Water Resources Research

RESEARCH ARTICLE
10.1029/2025WR040224

Key Points:

- A novel copula regression method is developed to predict sediment concentration and load using streamflow data
- The method is cross validated using measured data with very good statistical metrics
- The method is a time-saving and cost-effective alternative as compared to SWAT model

Correspondence to:
Y. Ouyang,
ying.ouyang@usda.gov

Citation:
Ouyang, Y. (2026). A copula-based regression method to predict sediment concentration and load in rivers using streamflow data. *Water Resources Research*, *62*, e2025WR040224. https://doi.org/10.1029/2025WR040224

Received 21 FEB 2025
Accepted 8 AUG 2026

Author Contributions:
Conceptualization: Ying Ouyang
Data curation: Ying Ouyang
Formal analysis: Ying Ouyang
Methodology: Ying Ouyang
Resources: Ying Ouyang
Validation: Ying Ouyang
Writing – original draft: Ying Ouyang
Writing – review & editing: Ying Ouyang

A Copula-Based Regression Method to Predict Sediment Concentration and Load in Rivers Using Streamflow Data

Ying Ouyang[1]

[1]USDA Forest Service, Center for Bottomland Hardwoods Research, Southern Research Station, Mississippi State, MS, USA

Abstract Pollution and deposition of sediments in rivers and streams are critical environmental, ecological, navigational, and recreational concerns. While several well-known watershed models such as SWAT (Soil and Water Assessment Tool) and HSPF (Hydrological Simulation Program-FORTRAN) are applied to predict sediment concentrations and loads in rivers and streams, their application requires the acquisition of a large amount of climatic, GIS, land use, and watershed input data, which is very time-consuming and cost-prohibitive. This study developed a copula-based regression method to predict sediment concentrations and loads in rivers using streamflow data, which is difficult to achieve using traditional methods. Predicted sediment concentrations from the method were verified and validated by field measurements from three US Geological Survey (USGS) gauge stations across the US, as well as by statistical metrics. Prerequisites, advantages, uncertainties, and limitations of the method were presented and discussed. Results revealed that sediment loads were not proportional to watershed drainage area, indicating that land use and anthropogenic activities also play an important role in sediment load. Overall, no significant increasing or decreasing trends of annual sediment loads were found at study sites in New York and Florida over the 11-year period from 2011 to 2021. This study suggests that the copula-based regression approach, which is time-saving and cost-effective compared to the traditional watershed models, is a promising alternative for predicting sediment concentrations and loads using streamflow data when a good dependence (or correlation) exists between the sediment content and streamflow after copula transformation.

Plain Language Summary Sediment pollution and deposition in surface waters pose significant environmental, ecological, and recreational challenges. While several well-known watershed models have been used to predict sediment concentrations and loads in streams and rivers, their reliance on large amounts of input data and calibration efforts limits their usage. This study develops a novel method (or copula-based regression method) in a time-saving and cost-effective manner to predict sediment concentration and load using streamflow data, which is difficult, if not impossible, to achieve using traditional methods. This method also has potential for estimating pollution of other water quality constituents in streams by scientists, engineers, researchers, and water resource managers around the world.

1. Introduction

Transport and loading of sediments contaminated with pollutants are serious environmental and ecological concerns worldwide (Ouyang et al., 2003), while the deposition of sediments in surface waters can significantly alter hydrological channels and negatively affect aquatic life (Ouyang, 2022a). Inappropriate agricultural, industrial, and domestic activities have resulted in sediment contamination and deposition in rivers and streams (Almasalmeh et al., 2022; Noe et al., 2020). While three major water quality models, namely the HSPF (Hydrological Simulation Program-FORTRAN), SPARROW (SPAtially Referenced Regressions on Watershed attributes), and SWAT (Soil and Water Assessment Tool), are currently used to predict the transport and loading of sediments in rivers and streams (Arnold et al., 2012; Bicknell et al., 2001; Schwarz et al., 2006), their use depends on extensive model calibration and validation using multi-year, continuously observed daily sediment data. However, these data are unavailable for most watersheds around the world, creating difficulties in applying the models. Additionally, the development of these models requires substantial climatic, land use, and watershed input parameters, and GIS data, making them time-consuming and labor-intensive.

In the past few decades, the Load Estimator (LOADEST) model has been applied to estimate sediment loads in streams and rivers (Runkel et al., 2004). One major limitation of LOADEST is the significant effort required to arrange input data into a specific format. To overcome this hurdle, a web-based interface was developed by Park et al. (2015). Although this web-based LOADEST was very handy for estimating water quality constituent loads, this web tool is no longer available for unknown reasons. Other drawbacks of LOADEST are: (a) significant biases when the relationship between the log-transformed water quality constituent concentration and streamflow is not strong; (b) considerable differences in the seasonal relationship shape; and (c) extremely high residuals (Hirsch, 2014). In addition, LOADEST may not be able to predict peaks and valleys of constituent concentrations as the regression equations used in LOADEST are not used to capture streamflow variations. These drawbacks make it difficult to apply LOADEST for estimating sediment load under high and low streamflow during wet and drought conditions.

To circumvent these drawbacks, Ouyang (2022a) developed a novel tool for gap-filling daily sediment load based on sparse and discontinuous measured data using the flow-weighted approach. This tool was rigorously validated with observed data from U.S. Geological Survey (USGS) gauge stations across the United States. The flow-weighted approach effectively predicted daily sediment loads when a good linear correlation was present between measured seasonal sediment loads and measured seasonal stream discharges, which is a prerequisite for using the tool. Five out of six selected USGS gauge stations used in the study met this prerequisite, proving the tool's usefulness for filling daily sediment load gaps. The author further stated that the tool could also be applied to predict other water quality constituent loads based on limited measured data. While the tool is valuable for addressing gaps in daily sediment loads when limited field data are available, it is difficult to predict daily sediment loads for periods without any measured sediment data.

To this end, a new method, the copula-based regression method, was developed in this study. A copula is multivariate statistical approach for building relationships among random variables, which is often challenging to achieve with traditional methods. Since its development by Sklar (1959), copula has gained widespread acceptance in scientific applications (Alizadeh et al., 2018; Frees & Valdez, 2008; Ouyang, 2022b). Currently, copula regression is being employed to predict one variable from other variables (Cote et al., 2019; Masarotto & Varin, 2017). Compared to ordinary least squares and generalized linear regression methods, copula regression is advantageous because it does not impose restrictions on probability distributions (Parsa & Klugman, 2011). In recent years, copula regression has been used to fit malaria incidence data (Masarotto & Varin, 2017), check property and casualty risks (Cote et al., 2019), and predict tree sap flow based on vapor pressure deficit (Ouyang & Sun, 2024). These copula-based regressions provide useful insights into establishing correlations among random variables, which traditional regressions cannot easily achieve.

The goal of this study was to predict sediment concentration and load based on streamflow data using copula-based regression method, which is challenging if not impossible, to achieve with traditional methods. The specific objectives were to: (a) present the step-by-step procedures for developing the copula-based regression method; (b) validate the method using measured data; and (c) apply the method to predict sediment concentrations and loads in rivers using streamflow data for periods when no measured sediment data are available. Prerequisites, advantages, uncertainties, and limitations of the method were discussed.

2. Materials and Methods

2.1. Data Acquisition

In this study, three U.S. Geological Survey (USGS) gauge stations were chosen to obtain sediment and streamflow data (Figure 1). The selected stations are: (a) Station #01358000 (drainage area = 20,953 km^2) located on the Hudson River, New York, (b) Station #02359170 (drainage area = 49,727 km^2) located on the Apalachicola River near Sumatra, Florida, and (c) Station #13342500 (drainage area = 24,042 km^2) located on the Clearwater River, Idaho. These three USGS gauge stations were chosen because they have continuous daily streamflow data, sparse sediment measurements, and represent diverse geographical locations across the United States. Units for the data obtained from the USGS gauge stations are typically in mg L^{-1} for total solid, dissolved solid and sediment contents and m^3 s^{-1} for streamflow.

Prior to performing copula regression, the correlations between streamflow and sediment content were investigated. Figure 2 shows that poor correlations exist between the streamflow and sediment content for all three

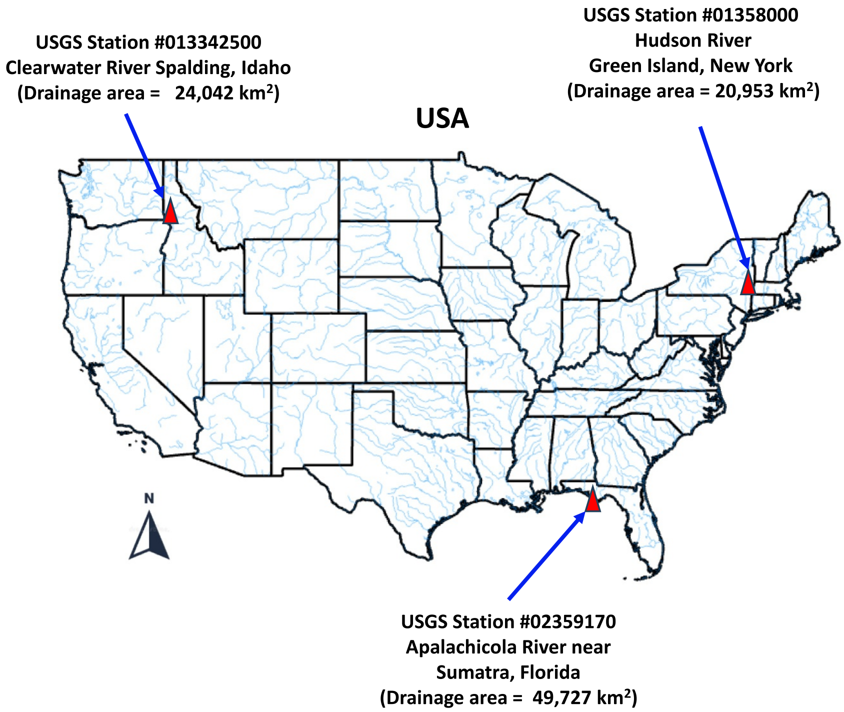

Figure 1. Locations of the study sites, showing the three US Geological Survey gauge stations.

stations. It is therefore difficult to predict sediment concentrations based on streamflow data using these simpler linear equations and thus a copula regression was employed.

2.2. Copula Regression

Traditionally, copulas are developed using inversion (elliptical copulas), generation (Archimedean copulas), and extreme value methods (Nelsen, 2006). These methods rely on a small number of parameters and lack flexibility in high dimensions. In recent years, vine copulas have emerged as highly flexible alternatives to overcome these limitations (Aas et al., 2009). While there are many copula families with numerous copula functions, five of the most commonly used copulas in literature, namely the BB8 (Joe-Frank), Clayton, Frank, Gumbel, and Normal copulas, were selected for analysis in this study. Using the Vinecopulas package in R, the best copula among the five was then chosen to develop the copula-based regression equations for the three USGS gauge stations, allowing for the prediction of sediment concentration using streamflow data. A flowchart for the copula analysis is shown in Figure 3, and the detailed steps are as follows:

1. *Frequency Distribution*. The frequency distributions of streamflow and sediment concentration were determined by histogram analysis. As shown in Figure 4, streamflow and sediment data followed a Gamma distribution, which was used as the marginal distribution function when developing bivariate distributions for the five copulas. In copula analysis, the probability distribution of all variables is termed the joint distribution, whereas the probability distribution of one variable is denoted as a marginal distribution. The dependence (or correlation) between two variables is determined by their marginal distributions. Two variables are dependent if and only if their joint distribution function is not equal to the product of their marginal distribution functions.
2. *Best Copula Selection*. In this study, Kendall's tau (τ) was employed to evaluate the correlation (or dependence) between observed streamflow and sediment concentration at a significance level of $p \leq 0.05$. If $\tau = 0$, no relationship exists. If $\tau = 1$ (or -1), a perfect positive (or negative) relationship is present. Based on the τ and p values, the normal copula was determined to be the best choice, as it showed the highest τ and lowest p values (Table 1) among the five copulas at all three study sites. The Normal copula is calculated as follows (Renard & Lang, 2007):

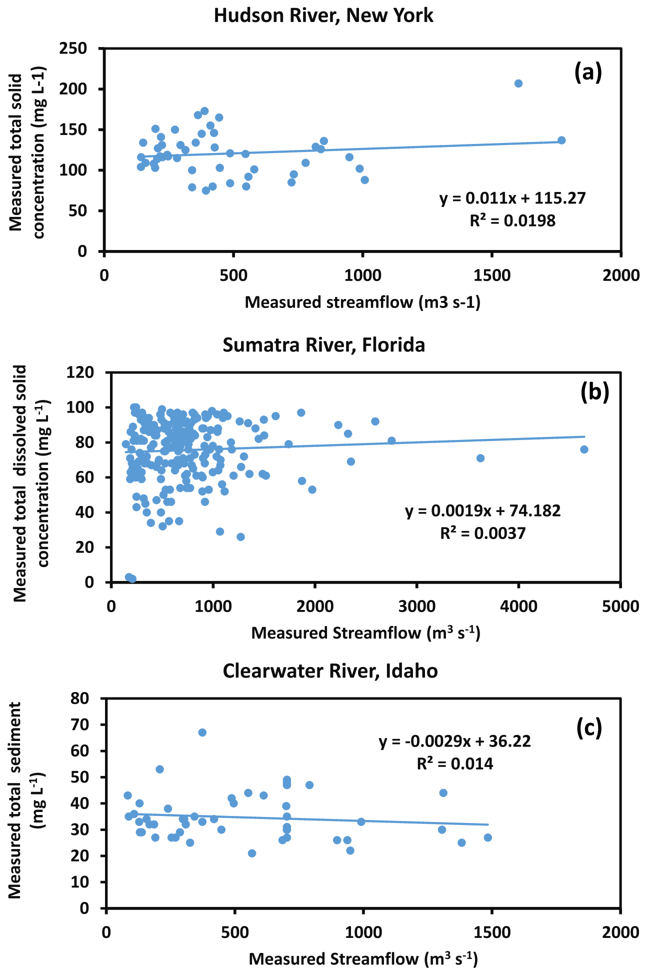

Figure 2. Plots of measured sediment content against measured streamflow for the three study sites used in this study.

$$C_N(u_1, u_2) = \Psi_2\left[\psi^{-1}(u_1), \psi^{-1}(u_2)\right], \tag{1}$$

where C_N is the Normal copula, u is the variable, Ψ is the cumulative distribution function (cdf) and ψ is the cdf of the standard normal distribution at (0, 1). The Normal copula allows the modeling of dependence in high dimensions with only one parameter. The correlation between two variables in the Normal copula was calculated using the tau value:

$$\tau = \frac{2}{\pi}\arcsin\theta, \tag{2}$$

where τ is tau value and θ is Pearson's correlation. The copula analysis was performed using R script in this study.

3. *Verification of Normal Copula*. The selected Normal copula was used to predict sediment concentration based on randomly generated streamflow data for all three study sites. The correlation between the generated streamflow and the predicted sediment concentration was verified using Kendall's τ at $p \leq 0.05$.
4. *Generation of Streamflow and Sediment Data*. Following verification, the Normal copula was employed to generate 500 random data points for both streamflow and sediment concentration. These data points were sufficient to establish a copula-based regression equation.
5. *Establishment and Validation of Copula-Based Equation*. The randomly generated streamflow and sediment concentration data were used to develop a regression equation. The predicted sediment concentrations from the equation were validated against measured data using statistical metrics such as coefficient of determination (R^2), normalized root mean-square error (nRMSE), Nash Sutcliff efficiency (NSE), and percent bias (PBIAS). The nRMSE is calculated as (Taebi & Mansy, 2017):

$$\mathbf{nRMSE} = \frac{\mathbf{1}}{\overline{\boldsymbol{O}}}\left(\sqrt{\frac{\sum_{i=1}^{n}(\boldsymbol{O}_i - \boldsymbol{S}_i)^2}{\boldsymbol{n}}}\right), \tag{3}$$

where O_i is the observations, S_i is the model predictions, $\overline{O}$ is the average of observations, and n is the total number of observations. The NSE is given as (Nash & Sutcliffe, 1970):

$$\mathbf{NSE} = \mathbf{1} - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}, \tag{4}$$

where NSE = 1 represents a perfect fit, NSE > 0.75 indicates a very good fit, 0.36 ≤ NSE ≤ 0.75 represents a reasonable fit, and NSE < 0.36 is considered an unsatisfactory fit (Krause et al., 2006). The PBIAS is estimated as (Rubin, 1976):

$$\mathbf{PBIAS} = \frac{\sum_{i=1}^{n}(O_i - S_i)}{\sum_{i=1}^{n} O_i}\,\mathbf{100}. \tag{5}$$

A PBIAS of 10% or less is acceptable in hydrological modeling analysis (Golmohammadi et al., 2014).

6. *Application of Copula-Based Regression Equations*. The established regression equations was used to predict sediment concentrations and loads based on measured streamflow data at the three USGS gauge stations.

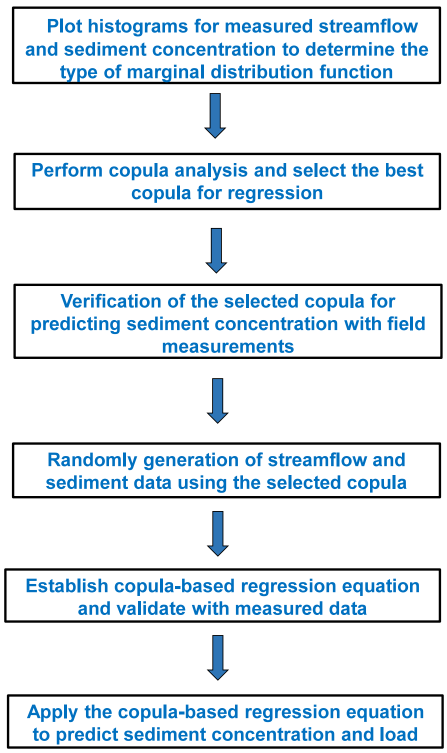

Figure 3. A flowchart used for copula-based regression analysis.

To develop scientific confidence, sediment concentration and load predictions from the copula method were further compared with those from SWAT model. The SWAT is a widely used model for simulating hydrological processes and water quality in large and complex watersheds affected by surficial processes such as land use, management, and climate variability (https://swat.tamu.edu/). In this study, the SWAT model was developed to predict sediment load in the Apalachicola River Basin, Florida (Figure 1). The primary reason for choosing this basin is that there are sufficient measured sediment data for SWAT calibration. Additionally, the hydrological processes for this basin were previously simulated (Ouyang et al., 2022) using the HAWQS (Hydrologic and Water Quality System) model (https://hawqs.tamu.edu/). The HAWQS model is a customized version of SWAT. The sediment load simulation was activated in the HAWQS model and calibrated using the measured total dissolved solid data for a period from 1993 to 2007.

3. Results

3.1. Copula Selection and Regression

Five commonly used copulas, namely BB8, Clayton, Frank, Gumbel, and Normal copulas, were employed to fit the measured streamflow and sediment content data. Based on Kendall's τ and p values, the Normal copula was identified as the best fit for the data across the three study sites. Specifically, the Normal copula had the highest τ and lowest p-values (Table 1), making it the optimal choice for developing copula-based regression equations between streamflow and sediment concentration. Figure 5 compares the simulated sediment concentrations using the randomly generated streamflow data obtained from the Normal copula with field measurements. Results showed that the copula simulated sediment concentrations matched the observed sediment concentrations well for the study sites in New York ($\tau = 0.611$ and $p = 0.05$) and Florida ($\tau = 0.591$ and $p = 0.028$) but not for the study site in Idaho ($\tau = 0.59$ and $p = 0.106$) because the value of p for this site was >0.05.

Using the data randomly generated by the Normal copula, the linear regression equations between streamflow and sediment concentration were developed for the three study sites (Figure 6):

$$Y_{\text{Sediment Conc}} = 0.0626X_{\text{Streamflow}} + 90.443, \tag{6}$$

for the Hudson River, New York,

$$Y_{\text{Sediment Conc}} = 0.0244X_{\text{Streamflow}} + 56.858, \tag{7}$$

for the Apalachicola River, Florida, and

$$Y_{\text{Sediment conc}} = 0.0605X_{\text{Streamflow}} - 1.0796, \tag{8}$$

for the Clearwater River, Idaho. These linear regression equations were not possible to obtain using the measured data (Figure 2) but were feasible after copula transformation (Figure 6). While there are no physical interpretations about the coefficients and constants, the above three copula-based regression equations were used to predict sediment concentrations and loads based on measured streamflow data. It should be pointed out that Equation 8 is not advisable to use because it did not meet the dependence (or correlation) criteria of copula analysis, as the p value was >0.05, although the statistical metrics appeared to be reasonable (Figure 6c). Equation 8 is only used to demonstrate how the copula-based regression fails when it does not meet the dependency criteria.

10.1029/2025WR040224

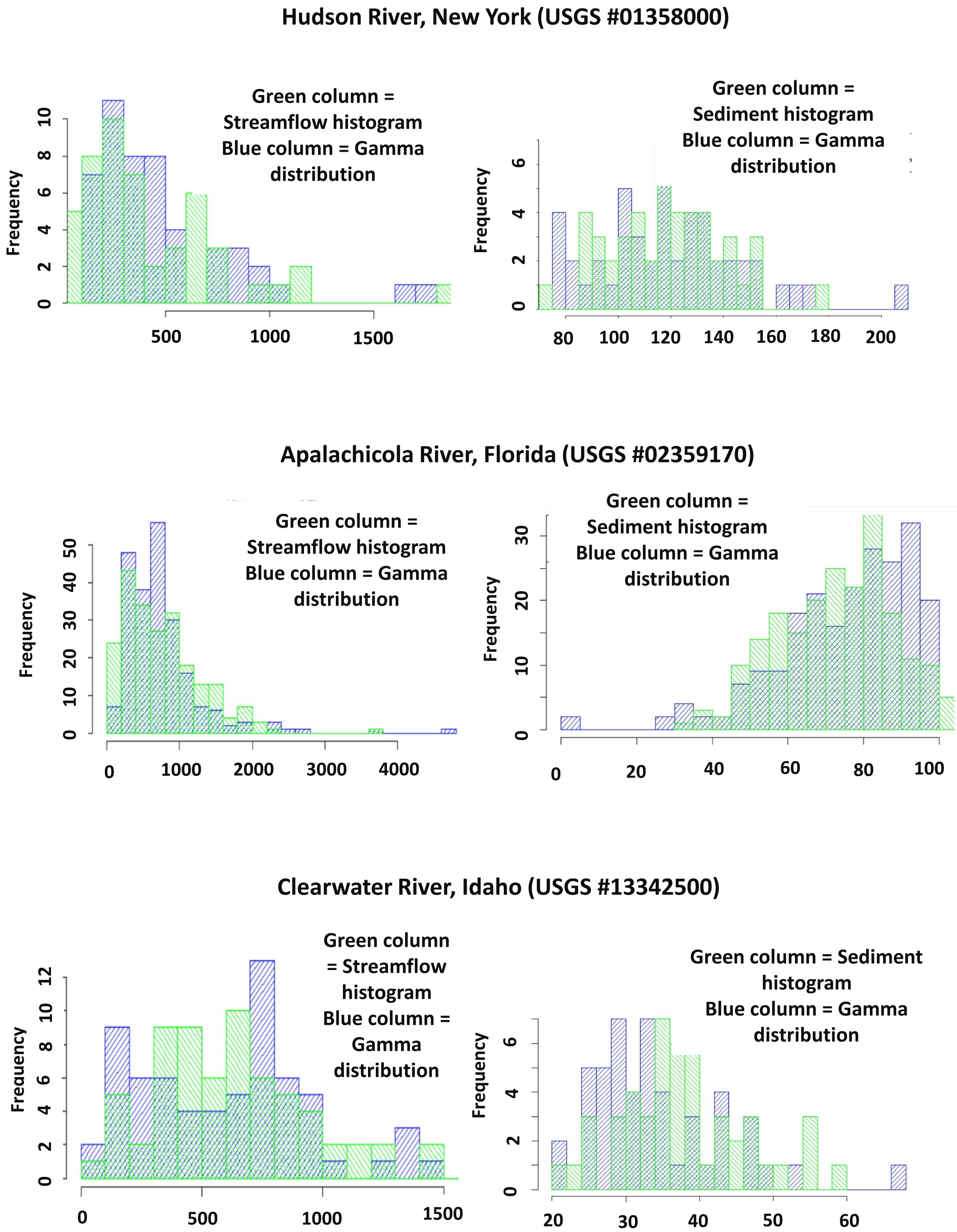

Figure 4. Histogram plots of observed streamflow and sediment data from the three USGS gauge stations, showing a Gamma type of distribution.

3.2. Validation of Copula-Based Prediction

While the selection of the Normal copula was verified using Kendall's τ values at $p \leq 0.05$, further validation of the sediment concentration predictions was performed using Equations 6–8 against measured data. This validation provided strong scientific confidence in applying the copula-based regression method. Comparisons between the copula-based regression predictions and field measurements for sediment (or total dissolved solids) concentrations are shown in Figure 7. The statistical metrics ranged from 0.92 to 0.96 for R^2, from 0.09 to 0.97 mg L^{-1} for nRMSE, from 0.416 to 3.294 for NSE, and from 0.0004 to 129 for PBIAS, and were <0.001 for

10.1029/2025WR040224

Table 1
Comparisons of the Kendall's τ and p Values Among the BB8, Clayton, Frank, Gumbel, and Normal Copulas for the Three Study Sites

	Hudson River, New York		Apalachicola River, Florida		Clearwater River, Idaho	
Copula	τ	p-value	τ	p-value	τ	p-value
BB8	0.331	0.723	0.332	0.558	0.331	0.897
Clayton	0.213	0.833	0.202	0.828	0.218	0.843
Frank	0.157	0.930	0.160	0.981	0.169	0.963
Gumbel	0.036	0.771	0.070	0.983	0.037	0.921
Normal	0.611	0.05	0.591	0.028	0.590	0.106

Note. These values were obtained by fitting the observed streamflow and sediment load with the copulas.

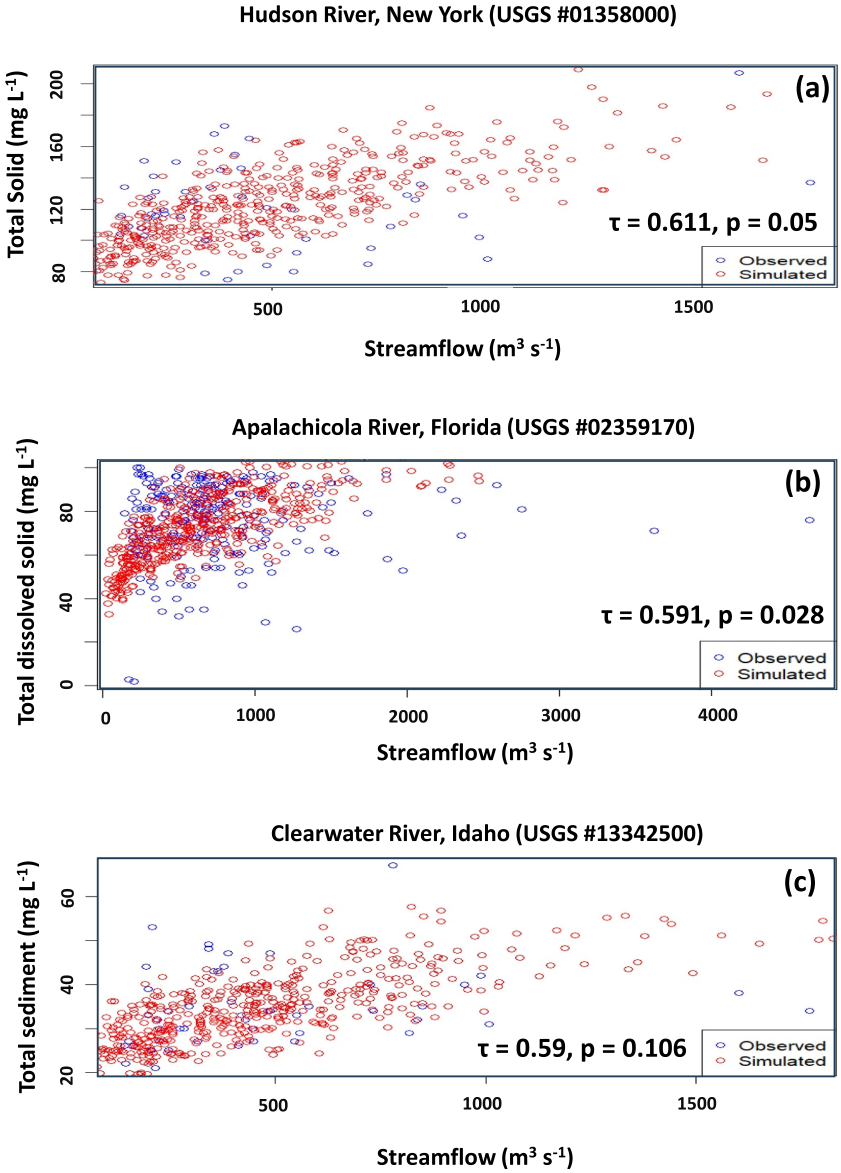

Figure 5. Comparison of the copula simulated stream flows and sediment contents with the field observations at the three study sites.

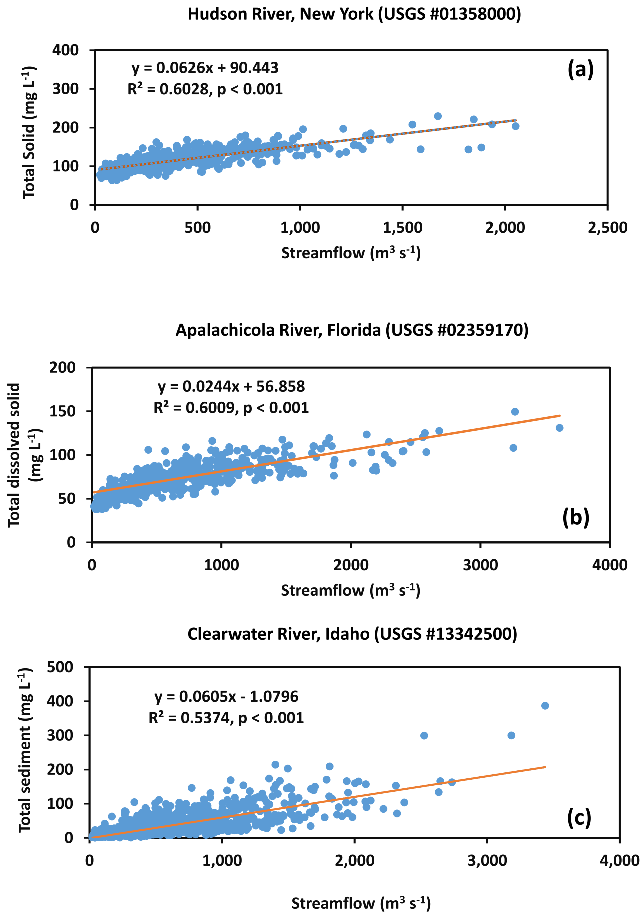

Figure 6. Copula-based regressions between streamflow and sediment contents at the three study sites.

p-value. While these statistical matrices were reasonable for applying the copula-based regressions to predict the sediment concentrations using streamflow data for the study sites in New York and Florida, the intercept in the equation ($Y_{prediciton} = 0.457X_{measurement}$) and NSE value (3.294) for the study site in the Clearwater River, Idaho were not acceptable. The prediction only accounted for 45.7% of the measurements and the NSE value was outside the acceptable range.

Figure 8 presents time series comparisons between predicted and measured sediment concentrations, showing that the copula-based regressions reasonably captured peaks and valleys in sediment concentrations for the study sites in New York and Florida while overestimating the sediment concentrations for the study site in Idaho. Therefore, Equation 8 was not applied to predict sediment concentrations and loads based on streamflow data for this study site.

3.3. Comparison With SWAT Model

Comparison of the copula and SWAT predicted daily sediment loads with the measured sediment loads at the basin outlet in the Apalachicola River, Florida over the period from 2008 to 2020 is presented in Figure 9. The predicted daily sediment loads were obtained by first predicting the sediment concentrations using Equation 7 and multiplying the daily concentrations with daily streamflow values.

While both the copula and SWAT methods predicted the daily sediment loads reasonably well, differences in predicting daily sediment loads between the two methods were discernible. As shown in Figure 9, the copula method had higher values of R^2 (0.861 for copula and 0.552 for SWAT) and NSE (0.86 for copula and 0.13 for SWAT) and lower nRMSE (0.276 for copula and 0.695 for SWAT), indicating that the copula method outperformed the SWAT model in predicting daily sediment loads at the basin outlet in the Apalachicola River, Florida.

3.4. Application of Copula-Based Regression

Predicted daily sediment concentrations from 2011 to 2021 at the two study sites are shown in Figure 10. These predictions were accomplished using the copula-based regression equations (Equations 6 and 7). It should be noted that no measured sediment data were available for the Hudson River, New York during this period, whereas some sparse and discontinued measured sediment data were available for the Apalachicola River, Florida. While the predicted sediment concentrations varied with study sites and times, the average sediment concentration over the period was in the order: Hudson River (118 mg L^1) > Apalachicola River (70 mg L^{-1}).

Predicted annual sediment loads over an 11-year period from 2011 to 2021 at the two study sites are shown in Figure 11. The annual sediment loads ranged from 1.2E+6 to 2.8E+6 tons for the Hudson River, New York and 0.5E+6 to 1.8E+6 tons for the Apalachicola River, Florida. Average annual sediment load was higher for the Hudson River (1.8E+6 tons) than for the Apalachicola River (1.4E+6 tons). While there was an increasing trend of annual sediment loads at the Hudson River, New York at $\tau = 0.289$, $p = 0.283$ and a decreasing trend of annual sediment loads at the Apalachicola River, Florida at $\tau = -0.244$, $p = 0.371$ (Figure 11), these trends were not significant at $p \leq 0.05$ based on the Mann-Kendell test.

4. Discussion

4.1. Copula Method Selection and Validation

As presented in Figure 2, there were no good correlations between streamflow and sediment concentration for the three study sites because the values of R^2 were very poor. Therefore, it is not advisable to predict sediment

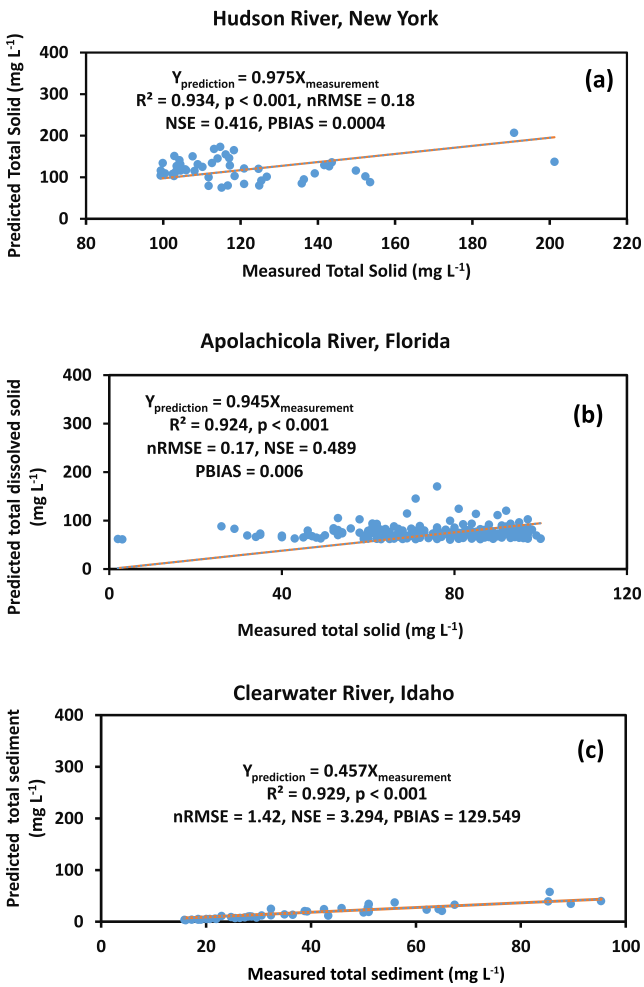

Figure 7. Validation of the copula simulated sediment contents with field observations at the three study sites.

concentrations and loads based on streamflow data using traditional methods (e.g., simpler linear regression) in this study. Normally, predicting sediment concentration and load in rivers using measured streamflow data are difficult because these two variables often lack a good linear correlation in many watersheds worldwide. These occur because of spatial and temporal variations of sediment contents in rivers, resulting from variations of watershed conditions. However, with the adoption of copula-based regression, it may be feasible to predict sediment concentrations using streamflow data as presented in Figure 6.

The copula-based regression process involves selecting the best copula based on analyzing measured streamflow and sediment concentration histograms, selecting the best copula among the five commonly used copulas by checking Kendall's τ and p values when comparing the copulas' predictions to field measurements, verifying the selected copula with another set of measured data, and finally developing a copula regression equation between streamflow and sediment concentration using data randomly generated by the selected copula.

Based on the statistical metrics listed in Table 1, Normal Copula was selected as it has highest τ and lowest p values for the basin in Hudson River, New York and Apalachicola River, Florida. While the highest τ value was also found for Normal Copula in Clearwater River, Idaho, the p value (0.106) was not at the significant level of $\leq$0.05. Therefore, it is normally not recommended to apply the copula method for this study site. In general, a key assumption of applying copula-based regression to predict sediment concertation is the existence of a good correlation (dependence) between measured sediment content and streamflow after copula transformation. This correlation was determined using Kendall's τ at $p \leq 0.05$. According to Schober et al. (2018), correlations are weak at $\tau < 0.39$, moderate at $0.4 \leq \tau \leq 0.69$, and strong at $\tau > 0.7$. Based on the τ and p values shown in Figure 5, the correlations between sediment concentration and streamflow generated by the Normal copula at two study sites (Figures 5a and 5b) are acceptable. An elaborate discussion of scientific basis for the τ criteria can be found elsewhere (Joe, 2014; Nelsen, 2006; Schober et al., 2018).

To develop confidence in applying the copula-based regression equations, a validation of the predicted sediment concentrations with field measurements was also performed (Figures 7 and 8). Given the reasonable statistical metrics (Figure 7) and peak and valley visualization (Figure 8), the copula-based regression method has proven to be a valuable tool for predicting sediment concentrations using measured streamflow data for the basins in Hudson River, New York and Apalachicola River, Florida, a task that traditional regression methods may be difficult to accomplish. It should be noted that the predicted sediment concentrations for the basin in Clearwater River, Idaho were not acceptable as the value of NSE (3.294) was unrealistic. The copula-based regression (Equation 8) for this basin overestimated the sediment concentrations as compared to those of the field measurements. This further confirmed that the copula method was not appropriate for this study site because the copula transformation of the measured sediment concentration and streamflow data did not meet the dependency criteria of copula analysis at $p \leq 0.05$.

4.2. Copula Versus SWAT

To develop users' confidence, the copula method established in this study was compared with the SWAT model (Figure 9). In general, both the copula and SWAT well predicted daily sediment loads at the basin outlet in the Apalachicola River, Florida over the period from 2008 to 2020 when compared with the field measurements as indicating by the statistical metrics. In particular, the copula method was somewhat better than the SWAT model

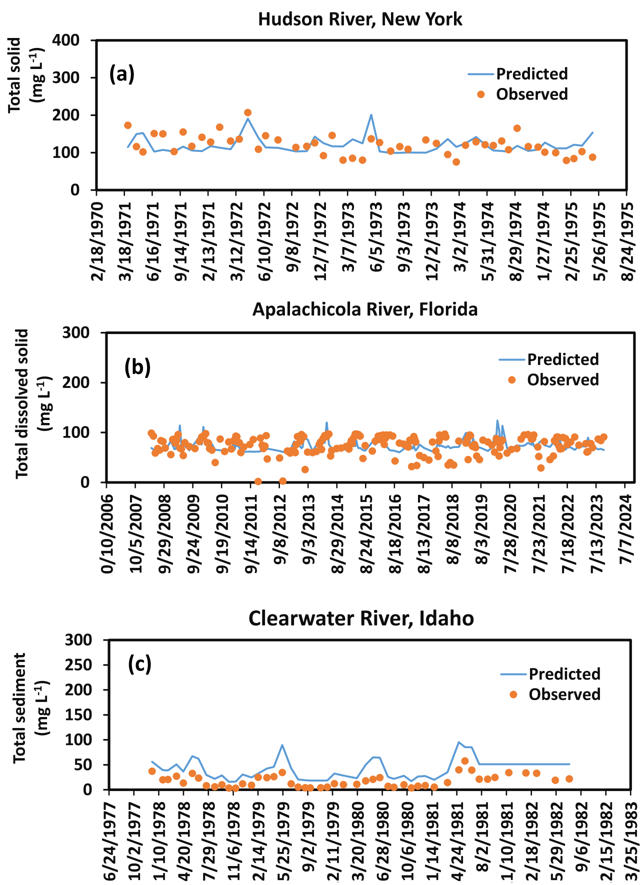

Figure 8. Time series comparison of the copula simulated sediment contents with field observations at the three study sites.

for this study site because the former had higher values of R^2 and NSE than those of the latter (Figure 9). A major advantage for using the copula method is that the method only uses streamflow data, which is cost-effective and timesaving. Complex watershed models such as HSPF and SWAT are used to predict sediment loads in rivers. However, these models require extensive input parameters, climatic data, and GIS resources, making them time-consuming and labor-intensive to develop and apply. They also require calibration and validation for both hydrological and sediment components, further adding to the complexity. The copula-based regression approach, on the other hand, is time-saving and cost-effective, providing a viable alternative for sediment concertation and load predictions. Furthermore, the copula method could be extended to predict other water quality constituents (e.g., nitrogen and phosphorus) using streamflow data if a suitable correlation (or dependence) is established after copula transformation. Further research is therefore warranted to explore this potential.

In the past, the LOADEST model was applied to estimate sediment load in streams and rivers. In addition to the model's limitations such as significant biases when the relationship between the log-transformed water quality constituent concentration and streamflow is not strong and considerable differences in the seasonal relationship shape (Hirsch, 2014), this decade old DOS-version model from US Geological Survey is not compatible with recent computer system. While a web-based LOADEST model was published online by Park et al. (2015), the web tool is no longer available for some reason. There is a also a web-based tool for LOADEST developed by the Department of Engineering, Purdue University (https://engineering.purdue.edu/WaterDST/StandaloneTool/References.html). This tool does not work either. Furthermore, the LOADEST model is primarily used to predict sediment load rather than sediment concentration, while the copula method predict both sediment concertation and load.

4.3. Copula Application, Sensitivity, Uncertainty, and Limitation

Variations in annual average sediment loads with study sites occurred depending on watershed conditions and anthropogenic activities. It is worth noting that the annual average sediment load may not be always proportional to watershed drainage area. For example, the drainage area and annual average sediment load were, respectively, 20,953 km^2 and 1.8E+6 tons at the study site in Hudson River, New York but were, respectively, 49,727 km^2 and 1.4E+6 tons at the study site in Apalachicola River, Florida. In other words, the drainage area in Hudson River watershed was about 2.4 times less than in the Apalachicola River watershed, whereas the annual sediment load in Hudson River basin was about 28% higher than in the Apalachicola River basin. Results indicated that land use and anthropogenic activities could also play an important role in annual average sediment load (Chakrapani, 2005). Overall, no significant trends in sediment loading were observed over the 11-year period from 2011 to 2021 (Figure 11), implying that sediment loads remained relatively unchanged at these study sites.

In Normal copula analysis, only input parameter can be changed to determine the sensitivity of dependency (or τ value) between two variables (e.g., sediment concentration and streamflow for this study) is the Pearson correlation coefficient θ as shown in Equation 2. This is an advantage in using Normal copula, especially for high dimension analysis. It has been reported that the dependence is weak at $\theta < 0.39$, moderate at $0.40 \le \theta \le 0.69$, and strong at $0.70 < \theta \le 1.0$ (Schober et al., 2018). Thus, a sensitivity analysis was performed to inspect the variations of τ value by changing the Pearson correlation coefficient. Results revealed that the changes in θ value from 0.5 to 1.0 increased the τ value from 0.57 to 0.59 for the Apalachicola River Basin, Florida and 0.60 to 0.61 for the Hudson River Basin, New York, indicating the Normal Copula was not sensitive to the Pearson correlation coefficient in this study.

10.1029/2025WR040224

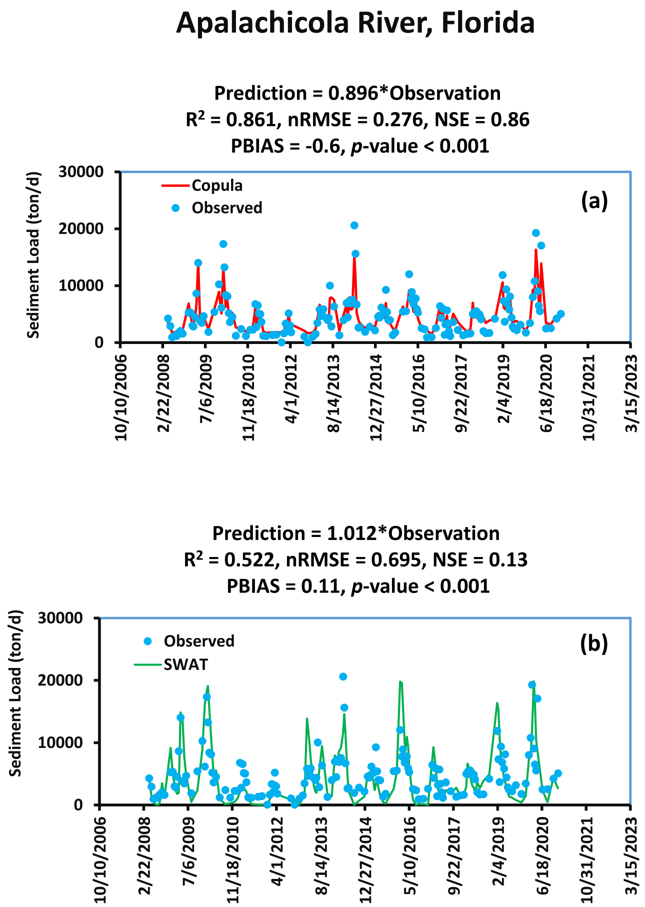

Figure 9. Comparisons of copula and SWAT predicted daily sediment loads with measured ones.

Uncertainty analysis is essential in hydrological and water-quality modeling by assessing how variability in input data, model parameters, and structural assumptions propagates through predictive models and influences the reliability of their outputs. Common approaches include Monte Carlo simulation, GLUE, Bayesian inference, and bootstrap resampling (Farzana, 2023). In this study, a bootstrap-based uncertainty framework was applied to evaluate: (a) the Normal copula's Pearson correlation coefficient, and (b) the propagated predictive uncertainty of sediment-concentration estimates derived from the Normal copula.

Figure 12 shows the Pearson correlation coefficient (θ) uncertainties and prediction-interval bounds for the study sites in New York and Florida. The point estimates of the Pearson correlation coefficients were 0.787 for New York and 0.808 for Florida, indicating strong positive dependence between streamflow and sediment concentration. The corresponding 95% confidence intervals were narrow—0.752 to 0.820 for New York and 0.775 to 0.836 for Florida—reflecting differences of only 0.068 and 0.061, respectively. These narrow intervals indicate stable copula parameter estimates with low sampling uncertainty.

The uncertainty-propagation results further show that majority of predicted sediment concentrations fall within the 95% prediction interval. Fewer than 4% of predictions lie outside the interval at both study sites, which is consistent with expectations, because approximately 5% of observations are naturally expected to fall outside a 95% prediction band (Farzana, 2023). This alignment suggests that the Normal copula adequately captures the variability in sediment concentrations and provides reliable predictive performance for both locations.

Nonlinear or hysteretic sediment content-streamflow relationships can arise when sediment concentration is asynchronous to streamflow during extreme weather events such as thunderstorms and hurricanes. Under these conditions, plots of sediment concentration versus streamflow may display looping patterns, including clockwise hysteresis (peak sediment concentration preceding peak streamflow or a “first flush” effect) and counterclockwise hysteresis (peak sediment concentration lagging peak streamflow). The direction and magnitude of hysteresis depend on factors such as sediment availability, flow intensity, and the distance between sediment source and stream location (Hamshaw et al., 2018; Long et al., 2024). Such situations may produce a low τ value ($\leq$0.4) and a high p value (>0.05) when assessing the sediment and streamflow relationship during copula selection (see Step 2 of Section 2.2), indicating that the proposed copula method may not be applicable or may require further investigation.

It should be aware that sparse measured sediment and streamflow data are required as a prerequisite to establish a copula-based regression equations for applying the method developed in this study. If the τ value is $\leq$0.4 and p value is >0.05 when comparing the copula predictions with field measurements, it is not advisable to apply the copula method to avoid uncertainties and failure as demonstrated for the study site in Clearwater River Basin, Idaho. Additionally, the method has not been applied to event-driven conditions such as hurricanes, floods, droughts, and snowmelt because no such measured data are available for the study sites. Song et al. (2024) applied the time-varying copula-based approach to assess extreme rainfall and high-water level in the Qinhuai River Basin and Yangtze River, China. These authors argued that the time-varying copula-based approach could provide a flood risk assessment framework. Therefore, further study is warranted to tackle this issue using Normal copula and time-varying copula-based approaches when the measured streamflow and sediment data are available under such extreme weather conditions. Furthermore, while the copula method was used to predict sediment content based on streamflow data for a single gauge, it can be extended for multiple gauges within a basin.

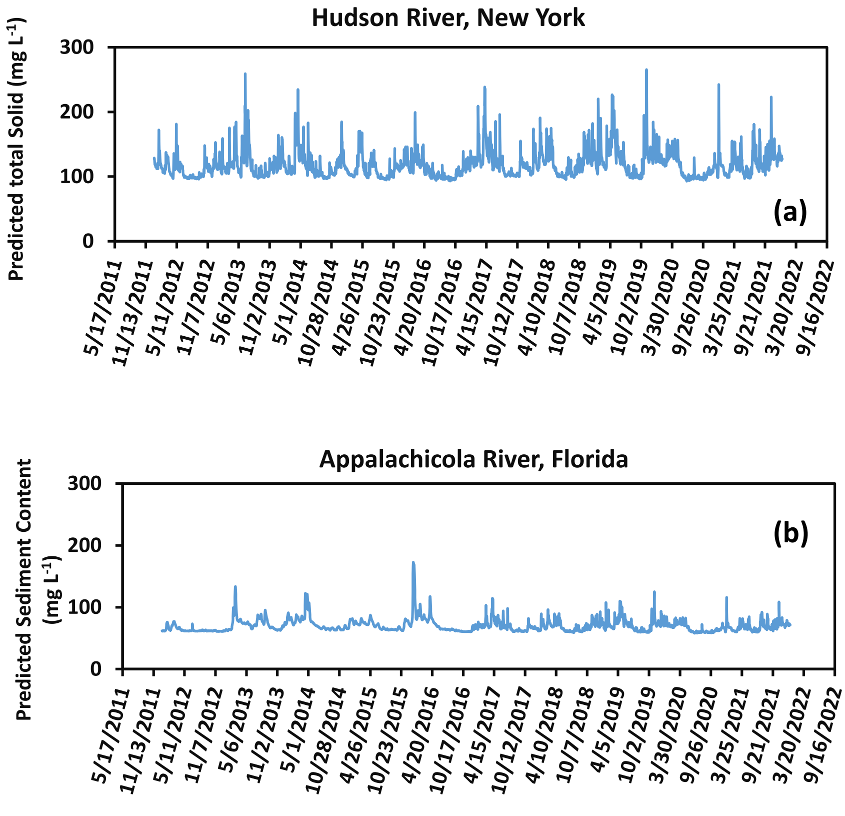

Figure 10. Predicted daily sediment contents at the two study sites.

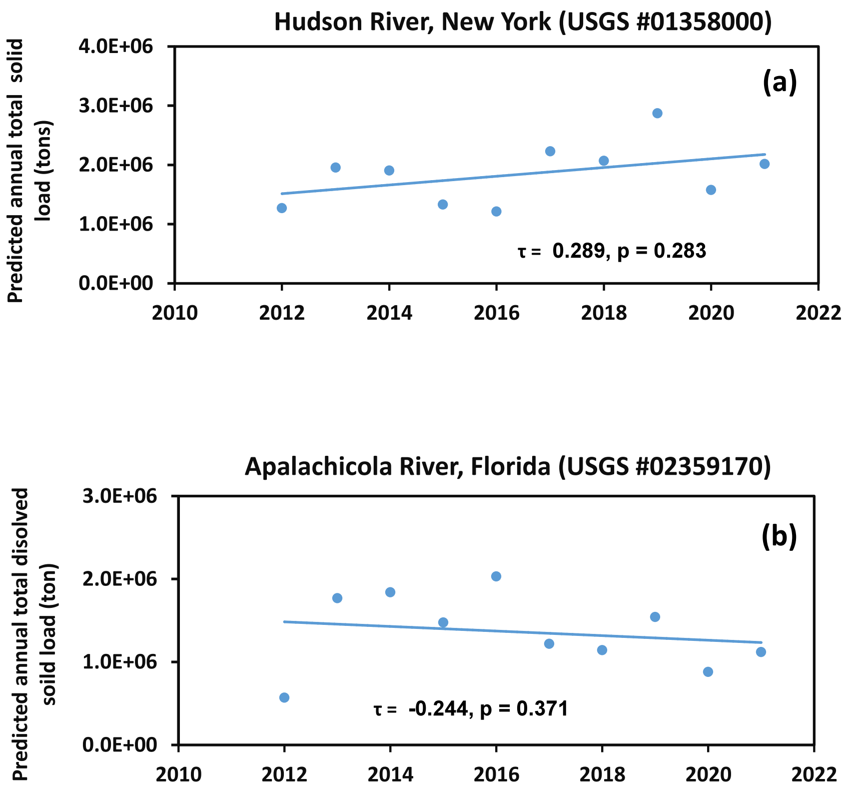

Figure 11. Predicted annual sediment contents at the two study sites.

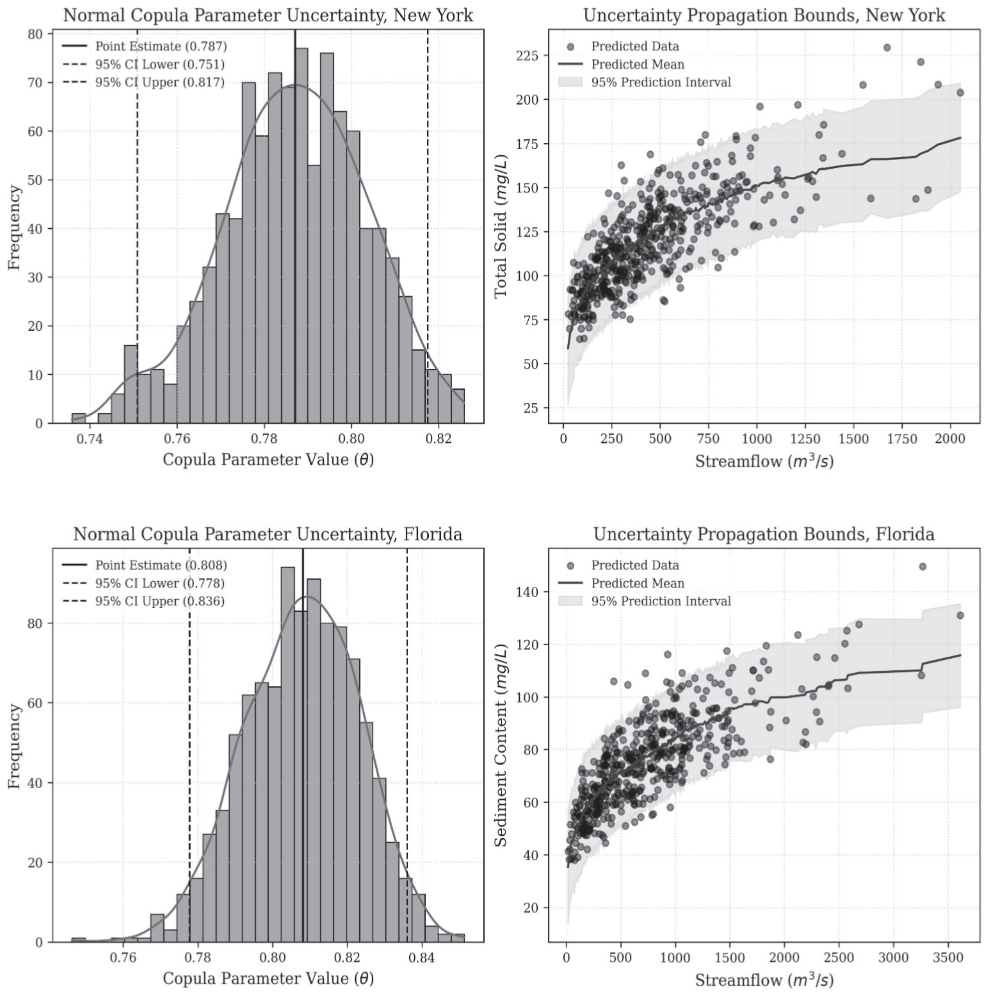

Figure 12. Uncertainty analysis of the two study sites.

5. Conclusions

Sediment pollution in surface waters poses significant environmental and ecological challenges. While watershed models have been used to predict sediment loads in rivers, their reliance on vast amounts of input data and calibration efforts limits their efficiency. Additionally, the lack of good linear correlations between streamflow and sediment load in many watersheds complicates predictions using traditional methods.

To address these challenges, this study developed an alternative approach using copula-based regression. The process included selecting the best copula, verifying it against field measurements, and developing regression equations for predicting sediment concentrations and loads. Validation with field measured data showed very good statistical performance, highlighting the copula-based regression method as a robust tool for predicting sediment concentrations and loads based on streamflow data.

Compared to the SWAT model, the copula-based regression method is more time-saving and cost-effective. Additionally, this method has potential to predict other water quality constituents, provided there is a good correlation (dependence) validating by the τ and p values after copula transformation.

Furthermore, while the copula-based regression method is a promising approach for estimating sediment concentration and load using streamflow data, further study is warranted to test its applicability by comparing with machine learning approaches. Although real-time predictions of sediment concentration and load in streams are not considered in this study, interested readers can apply the copula-based method to predict real-time sediment concertation and load if the real-time sediment and streamflow data is available.

Conflict of Interest

The author declares no conflicts of interest relevant to this study.

Availability Statement

The original data sets used for analysis are publicly available by USGS as described in Section 2.1. More specially, the data for the Hudson River at Green Island, New York was downloaded from https://waterdata.usgs.gov/nwis/uv?site_no=01358000&legacy=1, for the Apalachicola River near Sumatra, Florida was downloaded from https://waterdata.usgs.gov/nwis/uv?site_no=02359170&legacy=1, and for the Clearwater River at Spalding, Idaho was downloaded from https://waterdata.usgs.gov/nwis/uv?site_no=13342500&legacy=1. The example R script and input data for copula analysis can be download at: https://figshare.com/s/f91fc37dee7b3d8e5773.

Acknowledgments
None.

References

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Mathematics & Economics*, *44*(2), 182–198. https://doi.org/10.1016/j.insmatheco.2007.02.001

Alizadeh, H., Mousavi, S. J., & Ponnambalam, K. (2018). Copula-based chance-constrained hydro-economic optimization model for optimal design of reservoir-irrigation district systems under multiple interdependent sources of uncertainty. *Water Resources Research*, *54*(8), 5763–5784. https://doi.org/10.1029/2017wr022105

Almasalmeh, O., Saleh, A. A., & Mourad, K. A. (2022). Soil erosion and sediment transport modelling using hydrological models and remote sensing techniques in Wadi Billi, Egypt. *Modeling Earth Systems and Environment*, *8*(1), 1215–1226. https://doi.org/10.1007/s40808-021-01144-1

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., et al. (2012). Swat: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1491–1508.

Bicknell, B. R., Imhoff, J. C., Kittle, J. L., Jr., Jobes, T. H., Donigian, A. S., Jr., & Johanson, R. (2001). *Hydrological simulation program-Fortran: HSPF version 12 user's manual* (Vol. 845). AQUA TERRA Consultants, Mountain View, California.

Chakrapani, G. J. (2005). Factors controlling variations in river sediment loads. *Current Science*, 569–575.

Cote, M. P., Genest, C., & Omelka, M. (2019). Rank-based inference tools for copula regression, with property and casualty insurance applications. *Insurance Mathematics & Economics*, *89*, 1–15. https://doi.org/10.1016/j.insmatheco.2019.08.001

Farzana, S. Z. (2023). Uncertainty in hydrological modelling: A review. *International Journal of Hydrology Research*, *8*(1), 1–13. https://doi.org/10.18488/ijhr.v8i1.3297

Frees, E. W., & Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, *103*(484), 1457–1469. https://doi.org/10.1198/016214508000000823

Golmohammadi, G., Prasher, S., Madani, A., & Rudra, R. (2014). Evaluating three hydrological distributed watershed models: MIKE-SHE, APEX, SWAT. *Hydrology-Basel*, *1*(1), 20–39. https://doi.org/10.3390/hydrology1010020

Hamshaw, S. D., Dewoolkar, M. M., Schroth, A. W., Wemple, B. C., & Rizzo, D. M. (2018). A new machine-learning approach for classifying hysteresis in suspended-sediment discharge relationships using high-frequency monitoring data. *Water Resources Research*, *54*(6), 4040–4058. https://doi.org/10.1029/2017wr022238

Hirsch, R. M. (2014). Large biases in regression-based constituent flux estimates: Causes and diagnostic tools. *Journal of the American Water Resources Association*, *50*(6), 1401–1424. https://doi.org/10.1111/jawr.12195

Joe, H. (2014). *Dependence modeling with copulas*. CRC Press.

Krause, T., Beck, E. V., Cherkaoui, R., Germond, A., Andersson, G., & Ernst, D. (2006). A comparison of Nash equilibria analysis and agent-based modelling for power markets. *International Journal of Electrical Power and Energy Systems*, *28*(9), 599–607. https://doi.org/10.1016/j.ijepes.2006.03.002

Long, Y., Lei, M., Li, T., Xiao, P., Liu, S., Xu, J., et al. (2024). Suspended sediment-discharge hysteresis characteristics and controlling factors in a small watershed of southern China. *Catena*, *243*, 108198. https://doi.org/10.1016/j.catena.2024.108198

Masarotto, G., & Varin, C. (2017). Gaussian Copula regression in R. *Journal of Statistical Software*, *77*(8), 1–26. https://doi.org/10.18637/jss.v077.i08

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Noe, G. B., Cashman, M. J., Skalak, K., Gellis, A., Hopkins, K. G., Moyer, D., et al. (2020). Sediment dynamics and implications for management: State of the science from long-term research in the Chesapeake Bay watershed, USA. *WIREs Water*, *7*(4), e1454. https://doi.org/10.1002/wat2.1454

Ouyang, Y. (2022a). A gap-filling tool: Predicting daily sediment loads based on sparse measurements. *Hydrology-Basel*, *9*(10), 181. https://doi.org/10.3390/hydrology9100181

Ouyang, Y. (2022b). A hybrid of copula prediction and time series computation to estimate stream discharge based on precipitation data. *Journal of the American Water Resources Association*, *58*(3), 471–484. https://doi.org/10.1111/1752-1688.13003

Ouyang, Y., Grace, J. M., Parajuli, P. B., & Caldwell, P. V. (2022). Impacts of multiple hurricanes and tropical storms on watershed hydrological processes in the Florida panhandle. *Climate*, *10*(3), 42. https://doi.org/10.3390/cli10030042

Ouyang, Y., Higman, J., Campbell, D., & Davis, J. (2003). Three-dimensional kriging analysis of sediment mercury distribution: A case study. *Journal of the American Water Resources Association*, *39*(3), 689–702. https://doi.org/10.1111/j.1752-1688.2003.tb03685.x

Ouyang, Y., & Sun, C. (2024). A copula approach for predicting tree sap flow based on vapor pressure deficit. *Forests*, *15*(4), 695. https://doi.org/10.3390/f15040695

Park, Y. S., Engel, B. A., Frankenberger, J., & Hwang, H. (2015). A web-based tool to estimate pollutant loading using LOADEST. *Water Science*, *7*(9), 4858–4868. https://doi.org/10.3390/w7094858

Parsa, R. A., & Klugman, S. A. (2011). Copula regression. *Variance Advancing and Science of Risk*, *5*(1), 45–54. https://doi.org/10.66573/001c.84912

Renard, B., & Lang, M. (2007). Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, *30*(4), 897–912. https://doi.org/10.1016/j.advwatres.2006.08.001

Rubin, D. B. (1976). Multivariate matching methods that are equal percent bias reducing. 1. Some examples. *Biometrics*, *32*(1), 109–120. https://doi.org/10.2307/2529342

Runkel, R. L., Crawford, C. G., & Cohn, T. A. (2004). Load estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ane.0000000000002864

Schwarz, G. E., Hoos, A. B., Alexander, R., & Smith, R. (2006). The SPARROW surface water-quality model: Theory, application and user documentation.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges (pp. 229–231).

Song, M., Zhang, J., Liu, Y., Liu, C., Bao, Z., Jin, J., et al. (2024). Time-varying copula-based compound flood risk assessment of extreme rainfall and high water level under a non-stationary environment. *Journal of Flood Risk Management*, *17*(4), e13032. https://doi.org/10.1111/jfr3.13032

Taebi, A., & Mansy, H. A. (2017). Time-frequency distribution of seismocardiographic signals: A comparative study. *Bioengineering*, *4*(2), 32. https://doi.org/10.3390/bioengineering4020032