareto,
**GEV: Generalized Extreme Value.

Fig. 3I). These results display that the proposed modeling framework can closely track the SSL levels, with highly likely ranges of prediction (red section of the conditional marginal distribution) covering the majority of observed SSLs (black dots). It is also noticeable in this figure that the relationship between SSL and discharge volume is highly stochastic. See Fig. 3E, for example, in which a similar discharge volume ($\sim 0.4\,km^3$) can yield hugely different SSLs ($\sim 2\times10^4 - 10^5$).

The sedimentation processes are, indeed, dependent on several factors, including the river shape, the distribution of sediment particle size, the heterogeneous spatio-temporal characteristics of the watershed, climate forcing, availability of sediment load, impacts of wildfires, river flow energy to transport suspended sediment, land cover and use, and human activities, among others (Asselman et al., 2003; Garcia, 2008; Gupta, 2008; Hall et al., 2010; Kuhnle et al., 1996). While the impacts of such factors are captured in the constructed underlying data generating PDF, our parsimonious framework condition SSL only on the discharge level, and presents the impacts of the other determining factors as the uncertainty ranges around the most likely prediction. In fact, the hydrologic and hydraulic forces, as represented by the proxy variable of discharge volume, play a major role in generating and transporting sediment loads; and hence our simplification is quite pragmatic. Future studies will focus on extending the conditional framework to include states of the system such as burned area and seasonality impacts.

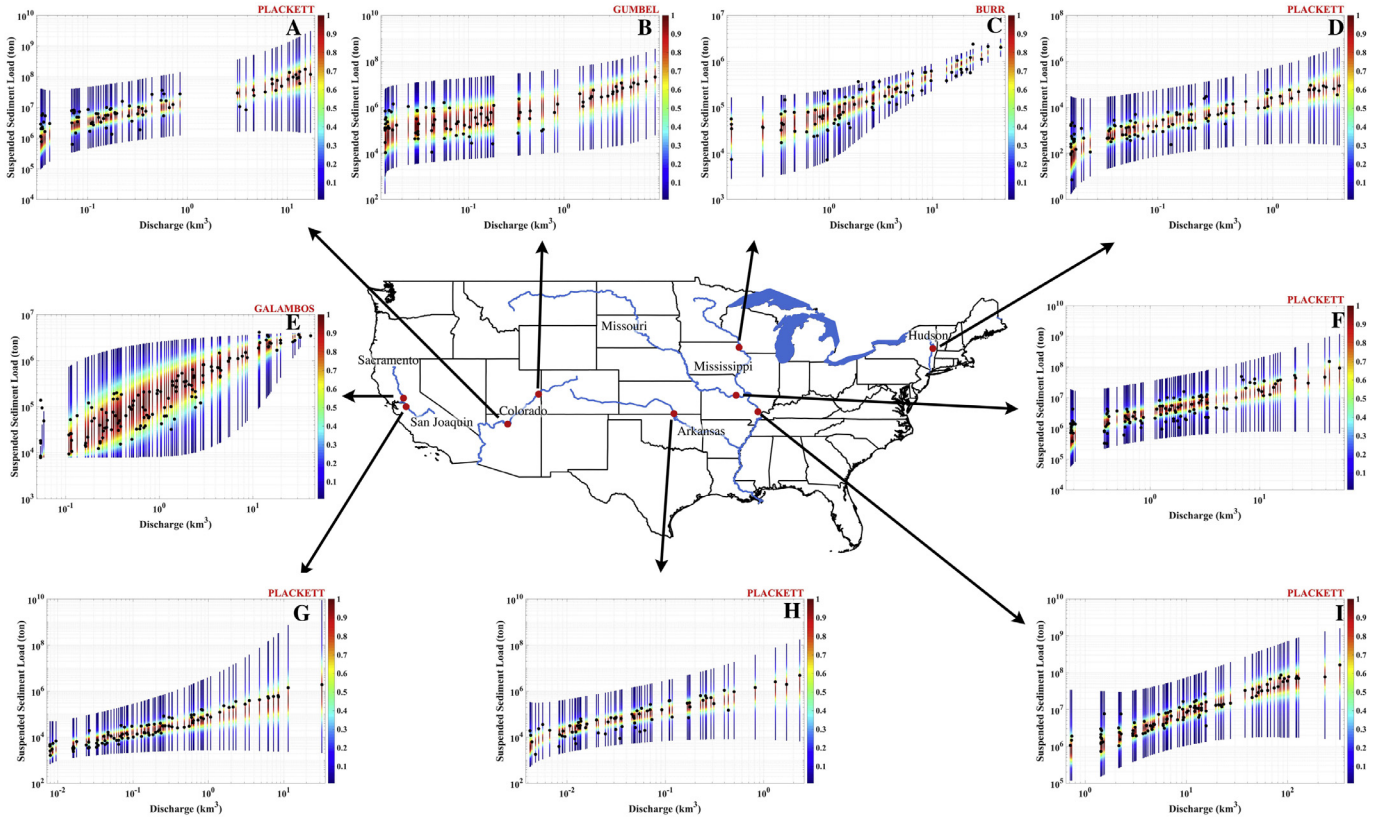

Fig. 3. Model predictions of SSL based on high flow volumes for the calibration data. Each vertical line represents the conditional marginal distribution of SSL, color-coded based on the density values. Red sections present higher densities, whereas blue sections display lower density levels. Black dots display the observed SSL level for each discharge volume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To elaborate the concept of model prediction in terms of conditional marginal distribution, Fig. 4 displays the estimated distribution of SSL at a discharge volume of 1 km^3 for all the studied stations. In other words, this figure shows the probabilistic prediction of suspended sediment delivery, as defined by accumulated suspended sediment load passing a specific cross section of a river (Garcia, 2008), given a high flow event with 1 km^3 volume. The suspended sediment delivery at the lower station (A) on the Colorado river, for example, is one order of magnitude larger than that of the upper station (B) for a high flow event of 1 km^3 in volume (0.5×10^6 vs 0.5×10^7 ton). Different factors play into this behavior. Station A is downstream of station B, and some branches of the Colorado river such as the Green river discharge significant quantities of sediment to the river in between these two cross sections (Andrews, 1991). Moreover, although upstream sediment load could be deposited along the river bed between points A and B, likelihood of such behavior is minimal given the high energy level of the high flow events. In fact, not only they have enough energy to transport upstream sediment loads to the downstream, but also flood events could prompt suspension of the bed materials (Garcia, 2008). Note that the study period for station A is bound between 1930 and 1962, before the construction of the Glen Canyon Dam (Lake Powell), which significantly changed the behavior of the Colorado river.

The conditional marginal distributions also provide insights into the likelihood of SSL levels. Highest likely SSL level for the Missouri river (F), for example, has a density of $\sim$0.95, which holds a higher confidence degree compared to that of the Arkansas river (H) with a density of $\sim$0.4. Higher confidence is associated with lower variance in the distributions. Indeed, the SSL distribution of the Missouri river (F) is least scattered compared to the other cases. Note that the scatter of SSL distributions might significantly change at a different high flow volume level (see Fig. 3C, for example). Furthermore, while most of the predictive distribution forms are close to Gaussian, those of the upper Mississippi river (C) and the Hudson river (D) are skewed to the left, and the SSL distribution of the middle Mississippi river (I) is skewed to the right.

Finally, Fig. 5 depicts the conditional marginal distributions of SSL for the out-of-sample evaluation data for the nine stations of this study. These results confirm our previous claim (Table 2) that the performance of the proposed model is consistent for the calibration and the evaluation sections (compare Figs. 4 and dummyTXdummy- 6). An interesting observation here is that the uncertainty span of SSL predictions for smaller and larger high flow events (in terms of discharge volumes for each station) are considerably higher than that of the medium range high flow events. The smaller events might include flash floods as a result of relatively large quantities of precipitation over a short period of time, sediment generation of which is highly dependent on the seasonality of the event as well as the availability of sediment for erosion. For example, majority of the Hudson river's (D) watershed area ($\sim$68%, Phillips and Hanchar (1996)) is covered by forest, which during spring and summer yields lower sediment load and during fall yields more. Another example is the Colorado river (A, B) watershed which experiences significant number of wildfires throughout the dry summer and fall months. This can potentially prompt generation of large quantities of sediment with precipitation over the burned area (Wohl et al., 1998). One potential approach is to perform this study at seasonal scale to diminish such impacts, but the drawback is that some seasons may not hold a long enough record to do the analysis.

The larger (extreme) events are generally difficult to model in the field of environmental science (Sadegh and Vrugt, 2013; Sadegh et al., 2016; Sadegh and Vrugt, 2014). Indeed, these events are rare, and usually there is not enough data to sufficiently constrain the components of the model associated with them. Results for the upper Mississippi (C) and Sacramento (E) rivers, however, contradict this finding. The upper

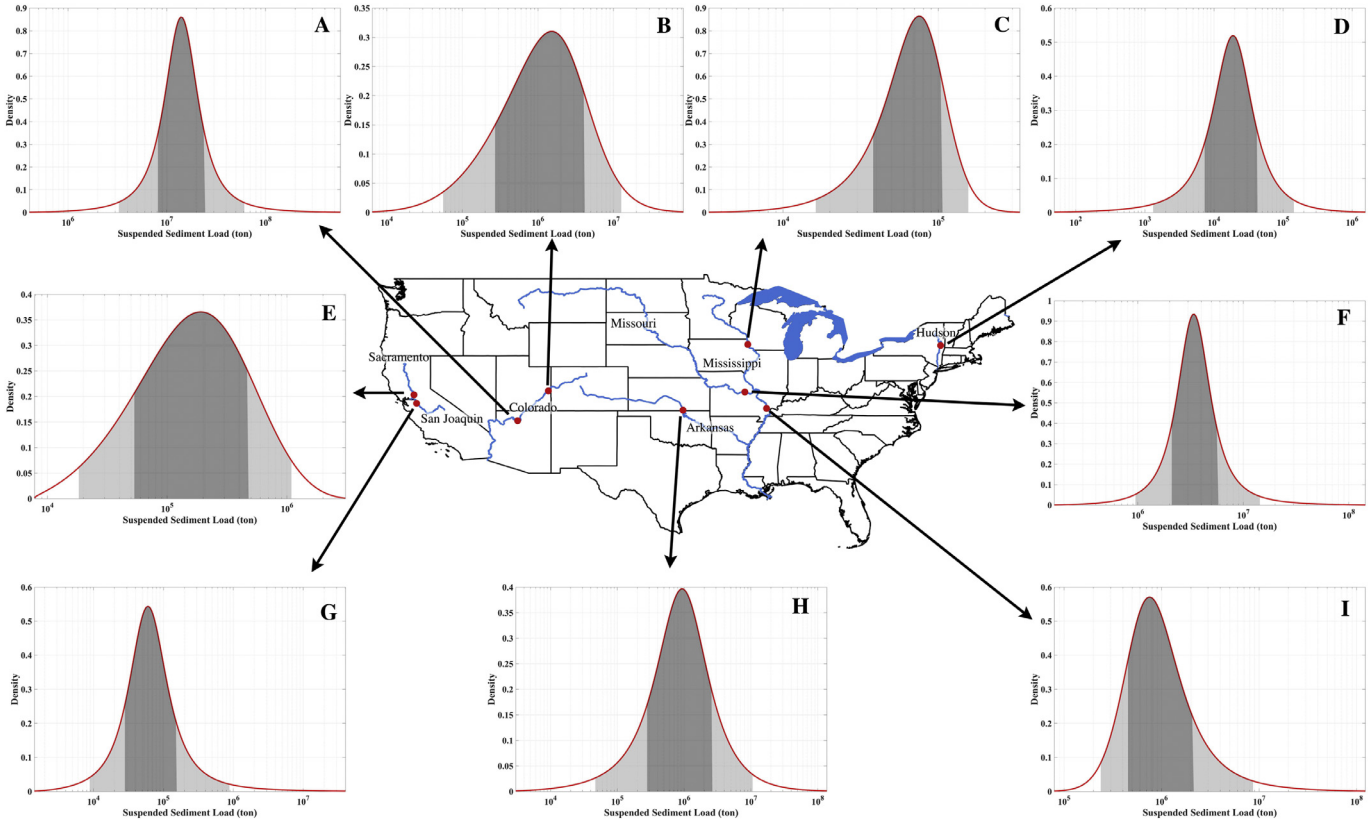

Fig. 4. Conditional marginal distributions of SSL for a given high flow volume of 1 km^3. The light gray section on each distribution presents the 95% uncertainty bounds, whereas the dark gray region shows the 68% predictive uncertainty ranges.

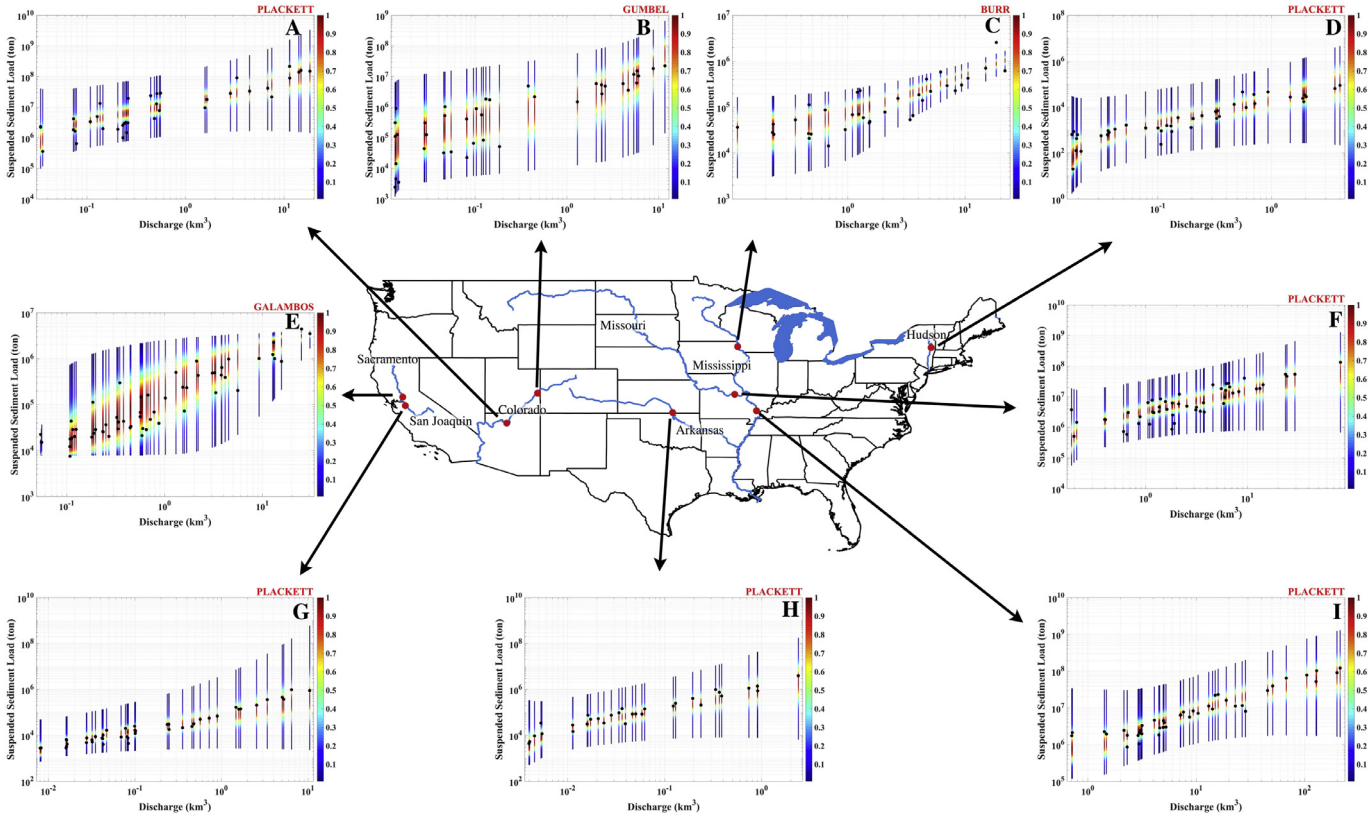

Fig. 5. Model predictions of SSL for different high flow volumes in the evaluation data. Each vertical line represents the conditional marginal distribution of SSL, color-coded based on the density values. Red sections present higher densities, whereas blue sections display lower density levels. Black dots display the observed SSL level for each discharge volume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Mississippi river holds a peculiar characteristic, with a watershed and a river bed that can store and remobilize sediment with a lag-time of several years to decades (Initiative, 1993). This basin-wide property makes it very difficult to model the sedimentation processes in this section of the Mississippi river. However, it is the choice of the Burr copula model that dominates the uncertainty ranges for prediction of SSL for larger high flow events in this case. Figures S31-35 (supplementary materials) show the uncertainty bounds of the marginal distributions of SSL for large flood events are indeed very wide using Gaussian, Clayton, Frank, Gumbel and AMH copulas. As previously discussed, we select the copula model to maximize the NSE performance metric for the best SSL prediction without recourse to the uncertainty levels. This depicts the importance of the choice of model, as well as the significance of the selection metric.

The Sacramento river's (Fig. 5E) sediment yield is sourced from agricultural activities in the watershed, the bankline shifts of the river, and the wildfire induced sediment load, among others (Hall et al., 2010). We hypothesize that the larger high flow events are bounded by the sediment availability in the watershed, and hence show a lower uncertainty range for the prediction of SSL. This hypothesis is, however, to be tested in future detailed studies. Our relative confidence in such hypothesis stems from the conditional marginal distributions of SSL that are skewed to the left for the large high flow events ($>10\,km^3$). The smaller high flow events ($\sim<0.2\,km^3$) for this river, on the contrary, have conditional marginal distributions of SSL that are skewed to the right. This might suggest the availability of sediment load but the lack of energy in the "small" high flow event to transport the sediment load. The SSL's conditional distribution form for the mid-sized high flow events (0.2–$10\,km^3$) in the Sacramento river (E), as well as for most of the high flow events for other rivers is close to normal.

One remark is due here that the proposed framework requires the underlying data generating process to be stationary. Indeed, most of the environmental modeling practices assume, by default, that the behavior of the studied processes are time-invariant, and vary around a constant mean value with a fixed variance and serial correlation structure (Clarke, 2007). This assumption is convenient and opens an arsenal of statistical and numerical frameworks for environmental modeling (Sadegh et al., 2015). We have carefully examined the selected rivers for stationarity assumption in the period of study. Indeed, none of these cases have undergone significant development in the period of observation. However, time-series of the Sacramento (E) and San Joaquin (G) rivers might suggest that some watershed management practices have been employed in the recent decades (Fig. 2). While we don't have physical evidence to prove these two rivers has endured nonstationarity, we hypothesize the wide uncertainty ranges of the SSL predictions for these two rivers (average 95% predictive uncertainty spread of $\sim$2.6 and $\sim$9 $\times$ average annual sediment yield for the Sacramento and San Joaquin rivers, Table 2) might be partly influenced by the anthropogenic interventions in the sediment generation processes.

To wrap up, sedimentation processes are stochastic in nature, and a stochastic framework is hence required to adequately capture the system behavior and explain the underlying uncertainties. In this paper, we propose a probabilistic modeling scheme that can accurately and precisely predict the suspended sediment load and its uncertainty ranges in alluvial rivers. The proposed approach is a data-driven model, which requires long-term observations of SSL and discharge. This drawback is, however, shadowed by relaxing the need for detailed physical characteristics of the watershed and by diminishing the computational burden of physics-based models. Modeling the sediment load in rivers is of paramount importance in the field of environmental engineering and science (Belmont et al., 2011), as moderate concentrations of fine sediment deliver important nutrient loads to the aquatic biota and the downstream agricultural sector (Mohajeri et al., 2016), but excessive concentrations of sediment inflict detrimental environmental impacts (Heppell et al., 2009; Wood and Armitage, 1997). Indeed, excessive nutrients may prompt algal blooms, reduce dissolved oxygen, harm aquatic life and limit the recreational services of the bodies of water (Wohl et al., 1998). Sediments also transport pesticides, toxicants, salts and heavy metals to the downstream river and lake habitats (Mussetter Engineering, 2000).

5. Conclusion

Accurate modeling of sediment transport, and characterizing its underlying uncertainties, in alluvial rivers has historically been a challenge. Sediment is not only a non-point source of pollution, but it also transports pesticides, toxicants and heavy metals. Numerical models of sediment transport are computationally demanding and require large amounts of data that are often not available. Here, we have introduced a stochastic framework, rooted in the multivariate probability theory and Bayesian Network, to model Suspended Sediment Load (SSL) in rivers. The majority of the annual sediment load of rivers (70-90%) is transported in high flow events and hence, the SSL and discharge volume are highly interdependent. We exploit their dependence structure to estimate SSL based on discharge volume as the predictor. This event-based approach approximates the conditional marginal distribution of SSL for any given high flow volume. The proposed model characterizes the likelihood of the SSL prediction, and pinpoints the modeling uncertainties.

The presented framework relaxes the need for detailed information about the physical characteristics of the watershed and river system, and employs a parsimonious model to describe the highly nonlinear relationship between discharge and SSL. We have tested this approach for nine USGS stations, the results of which highlight the capability of this model to describe the probability distribution function of SSL for different discharge values in major rivers of the US. The proposed methods offers uncertainty bounds of the SSL predictions which shed light on the stochastic nature of the sediment transport processes.

Acknowledgements

This study was partially supported by National Science Foundation (NSF) grant (CMMI-1635797).

The data used in this study is obtained from sediment portal of the US Geological Survey, available at https://cida.usgs.gov/sediment/.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2018.06.006.

References

Andrews, E.D., 1991. Sediment transport in the colorado river basin. In: Colorado River Ecology and Dam Management: Proceedings of a Symposium, May 24–25, 1990, Santa Fe, New Mexico, pp. 54–74.

Asselman, N.E., Middelkoop, H., Van Dijk, P.M., 2003. The impact of changes in climate and land use on soil erosion, transport and deposition of suspended sediment in the river rhine. Hydrol. Process. 17 (16), 3225–3244.

Aytek, A., Kişi, Ö., 2008. A genetic programming approach to suspended sediment modelling. J. Hydrol. (Amst) 351 (3), 288–298.

Bagnold, R., 1962. Auto-suspension of transported sediment; turbidity currents. Proc. R. Soc. Lond. A Math Phys. Sci. 265 (1322), 315–319.

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., et al., 2011. Large shift in source of fine sediment in the upper mississippi river. Environ. Sci. Technol. 45 (20), 8804–8810.

Berlamont, J., Ockenden, M., Toorman, E., Winterwerp, J., 1993. The characterisation of cohesive sediment properties. Coastal Eng. 21 (1–3), 105–128.

Best, J.L., Bristow, C.S., 1993. Braided rivers. no. 75. Amer Assn Petroleum Geologists.

Bowman, A.W., Azzalini, A., 1997. Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations, 18. OUP Oxford.

Buendia, C., Bussi, G., Tuset, J., Vericat, D., Sabater, S., Palau, A., Batalla, R., 2016. Effects of afforestation on runoff and sediment load in an upland mediterranean catchment. Sci. Total Environ. 540, 144–157.

Bulygina, N., Gupta, H., 2009. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. Water Resour. Res. 45 (W00B13), 1–20. https://doi.org/10.1029/2007WR006749.

Bulygina, N., Gupta, H., 2010. How bayesian data assimilation can be used to estimate the mathematical structure of a model. Stochastic Environ. Res. Risk Assessment 24 (6), 925–937.
Bulygina, N., McIntyre, N., Wheater, H., 2009. Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis. Hydrol. Earth Syst. Sci. 13 (6), 893–904.
Bulygina, N., McIntyre, N., Wheater, H., 2011. Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Water Resour. Res. 47 (2).
Chalov, S.R., Jarsjö, J., Kasimov, N.S., Romanchenko, A.O., Pietroń, J., Thorslund, J., Promakhova, E.V., 2015. Spatio-temporal variation of sediment transport in the selenga river basin, mongolia and russia. Environ. Earth Sci. 73 (2), 663–680.
Clarke, R.T., 2007. Hydrological prediction in a non-stationary world. Hydrol. Earth Syst. Sci. Discuss. 11 (1), 408–414.
Cobaner, M., Unal, B., Kisi, O., 2009. Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. J. Hydrol. (Amst) 367 (1), 52–61.
Garcia, M., 2008. Sedimentation engineering: processes, measurements, modeling, and practice. American Society of Civil Engineers.
Gupta, A., 2008. Large rivers: Geomorphology and Management. John Wiley & Sons.
Hall, B., Andrey, S., Lea, A., Ron, C., Bryce, C., 2010. Comprehensive geomorphic and sedimentation analyses of lower sacramento river for flood management, erosion mitigation and habitat enhancement design. Joint Federal Interagency Conference. Las Vegas, NV June 27 - July 1, 2010.
Heppell, C., Wharton, G., Cotton, J., Bass, J., Roberts, S., 2009. Sediment storage in the shallow hyporheic of lowland vegetated river reaches. Hydrol Process. 23 (15), 2239–2251.
Initiative, U.M.R.W.Q., 1993. Report of sedimentation workshop. Technical Report. Upper Mississippi River Basin Association.
Joe, H., 2014. Dependence modeling with copulas. CRC Press.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1993. Continuous univariate distributions, vol. 1. Wiley-Interscience, Hoboken, NJ.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous univariate distributions, vol. 2. Wiley-Interscience, Hoboken, NJ.
Kisi, O., Ozkan, C., Akay, B., 2012. Modeling discharge–sediment relationship using neural networks with artificial bee colony algorithm. J. Hydrol. (Amst) 428, 94–103.
Kleinhans, M.G., van Rijn, L.C., 2002. Stochastic prediction of sediment transport in sand–gravel bed rivers. J. Hydraul. Eng. 128 (4), 412–425.
Kuhnle, R.A., Bingner, R.L., Foster, G.R., Grissinger, E.H., 1996. Effect of land use changes on sediment transport in goodwin creek. Water Resour. Res. 32 (10), 3189–3196.
Leopold, L.B., 1968. Hydrology for urban land planning: A guidebook on the hydrologic effects of urban land use. GEOLOGICAL SURVEY CIRCULAR 554, Washington DC.
Lisle, I., Rose, C., Hogarth, W., Hairsine, P., Sander, G., Parlange, J.-Y., 1998. Stochastic sediment transport in soil erosion. J. Hydrol. (Amst) 204 (1–4), 217–230.
Lovejoy, S.B., Lee, J.G., Randhir, T.O., Engel, B.A., 1997. Research needs for water quality management in the 21st century a spatial decision support system. J. Soil Water Conserv. 52 (1), 18–22.
Madadgar, S., AghaKouchak, A., Farahmand, A., Davis, S.J., 2017. Probabilistic estimates of drought impacts on agricultural production. Geophys. Res. Lett. 44 (15), 7799–7807.
Madadgar, S., Moradkhani, H., 2013. A bayesian framework for probabilistic seasonal drought forecasting. J. Hydrometeorol. 14 (6), 1685–1705.
Malmon, D.V., Dunne, T., Reneau, S.L., 2002. Predicting the fate of sediment and pollutants in river floodplains. Environ. Sci. Technol. 36 (9), 2026–2032.
Mazdiyasni, O., AghaKouchak, A., Davis, S.J., Madadgar, S., Mehran, A., Ragno, E., Sadegh, M., Sengupta, A., Ghosh, S., Dhanya, C., et al., 2017. Increasing probability of mortality during indian heat waves. Sci. Adv. 3 (6), e1700066.
Mohajeri, S.H., Righetti, M., Wharton, G., Romano, G.P., 2016. On the structure of turbulent gravel bed flow: implications for sediment transport. Adv. Water Resour. 92, 90–104.
Morgan, R., Quinton, J., Smith, R., Govers, G., Poesen, J., Auerswald, K., Chisci, G., Torri, D., Styczen, M., 1998. The european soil erosion model (eurosem): a dynamic approach for predicting sediment transport from fields and small catchments. Earth Surf. Processes Landforms 23 (6), 527–544.
Morris, G.L., Fan, J., 1998. Reservoir Sedimentation Handbook: Design and Management of Dams, Reservoirs, and Watersheds for Sustainable Use. McGraw Hill Professional.
Mussetter Engineering, 2000. Inc., J. . S. A. I. Sacramento and San Joaquin River Basins Comprehensive Study, California. Technical Report. U.S. Army Corps of Engineers.
Nelsen, R.B., 2007. An Introduction to Copulas. Springer Science & Business Media.
Parker, G., Ikeda, S., 1989. River meandering. AGU Water Resources Monograph Board, Washington DC.
Phillips, P.J., Hanchar, D.W., 1996. Water-quality assessment of the hudson river basin in new york and adjacent states. US Geological Survey Report 96, 1–76.
Rankl, J.G., 2004. Relations between total-sediment load and peak discharge for rainstorm runoff on five ephemeral streams in wyoming. US Department of the Interior, US Geological Survey.
Russel, S., Norvig, P., 2003. Artificial Intelligence: A Modern Approach. EUA: Prentice Hall.
Sadegh, M., Majd, M.S., Hernandez, J., Haghighi, A.T., 2018a. The quest for hydrological signatures: effects of data transformation on bayesian inference of watershed models. Water Resour. Manage. 1–15.
Sadegh, M., Moftakhari, H., Gupta, H.V., Ragno, E., Mazdiyasni, O., Sanders, B., Matthew, R., AghaKouchak, A., 2018b. Multi-hazard scenarios for analysis of compound extreme events. Geophys. Res. Lett..
Sadegh, M., Ragno, E., AghaKouchak, A., 2017. Multivariate copula analysis toolbox (mvcat): describing dependence and underlying uncertainty using a bayesian framework. Water Resour. Res. 53 (6), 5166–5183. https://doi.org/10.1002/2016WR020242.
Sadegh, M., Vrugt, J., 2013. Bridging the gap between glue and formal statistical approaches: approximate bayesian computation. Hydrol. Earth Syst. Sci. 17 (12).
Sadegh, M., Vrugt, J., Gupta, H., Xu, C., 2016. The soil water characteristic as new class of closed-form parametric expressions for the flow duration curve. J. Hydrol. (Amst) 535, 438–456.
Sadegh, M., Vrugt, J.A., 2014. Approximate Bayesian computation using markov chain monte carlo simulation: dream (abc). Water Resour. Res. 50 (8), 6767–6787.
Sadegh, M., Vrugt, J.A., Xu, C., Volpi, E., 2015. The stationarity paradigm revisited: hypothesis testing using diagnostics, summary metrics, and dream (abc). Water Resour. Res. 51 (11), 9207–9231.
Schumm, S.A., Harvey, M.D., Watson, C.C., 1984. Incised Channels: Morphology, Dynamics, and Control. Water Resources Publications.
Sherriff, S.C., Rowan, J.S., Fenton, O., Jordan, P., Melland, A.R., Mellander, P.-E., Huallachain, D.O., 2016. Storm event suspended sediment-discharge hysteresis and controls in agricultural watersheds: implications for watershed scale sediment management. Environ. Sci. Technol. 50 (4), 1769–1778.
Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Université Paris 8.
Thulasiraman, K., Swamy, M., 1992. 5.7 acyclic directed graphs. Graphs: Theory and Algorithms 118.
Van Rijn, L.C., 1984. Sediment transport, part i: bed load transport. J. Hydraul. Eng. 110 (10), 1431–1456.
Van Rijn, L.C., et al., 1993. Principles of sediment transport in rivers, estuaries and coastal seas, 1006. Aqua publications Amsterdam.
Vanoni, V., 2006. Sedimentation engineering: Classic edition, pp. ixii.
Walling, D.E., 1999. Linking Land Use, Erosion and Sediment Yields in River Basins. In: Man and River Systems. Springer, pp. 223–240.
Warner, J.C., Sherwood, C.R., Signell, R.P., Harris, C.K., Arango, H.G., 2008. Development of a three-dimensional, regional, coupled wave, current, and sediment-transport model. Comput. Geosci. 34 (10), 1284–1306.
Wiberg, P.L., Smith, J.D., 1987. Calculations of the critical shear stress for motion of uniform and heterogeneous sediments. Water Resour. Res. 23 (8), 1471–1480.
Wohl, E.E., McConnell, R., Skinner, J., Stenzel, R., 1998. Inheriting our past: river sediment sources and sediment hazards in colorado. Water in the balance; no. 7.
Wolman, M.G., 1967. Two problems involving river channel changes and background observations. Quantitative geography. Part II: Physical and cartographic topics 67–107.
Wood, P.J., Armitage, P.D., 1997. Biological effects of fine sediment in the lotic environment. Environ. Manage. 21 (2), 203–217.
Yang, C.T., 1996. Sediment Transport: Theory and Practice. MCGRAW-HILL BOOK CO,(USA). 1996..