

Article

Evaluation and Validation of Estimated Sediment Yield and Transport Model Developed with Model Tree Technique

Woochul Kang ^{*,†} , Kyungsu Lee [†]  and Eun-kyung Jang [†] 

Department of Hydro Science and Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang 10285, Korea; kidhan28@kict.re.kr (K.L.); jang@kict.re.kr (E.-k.J.)

* Correspondence: kang@kict.re.kr; Tel.: +82-10-8505-0471

† These authors contributed equally to this work.

Abstract: This study evaluated the applicability of existing sediment yield and transport estimation models developed using data mining classification and prediction techniques and validated them. Field surveys were conducted by using an acoustic Doppler current profiler and laser in situ scattering and transmission at measuring points in the main stream of the Nakdong River located where the tributaries of the Geumho, Hwang, and Nam Rivers join. Surveys yielded estimations of water velocity, discharge, and suspended sediment concentrations were measured. In contrast with models based on the general watershed characteristics factors, some models based on hydraulic explanatory flow variables demonstrated an excellent predictability. This is because the selected submodels for validation, which provided excellent prediction results, were based on a large number of calibration data. It indicates that a sufficient number of reliable data is required in developing a sediment yield estimation model using data mining. For practical applications of data mining to extant sediment yield estimation models, comprehensive considerations are required, including the purpose and background of model development, and data range. Furthermore, the existing models should be periodically updated with the consideration of temporal and spatial lumping problems.



Citation: Kang, W.; Lee, K.; Jang, E.-k. Evaluation and Validation of Estimated Sediment Yield and Transport Model Developed with Model Tree Technique. *Appl. Sci.* **2022**, *12*, 1119. <https://doi.org/10.3390/app12031119>

Academic Editor: Kelin Hu

Received: 2 December 2021

Accepted: 14 January 2022

Published: 21 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: data mining; model tree; sediment yield; sediment transport; specific degradation

1. Introduction

Soil erosion occurs when the shear force of the overland flow is greater than soil cohesion. The sediments generated by weathering are transported by natural media and deposited in rivers before or after reaching them. In this process, the sediments have a direct impact on the hydraulic function and cause multiple problems from the aspect of river maintenance and management [1]. For example, the sediment transfer can cause rise of flood, decrease stability of the embankment, and headward erosion [2,3]. It can also cause infrastructure instability and changes in the ecological environment [4,5]. Therefore, for the design, operation, and management of ecofriendly and sustainable rivers, it is instrumental to have a sufficient understanding of the discharge of inflow and sediment transport and quantitatively identify and predict the sediment yield [6,7]. To solve erosion and sediment related problems, various models have been proposed for the estimation of the sediment yield or sediment transport rate [8–10]. Especially, numerous mathematical formulas based on statistical methods have focused on sediment yield quantification [11–14]. However, with the present level of knowledge, elucidation, and accurate quantitative prediction of the complex process of sediment generation, transport, and deposition that must consider various factors, such as geography, topography, geology, pedology, climatology, forestry, and hydraulics of rivers and basins, remain highly challenging [1].

The developed models for sediment yield are generally classified as physical, conceptual, and empirical, and an appropriate model is used in consideration of the purpose and conditions [7]. In the case of the empirical model used for prediction of sediment

yield and its validation in this study, the relationship between various characteristics and sediment observation results were used. To develop an empirical model with excellent predictability, reliable measurement results are required. Even though the models were developed using reliable measurements, there are cases in which various empirical models that have been developed to date show a considerable difference even when subjected to the same gauging points and identical hydraulic conditions. Recently, active research has been conducted for the development of a sediment yield estimation model with excellent prediction performance using data mining techniques that allows systematic and automatic derivation of statistical rules and patterns from big data composed of various variables. Jain [15] estimated the sediment concentration using an artificial neural network (ANN) model, and Lin and Namin [16] used an ANN model and numerical analysis in combination for analysis of suspended sediment transport. Nagy et al. [17] used an ANN model for sediment concentration estimation and compared the results with the values derived from existing formulas. Bhattacharya et al. [18] performed comparative evaluation using an ANN and a model trees method for the estimation of sediment yield. Khan et al. [19] employed ANN and wavelet-cum-ANN methods to estimate suspended sediment loads in the Elbe River in Germany. In addition, Nhu et al. [20] presented erosion susceptibility mapping through the development of a hybrid model that simultaneously utilized various data mining techniques. In South Korea, Jang [13] developed a model to predict the total sediment load through the relationship between the river channel and hydraulic factors using the model tree method, one of the data mining techniques. Additionally, using the same technique, Kang et al. [14] developed a model based on various watershed characteristics to predict specific degradation (SD).

A data mining technique could be one of the approaches to predict sediment yield; however, this method only focuses on the accuracy of prediction. Therefore, the models based on this technique should consider appropriate hydraulic and hydrogeomorphic conditions, development purpose, and the range of data used. Prior studies related to sediment yield and transport estimation model using data mining provided outstanding predictive results, but there were insufficient considerations for model development processes, application, and validation. In this study, considering the reported and analyzed results that more than 99% of the total sediment load is transported as suspended sediment in the main stream of Nakdong River, this study aimed to perform validation for sediment yield prediction models that were developed using data mining techniques for rivers in South Korea. Additionally, guidelines for the development and application of the model using the data mining technique were proposed by evaluating the existing models.

2. Materials and Methods

2.1. Study Site

The main rivers in South Korea include the Han, Nakdong, Geum, Seomjin, and Yeongsan Rivers (Figure 1a,b).

These rivers generally flow from east to west (except the Nakdong River) due to the topographic characteristics of the country, whereby 70% of the national land is composed of mountains, with most of the steep mountains located in the eastern part of the country. In particular, given that most of the main streams of the rivers have relatively gentle slopes, long lengths, and a large watershed area, they show typical characteristics of alluvial rivers owing to sedimentation. Rainfall is concentrated in the rainy season, thus exhibiting a large seasonal variation of precipitation. Due to the topographical characteristics with many mountainous areas, the flow of water quickly concentrates in the rivers. Thus, the coefficient of flow fluctuation (maximum discharge/minimum discharge in certain year) is very high. Owing to these characteristics, the discharge is not stable, thus utilizing various facilities, such as dams, reservoirs, embankments, and weirs for flood control and water resources. In the case of sediments, there is a considerable outflow of weathered soil from granite mountains due to intensive heavy rainfall events; however, after sedimentation

in various facilities and rivers, a relatively small volume of sediment is transported in the main streams as suspended sediment is deposited in various facilities [13,20–22].

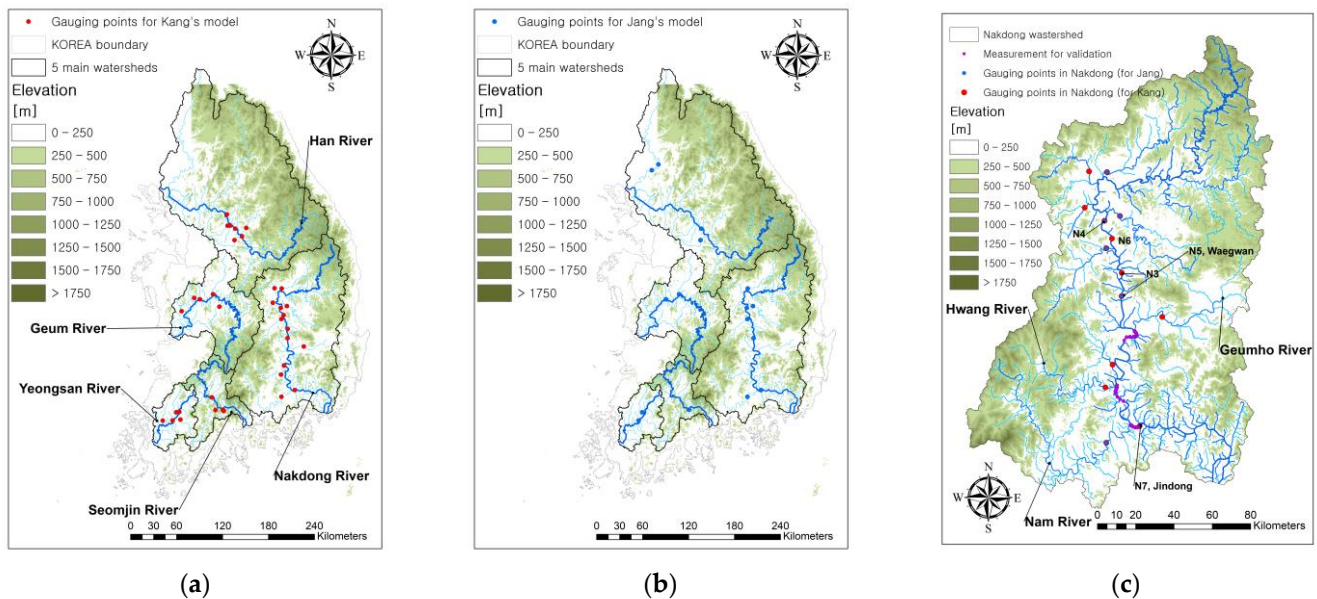


Figure 1. South Korea and (a) gauging points of calibration data for the model by Kang et al. [14], (b) gauging points of calibration data for Jang's model [13], and (c) measurement points of validation data in the Nakdong River.

In South Korea, the following project/surveys were undertaken in the 2000s: a nationwide river basin survey project, a national project of hydrological survey, and a nationwide sediment yield survey at full scale. Based on these surveys, there has been active research related to sediment for various rivers in South Korea. In the case of the model developed by Kang, which is one of the models that needs to be validated in this study, the SD results of 34 gauging points from 2005 to 2015 were used as response variables (Figure 1a). SD is the ratio of the sediment yield divided by the watershed area, which corresponds to the annual specific sediment yield generated and transported at the basic scale. Kang calculated the total sediment load using the modified Einstein's procedure (MEP), and the series expansion modified Einstein's procedure (SEMEP) for the 1962 and 1808 sediment yield measurements, respectively. The SD values were then calculated using daily discharge data for 10 years and the flow duration and sediment rating curve (FD–SRC) method [20]. In the case of total sediment load calculated using MEP and SEMEP, the difference was up to 25%, and the total sediment load estimated by MEP was larger than that obtained by SEMEP (Figure 2a). The details of the difference between the two methodologies used for estimating the total sediment load can be found in Shah–Fairbank [23] and Yang and Julien [24].

Jang [12] used sediment measurement results obtained from 14 gauging points from 2006 to 2012, and 540 total sediment load results calculated using MEP (Figure 1b). All data apart from some of the sediment yield measurement result in Gunnam and Jeokseong located in the Han River, and the measurements of the International Hydrological Program (IHP) were used for the development of the model proposed by Kang [13] (Figure 2b). In terms of data for model validation, a total of 82 suspended sediment concentrations (SSCs) were measured at 30 measuring points of the main stream of the Nakdong River from 2016 to 2019 using LISST-100X, and the total suspended sediment load was calculated based on the measured results (Figure 1c). The models validated in this study were developed based on direct or indirect using of the total sediment load outcomes, but data for validation were based on suspended sediments, which is the typical mode of transportation in South Korean rivers [25]. Specifically, Julien [26] suggested that the primary mode of transport

of sediments is suspended when $u_*/\omega > 5$. As a result of the analysis of the calibration data used for the model development, it was found that u_*/ω for the calibration data for model development is typically high, particularly in the main stream of Nakdong rivers (Figure 3a).

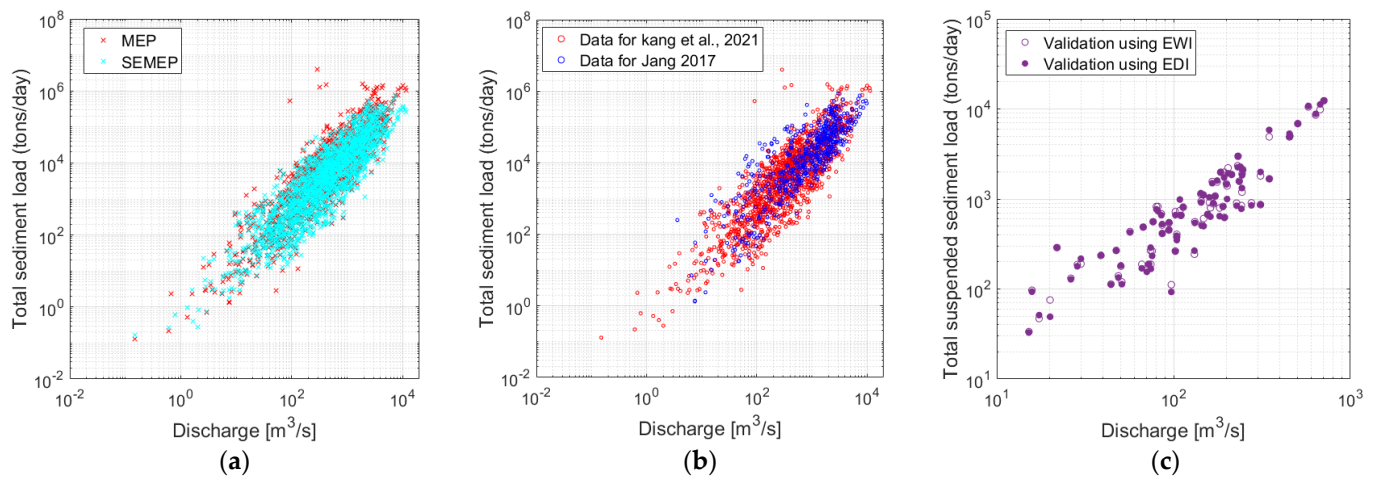


Figure 2. (a) Total sediment load based on the modified Einstein’s procedure (MEP) and the series expansion modified Einstein’s procedure (SEMEP) (calibration data for the model by Kang et al. [14]), (b) total sediment load for Jang’s model [13], and (c) total suspended sediment load of validation data from equal-discharge-increment method (EDI) and the equal-width-increment method (EWI).

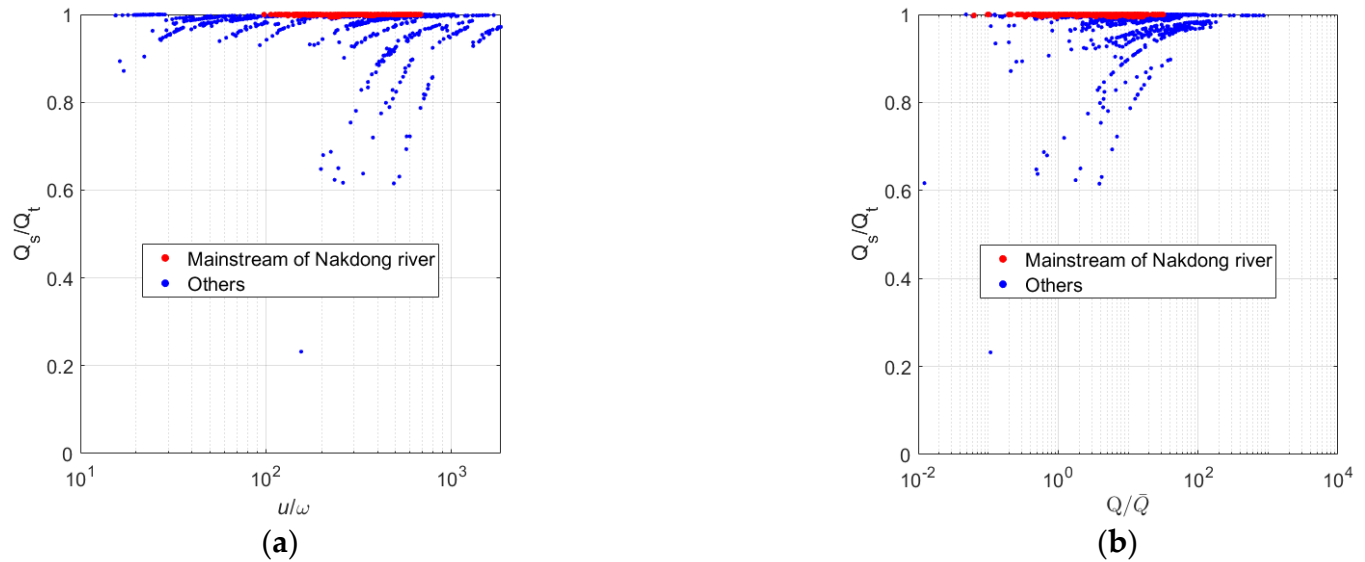


Figure 3. Relationship between the ratio of suspended sediment to total sediment load based on calibration data and (a) u_*/ω and (b) the ratio of daily discharge to mean annual discharge, Q/\bar{Q} .

In Figure 3b, some results for gravel and cobble-bed river (tributary) yielded a low value of Q_s/Q_t in high-water discharge cases (the ratio of daily discharge to mean annual discharge, $Q/\bar{Q} > 1$). However, given that the main stream of Nakdong River, which is a sandy river, yielded a high Q_s/Q_t value (average value was 0.99) regardless of the discharge condition (Q/\bar{Q}), in this study, the 82 total suspended sediment load results were assumed as the total sediment load. Based on the assumption, validation, and analysis were performed on the existing models using data mining technique.

2.2. Existing Models Using Model Tree Technique of Data Mining

The representative data mining techniques include ANNs, regression, or logistic regression analysis, and decision trees [15]. Among these methods, a decision tree method was used as it can derive a model without the need for additional training groups. Given that there are no hidden layers, the user's opinion can be involved in all steps of the pruning process. The model tree technique of decision trees used for the model development is based on a data classification procedure with similar characteristics and separating them into subgroups based on the characteristics of the data. In particular, because the method is the most clearly defined among several model development methods, it allows easy interpretation of the results and is less affected by outliers in comparison. It could be considered as a suitable method for estimating sediment discharge based on the strong correlations between several physical quantities.

Kang [14] developed a model using the model tree technique to identify factors that affect the generation, transport, and deposition of sediments in watersheds and to compare sediment yields between watersheds. The model uses 32 explanatory variables related to watershed characteristics, rainfall data, soil type, land use, and bed materials, which are frequently used in hydrological and hydraulic models. The model (M1) was developed based on the calculated SD values using the MEP and FD-SRC methods and consisted of five submodels defined based on the respective conditions of elevation at the middle relative area of the hypsometric curve (*hyp*), main stream length (Main), percentage of urban (*U*), and percentage of wetland and water (WW) (Table 1).

Table 1. Application conditions and determination coefficients of Model 1 (M1), specific degradation (SD) values based on modified Einstein's procedure (MEP)).

Model	Submodel	Conditions				Equations
		<i>hyp</i> (m)	Main (km)	WW (%)	<i>U</i> (%)	
M1	E1	≤187	-	-	-	$661 \times U^{0.55} \times WW^{-0.34} \times hyp^{-0.44}$
	E2	>187	≤265	≤2.63	≤3.11	$109 \times U^{0.63} \times WW^{-0.26} \times hyp^{-0.23}$
	E3			>2.63	>3.11	$112 \times U^{0.64} \times WW^{-0.26} \times hyp^{-0.23}$
	E4			>2.63	-	$113 \times U^{0.58} \times WW^{-0.27} \times hyp^{-0.23}$
	E5		>265	-	-	$101 \times U^{0.56} \times WW^{-0.18} \times hyp^{-0.23}$

Another model (M2) developed using SEMEP predicts SD values based on elevation at the middle relative area of the hypsometric curve (*hyp*), mean annual precipitation (*P*), watershed form factor (FF), and percentage of urban (*U*) (Table 2).

Table 2. Application conditions and determination coefficients of model 2 (M2, SD values based on SEMEP).

Model	Submodel	Conditions				Equations
		<i>hyp</i> (m)	<i>P</i> (mm)	FF (-)	<i>U</i> (%)	
M2	E1	≤187	-	-	-	$4.9 \times 10^{-4} \times P^{1.89} \times U^{0.38} \times hyp^{-0.39}$
	E2	>187	≤1133	-	-	$3.2 \times 10^{-5} \times P^{2.05} \times U^{0.39} \times hyp^{-0.21}$
	E3		>1133	≤0.33	-	$8.3 \times 10^{-4} \times P^{1.61} \times U^{0.41} \times hyp^{-0.21}$
	E4			>0.33	≤2.61	$8.5 \times 10^{-4} \times P^{1.61} \times U^{0.43} \times hyp^{-0.21}$
	E5				>2.61	$8.7 \times 10^{-4} \times P^{1.61} \times U^{0.43} \times hyp^{-0.21}$

Jang [13] developed a model to derive the optimal sediment transport estimation for identification and accurate prediction of the sediment transport characteristics of rivers in

South Korea. The model used depth (h), bed slope (S), width (W), flow velocity (V), and the median particle diameters (d_{50}) of bed materials were used as explanatory variables (given that the variables with constant values under the same fluid and bed material conditions were excluded, e.g., settling velocity, gravitational acceleration, and geometric mean of sediment particle diameter) among the existing hydraulic parameters related to flow, fluid, and grain properties frequently used in sediment transport formulas related to flow properties in the alluvial rivers. In the case of the proposed model, all the models (six cases in total) are listed in Tables 3–8 based on considerations of the condition of use of all of the five variables and the cases associated with the exclusions of single variables, respectively [15].

Table 3. Application conditions and determination coefficients of model 3 (M3).

Model	Submodel	Conditions				Equations
		V (m/s)	H (m)	d_{50} (mm)	W (m)	
M3	E1	≤ 0.25	-	-	-	$1131.27 \times h^{1.5975} \times S^{0.0981} \times V^{1.5674} \times d_{50}^{-0.438}$
	E2	$0.25 < V \leq 0.51$	-	-	-	$1866.46 \times h^{1.4079} \times S^{0.1057} \times V^{1.0096} \times d_{50}^{-0.1911}$
	E3	$0.51 < V \leq 0.96$	≤ 2.23	-	-	$152.32 \times h^{0.5052} \times S^{0.2124} \times W^{0.9278} \times V^{3.1359} \times d_{50}^{-0.0166}$
	E4		> 2.23	≤ 0.42	-	$54100.57 \times h^{0.6527} \times S^{0.2418} \times V^{1.7366} \times d_{50}^{0.1959}$
	E5			≤ 310.44	≤ 310.44	$22285.69 \times h^{0.6723} \times S^{0.0758} \times V^{2.2532} \times d_{50}^{0.1067}$
	E6			> 0.42	> 310.44	$18760.33 \times h^{1.3872} \times S^{0.1598} \times V^{1.5523} \times d_{50}^{0.1067}$
	E7		-	-	≤ 278.38	$4415.16 \times h^{1.3998} \times S^{0.004} \times W^{-0.1107} \times V^{1.9722} \times d_{50}^{-0.0293}$
	E8	> 0.96	-	≤ 0.66	> 278.38	$20.33 \times h^{0.2358} \times S^{0.004} \times W^{1.3354} \times V^{2.6613} \times d_{50}^{1.2784}$
	E9	-	-	> 0.66	-	$1.21 \times h^{1.1201} \times S^{0.004} \times W^{1.5448} \times V^{1.3151} \times d_{50}^{-0.5361}$

Table 4. Application conditions and determination coefficients of Model 4 (M4).

Model	Submodel	Conditions				Equations
		W (m)	H (m)	S (m/m)	d_{50} (mm)	
M4	E1	$180.73 < W \leq 279.78$	≤ 0.81	≤ 0.000184	-	$74083248.61 \times h^{2.0003} \times S^{1.1554} \times W^{-0.2537} \times d_{50}^{0.145}$
	E2			$0.000184 < S \leq 0.000292$	-	$23004490.64 \times h^{2.0003} \times S^{0.952} \times W^{-0.2537} \times d_{50}^{0.145}$
	E3			> 0.000292	-	$1256.39 \times h^{1.6431} \times S^{-0.0773} \times W^{-0.1672} \times d_{50}^{0.145}$
	E4	≤ 180.73	$0.81 < h \leq 1.34$	-	-	$1669.37 \times h^{-0.4311} \times S^{0.1731} \times W^{0.2355} \times d_{50}^{0.145}$
	E5	-		-	$2895.46 \times h^{0.8474} \times S^{0.1656} \times W^{0.2036} \times d_{50}^{0.145}$	
	E6	$180.73 < W \leq 279.78$	$1.34 < h \leq 3.92$	-	-	$382.26 \times h^{3.9308} \times S^{0.0918} \times W^{0.1515} \times d_{50}^{3.0763}$
	E7		$3.92 < h \leq 4.61$	≤ 0.000263	-	$10.35 \times h^{5.4007} \times S^{0.0918} \times W^{-0.3817} \times d_{50}^{0.2251}$
	E8		> 4.61	-	-	$92.95 \times h^{3.9622} \times S^{0.0918} \times W^{-0.1703} \times d_{50}^{0.3587}$
	E9	> 0.000263		-	$4199600.67 \times h^{1.9452} \times S^{1.038} \times W^{0.012} \times d_{50}^{-0.5381}$	
	E10	> 279.78	≤ 2.8	-	≤ 0.66	$2.37 \times 10^{-25} \times h^{0.461} \times S^{1.1506} \times W^{12.8474} \times d_{50}^{0.1335}$
	E11		> 2.8	-		$0.03 \times h^{1.453} \times S^{0.133} \times W^{2.1839} \times d_{50}^{0.1509}$
	E12		≤ 5.97	-	> 0.66	$16.24 \times h^{0.2934} \times S^{0.0495} \times W^{1.4427} \times d_{50}^{0.1509}$
	E13		> 5.97	-		$2.65 \times 10^{-05} \times h^{2.2029} \times S^{0.4417} \times W^{3.6582} \times d_{50}^{0.1509}$

Table 5. Application conditions and determination coefficients of Model 5 (M5).

Model	Submodel	Conditions				Equations
		V (m/s)	W (m)	d ₅₀ (mm)	S (m/m)	
M5	E1	≤ 0.51	-	-	-	$2536.40 \times S^{0.0079} \times W^{0.0091} \times V^{1.78} \times d_{50}^{-0.2791}$
	E2	$0.51 < V \leq 0.96$	≤ 236.51	≤ 0.76	-	$165777.2 \times S^{0.2737} \times W^{-0.1202} \times V^{2.4127} \times d_{50}^{-0.8117}$
	E3		> 0.76	-	$2692.43 \times S^{0.267} \times W^{0.5352} \times V^{3.5318} \times d_{50}^{-0.6276}$	
	E4		≤ 0.42	-	$391.86 \times S^{0.0079} \times W^{0.5971} \times V^{1.5285} \times d_{50}^{-0.3072}$	
	E5		> 0.42	-	$8.49 \times 10^{-04} \times S^{0.0079} \times W^{2.938} \times V^{0.5975} \times d_{50}^{-0.3072}$	
	E6		≤ 0.79	-	$212266.96 \times S^{0.3302} \times W^{0.0219} \times V^{2.5851} \times d_{50}^{-0.0746}$	
	E7	$1.2 < V \leq 1.95$	≤ 278.38	-	-	$45922.02 \times S^{0.121} \times W^{0.0429} \times V^{3.018} \times d_{50}^{0.3928}$
	E8			> 0.79	-	$9024.51 \times S^{0.4256} \times W^{0.5475} \times V^{2.2057} \times d_{50}^{0.3928}$
	E9			-	≤ 0.000358	$68851 \times S^{0.2539} \times W^{-0.247} \times V^{3.1039} \times d_{50}^{-0.3761}$
	E10	> 1.95	-	> 0.000358	$1690.37 \times S^{-0.0373} \times W^{0.2} \times V^{2.8551} \times d_{50}^{-0.3761}$	
	E11	> 0.96	$278.38 < W \leq 311.69$	≤ 0.66	-	$32.78 \times S^{0.0411} \times W^{1.3535} \times V^{2.657} \times d_{50}^{1.2195}$
	E12		> 311.69	> 0.66	-	$968.65 \times S^{0.0735} \times W^{0.7367} \times V^{0.785} \times d_{50}^{-0.0452}$
	E13		> 0.66	-	$1351 \times S^{0.3632} \times W^{1.1942} \times V^{0.4309} \times d_{50}^{-0.0452}$	

Table 6. Application conditions and determination coefficients of Model 6 (M6).

Model	Submodel	Conditions			Equations
		V (m/s)	h (m)	d ₅₀ (mm)	
M6	E1	≤ 0.96	-	-	$11605.1 \times h^{1.6891} \times S^{0.2044} \times V^{2.0979} \times d_{50}^{-0.0166}$
	E2	> 0.96	≤ 2.74	-	$3298.09 \times h^{1.2913} \times S^{0.004} \times V^{2.1405} \times d_{50}^{-1.1727}$
	E3		> 2.74	≤ 0.72	$137902.2 \times h^{0.286} \times S^{0.1815} \times V^{2.8344} \times d_{50}^{0.919}$
	E4			> 0.72	$6624.3 \times h^{1.4408} \times S^{0.004} \times V^{1.2056} \times d_{50}^{-0.7407}$

Table 7. Application conditions and determination coefficients of Model 7 (M7).

Model	Submodel	Conditions				Equations
		V (m/s)	H (m)	d ₅₀ (mm)	W (m)	
M7	E1	≤ 0.51	-	-	-	$1535.48 \times h^{1.4971} \times V^{2.0023} \times d_{50}^{-0.3761}$
	E2	0.51 < V ≤ 0.96	≤ 2.23	-	-	$36.83 \times h^{0.4807} \times W^{0.9002} \times V^{3.5055} \times d_{50}^{-0.0133}$
	E3		> 2.23	≤ 0.42	-	$435.37 \times h^{0.6409} \times W^{0.4972} \times V^{1.7106} \times d_{50}^{0.2207}$
	E4			> 0.42	≤ 310.44	$12383.25 \times h^{0.6574} \times V^{2.3264} \times d_{50}^{0.1225}$
	E5	> 0.96			> 310.44	$5570.83 \times h^{1.3264} \times V^{1.6255} \times d_{50}^{0.125}$
	E6				≤ 278.38	$4272.69 \times h^{1.4001} \times W^{-0.1107} \times V^{1.9732} \times d_{50}^{-0.0269}$
	E7				> 278.38	$19.67 \times h^{0.2361} \times W^{1.3354} \times V^{2.6623} \times d_{50}^{1.2808}$
	E8	> 0.96			> 278.38	$1.17 \times h^{1.1204} \times W^{1.5448} \times V^{1.3161} \times d_{50}^{-0.5337}$

Table 8. Application conditions and determination coefficients of Model 8 (M8).

Model	Submodel	Conditions		Equations
		V (m/s)	W (m)	
M8	E1	≤ 0.96	-	$10883.47 \times h^{1.6921} \times S^{0.1987} \times W^{0.0036} \times V^{2.0957}$
	E2	> 0.96	≤ 278.38	$4208.66 \times h^{1.4047} \times S^{0.9002} \times W^{-0.1067} \times V^{1.9567}$
	E3		> 278.38	$0.12 \times h^{0.7149} \times S^{0.0951} \times W^{2.0888} \times V^{1.5412}$

In the model development process based on the use of the model tree, the minimum number of data points in the subsections of the model tree and the standard deviation reduction rate calculated after grouping for each variable must be specified. Witten and Frank [27] proposed that the subsections of the model tree can be established without being restricted by the minimum number of data and indicated that the minimum number of data for multiple regression analysis is four. Considering this and the utility of the formula, if a standard deviation $\geq 5\%$ was not obtained after classification, the model development was terminated. Examination of all the submodels, which vary with respect to the conditions of the model (M1 and M2) proposed by Kang, the index values that corresponded to the explanatory variables exhibited consistency, and the correlations that considered the physical meaning of the applicable explanatory variables and SD values were identified. For example, with the developed model, it was argued that an increase in the mean annual precipitation will increase the sediment yield generated and transported. When the percentage of WW in the watershed increases, there will be increased chances of deposition of the transported sediment, thus properly reflecting the characteristics of the decrease in SD values.

2.3. Validation Data Measurement Method

In this study, an acoustic Doppler current profiler (ADCP, River Surveyor M9, Sontek, San Diego, CA, USA) and LISST were used for the hydraulic survey and SSC measurements of the target measuring points. The instruments were attached to the side of the boat, and the boat moved in a direction perpendicular to the flow at a speed of ≤ 1 m/s to perform the measurement (Figure 4).



(a)

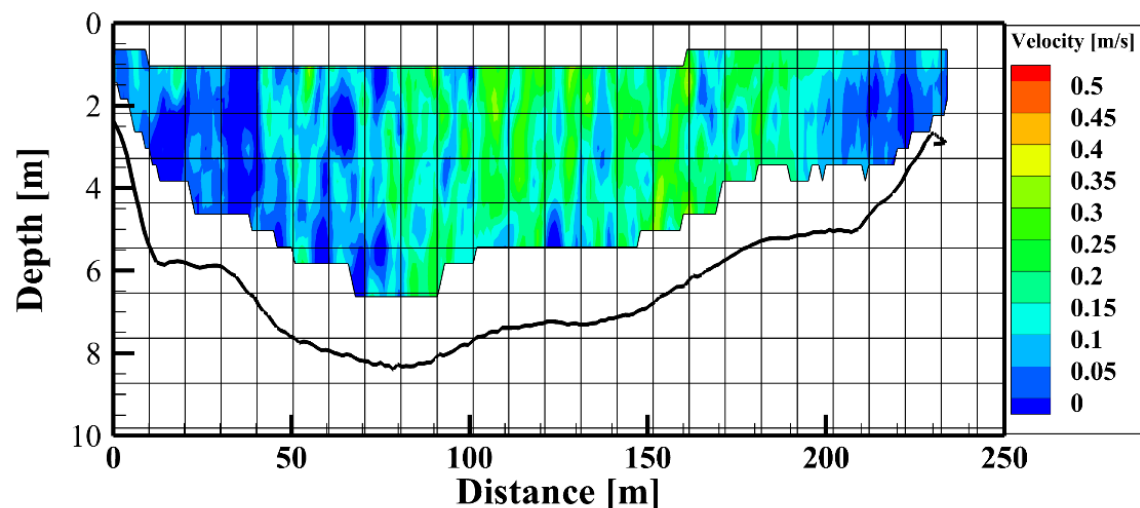


(b)

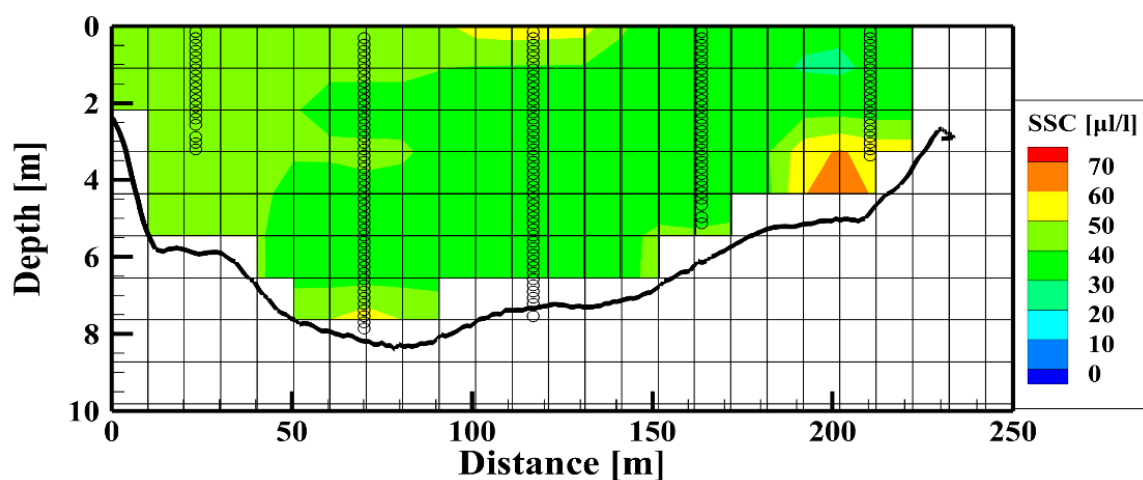
Figure 4. (a) Field measurements and (b) measurement equipment (LISST-100X was attached to the boat).

Regarding the measurement items, ADCP was used to measure the water depth and three-dimensional water velocity (Figure 5a). Continuous measurements of vertical SSC were conducted at a rate of one per second by fixing the LISST to a wire reel at the measuring points (3–5 points) designated in the river width direction for each section (Figure 5b). The location information of the corresponding measuring points was acquired using the GPS of

ADCP. ADCP has the advantage of efficient and quick measurements of the water depth and velocity compared with the existing point-based velocity measurement method, but there are unmeasured areas near the water surface, near the riverbed, and at the edges of the river. The values for these areas are estimated using the measured data through indirect extrapolation. In the LISST-100X, which was used to obtain the SSC, volume concentration was measured according to the particle size based on light scattering, and the beam attenuation by the suspended particles was calculated through the transmittance of the unscattered light. In addition, the measurable range of the suspended sediment particle size ranged from 1.25 to 250 μm , and measurements of the SSC, particle size distribution (PSD), beam attenuation, water depth, and water temperature can be performed. To estimate the total suspended sediment load of the river section from the measured results, the following two methods were used: (1) the equal-discharge-increment method (EDI) and (2) the equal-width-increment method (EWI) (Figure 2c). The EDI method requires the SSC at the center of the EDI section, and the EWI method requires the SSC at each measuring line separated by equal intervals. Grids (10×1 m) were generated that subdivided the section in the direction of river width and water depth, and the measured water velocity and SSC were interpolated to obtain the distribution (Figure 5a,b).

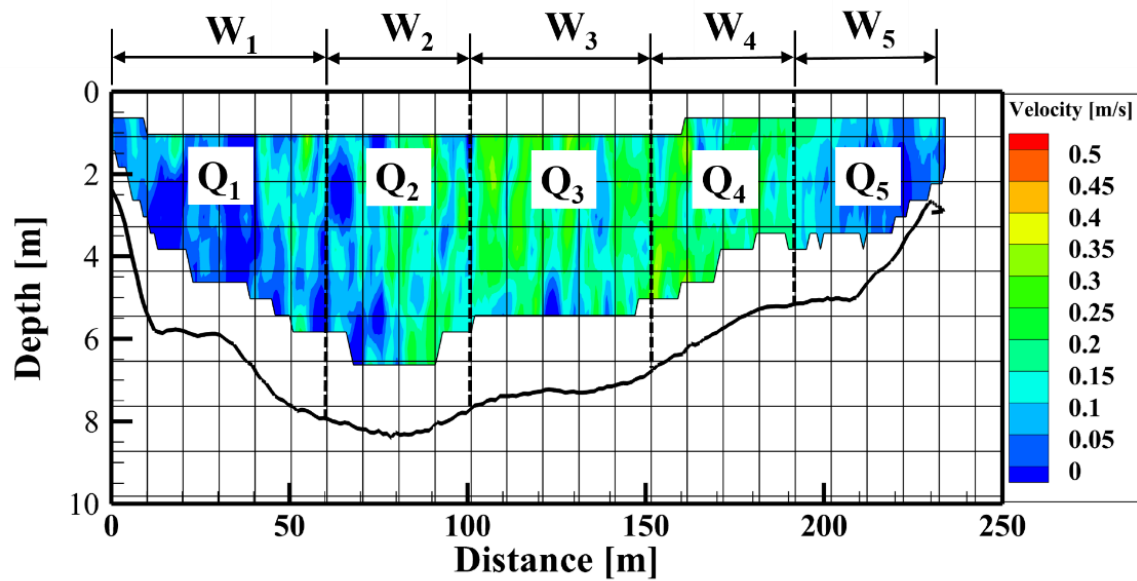


(a)



(b)

Figure 5. Cont.



(c)

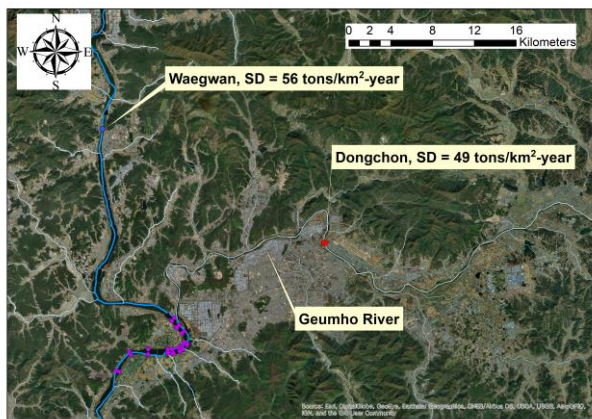
Figure 5. Estimating procedures for total suspended load from measurement using LISST-100X. (a) Measured water velocity by ADCP, (b) two-dimensional (2D) distribution results from the measured SSC by LISST-100X, and (c) 2D distribution outcome based on the EDI method used for the estimation of total suspended sediment load.

The average SSC of the EDI section (Figure 5c) and the EWI section were calculated from the generated two-dimensional (2D) section to calculate the total suspended sediment load for each section. In the EDI method, the accumulated discharge in each section is the same, but the width is different. In the case of the EDI method, it is possible to analyze the SSCs for each section, and consideration of outliers is possible. In the applicable measuring points, the EDI method provides more accurate results compared with the EWI method. Therefore, in this study, the total suspended sediment load estimated based on the EDI method was used [28].

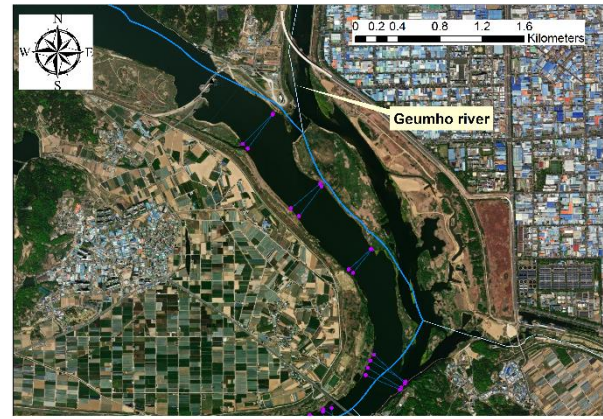
3. Results

3.1. Validation of Existing Models M1 and M2

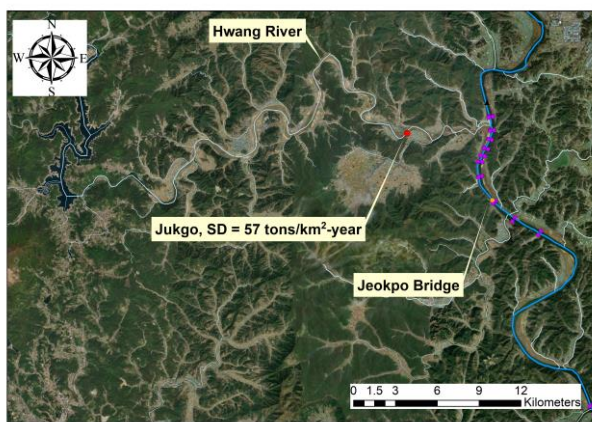
Given that the model proposed by Kang [14] used various watershed characteristics as explanatory variables, the latter must be obtained based on GIS analysis, and the target watersheds must first be determined through watershed delineation. Tributaries of various sizes flow in the main stream of the Nakdong River, and in the case of the total sediment load used for validation, measurements were acquired at upstream and downstream of the confluence of the Geumho River (watershed area = 2053 km²), Hwang River (watershed area = 1340 km²), and Nam River (watershed area = 3644 km²) in the continuous main stream of the Nakdong River (Figure 1c). The SD values of gauging points located in the Geumho (Dongchon), Hwang (Jukgo), and Nam Rivers (Jeongam) used for the development of M1 and M2 models were 49.7, 57.2, and 33.7 tons/km²·year, respectively (Figure 6).



(a)



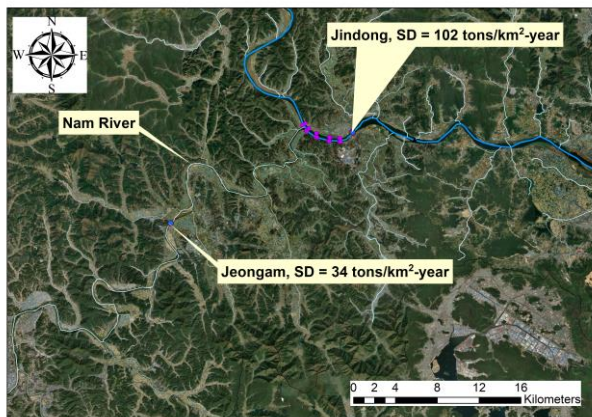
(b)



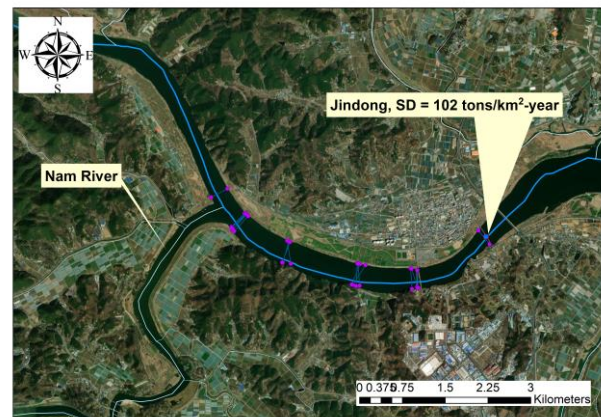
(c)



(d)



(e)



(f)

Figure 6. Measurement locations of validation data near the (a,b) Geumho, (c,d) Hwang, and (e,f) Nam River confluences.

Figure 7 shows the comparative results among different measurement points, which indicate big differences between upstream and downstream after tributary merging of Hwang River and Nam River.

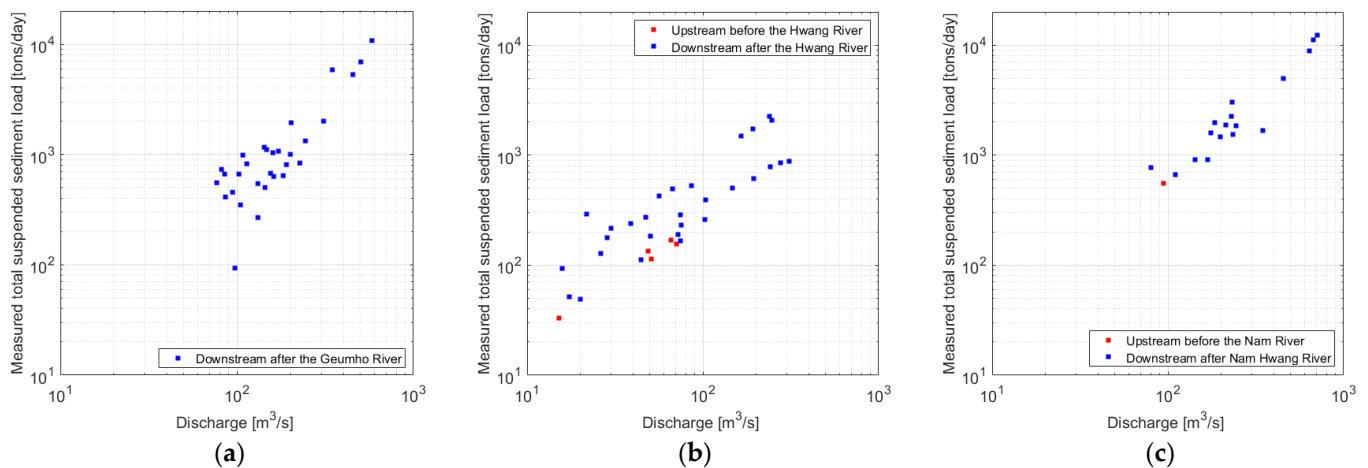


Figure 7. Measured total suspended sediment load at (a) Geumho, (b) Hwang, and (c) Nam River.

In case of the Geumho River, the confluence was very complex (Figure 6b). Given that most of the inflow sediments were deposited downstream of the confluence, only the watershed after the confluence (W1) was considered. Taking into account the measurement results, the watersheds set for validation of M1 and M2 include those at upstream and downstream locations those of the Hwang River confluence (W2 and W3) and those of the Nam River merging (W4 and W5). Table 9 lists the explanatory variables and predicted SD values for the five watersheds for model validation.

Table 9. Validation result of models M1 and M2.

Watershed	A	hyp	Main	WW	U	P	FF	M1 (SD-MEP)	M2 (SD-SEMEP)
	(km ²)	(m)	(km)	(%)	(%)	(mm)	(-)	(tons/km ² ·year)	(tons/km ² ·year)
W1 (after Geumho River)	13,926.19	307.7	320.1	2.38	4.59	1142.24	0.44	53.93	42.67
W2 (before Hwang River)	15,070.54	304.8	363.2	2.82	4.68	1198.33	0.40	53.38	44.54
W3 (after Hwang River)	16,405.51	301.5	366.1	3.12	4.87	1216.16	0.36	53.74	46.45
W4 (before Nam River)	16,973.5	295.7	389.8	3.32	5.12	1257.5	0.36	54.94	50.24
W5 (after Nam River)	18,064.13	292.5	395.6	3.56	5.32	1283.65	0.28	55.61	50.21

Considering the range of SD values of 44, 6, 21, 56, and 102 tons/km²·year at the five gauging points located in the main stream of the Nakdong River (N4, N6, N3, N5, and N7 in Figure 1c) used for calibration for model development, the predicted SD values were judged to be appropriate [14]. However, considering that the SD values of Waegwan (N7) and Jindong (N12) adjacent to the measurement points were 56 and 102 tons/km²·year, the results predicted by M1 and M2 did not show a considerable difference between Geum, Hwang, and Nam River, respectively. In addition, the models could not predict the different total sediment load between upstream and downstream of confluences. This is because the two models (M1 and M2) were developed to compare sediment yields between different watersheds. To be specific, many characteristics for watershed including tributaries did not change much when the change in the watershed area was small due to the tributaries or location of gauging points.

According to previous studies, Yang and Julien [24] reported that SEMEP was capable of yielding reasonable and accurate estimations of the total sediment load. Kang [14] also argued that M2 developed through the total sediment load predicted using SEMEP presented superior results compared with M1. However, the results of M1 developed using the MEP results are considered to show better predictability. The reason for the

contradictory results from the previous studies is thought to be the spatial lumping problem caused by the fact that the sediment yield at the gauging points at which the measurements were acquired for validation was larger compared with that generated and transported by other main river streams and by the temporal lumping problem between validation and calibration data.

3.2. Validation of Existing Models M3 through M8

For the validation of the six models proposed by Jang [13], the measured results obtained by ADCP were used in the case of the depth (h), the river width (W), and the average velocity (V). The bed slope (S) was calculated by dividing the difference in the thalweg elevation at each point (measured using ADCP) by the river distance between successive points. These distances were calculated with the use of the Korea Reach File (KRF version 3.0 (Sejong, Korea)) provided by the Ministry of Environment for delineating stream network. For the median particle diameter of bed materials (d_{50}), basic reference data values for river management were used. Model validation was performed using 58 measurements, including the median particle diameter of bed materials among a total of 82 measurements. Among the 58 measured data, the Geumho (25 data) and Nam Rivers (14 data) yielded no data for median particle diameters of the bed materials before the confluence. The comparison was possible only in the case of the Hwang River (two data at upstream of confluence and 17 at downstream). Figure 8 shows the validation results of the models proposed by Jang [13] using the explanatory variables.

In the case of the models, relatively accurate prediction results were presented comparison with the measured results, except for M4. In particular, the measured results obtained from the Geumho and Nam Rivers have accurately predicted the characteristics, such as the value of the total sediment load, which was large in the vicinity of the Hwang River confluence, and the value of the total sediment load, which was smaller in regions before (rather than after) the Hwang River confluence. However, the predicted results of some models (M4 and M5) showed a low level of reliability (root-mean-square error (RMSE) > 3000 ton/km²·year). This is thought to be caused by the use of overly subdivided formulas compared with other models, thus leading to low predictability. In particular, the predicted total sediment load value of M4 was highly inaccurate even when compared with the results of M5. This is likely caused by the problem of determining submodels used for prediction (Figure 9).

In the case of explanatory variables for 58 hydraulic factors used for model validation, the conditions are relatively similar owing to the location in the main stream of the Nakdong River. However, unlike other models, which used the water velocity with small coefficient of variations as the first classification criteria, M4 determined submodels based on the river width and water depth with relatively large variation. This led to a lack of consistency in the index values that corresponded to the explanatory variables of the subformulas developed based on inadequate explanatory variables, and thus resulted in inaccurate prediction results (Table 10).

Table 10. Simple statistical analysis outcomes for explanatory variables for M3–M8.

Statistical Factor	Water Depth (h)	Bed Slope (S)	Width (W)	Velocity (V)	Median Diameter (d_{50})
average	5.50	0.001002	333.36	0.10	0.96
variance	1.26	0.000001	4708.82	0.01	1.04
standarddeviation	1.12	0.001208	68.62	0.08	1.02
coefficient of variation	4.90	0.83	4.86	1.27	0.94

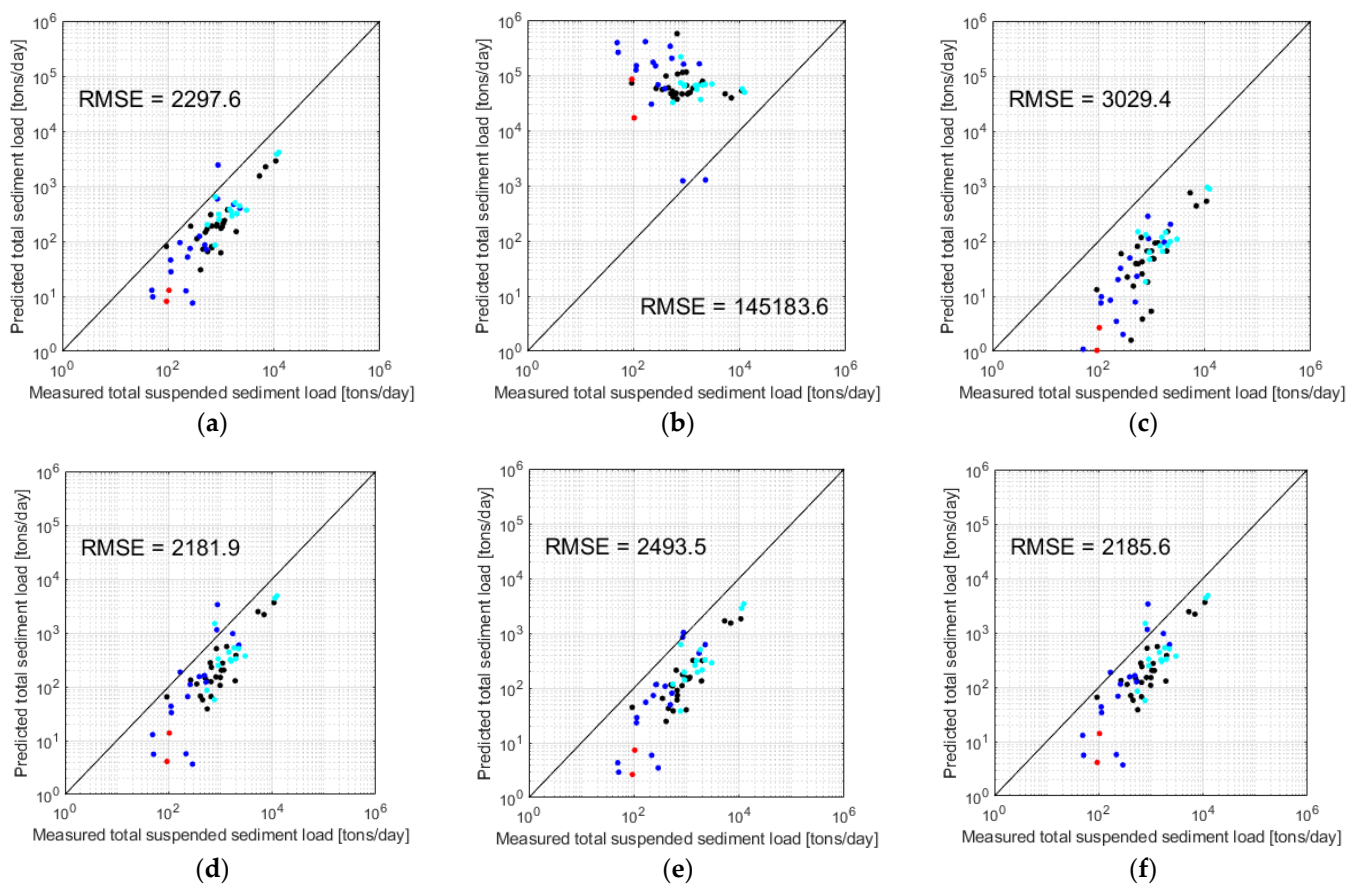


Figure 8. Predicted total sediment loads from models (a) M3, (b) M4, (c) M5, (d) M6, (e) M7, and (f) M8. The black circles represent total sediment load for Geumho River, red circles represent total sediment load for down stream of Hwang River confluence, blue circles represent total sediment load for upstream of Hwang River confluence, and the cyan circles represent total sediment load for Nam River. (In this figure, the measured total suspended sediment load is assumed to be the measured total sediment load).

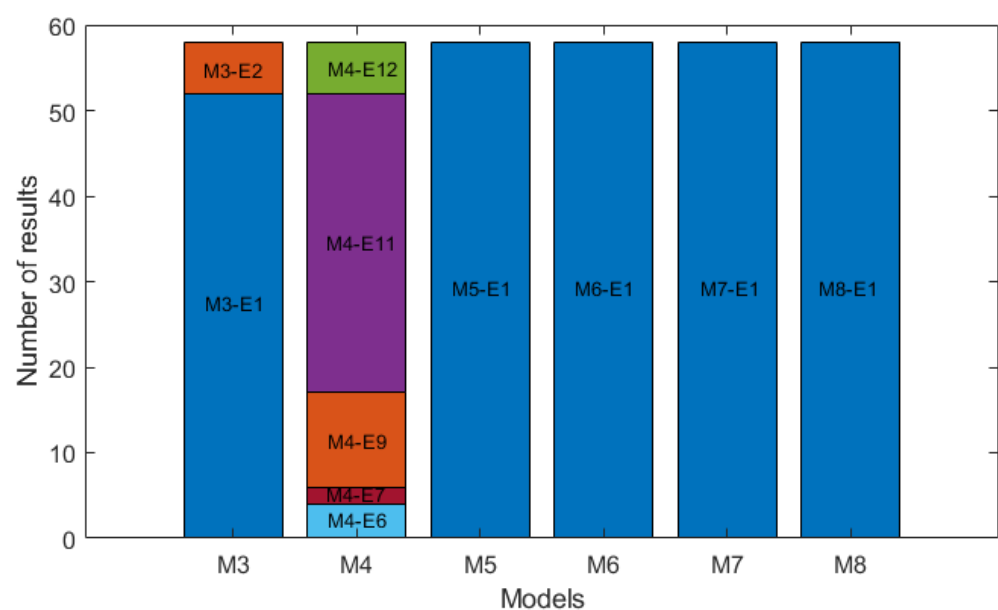


Figure 9. Selected submodels (equations) for validation (M3–M8).

4. Discussion

In case of models M1–M2 and M3–M8, direct comparison is not possible because different results (SD vs. total sediment load) are presented. To obtain SD values for comparison between the two sets of models, the total sediment load values predicted based on the models M3, M6, M7, and M8 and daily discharge data from 2010 to 2020 in the gauging points of Jeokpo Bridge and Jindong (Figure 6b,c) located downstream of the Hwang River confluence and Nam River confluence, respectively, were used to present flow duration curves (Figures 10 and 11). Additionally, the sediment rating curves presented using the predicted total sediment load results are shown in Figures 9 and 10. M4 and M5 were excluded from comparisons between the models because their predictability was low.

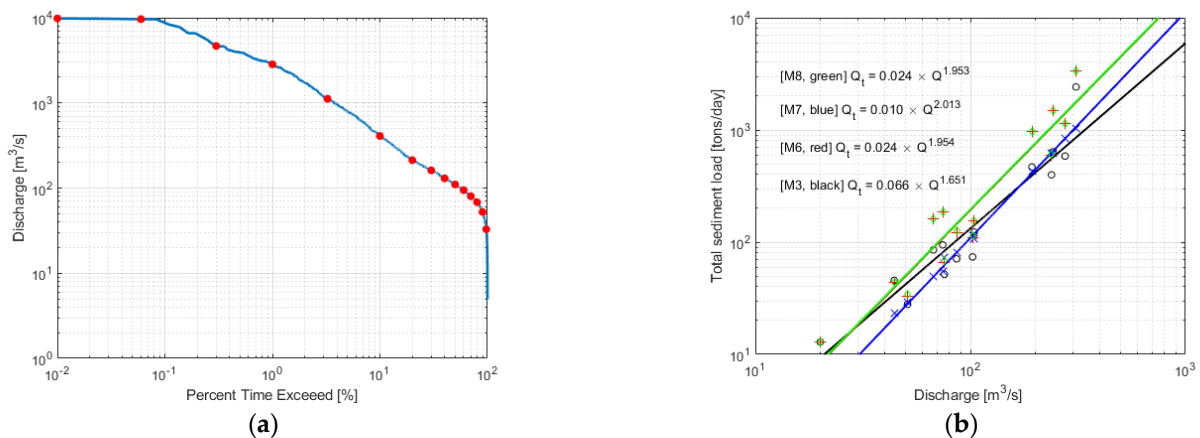


Figure 10. (a) Flow duration and (b) sediment rating curves from the predicted total sediment load from models M3, M6, M7, and M8 at the Jeokpo Bridge.

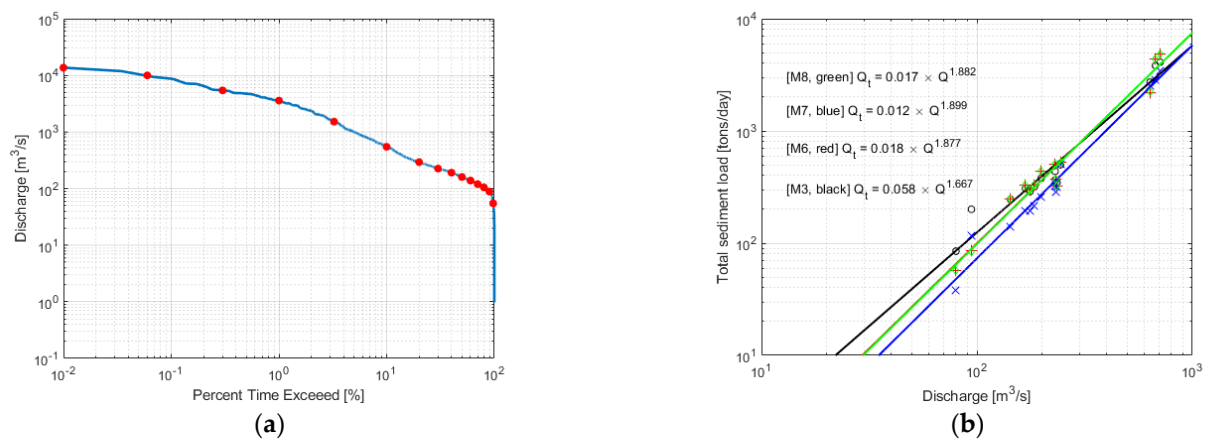


Figure 11. (a) Flow duration and (b) sediment rating curves from the predicted total sediment load from models M3, M6, M7, and M8 at Jindong.

Table 11 shows the SD values estimated using the predicted results of M3, M6, M7, and M8, and the FD–SRC method.

Table 11. SD at Jindong and Jeokpo from the predicted total sediment load.

Site	Model	SD	Model	SD	Model	SD	Model	SD
Jeokpo	M3	36.31	M6	61.41	M7	47.95	M8	62.01
Jindong	M3	31.44	M6	118.53	M7	79.91	M8	120.58

Considering Jindong ($SD = 102 \text{ ton/km}^2 \cdot \text{year}$) used as calibration data for the model development and Waegwan ($SD = 56 \text{ ton/km}^2 \cdot \text{year}$) located in the main stream of the Nakdong River near the validation section, the predictabilities of M6 and M8 yielded the best performances. This contradicts the expectation that models with more submodels will produce superior prediction results, thus showing that in fact, simplified models (with fewer submodels) yielded superior prediction results. The reason M6 and M8 yielded the best predictive results is attributable to the fact that the submodels directly used for validation (M6–E1 and M8–E1) were developed using 227 calibration data (Table 12).

Table 12. Number of calibration data used to develop the models and submodels.

Equation	M1	M2	M3	M4	M5	M6	M7	M8
E 1	7	7	31	5	60	227	60	227
E 2	9	6	29	9	30	112	66	127
E 3	8	8	66	28	56	87	33	186
E 4	2	5	33	21	26	114	54	
E 5	8	8	54	27	55		14	
E 6			14	59	12		127	
E 7			127	27	8		87	
E 8			87	24	71		99	
E 9			99	104	15			
E 10				27	21			
E 11				102	87			
E 12				84	34			
E 13				24	65			
Sum	34	34	540	540	540	540	540	540

In the cases of M1 and M2, the models were developed using a small number of samples (34); however, the response variable was developed using the SD values calculated considering total sediment load and flow conditions, and the SD values (ranges for calibration data: $10\text{--}1000 \text{ tons/km}^2 \cdot \text{year}$) did not exhibit a large variation unlike the total sediment load results (range for calibration data: $1\text{--}1,000,000 \text{ tons/day}$). Thus, these models yielded consistent results.

In addition, in this study, in the case of models M3–8 developed based on data from 2006 to 2012, considering that the data directly measured for validation were obtained from 2016 to 2019, a temporal lumping problem, which is one of the drawbacks of an empirical model, was expected. However, in the case of models M6 and M8, contrary to the expectation, a superior prediction performance was obtained compared to models M1 and M2, which were developed based on the data obtained from 2005 to 2015. This is considered to be attributed to a combination of reasons. First, as calibration data, M1 and M2 used most of the data from other rivers, unlike the models M3–M8, which mostly used the sediment measurement outcomes for Nakdong River. Second, the data measured for validation are shown to be affected by the spatial lumping problem with high-sediment yield value compared with other data used for model development located in the main Nakdong River. In addition, in this study, there were no measurements for Jindong and Waegwan from 2016 onward. Thus, the analysis was not possible, but it is expected that there will be an effect of the temporal lumping problem owing to changes in the river environment. Considering this aspect, data-based empirical models for sediment yield predictions must be updated periodically in consideration of the spatial and lumping problem.

Although the presence of spatial and temporal lumping problems was confirmed in the case of the data mining technique based on the results of this study, the existing models

showed sufficiently reliable prediction results except for some cases (M4). These results showed that unless there was a drastic change in the river environment, the sediment yield did not change significantly. Therefore, in the development of a sediment yield estimation model using data mining techniques, acquiring a sufficient amount of high-quality data of sediment yield measurement is instrumental.

5. Conclusions

In this study, validation and evaluation of the measured data were performed for the sediment yield estimation models developed using the classification and prediction techniques of data mining. The latter was capable of generating accurate classification and prediction results for rivers in South Korea, as follows: (1) models developed using the general watershed characteristic factors that affect the sediment and 34 SD values calculated based on the FD–SRC method (M1 using MEP for total sediment load, and M2 using SEMEP for total sediment load) and (2) models (M3–M8) developed using hydraulic explanatory variables related to flow and 540 total sediment load results estimated through MEP. As a result of the analysis, that most of the sediments were confirmed to be transported in the form of suspended sediments ($Q_s/Q_t > 99\%$) in the mainstream of Nakdong River. For model validation, flow characteristics were measured using ADCP, and the SSCs of 82 point were measured using laser diffraction equipment at the main stream of the Nakdong River at the site at which the tributaries of Geumho, Hwang, and Nam Rivers inflowed in the main stream. Based on the analyzed results, the total suspended sediment load estimated using the EDI method was assumed to be the total sediment load, and the model was validated accordingly.

Models M1 and M2 provided relatively reliable prediction results, considering the calibration SD values of other points located in the main stream of the Nakdong River. Although the models were developed using a small number of results (34), a sufficient number of total sediment loads (1962 and 1808) was considered at various flow conditions to calculate the SD values. Additionally, the consistent predicted results were obtained from the consistent index values, which corresponded to the explanatory variables; accordingly, the two models yielded reasonable and consistent prediction results. However, the accurate prediction of the change in sedimentation caused by the inflow of relatively small tributaries in the continuous main stream with a large watershed area was not possible due to the purpose of model development. In addition, unlike the results of previous studies, model M2 was developed based on the total sediment load calculated according to the SEMEP and yielded superior prediction results, and model M1 based on the total sediment load calculated according to the MEP, which yielded better prediction results owing to spatial and temporal lumping problems. In the cases of models M3–M8, which were developed to predict the total sediment transport of rivers in South Korea, a good prediction performance was demonstrated for changes in the sediment yield owing to the inflow of tributaries compared with the models, which used watershed characteristics as explanatory variables. However, there was a significant difference in the prediction results depending on the model structure. Specifically, validation was performed with the results measured in the main stream of the Nakdong River with similar characteristics, but the determined submodels of M4 were based on explanatory variables with relatively large coefficients of variation and thus exhibited a low prediction performance ($RMSE = 145,134 \text{ tons/km}^2\cdot\text{year}$).

For comparison between models with other explanatory variables and response variables, SD values were calculated by applying the FD–SRC method based on the prediction results of the models developed by Jang and the 10-year daily discharge data in the vicinity of the gauging points. In the result, models (M6 and M8) with simplified submodel classifications based on low-coefficient variations yielded highly accurate prediction results, which reflected the differences due to the inflow of tributaries. The reason based on which the models showed excellent prediction results is that M3–M8 used a large number of sediment load measurements in Nakdong River as calibration data. It means that the number of

calibration data used for developing the sediment yield estimation model using the data mining technique is thought to have the most significant impact on the predicted results.

The results presented herein confirm that spatial and temporal lumping problems were still present when the sediment yield estimation model was developed using data mining techniques, as was the case of other empirical models. However, given that the technique enables more accurate classification and prediction, real-world applications for river planning and management will be possible if periodic updates are made according to the latest measurement of sediment yield and sediment results of additional gauging points considering various conditions. In summary, if in-depth consideration is given to the background of the model, development purpose, changes in the river environment, and the range of data used, the sediment yield and transport estimation using data mining techniques can serve as a novel method of sediment yield estimation. In addition, it is expected that the low reliability, uncertainty, or limitations of existing methods in use can be improved significantly by acquiring a large number of reliable, high-quality sediment yield of measurement data.

Author Contributions: Conceptualization, W.K.; formal analysis, W.K., E.-k.J. and K.L.; writing—original draft, W.K., E.-k.J. and K.L.; writing—review and editing, W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure, and Transport (Grant 21AWMP-B121100-06). This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. 2021R1C1C101040411).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research was supported by ‘National Research Council of Science & Technology (NST)’—‘Korea Institute of Civil Engineering and Building Technology’ Postdoctoral Fellowship Program for Young Scientists at KICT in South Korea. This work utilized an experimental infrastructures at River Experiment Center of Korea Institute of Civil Engineering and Building Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Julien, P.Y. *River Mechanics*, 2nd ed.; Cambridge University Press: Colorado, CO, USA, 2018.
2. Bizzi, S.; Dinh, Q.; Bernardi, D.; Denaro, S.; Schippa, L.; Soncini-Sessa, R. On the control of riverbed incision induced by run-of-river power plant. *Water Res. Res.* **2015**, *51*, 5023–5040. [[CrossRef](#)]
3. Zhang, J.L.; Shang, Y.Z.; Liu, J.Y.; Lu, J.; Wei, S.T.; Wan, Z.W.; Luo, Q.S.; Chen, C.X.; Tong, L.; Wang, Q.; et al. Optimisation of reservoir operation mode to improve sediment transport capacity of silt-laden rivers. *J. Hydrol.* **2021**, *594*, 125951. [[CrossRef](#)]
4. Amasi, A.; Wynants, M.; Blake, W.; Mtei, K. Drivers, Impacts and Mitigation of Increased Sedimentation in the Hydropower Reservoirs of East Africa. *Land* **2021**, *10*, 638. [[CrossRef](#)]
5. Areu-Rangel, O.S.; Bonasia, R.; Di Traglia, F.; Del Soldato, M.; Casagli, N. Flood Susceptibility and Sediment Transport Analysis of Stromboli Island after the 3 July 2019 Paroxysmal Explosion. *Sustainability* **2020**, *12*, 3268. [[CrossRef](#)]
6. Bhatti, M.T.; Ashraf, M.; Anwar, A.A. Soil Erosion and Sediment Load Management Strategies for Sustainable Irrigation in Arid Regions. *Sustainability* **2021**, *13*, 3547. [[CrossRef](#)]
7. Wheater, H.S.; Jakeman, A.J.; Beven, K.J. Progress and directions in rainfall-runoff modeling. In *Modeling Change in Environmental Systems*; Jakeman, A.K., Beck, M.B., McAleer, M.J., Eds.; John Wiley & Sons: Chichester, UK, 1993; pp. 101–132.
8. Kulsoontornrat, J.; Ongsomwang, S. Suitable Land-Use and Land-Cover Allocation Scenarios to Minimize Sediment and Nutrient Loads into Kwan Phayao, Upper Ing Watershed, Thailand. *Appl. Sci.* **2021**, *11*, 10430. [[CrossRef](#)]
9. Zhang, Q.G.; Huang, R.Q.; Liu, Y.X.; Su, X.P.; Li, G.Q.; Nie, W.A. Physically-Based Geometry Model for Transport Distance Estimation of Rainfall-Eroded Soil Sediment. *Appl. Sci.* **2016**, *6*, 34. [[CrossRef](#)]
10. Bernardi, D.; Schippa, L. A mobile bed 1D numerical model to support Red River (VietNam) basin management. In Proceedings of the 7th International Conference on Fluvial Hydraulics, River Flow 2014, Lausanne, Switzerland, 3–5 September 2014; Schleiss, A.J., de Cesare, G., Franca, M.J., Pfister, M., Eds.; CRC Press/Balkema: Leiden, The Netherlands, 2014.

11. Stone, M.; Krishnappan, B.G.; Silins, U.; Emelko, M.B.; Williams, C.H.S.; Collins, A.L.; Spencer, S.A. A New Framework for Modelling Fine Sediment Transport in Rivers Includes Flocculation to Inform Reservoir Management in Wildfire Impacted Watersheds. *Water* **2021**, *13*, 2319. [[CrossRef](#)]
12. Ferreira, R.M.L.; Franca, M.J.; Leal, J.G.A.B.; Cardoso, A.H. Mathematical modelling of shallow flows: Closure models drawn from grain-scale mechanics of sediment transport and flow hydrodynamics. *Canad. J. Civil Eng.* **2009**, *36*, 1605–1621. [[CrossRef](#)]
13. Jang, E.K. Sediment Discharge Assessment for Rivers Using Model Tree in Data Mining. Ph.D. Thesis, University of Myongji, Yongin, Korea, 2017. (In Korean).
14. Kang, W.; Jang, E.K.; Yang, C.Y.; Julien, P.Y. Geospatial analysis and model development for specific degradation in South Korea using model tree data mining. *CATENA* **2021**, *200*, 105142. [[CrossRef](#)]
15. Jain, S.K. Development of integrated sediment rating curves using ANNs. *J. Hydraul. Eng.* **2001**, *127*, 30–37. [[CrossRef](#)]
16. Lin, B.; Namin, M.M. Modelling suspended sediment transport using an integrated numerical and ANNs model. *J. Hydraul. Res.* **2005**, *43*, 302–310. [[CrossRef](#)]
17. Nagy, H.M.; Watanabe, K.A.N.D.; Hirano, M. Prediction of sediment load concentration in rivers using artificial neural network model. *J. Hydraul. Eng.* **2002**, *128*, 588–595. [[CrossRef](#)]
18. Bhattacharya, B.; Price, R.; Solomatine, D. Machine learning approach to modeling sediment transport. *J. Hydraul. Eng.* **2007**, *133*, 440–450. [[CrossRef](#)]
19. Khan, M.A.; Stamm, J.; Haider, S. Assessment of soft computing techniques for the prediction of suspended sediment loads in rivers. *Appl. Sci.* **2021**, *11*, 8290. [[CrossRef](#)]
20. Nhu, V.H.; Janizadeh, S.; Avand, M.; Chen, W.; Farzin, M.; Omidvar, E.; Shirzadi, A.; Shahabi, H.; Clague, J.J.; Jaafari, A.; et al. GIS-based gully erosion susceptibility mapping: A comparison of computational ensemble data mining models. *Appl. Sci.* **2020**, *10*, 2039. [[CrossRef](#)]
21. Kang, W.; Yang, C.Y.; Lee, J.; Julien, P.Y. Sediment yield for ungauged watersheds in South Korea. *KSCE J. Civ. Eng.* **2019**, *23*, 5109–5120. [[CrossRef](#)]
22. Yoon, B.; Woo, H. Sediment problems in Korea. *J. Hydraul. Eng.* **2000**, *126*, 486–491. [[CrossRef](#)]
23. Shah-Fairbank, S.C. Series Expansion of the Modified Einstein Procedure. Ph.D. Thesis, Colorado State University, Colorado, CO, USA, 2009.
24. Yang, C.Y.; Julien, P.Y. The ratio of measured to total sediment discharge. *Int. J. Sediment Res.* **2019**, *34*, 262–269. [[CrossRef](#)]
25. Yang, C.Y.; Kang, W.; Lee, J.H.; Pierre, Julien, P.Y. Sediment regimes in South Korea. *River Res. Appl.* **2021**. [[CrossRef](#)]
26. Julien, P.Y. *Erosion and Sedimentation*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2010.
27. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2005.
28. Lee, K.S. A Study on the Spatial Distribution of Suspended Sediment Considering Channel Flow Characteristics. Ph.D. Thesis, Changwon National University, Changwon, Korea, 2021. (In Korean).