Article

Principal Component and Multiple Regression Analyses for the Estimation of Suspended Sediment Yield in Ungauged Basins of Northern Thailand

Piyawat Wuttichaikitcharoen * and Mukand Singh Babel

Water Engineering and Management Program, School of Engineering and Technology, Asian Institute of Technology, P.O. Box 4, Klong Luang, Pathumthani 12120, Thailand; E-Mail: msbabel@ait.asia

* Author to whom correspondence should be addressed; E-Mail: piyawat@rmutl.ac.th; Tel.: +66-818-822-763; Fax: +66-25-246-425.

Abstract: Predicting sediment yield is necessary for good land and water management in any river basin. However, sometimes, the sediment data is either not available or is sparse, which renders estimating sediment yield a daunting task. The present study investigates the factors influencing suspended sediment yield using the principal component analysis (PCA). Additionally, the regression relationships for estimating suspended sediment yield, based on the selected key factors from the PCA, are developed. The PCA shows six components of key factors that can explain at least up to 86.7% of the variation of all variables. The regression models show that basin size, channel network characteristics, land use, basin steepness and rainfall distribution are the key factors affecting sediment yield. The validation of regression relationships for estimating suspended sediment yield shows the error of estimation ranging from −55% to +315% and −59% to +259% for suspended sediment yield and for area-specific suspended sediment yield, respectively. The proposed relationships may be considered useful for predicting suspended sediment yield in ungauged basins of Northern Thailand that have geologic, climatic and hydrologic conditions similar to the study area.

Keywords: suspended sediment; principal component analysis; multiple regression; soil erosion

1. Introduction

An estimation of suspended sediment yield is required for engineering practices that deal with improved land and water management practices in a river basin. The transport of sediment in rivers implies a series of negative effects, such as reservoir siltation and channel bed modification. Such effects may disturb the sediment balance in the basin. In particular, sediment that is eroded from sloping areas can accumulate in the river's network, thereby affecting channel water conveyance [1]. Moreover, several problems due to soil erosion, such as the loss of fine and nutrient-rich topsoil that reduces land productivity, as well as the pollution of surface water bodies, are evident [2–5]. The study of erosion and sediment yield has long established itself as an important area of hydrological research due to the economic significance of the processes involved.

Similar to other developing Southeast Asian countries, land degradation is a major problem in Thailand. This problem manifests itself in terms of the soil structure and its fertility deterioration, in particular for sloping land [6]. Cultivation on sloping areas influences the environment in terms of siltation, flash floods, poor crop yields, *etc.* [7]. The estimation of sediment yield is required in planning and designing water resource development projects, especially for studying the feasibility of a dam or a barrage, assessing sediment budgets and examining the delivery of sediment and contaminants to the estuarine or ocean system, which also provides a valuable means of studying the denudation process [8]. However, sediment data is rarely available due to the lack of monitoring. Erosion and sediment transport are complex phenomena, and these processes are affected by several factors, such as climatic and geomorphological conditions, land use, *etc.*

The approaches employed to estimate sediment yield can be divided into four main groups [9], namely: (1) the soil erosion and sediment delivery approaches, wherein estimated soil erosion rates are factored by a sediment delivery ratio, which is often based on basin characteristics; (2) the physically-based and/or distributed basin modeling approaches, wherein the movement of water and soil is estimated in a distributed way throughout the basin; (3) the models relating sediment concentration or the load to the river flow, wherein measured sediment concentration data is related to river flow characteristics; and (4) empirical models based on broad basin and climate descriptors, wherein sediment yield equations are derived from known basin characteristics. The soil erosion and sediment delivery approaches are usually based on the Universal Soil Loss Equation (USLE) [10] and the concept of the sediment delivery ratio (SDR) [11]. Although many combinations of erosion and sediment delivery modelling are available [12–16], they still require calibration and, thus, cannot be transferred from the study area to other catchments and environments. Moreover, USLE cannot be applied easily to non-agricultural land uses or to areas outside of the range of the original development and application [9].

The physically-based model describes the physical processes involved in the flow and transport of sediment, and these processes use the laws of the conservation of mass, momentum and sediment transport to explain the inherent processes; however, the physically-based model requires extremely onerous input data. When the input data is scarce, the large number of involved parameters may cause significant uncertainty in soil erosion estimates [1]. Furthermore, the simulation of sediment transport at the basin scale is still computationally very expensive. The models relating sediment concentration or load to river flow are most commonly used in practice. These models assume that river flow, rather than sediment supply, is the dominant factor in sediment yield. However, such models also require a

large amount of data to give realistic estimates of long-term average annual sediment yield. This approach is based on "what has happened" rather than "what may happen". Understanding sediment supply and transport processes is required to extrapolate their potential consequences during unmonitored future climate and/or land-use scenarios. The empirical model is based on limited knowledge of the processes and relies on the data describing input and output behavior. This method, however, is able to make abstractions and generalizations of the process and often complements the physically-based model [17].

Several authors have shown the effectiveness of statistical relationships, which allow one to estimate river sediment transport depending on easily available geomorphologic, hydrological and climatic parameters [1,18–33]. Sediment yield is controlled by factors that control erosion and sediment delivery, including local topography, soil properties, climate, vegetation cover, catchment morphology, drainage network characteristics and land use [26,28]. Langbein and Schumm [32] studied the relationship between mean annual precipitation and sediment yield in the United States, while Walling and Webb [31] concluded that no simple relationship exists between climate and sediment yield, because climate's effect on sediment load is very complex. Anderson [33] proposed three major groups of explanatory variable as being involved in relating sediment yield to watershed variables. These are the hydrologic event variables, the watershed conditions and land use variables, as well as the inherent watershed variables, such as area, geology and physiography. He also mentioned that sediment measuring device and its efficiency is also important in having accurate sediment measurements. Bray and Xie [29] identified six categories of variables that can be related to the processes associated with the generation and delivery of suspended sediment to the basin outlet in Canada, which are hydroclimatic conditions, basin topographic features, land surface features, soil characteristics, channel network features and human activities. Ciccacci *et al.* [30] and Grauso *et al.* [1] investigated the correlation between the sediment yield and some geomorphologic, hydrological and climatic parameters in Italy. They found a significant relationship between average yearly sediment yield per unit watershed area and the drainage density. Restrepo *et al.* [23] developed a multiple regression model for estimating the sediment yield in a South American watershed. They reported six catchment variables that predict sediment yield, including runoff, precipitation, precipitation peakedness, mean elevation, mean water discharge and relief, while the mean annual runoff is the dominant control factor. Syvitski and Milliman [22] provided a description of factors influencing the estimation of sediment loads from rivers, which are drainage area size, basin relief, geologic condition, climate and vegetation cover. They successfully estimated the long-term flux of sediment delivered by rivers to the coastal zone (488 global rivers) by the BQART model, which is influenced by geomorphic and tectonic characteristics, geography, geology and human activities. Recently, Cohen *et al.* [19] introduced a comprehensive global fluvial sediment predictor named WBMsed (Water Balance Model with sediment), a distributed global-scale riverine sediment flux model. The major important inputs for the model are anthropogenic factors, ice cover, lithology, reservoir sediment trapping, drainage area size, maximum basin relief, daily temperature and daily discharge.

The statistical method for reducing a large number of interrelated variables into a smaller number of dominant variables is called principal components analysis (PCA) and has been used in many areas of scientific research [17,34–40]. Recently, Tayfur *et al.* [41] investigated sediment load prediction and generalization from the laboratory scale to the field scale using principle component analysis (PCA) in

conjunction with data-driven methods of artificial neural networks and genetic algorithms. In spite of these several uses, there is a disadvantage to PCA: the interpretability of the second and higher components may be limited. For this reason, Varimax rotation is applied to the PCA's solution to enhance the interpretability of the components by maximizing a simple structure. An alternative rotational approach is known as the independent component analysis (ICA) [42–44], which finds a linear representation of non-Gaussian data, so that the components are statistically independent. Westra *et al.* [44] report that the PCA and Varimax rotations provide fairly accurate interpretations for global and local phenomena, respectively, while the interpretability of ICA results appears to be less successful.

The objectives of this study are to propose a complementary methodology that can be used in the prediction of suspended sediment yield in an ungauged basin (*i.e.*, one where the river flow data is unavailable) based on a data-driven modeling approach. The use of the PCA with Varimax rotation to identify the key factors affecting sediment yield and the use of multiple regression analysis to establish the relationships between suspended sediment yield and the basin's characteristics in terms of geomorphology and climate are also investigated.

2. Study Area

The study basin covers an area of 102,636 km^2 of Ping, Wang, Yom and Nan river basins in Northern Thailand. It is located between 15°30′ N and 20°00′ N latitudes and 98°00′ E and 101°30′ E longitudes (Figure 1). The Ping, Wang, Yom and Nan rivers are the main tributaries of the Chao Phraya River, the most important river of Thailand. These four tributaries originate from the Phi Pannam Mountain and course through mountainous areas before merging with each other in the alluvial plains of the Nakhon Sawan Province to form the Chao Phraya River.

Figure 1. The study area showing the locations of suspended sediment gauging stations.
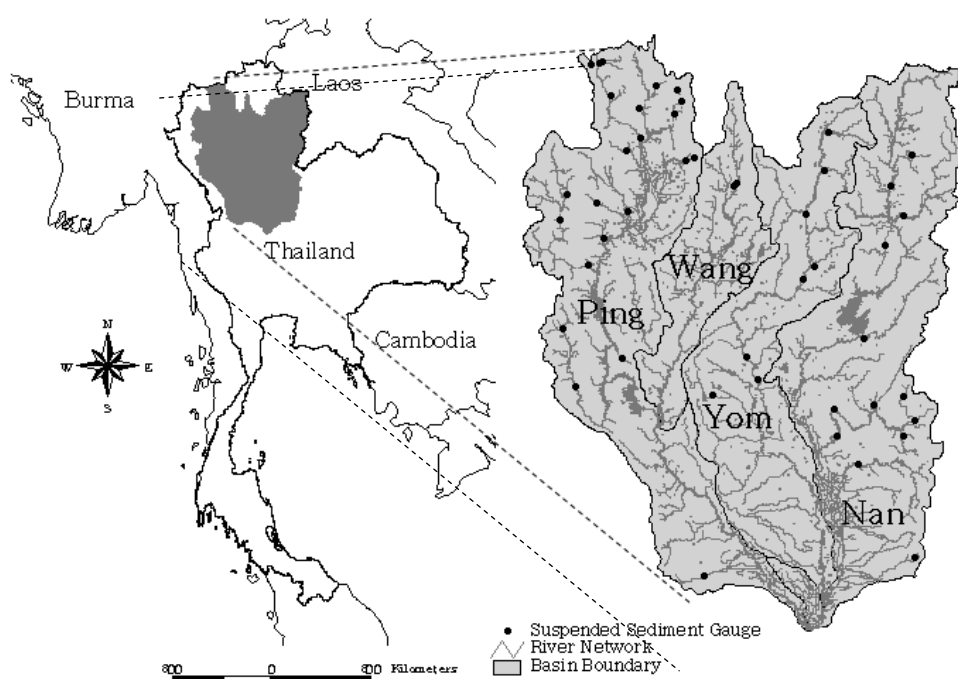
The study area is mountainous, with agriculturally productive valleys. The Ping, Wang, Yom and Nan rivers travel from north to south. The climate of the study area is dominated by seasonal monsoons. The rainy season that lasts from May to October is influenced by the southwest monsoon from the Indian Ocean and the depressions originating in the Pacific Ocean. The average monthly temperature ranges from 15 °C in December to 40 °C in April, except in high altitude locations. The study area can be classified as a tropical rainforest with high biodiversity. The general description of the study area [45] is presented in Table 1.

Table 1. A general description of the study area.

Basin Characteristic	Ping	Wang	Yom	Nan
Drainage area (km^2)	33,896	10,791	23,616	34,331
Main river length (km)	740	460	735	770
Forest area (percent)	73.66	76.07	49.68	45.14
Mean annual discharge (m$^3 \cdot$s^{-1})	276.59	51.26	115.95	381.07
Mean annual runoff (10^6 m$^3 \cdot$yr^{-1})	8725.30	1617.50	3656.60	12,014.80
Mean annual rainfall (mm$\cdot$yr^{-1})	1125	1099	1159	1241
No. of selected rain gauge stations	45	23	23	34
No. of selected suspended sediment gauging stations	22	1	4	10

In terms of soil erosion, Alford's report [46] on mountain watersheds informs us that the Chao Phraya river basin, in Northern Thailand, showed no evidence of a significant increase in sediment yield during the period extending from the late 1950s to the mid-1980s. However, the Northern region of Thailand is very vulnerable to soil erosion, due to its undulating topography, steep slopes and high rainfall. Due to rapid economic development and population growth in the area, the forest-covered land in this northern region decreased from 68.54% in 1961 to 54.27% in 2004 [47]. The most vulnerable area is steeply sloping land, which is under cultivation (more than 35% of sloping land). In recent times, human encroachment on forest areas in the upper part of the study area and land use changes with respect to agriculture have become problematic [48].

3. Framework of the Analysis

The overall study framework involves basin data collection, principal component analysis (PCA) and multiple regression analysis. The data used in the analysis were obtained from hydro-meteorological stations, topographic maps, soil maps and land use maps of the study area. PCA is employed to determine the most prominent variables, which are then used in multiple regression analysis. The details of the data compiled and the methodology employed are presented in the following sections.

3.1. Basic Data

3.1.1. Geomorphic Parameters

The topography of the study area was acquired as a 30-m digital elevation model (DEM) from the Geo-Informatics and Space Technology Development Agency (GISTDA). The 30-m DEM was aggregated to 150-m resolution. This aggregation was done for the further use of the DEM in the

physically distributed watershed model, Distributed Hydrology Soil Vegetation Model (DHSVM) [49], in the next phase of this research, which will be published in the near future. The characteristics of each of the sub-basins within the study area were then derived using HEC-GeoHMS 10.1 [50]. These characteristics include basin area, basin perimeter, basin length, basin slope, main channel length, distance between the basin outlet and a point on the stream nearest to the centroid of the basin area, total channel length, drainage density, basin relief, relief ratio, basin elongation and basin circularity. Most of the extracted basin areas and river networks match well with the existing GIS maps published by the Department of Water Resources (DWR) [51]. It is worth noting that a few sub-basins could not be delineated in areas that are relatively plains, which create difficulties in delineating the river and basin boundary. The river network's properties, namely the hierarchical anomaly index and the hierarchical anomaly density [1,21,30,52,53], were estimated based on the digitized river network derived from 1:50,000 topographical maps obtained from the Royal Thai Survey Department, which was satisfactorily compared to the existing river network [51]. To elaborate on the river network's properties, let us assume G as the number of first order streams necessary to make a drainage network perfectly ordered in a binary tree-shaped structure with streams of order K flowing into streams of order K + 1, and N is the number of first order streams present in the drainage network. The hierarchical anomaly index (DA) is given by the ratio of G to N, while the hierarchical anomaly density (GA) is the ratio of G to the basin area in square kilometers. These two parameters express the organization degree of drainage networks. Ciccacci *et al.* [30] and Grauso *et al.* [21] provide more details of these two parameters.

3.1.2. Soil Properties

The soil map of the study area was drawn using the Soil Program software [54], which derives 5-min resolution (about 10 km) soil data from the World Inventory of Soil Emission Potentials (WISE) pedon database [55], developed by the International Soil Reference and Information Centre (ISRIC) and the FAO-UNESCO Digital Soil Map of the World [56]. In this study, the soil clay content as a percentage was extracted and used as the soil's representative property.

3.1.3. Land Use

The digital land use map, obtained from the Land Development Department (LDD) of the Royal Thai Government, was employed in this study. It was derived from Landsat 5 satellite imagery with 30-m resolution and the ground truth survey of 2000–2003. The forest and agricultural areas were extracted from each of the sub-basins to represent the land cover property used in the analysis.

3.1.4. Hydro-Meteorological and Sediment Data

The daily suspended sediment yield data, observed at 37 gauging stations operated by the Royal Irrigation Department (RID) and the Department of Water Resources (DWR) of the Royal Thai Government, were obtained as presented in Table 2. Daily suspended sediment data were calculated using the Sediment-discharge rating curve technique. The rating curve is derived using at least 20 measurement points per year for each station. The United States standard sampling method and

equipment e.g., depth integrating sampler (US DH-48, US DH-49 and US DH-59) or point integrating sampler (US-P-46, US-P-61, US-P-63 and US-P-50), are employed based on water depth and the accessibility of each measurement point [57]. The examples of sediment-discharge rating curve equations provided by RID for the year 2000 for Ping, Wang, Yom and Nan river basins are: $Q_S = 4.8767\ Q_W^{1.323}$ ($R^2 = 0.811$, 32 observation points at P.65 for Ping), where Q_S is suspended sediment load in tons day^{-1} and Q_W is river discharge in m$^3 \cdot$s^{-1}; $Q_S = 3.0996\ Q_W^{1.39}$ ($R^2 = 0.945$, 27 observation points at W.16A for Wang); $Q_S = 6.0248\ Q_W^{1.244}$ ($R^2 = 0.886$, 63 observation points at Y.34 for Yom); and $Q_S = 0.1792\ Q_W^{1.9793}$ ($R^2 = 0.922$, 31 observation points at N.42 for Nan); respectively. Since there are no existing water infrastructures upstream of the gauging stations, the data are free from the effects of regulating structures. The daily rainfall data, monitored by RID and the Thai Meteorological Department (TMD), were obtained from 125 stations located in the study area. Since land use data were available for 2000–2002, both sediment and rainfall data collected from 1995 to 2007 (based upon availability) were used in the study, assuming that the land use remains the same and that there are no major man-made changes taking place for the study data period.

The annual suspended sediment yield was calculated using the daily data for each of the selected sub-basins. Table 2 shows the general description of the suspended sediment gauging stations in this study. For climate characteristics, annual rainfall, wet season rainfall (May–October), dry season rainfall (November–April) and the precipitation concentration index [58] were estimated. The mean areal rainfall was estimated by the Thiessen polygon method using the ArcView ArealRain Extension [59]. In each of the sub-basins, the time series data—suspended sediment yield, annual rainfall, *etc.*—were averaged as long-term average data for further analysis. The glossary and summary statistics of the variables used in this study are given in Table 3.

3.2. Principle Component Analysis

Principal component analysis (PCA) was applied in this study to identify the factors influencing suspended sediment yield. The PCA is a method of data reduction that aims to identify a small number of derived variables from a larger number of original variables in order to simplify the subsequent analysis of the data [60,61]. Moreover, the PCA has been used in the present study as the preliminary step in the development of a prediction model [62]. The sequence of the main steps involved in the PCA, as applied by Halim *et al.* [38], were adapted and are described below:

(1) Selection of a set of basin characteristics and meteorological indicators for the study area. The initial set consisted of 17 basin characteristics and 4 climate factors (Table 3).

(2) Assessment of the suitability of data for the PCA using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy [63] and Bartlett's test of sphericity [64]. KMO tests the ratio of item correlations to partial item correlations. If the partials are similar to the raw correlations, it means that the item does not share much variance with other items. This is a necessary criterion, since PCA assumes that common factors are the source of variance for the variables under investigation. The range of KMO is from 0.0 to 1.0; however, the score of 0.50 is suggested as the minimum value for a good PCA [65]. Bartlett's test of sphericity checks for the hypothesis that the correlation matrix is an identify matrix, which means that all of the variables are uncorrelated. The significance value for this analysis led us to reject the null

hypothesis and conclude that there are correlations in the data set that are appropriate for the PCA. The score from Bartlett's test of sphericity with significance at 95% ($p < 0.05$) is considered appropriate for the PCA [61]. In addition, Tabachnick and Fidell [66] also recommend that, for the PCA, the correlation matrix should show at least some correlations, with the correlation coefficient being greater than or equal to 0.30.

(3) Determination of dominant factors. The PCA with Varimax rotation is performed to identify the principal components (PCs) or subsets from a larger data set. For selecting the dominant factors, Kaiser's criterion or the eigenvalues rule, *i.e.*, only components with eigenvalues of 1.0 or more are retained for further investigation [38,67,68], was employed.

3.3. Regression Analysis

The regression relationships between suspended sediment yield and the dominant factors obtained from the PCA, *i.e.*, biophysical and climate factors, were established using Equation (1). In order to avoid the negative lower boundary of estimation, a logarithmic transformation was used. The regression coefficients were obtained by ordinary least squares linear regression on logarithms of response and predictor variables. Finally, a back-transformed relationship was obtained in the form [62]:

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} ... X_p^{\beta_p} \tag{1}$$

where Y is the response variable (suspended sediment yield in this study), X_1, X_2, …, X_p are the predictor variables (the factors influencing suspended sediment yield) and β_0, β_1, β_2, …, β_p are constants derived by the multiple linear regression analysis. The most commonly used procedure for selecting the best regression equation is stepwise linear regression analysis (using an F probability of 0.05 for the selected factor), as described by Landau and Everitt [60], was performed using SPSS for Windows Release 11.5.

Generally, the size of the drainage area is an important factor for both suspended sediment yield and area-specific suspended sediment yield. The relationship between the size of the drainage area and suspended sediment yield is complicated by many other factors, such as rainfall, plant cover, texture of the sediment and land use [69]. In order to evaluate the effect of each dominant factor in predicting suspended sediment yield in various categories of basin sizes [70] and based on the available data, regression models were generated based on 4 groups of data: (1) 7 sub-basins with a drainage area of less than 100 km^2 (small basins); (2) 15 sub-basins with a drainage area of more than 100 km^2, but less than 1000 km^2 (medium basins); (3) 8 sub-basins with a drainage area of more than 1000 km^2 (large basins); and (4) all 37 sub-basins irrespective of drainage area size.

Table 2. List of suspended sediment gauging stations in this study.

No. *	Station Code	Station Name or Location	Data Period	Basin Area ** (km^2)	Avg. Annual Rainfall (mm·yr^{-1})	Avg. Annual Runoff (mm·yr^{-1})	Avg. Annual Sediment (ton km^{-2}·yr^{-1})
	Ping						
1	**060201**	**Nam Mae Mae at Ban Mae Na *****	**1995–2007**	**47.54**	**1292**	**482**	**53.84**
2	060202	Nam Mae Pam at Sop Huai Mae Mat (Down Stream)	1995–2007	205.63	1034	295	120.36
3	060301	Nam Mae Ngat at Ban Teen That	1997–2007	85.28	969	535	26.88
4	060302	Nam Mae Saluam at Ban Thung Ku	1997–2007	43.65	1152	286	50.43
5	060401	Huai Mae Hat at Ban Na Mon	1997–2004	75.02	1036	356	18.05
6	060406	Nam Mae Taeng at Ban San Pa Sak (Upstream)	1997–2004	875.05	1170	291	64.90
7	**060602**	**Nam Mae Rim at Ban Nong Gai**	**1995–2004**	**163.35**	**1053**	**308**	**24.73**
8	060701	Nam Mae Wan at Aban Mae Wan	1998–2005	47.66	1291	728	60.96
9	060704	Huai Ma Klaing at Ban Pa Maing Pang Bong	2000–2005	5.02	1425	932	122.34
10	060804	Nam Mae Sapok at Ban Mae Sapok (Upstream)	1996–2006	36.41	1352	352	20.97
11	060808	Nam Mae Khan at Ban Piang	1995–2007	1199.95	1318	181	53.82
12	061202	Nam Mae Mu at Ban Mae Mu	1995–2006	67.10	1350	373	13.99
13	061302	Nam Mae Chaem at Bana Kong Kan	1996–2007	2055.65	1152	309	134.58
14	061501	Nam Mae Tun at Ban Pa Kha	1995–2007	1588.79	995	454	140.61
15	P14	Nam Mae Chaem at Kaeng Ob Luang Chiang Mai	2000–2005	3828.31	1256	274	216.25
16	**P24A**	**Nam Mae Klang at Pracha Uthit Chiang Mai**	**1997–2005**	**449.33**	**1051**	**316**	**49.58**
17	P35	Khlong Khlung at Ban Pang Wai Kamphaeng	1995–2001	745.22	1406	468	73.51
18	P4A	Nam Mae Taeng at Ban Sanmahaphon Chiang Mai	1995–2007	1892.48	1161	180	44.84
19	P56A	Nam Mae Ngat at Ban Sahakhon Chiang Mai	2000–2007	546.39	1082	346	55.33
20	P64	Nam Mae Tun at Highway Bridge Chiang Mai	1997–2002	494.51	1021	419	99.19
21	P65	Nam Mae Teang at Ban Muang Pog Chiang Mai	1995–2001	233.71	1041	398	45.81
22	P70	Nam Mae Taeng at Ban Huai Khrai Chiang Mai	1997–2000	173.14	1049	324	110.05
	Wang						
23	W16A	Nam Wang at Ban Hai Lampang	2000–2007	1330.07	1,299	243	74.91
	Yom						
24	Y24	Nam Pi at Highway Bridge Phayao	1997–2005	591.19	1,251	233	44.12

Table 2. *Cont.*

No. *	Station code	Station name or location	Data period	Basin area ** (km²)	Avg. annual rainfall (mm·yr⁻¹)	Avg. annual runoff (mm·yr⁻¹)	Avg. annual sediment (ton km⁻²·yr⁻¹)
25	**Y26**	**Nam Mae Mok at Ban Mae Phu Lampang**	**1998–2006**	**787.01**	**1222**	**253**	**13.93**
26	Y34	Nam Mae Lai at Ban Mae Lai Phrae	1997–2002	333.23	1295	306	32.58
27	Y36	Mae Nam Khuan at Ban Pa Kha Phayao	2000–2005	851.54	1285	448	78.34
	Nan						
28	N22	Khwae Noi at Ban Yang Phitsanulok	1997–2006	4648.48	1340	443	125.12
29	**N24**	**Nam Khek at Ban Wang Nok Phitsanulok**	**1998–2006**	**1816.70**	**1383**	**524**	**149.53**
30	**N40**	**Khwae Noi at Ban Nong Bon Phitsanulok**	**1999–2004**	**4180.45**	**1420**	**507**	**156.74**
31	N42	Nam Wa at Ban Hat Khao Nan	1997–2002	2085.91	1316	1013	255.52
32	N49	Nam Yao at Highway Bridge Nan	2000–2005	157.66	1456	986	160.89
33	N53	Khlong Butsabong at Ban Huai Tum Phetchabun	1999–2006	100.78	1472	577	342.75
34	**N58**	**Nam Fua at Ban Kok Muang Phitsanulok**	**2001–2006**	**296.53**	**1263**	**437**	**301.56**
35	N59	Lam Nam Khan at Ban Na Chan Phitsanulok	2001–2006	404.37	1,336	551	186.64
36	N63	Nam Haeng at Highway Bridge Nan	1997–2005	739.96	1,112	187	70.46
37	N69	Nam Nan at Ban Na Thung Yai Phitsanulok	2000–2004	168.75	1227	633	96.81

Notes: * Stations No. 1 to 14 belong to Department of Water Resources (DWR), and those from No. 15 to 37 belong to Royal Irrigation Department (RID); ** the basin area is extracted from DEM; *** bold stations and data are not used in developing the regression model and are used for validating the developed model.

Table 3. Glossary and summary statistics for the sub-basin's characteristics.

Sub-basin characteristic	Unit	Description	Sub-basin range		
			95% l.l.*	Median	95% u.l.**
Basin characteristics					
AA	percent	Agricultural area	0.01	16.44	64.68
AREA	km^2	Basin area	33.27	449.33	4227.26
BC	-	Basin circularity	0.158	0.271	0.395
BE	-	Basin elongation	0.484	0.751	1.085
BL	km	Basin length	8.20	26.05	91.51
BP	km	Basin perimeter	37.05	154.80	570.45
BR	m	Basin relief	374.90	1010.00	1680.80
BS	-	Basin slope	0.006	0.023	0.089
DA	-	Hierarchical anomaly index	0.235	1.025	2.341
DD	km^{-1}	Drainage density	0.945	1.677	2.165
FA	percent	Forest area	33.15	82.20	97.64
GA	-	Hierarchical anomaly density	0.17	0.94	1.96
LC	km	Distance from basin outlet to a point on the stream nearest the centroid of the basin area	6.53	18.83	85.81
MCL	km	Main channel length	10.90	41.66	175.77
RR	-	Relief ratio	0.011	0.038	0.128
TCL	km	Total channel length	54.67	584.44	6168.68
TSCC	percent	Top soil clay content	24.51	25.07	27.20
Climate					
AR	mm·yr^{-1}	Annual rainfall	992.48	1256.34	1457.77
DSR	mm·yr^{-1}	Dry season rainfall	113.17	154.75	195.10
PCI	-	Precipitation concentration index	15.29	16.75	18.26
WSR	mm·yr^{-1}	Wet season rainfall	855.23	1106.34	1333.35
Sediment					
ASSY	ton km^{-2}·yr^{-1}	Area-specific suspended sediment yield	13.98	73.51	305.67
SSY	ton yr^{-1}	Suspended sediment yield	749	30,232	672,487

Notes: The overall data is 37 samples from 37 sub-basins; * lower limit; ** upper limit.

3.4. Model Validation

From the 37 samples (37 selected stations), 30 samples were used for the multiple regression model's development, while the remaining 7 samples were randomly excluded, based on the drainage area size, for the validation of the model. Additionally, a method called the jack-knife technique [62] was applied to examine the validity of developed regression models. This technique is generally performed by excluding one sub-basin from the total sub-basins. After that, regression having the same form as that of the general model was fitted using the all-but-one sub-basins, and the suspended sediment yield (SSY) or area-specific suspended sediment yield (ASSY) of the left out sub-basin was estimated by the obtained regression model called the test model. The calculation procedure was repeated for all sub-basins, and the coefficient of determination of the test model was calculated. Furthermore, the Pearson product-moment correlation coefficient between the predicted SSY or ASSY from the general and test models was also calculated.

4. Results and Discussion

4.1. Factors Influencing Suspended Sediment Yield

In this study, prior to performing PCA, the suitability of data for analysis was assessed. There were 37 datasets of 23 variables consisting of 17 basin characteristics, four climate factors and two sediment related variables (Table 3). The cross-correlations among 23 variables are given in Table 4. The KMO score was 0.59 and Bartlett's test of sphericity showed significance at 95%, which reasonably supports the factorability of the cross-correlation. Additionally, the correlation matrix showed many correlation coefficients to be above 0.30. Therefore, factor analysis could be applied to reduce the number of factors in this study.

The PCA results based on the correlation matrix analysis with Varimax rotation indicate six principal components with eigenvalues greater than 1.00, which correspond to an overall cumulative variance of 86.7%. The order of significance of these variables is determined by the magnitude of their eigenvalues, as presented in Table 5.

The different variables considered in the PCA and their factor loadings within their respective PCs are presented in Table 6. It shows that the high weighted variables (factor loading ≥ 0.60) for PC1 consist of the total channel length (TCL), basin area (AREA), main channel length (MCL), distance from the basin outlet to a point on the stream nearest to the centroid of the basin area (LC), basin perimeter (BP), suspended sediment yield (SSY), basin length (BL), hierarchical anomaly index (DA), and hierarchical anomaly density (GA). PC2 consists of wet season rainfall (WSR), annual rainfall (AR) and hierarchical anomaly density (GA). PC3 consists of basin slope (BS), relief ratio (RR) and basin circularity (BC). PC4 consists of agricultural area (AA), forest area (FA) and area-specific suspended sediment yield (ASSY). PC5 consists of the precipitation concentration index (PCI) and dry season rainfall (DSR). Lastly, PC6 has basin elongation (BE) as a variable with high loading. From PC1 and PC4, it was seen that the suspended sediment yield (SSY) corresponds to the basin size (AREA), whereas area-specific suspended sediment yield (ASSY) corresponds to such land cover characteristics as forest area (FA) and agricultural area (AA), which imply that forest cover can reduce the erosion rate.

In addition, from Table 6, basin relief (BR) and top soil clay content (TSCC) show less commonality than the others, with scores of 0.319 and 0.645, respectively. These numbers suggest that a substantial portion of the variable's variances is not accounted for by these two factors, and these are considered as less closely related to other variables.

Table 4. Correlation matrix of the identified variables.

Variable	AREA	BP	BL	BS	MCL	LC	TCL	DD	BR	RR	BE	BC	TSCC	FA	AA	AR	WSR	DSR	PCI	GA	DA	SSY	ASSY
AREA	1.00	**0.97**	**0.82**	**−0.47**	**0.93**	**0.93**	**0.99**	**−0.33**	0.12	**−0.49**	0.09	**−0.62**	−0.07	−0.21	0.21	0.23	0.09	0.26	−0.03	**0.47**	**0.78**	**0.92**	0.29
BP		1.00	**0.88**	**−0.59**	**0.98**	**0.96**	**0.95**	**−0.43**	0.13	**−0.61**	0.04	**−0.75**	−0.01	−0.23	0.24	0.21	0.08	**0.32**	−0.05	**0.43**	**0.80**	**0.87**	0.29
BL			1.00	**−0.57**	**0.94**	**0.92**	**0.85**	−0.29	0.24	**−0.61**	**−0.32**	**−0.69**	−0.06	−0.01	0.02	0.11	−0.01	0.25	−0.05	**0.60**	**0.84**	**0.78**	0.29
BS				1.00	**−0.58**	**−0.53**	**−0.48**	**0.38**	0.05	**0.99**	−0.10	**0.76**	−0.16	0.29	**−0.31**	0.18	0.25	**−0.43**	0.22	**−0.35**	**−0.58**	**−0.38**	−0.12
MCL					1.00	**0.98**	**0.92**	**−0.40**	0.20	**−0.60**	−0.08	**−0.74**	−0.04	−0.16	0.16	0.19	0.08	0.29	−0.02	**0.48**	**0.83**	**0.86**	0.29
LC						1.00	**0.91**	**−0.39**	0.24	**−0.55**	−0.14	**−0.70**	−0.04	−0.14	0.15	0.23	0.11	0.23	0.02	**0.47**	**0.81**	**0.85**	0.27
TCL							1.00	−0.23	0.12	**−0.50**	0.03	**−0.59**	−0.09	−0.14	0.15	0.16	0.02	0.24	−0.06	**0.56**	**0.79**	**0.91**	0.28
DD								1.00	−0.08	**0.35**	−0.24	**0.56**	−0.20	**0.39**	**−0.40**	**−0.37**	**−0.35**	**−0.41**	0.06	**0.35**	−0.28	**−0.31**	−0.29
BR									1.00	0.07	−0.05	−0.12	0.11	−0.10	0.09	0.13	−0.06	0.25	0.07	0.19	0.15	−0.04
RR										1.00	−0.02	**0.76**	−0.18	0.27	−0.28	0.17	0.24	**−0.39**	0.21	**−0.38**	**−0.61**	**−0.40**	−0.14
BE											1.00	0.10	0.01	**−0.47**	**0.46**	0.01	−0.02	**0.41**	−0.24	**−0.30**	−0.15	0.01	0.04
BC												1.00	−0.22	0.29	**−0.30**	−0.10	−0.07	**−0.32**	0.07	−0.23	**−0.62**	**−0.54**	−0.27
TSCC													1.00	−0.13	0.09	0.12	0.19	**0.32**	−0.21	**−0.30**	−0.17	−0.10	0.17
FA														1.00	**−0.99**	**−0.35**	**−0.30**	−0.29	−0.04	0.28	−0.06	−0.20	**−0.55**
AA															1.00	**0.32**	0.27	0.29	0.05	−0.27	0.07	0.21	**0.55**
AR																1.00	**0.93**	0.19	0.25	**−0.36**	−0.02	0.25	**0.39**
WSR																	1.00	0.09	**0.33**	**−0.48**	−0.15	0.10	**0.33**
DSR																		1.00	**−0.67**	−0.01	0.25	0.26	0.25
PCI																			1.00	−0.29	−0.20	−0.04	−0.13
GA																				1.00	**0.74**	**0.48**	0.07
DA																					1.00	**0.76**	0.29
SSY																						1.00	**0.47**
ASSY																							1.00

Notes: Bold values indicate correlation coefficients above 0.30; Kaiser–Meyer–Olkin measure of sampling adequacy: 0.59; Bartlett's test of sphericity: 1347.44; significance: 0.000.

Table 5. Principal components (PCs) for basin characteristic and climate factors.

PCs	Eigenvalues	Variance (%)	Cumulative variance (%)
1	9.439	36.031	36.031
2	3.817	12.602	48.633
3	2.706	12.068	60.701
4	1.527	10.881	71.582
5	1.395	8.489	80.071
6	1.060	6.644	86.715

Table 6. Results of principal component analysis (Varimax rotated component matrix).

Factor	Eigenvectors						Commonalities
	PC1	PC2	PC3	PC4	PC5	PC6	
AREA	**0.955**	0.087	−0.101	0.087	0.009	−0.133	0.955
BP	**0.931**	0.099	−0.282	0.077	0.023	−0.093	0.972
BL	**0.891**	−0.001	−0.272	−0.045	0.015	0.258	0.937
BS	−0.427	0.278	**0.783**	−0.170	−0.137	0.059	0.924
MCL	**0.941**	0.091	−0.279	0.024	−0.004	0.018	0.972
LC	0.934	0.131	−0.245	0.004	−0.053	0.051	0.955
TCL	**0.962**	0.002	−0.068	0.060	0.029	−0.069	0.940
DD	−0.244	−0.512	0.525	−0.126	−0.109	0.279	0.702
BR	0.214	0.242	−0.024	−0.255	−0.312	0.227	0.319
RR	−0.448	0.282	**0.774**	−0.165	−0.134	−0.021	0.924
BE	−0.068	−0.004	0.031	0.296	0.270	**−0.864**	0.912
BC	−0.562	−0.085	**0.710**	−0.139	−0.036	−0.096	0.858
TSCC	−0.233	0.335	−0.448	0.037	0.466	0.242	0.645
FA	−0.056	−0.203	0.231	**−0.875**	0.019	0.284	0.944
AA	0.065	0.179	−0.239	**0.879**	−0.038	−0.285	0.948
AR	0.174	**0.868**	0.142	0.239	−0.022	0.026	0.863
WSR	0.023	**0.915**	0.094	0.184	−0.082	0.071	0.893
DSR	0.221	0.190	−0.273	0.097	**0.773**	−0.264	0.836
PCI	−0.060	0.268	0.011	0.027	**−0.903**	0.059	0.895
GA	**0.629**	**−0.612**	0.137	−0.083	0.118	0.338	0.925
DA	**0.867**	−0.200	−0.184	0.059	0.086	0.153	0.860
SSY	**0.922**	0.098	0.023	0.184	0.054	0.003	0.896
ASSY	0.283	0.247	0.086	**0.745**	0.255	0.317	0.869

Notes: Extraction method: principal component analysis, rotation method: Varimax with Kaiser normalization; bold values indicate highly correlated variables included in the PCs (>0.60); underlined values correspond to the first three highest factor loadings in the PCs.

To select prominent variables for subsequent regression analyses, the first three variables with the highest factor loadings and greater than 0.60 were selected as representative variables of each of the PCs. A threshold of 0.60 was used for identifying a reliable factor in this study [71]. Therefore, for PC1, the total channel length (TCL), basin area (AREA) and main channel length (MCL) were selected. For PC2, wet season rainfall (WSR), annual rainfall (AR) and hierarchical anomaly density (GA) were employed. For PC3, basin slope (BS), relief ratio (RR) and basin circularity (BC) were

used. Agricultural area (AA) and forest area (FA) were extracted from PC4, whereas area-specific suspended sediment yield (ASSY) was considered as the response variable in the regression analysis. For PC5, the precipitation concentration index (PCI) and dry season rainfall (DSR) were chosen. Finally, only basin elongation (BE) was considered from PC6. All 14 factors were assumed to be the forcing factors of suspended sediment yield with positive and negative effects, which can be used subsequently as predictor variables in regression analysis.

4.2. Regression Relationships to Estimate Suspended Sediment Yield

From the 37 samples (37 selected stations), 30 samples were used for the multiple regression model's development, while the remaining seven samples were randomly excluded, based on the drainage area size, for the validation of the model. The excluded stations comprise 060201 (47.54 km^2), 060602 (163.35 km^2), N58 (296.53 km^2), P24A (449.33 km^2), Y26 (787.01 km^2), N24 (1816.70 km^2), and N40 (4180.45 km^2) (shown in Table 2). The other 30 samples were used in the multiple regression analysis, which was performed using the selected 14 factors—TCL, AREA, MCL, WSR, AR, GA, BS, RR, BC, AA, FA, PCI, DSR and BE—as the predictor variables. SSY and ASSY were taken as response variables. The analysis was done using the stepwise regression technique [60] in each of the groups. The technique was applied based on the drainage area. To ensure that there is no multi-collinearity in the analysis [61], the result of the regression equations was finally inspected to ensure that there were no inter-correlations among the predictor variables. The sample adequacy criteria suggested by Haan [62] was also considered; he suggested that the sample number should be at least three- or four-times the number of predictor variables. The results of the multiple regression analysis are presented in Table 7.

Table 7. Results of the multiple regression analysis.

Basin area class	No. of sub-basins	Regression model	Equation No.	R^2	Standard error *
<100 km^2	7	SSY = 1947.97 GA$^{0.5408}$	(2)	0.643	584.38 ton yr^{-1}
	7	ASSY = 4127.26 BS$^{1.7451}$ GA$^{0.3772}$	(3)	0.979	7.76 ton km^{-2}·yr^{-1}
100 to 1000 km^2	15	SSY = 14,451.89 AREA$^{0.5549}$ FA$^{-0.6004}$	(4)	0.686	3311.64 ton yr^{-1}
	15	ASSY = 39,789.27 MCL$^{-1.0360}$ FA$^{-0.5627}$	(5)	0.835	9.77 ton km^{-2}·yr^{-1}
>1000 km^2	8	SSY = 0.5885 AREA$^{1.6857}$	(6)	0.624	71,303.96 ton yr^{-1}
		No Correlation for ASSY			
Overall					
	30	SSY = 28.74 AREA$^{1.1636}$	(7)	0.829	18,833.32 ton yr^{-1}
	30	ASSY = 0.0068 DSR$^{1.8506}$	(8)	0.078	13.93 ton km^{-2}·yr^{-1}

Note: * Standard error = [sum of square residual/(n − p)]$^{1/2}$, where n is the number of sample and p is the number of parameters to be estimated.

Based on the Stepwise regression analysis, ANOVA shows the significance to be less than 0.001 as per the F-test for all of the equations, while some cases have no relationships, because of statistical insignificances. Based on the coefficient of determination, R^2 and the standard error of estimation, it can be concluded that ASSY develops better relationships with the selected dominant factors than SSY in cases of a drainage area less than 1000 km^2. In contrast, the prediction of SSY is more reliable than

the prediction of ASSY in the case of basins with larger drainage areas (more than 1000 km^2). This implies that a larger basin contributes to complexity and uncertainty in ASSY modeling.

To evaluate the effect of predictor variables on suspended sediment amount, it was found that GA, AREA and FA contribute to SSY estimation, while BS, GA, MCL, FA and DSR contribute to ASSY estimation. Therefore, it was concluded that basin size, channel network characteristics, land use, basin steepness and rainfall distribution are the key factors affecting the amount of suspended sediment. For the medium-sized basins (100 to 1000 km^2), the regression relationships imply that less forest cover or more agricultural area contribute to more SSY and ASSY. The larger the basin size, the more is SSY. It was also found that BS, or basin slope characteristics, is the factor affecting ASSY for a drainage area of less than 100 km^2. A higher basin slope contributes to higher ASSY, only for basin size less than 100 km^2. Additionally, high dry season rainfall leads to a high amount of ASSY. This physically implies that the land use/cover (e.g., crops) during the dry season relatively enhance soil surface erosion compared to the wet season. Considering all data samples (irrespective of basin size), the relationships shown in Equations (7) and (8) revealed quite interesting results that the amount of suspended sediment depends on the basin size and dry season rainfall irrespective of the geomorphological conditions. Kazama *et al.* [72] also pointed out that, in the Mekong Basin, the suspended sediment transport is highly sensitive to particle size compared to the channel bed slope. This may mean that these regions (the study area and Mekong region, which are adjacent to each other) have a low influence of topography on sediment yield.

### 4.3. Regression Model Validation

The summary results of the jack-knife technique are presented in Table 8. Figures 2 and 3 elaborate the validation results of Equation (7). These results indicate that all general models have high correlation coefficient (R) between general and test models. However, some test models represented by Equations (2), (4) and (6) give a relatively low value of R^2 compared to Equations (3), (5) and (7). This also supports the result of regression analysis in the previous section that ASSY provides better correlations to the factors than SSY when the basin area is smaller than 1000 km^2. The results apparently show that Equation (8) is considered to be less reliable with very small values of R^2 for both general and test models, which might result in relatively higher error values. Thus, the use of Equation (8) is not recommended.

**Table 8.** Model validation results using the jack-knife technique.

Equation No.	R^2		Correlation Coefficient, R, between General and Test Model
	General Model	Test Model	
(2)	0.643	0.376	0.961
(3)	0.979	0.882	0.975
(4)	0.686	0.496	0.953
(5)	0.835	0.649	0.960
(6)	0.624	0.355	0.950
(7)	0.829	0.792	0.999
(8)	0.078	0.016	0.973

Figure 2. The general (for Equation (7)) and test models' correlation diagram of predicted *versus* observed SSY.
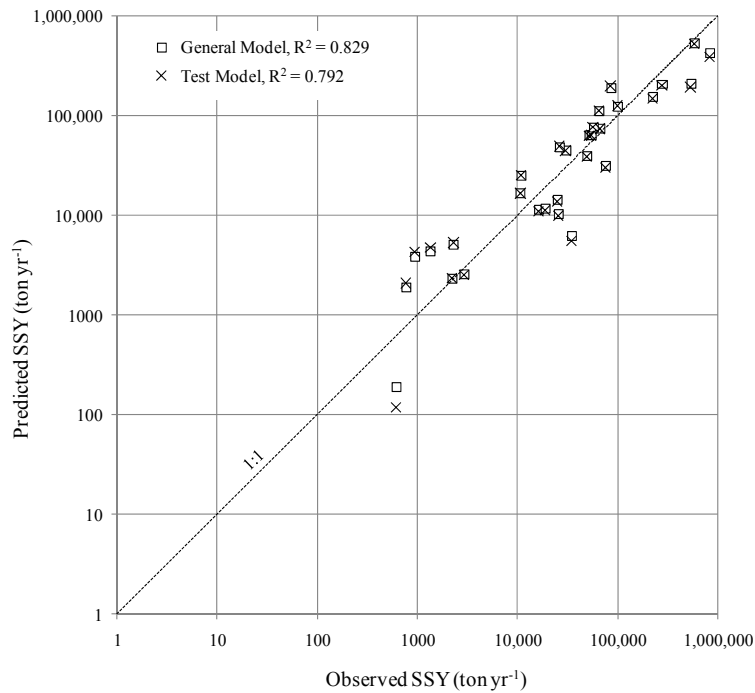
Figure 3. Scatter plot of the SSY results of the general model and Equation (7) *versus* the test models.
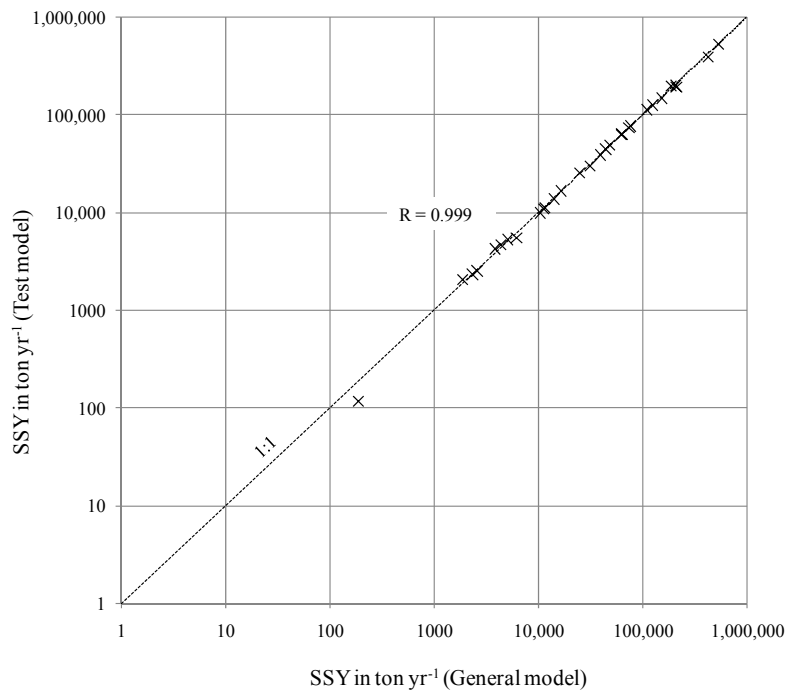
The seven stations that were left out initially were used for model testing. The validation results for SSY and ASSY are given in Tables 9 and 10, respectively. The graphical presentation is also consecutively shown in Figures 4 and 5 for SSY and ASSY, respectively. The validation results indicate that, in most cases, using the model in a particular group based on the drainage area size provides more accurate values than using a model developed from all data sets. The error of estimation ranges from −55% to +315% for SSY prediction and −59% to +259% for ASSY prediction

(Equations (7) and (8) are excluded). However, if Equations (7) and (8) are employed, the estimated error of SSY and ASSY will range from −76% to +514% (last column in Table 9) and −76% to 622% (last column in Table 10), respectively. Figures 4 and 5 also show that Equations (7) and (8) give a relatively higher error of estimations compared to the models developed for three classes of basin area.
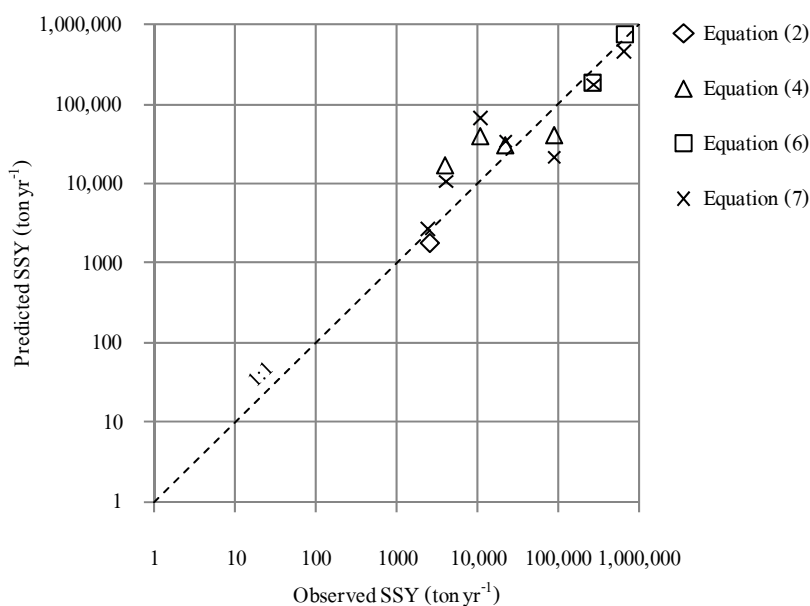
Table 9. Validation results for SSY prediction.

Station	Observed	Predicted SSY (Equation No.)							
	SSY	SSY (2)	% Error	SSY (4)	% Error	SSY (6)	% Error	SSY (7)	% Error
60201	2559.66	1798.08	−29.75	-	-	-	-	2570.21	0.41
60602	4039.89	-	-	16,762.50	314.92	-	-	10,807.36	167.52
N58	89,419.54	-	-	40,018.63	−55.25	-	-	21,629.09	−75.81
P24A	22,275.57	-	-	29,870.13	34.09	-	-	35,080.79	57.49
Y26	10,962.63	-	-	38,812.47	254.04	-	-	67,346.95	514.33
N24	271,656.94	-	-	-	-	183,678.25	−32.39	178,267.44	−34.38
N40	655,223.75	-	-	-	-	748,500.32	14.24	470,153.94	−28.25

Table 10. Validation results for ASSY prediction.

Station	Observed	Predicted ASSY (Equation No.)					
	ASSY	ASSY (3)	% Error	ASSY (5)	% Error	ASSY (8)	% Error
060201	53.84	43.48	−19.24			90.33	67.77
060602	24.73			88.72	258.74	52.75	113.28
N58	301.56			122.36	−59.42	73.62	−75.59
P24A	49.58			68.82	38.82	58.97	18.94
Y26	13.93			32.61	134.13	100.55	621.88
N24	149.53					66.75	−55.36
N40	156.74					93.97	−40.05

Figure 4. Validation results for SSY prediction.

Figure 5. Validation results for ASSY prediction.
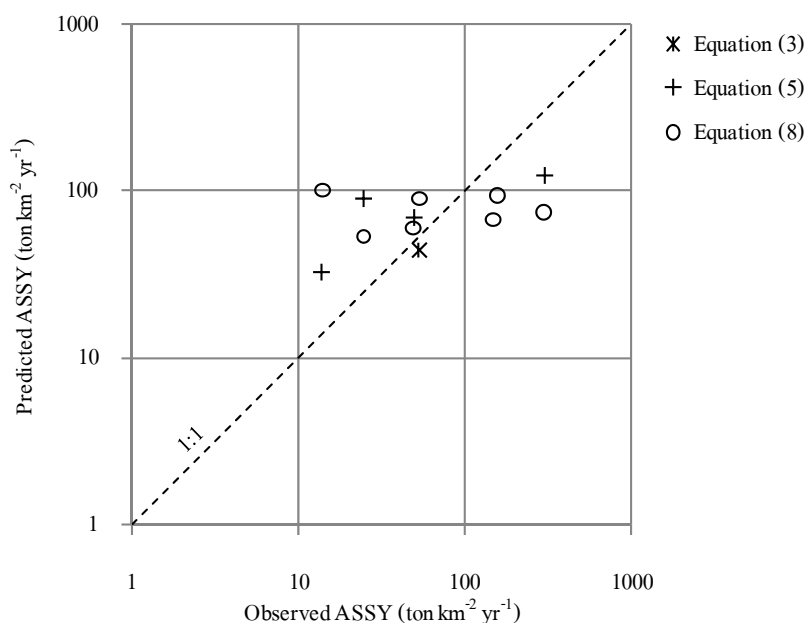
5. Conclusions

The investigation of factors affecting suspended sediment yield in the Ping, Wang, Yom and Nan river basins in Thailand, using principal component analysis, is presented in this study. From the principal component analysis, six components of dominant factors influencing suspended sediment yield were identified. These factors contribute to 86.7% of the total variance of all variables considered in the analysis. The dominant factors from each group were then taken as predictor variables in the successive multiple regression analysis to estimate suspended sediment yield and area-specific suspended sediment yield.

From the regression analysis, it was found that there are three factors that significantly affect suspended sediment yield. These factors are hierarchical anomaly density, basin area and forest area. On the other hand, there are five factors that significantly influence area-specific suspended sediment yield. These are basin slope, hierarchical anomaly density, main channel length, forest area and dry season rainfall. The regression models indicate better predictability of suspended sediment yield and area-specific sediment yield for basins with a drainage area of less than 1000 km^2.

A set of equations for predicting suspended sediment yield and area-specific suspended sediment yield for basin areas with different sizes within the error of estimation range was proposed. These equations may be used to estimate the expected sediment yield in ungauged basins in the planning and design of water and land development and conservation projects in the northern part of Thailand with easily determined dominant input variables. However, it should be noted that the error of estimation for suspended sediment is relatively high, which is partially due to uncertainties in the sediment sampling/measurement (especially during high discharges or flood events) and in developing the sediment-discharge rating curve equations. Additionally, these models were developed for the estimation of average annual suspended sediment in ungauged basins. Therefore, the application of the models for sediment yield on short time periods, such as event-based estimation, is not recommended, due to the hysteresis effect in sediment rating curves.

Since this is a data-driven approach, the availability of limited data may restrain its applicability. The proposed equations may further be tested in other basins, provided that there is adequate data with similar hydrological and geo-morphological conditions. It is to be noted here that the proposed models were developed under natural conditions, without a water regulating structure; hence, the use of models in other basins with infrastructure may not be warranted. In addition, the models were developed with a relatively short period of data; hence, it is suggested that the models may be updated with a long period of data and more stations, as well as for variable climate and land use conditions. This can be done by using variable data of land use (percent of forest area and agricultural area) and the climate characteristics (annual rainfall, dry and wet season rainfall, precipitation concentration index) and, then, re-generating a new set of regression models.

Acknowledgments

Author Contributions

Piyawat Wuttichaikitcharoen, being the doctoral student, conducted the research, including data collection, analysis and preparation of the manuscript. Mukand Singh Babel supervised the research, discussed the results and contributed to the finalization of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Grauso, S.; Fattoruso, G.; Crocetti, C.; Montanari, A. Estimating the suspended sediment yield in a river network by means of geomorphic parameters and regression relationships. *Hydrol. Earth Syst. Sci.* **2008**, *12*, 177–191.
2. Maidment, D.R. *Handbook of Hydrology*; McGraw-Hill, Inc.: New York, NY, USA, 1992.
3. Wani, S.P. Integrated Watershed Management for Sustaining Crop Productivity and Reducing Soil Erosion in Asia. In Proceedings of the 5th Management of Soil Erosion Consortium (MSEC) Assembly, Semarang, Central Java, Indonesia, 7–11 November 2001; Maglinao, A.R., Leslie, R.N., Eds.; International water Management Institute (IWMI): Semarang, Central Java, Indonesia.

4. Shi, Z.H.; Ai, L.; Fang, N.F.; Zhu, H.D. Modeling the impacts of integrated small watershed management on soil erosion and sediment delivery: A case study in the Three Gorges area, China. *J. Hydrol.* **2012**, *438–439*, 156–167.

5. Tetzlaff, B.; Friedrich, K.; Vorderbrügge, T.; Vereecken, H.; Wendland, F. Distributed modelling of mean annual soil erosion and sediment delivery rates to surface waters. *CATENA* **2013**, *102*, 13–20.

6. Kothyari, U.C. Sediment problems and sediment management in the indian sub-himalayan region. In *Sediment Problems and Sediment Management in Asian River Basins*; Walling, D.E., Ed.; IAHS Press: CEH Wallingford, Oxfordshire, UK, 2011; Volume 349.

7. Boonchee, S.; Sumalee, S.; Inthapan, P.; Rachadawong, S. *Management of Sloping Lands for Sustainable Agriculture*; Final Report of ASIALAND Network, Phase 4, International Water Management Institute: Bangkok, Thailand, 2002.

8. Walling, D.E.; Webb, B.W. Erosion and sediment yield: A global overview. In *Erosion and Sediment Yield: Global and Regional Perspectives*; Walling, D.E., Webb, B.W., Eds.; IAHS Press: Wallingford, UK, 1996; Volume 236, pp. 3–19.

9. White, S. Sediment yield prediction and modelling. *Hydrol. Process.* **2005**, *19*, 3053–3057.

10. Wischmeier, W.H.; Smith, D.D. *Predicting Rainfall Erosion Losses—A Guide to Conservation Planning*; USDA-SEA Agriculture Handbook No.537; US Department of Agriculture: Washington, DC, USA, 1978.

11. Walling, D.E. The sediment delivery problem. *J. Hydrol.* **1983**, *65*, 113–141.

12. Maner, S.B. Factors affecting sediment delivery rates in the Red Hills physiographic area. *Trans. Am. Geophys. Union* **1958**, *39*, 669–675.

13. Williams, J.R.; Berndt, H.D. Sediment yield computed with universal equation. *J. Hydraul. Div. ASCE* **1972**, *98*, 2087–2098.

14. Mou, J.; Meng, Q. *Sediment Delivery Ratio as Used in the Computation of Watershed Sediment Yield*; Chinese Society of Hydraulic Engineering: Beijing, China, 1980.

15. Van Oost, K.; Govers, G.; Desmet, P. Evaluating the effects of landscape structure on soil erosion by water and tillage. *Landsc. Ecol.* **2000**, *15*, 579–591.

16. Van Rompaey, A.; Krasa, J.; Dostal, T.; Govers, G. Modelling sediment supply to rivers and reservoirs in Eastern Europe during and after the collectivisation period. *Hydrobiologia* **2003**, *494*, 169–176.

17. Gurmessa, T.K.; Bárdossy, A. A principal component regression approach to simulate the bed-evolution of reservoirs. *J. Hydrol.* **2009**, *368*, 30–41.

18. Leh, M.; Bajwa, S.; Chaubey, I. Impact of land use change on erosion risk: An integrated remote sensing, geographic information system and modeling methodology. *Land Degrad. Dev.* **2013**, *24*, 409–421.

19. Cohen, S.; Kettner, A.J.; Syvitski, J.P.M.; Fekete, B.M. WBMsed, a distributed global-scale riverine sediment flux model: Model description and validation. *Comput. Geosci.* **2013**, *53*, 80–93.

20. Balthazar, V.; Vanacker, V.; Girma, A.; Poesen, J.; Golla, S. Human impact on sediment fluxes within the Blue Nile and Atbara river basins. *Geomorphology* **2013**, *180–181*, 231–241.

21. Grauso, S.; Pagano, A.; Fattoruso, G.; Bonis, P.D.; Onori, F.; Regina, P.; Tebano, C. Relations between climatic–geomorphological parameters and sediment yield in a mediterranean semi-arid area (Sicily, Southern Italy). *Environ. Geol.* **2008**, *54*, 219–234.

22. Syvitski, J.P.M.; Milliman, J.D. Geology, geography, and humans battle for dominance over the delivery of fluvial sediment to the coastal ocean. *J. Geol.* **2007**, *115*, 1–19.

23. Restrepo, J.D.; Kjerfve, B.; Herrnelin, M.; Restrepo, J.C. Factors controlling sediment yield in a major South American drainage basin: The magdalena river, colombia. *J. Geol.* **2006**, *316*, 213–232.

24. Liu, Q.Q.; Singh, V.P.; Xiang, H. Plot erosion model using gray relational analysis method. *J. Hydrol. Eng.* **2005**, *10*, 288–294.

25. Sharma, U.C.; Sharma, V. Mathematical model for predicting soil erosion by flowing water in ungauged watersheds. In *Erosion Prediciton in Ungauged Basins: Integrating Methods and Techniques*; Boer, D.D., Froehlich, W., Mizuyama, T., Pietroniro, A., Eds.; IAHS Press: Wallingford, UK, 2003; Volume 279, pp. 79–83.

26. Hovius, N. Controls on sediment supply by large rivers, relative role of eustasy, climate, and tectonism in continental rocks. *Relative Role of Eustasy, Climate, and Tectonism in Continental Rocks*; SEPM (Society for Sedimentary Geology): Tulsa, OK, USA, 1998; Volume 59, pp. 2–16.

27. Gurnell, A.; Hannah, D.; Lawler, D. Suspended sediment yield from glacier basins. In *Erosion and Sediment Yield: Global and Regional Perspectives*; Walling, D.E., Webb, B.W., Eds.; IAHS Press: Wallingford, UK, 1996; Volume 236, pp. 97–104.

28. Walling, D.E. Measuring sediment yield from river basin. In *Soil Erosion Research Methods*; Lal, R., Ed.; Soil and Water Conservation Society: Ankeny, IA, USA, 1994; pp. 39–80.

29. Bray, D.I.; Xie, H. A regression method for estimating suspended sediment yields for ungauged watersheds in Atlantic Canada. *Can. J. Civil Eng.* **1993**, *20*, 82–87.

30. Ciccacci, S.; Fredi, P.; Palmieri, E.L.; Pugliese, F. Indirect evaluation of erosion entity in drainage basins through geomorphic, climatic and hydrological parameters. In *International Geomorphology 1986 Part II*; Gardiner, V., Ed.; John Wiley & Son Ltd: Hoboken, NJ, USA, 1987.

31. Walling, D.E.; Webb, B.W. Patterns of sediment yield. In *Background to Hydrogeology*; Gregory, K.J., Ed.; Wiley: Chichester, UK, 1983; pp. 69–100.

32. Langbein, W.B.; Schumm, S.A. Yield of sediment in relation to mean annual precipitation. *Trans. Am. Geophys. Union* **1958**, *39*, 1076–1084.

33. Anderson, H.W. Relating sediment yield to watershed variables. *Trans. Am. Geophys. Union* **1957**, *38*, 921–924.

34. Brown, C.E. Use of principal-component, correlation, and stepwise multiple-regression analyses to investigate selected physical and hydraulic properties of carbonate-rock aquifers. *J. Hydrol.* **1993**, *147*, 169–195.

35. Pandzic, K.; Trninic, D. Principal component analysis of a river basin discharge and precipitation anomaly fields associated with the global circulation. *J. Hydrol.* **1992**, *132*, 343–360.

36. Hidalgo, H.G.; Piechota, T.C.; Dracup, J.A. Alternative principal components regression procedures for dendrohydrologic reconstructions. *Water Resour. Res.* **2000**, *36*, 3241–3249.

37. Bouvier, C.; Cisneros, L.; Dominguez, R.; Laborde, J.-P.; Lebel, T. Generating rainfall fields using principal components (pc) decomposition of the covariance matrix: A case study in mexico city. *J. Hydrol.* **2003**, *278*, 107–120.

38. Halim, R.; Clemente, R.S.; Routray, J.K.; Shrestha, R.P. Integration of biophysical and socio-economic factors to assess soil erosion hazard in the Upper Kaligarang watershed, Indonesia. *Land Degrad. Dev.* **2007**, *18*, 453–469.

39. Samani, N.; Gohari-Moghadam, M.; Safavi, A.A. A simple neural network model for the determination of aquifer parameters. *J. Hydrol.* **2007**, *340*, 1–11.

40. Al-Alawi, S.M.; Abdul-Wahab, S.A.; Bakheit, C.S. Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. *Environ. Model. Softw.* **2008**, *23*, 396–403.

41. Tayfur, G.; Karimi, Y.; Singh, V. Principle component analysis in conjuction with data driven methods for sediment load prediction. *Water Resour. Manag.* **2013**, *27*, 2541–2554.

42. Hyvärinen, A.; Oja, E. Independent component analysis: Algorithms and applications. *Neural Netw.* **2000**, *13*, 411–430.

43. Ikeda, S.; Toyama, K. Independent component analysis for noisy data—MEG data analysis. *Neural Netw.* **2000**, *13*, 1063–1074.

44. Westra, S.; Brown, C.; Lall, U.; Koch, I.; Sharma, A. Interpreting variability in global SST data using independent component analysis and principal component analysis. *Int. J. Climatol.* **2010**, *30*, 333–346.

45. *25 River Basins Report*; Royal Irrigation Department, Ministry of Agriculture and Cooperatives: Bangkok, Thailand, 2003. (In Thai)

46. Alford, D. Streamflow and sediment transport from mountain watersheds of the Chaophraya basin, Northern Thailand: A reconnaissance study. *Mt. Res. Dev.* **1992**, *12*, 257–268.

47. Tingting, L.V.; Xiaoyu, S.; Dandan, Z.; Zhenshan, X.; Jianming, G. *Assessment of Soil Erosion Risk in Northern Thailand*; Jun, C., Ed.; International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: Beijing, China, 2008.

48. Plangoen, P.; Babel, M.; Clemente, R.; Shrestha, S.; Tripathi, N. Simulating the impact of future land use and climate change on soil erosion and deposition in the Mae Nam Nan sub-catchment, Thailand. *Sustainability* **2013**, *5*, 3244–3274.

49. Wigmosta, M.S.; Vail, L.W.; Lettenmaier, D.P. A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.* **1994**, *30*, 1665–1679.

50. *HEC-GeoHMS Geospatial Hydrologic Modeling Extension User's Manual Version 10.1*; US Army Corps of Engineers, Hydrologic Engineering Center: Davis, CA, USA, 2013.

51. Department of Water Resources. *Standard Map of Main and Sub-River Basins Delineation of Thailand*; Sahamitr Printing & Publishing Co., Ltd.: Nonthaburi, Thailand, 2009. (In Thai)

52. Guarnieri, P.; Pirrotta, C. The response of drainage basins to the late quaternary tectonics in the Sicilian side of the Messina Strait (NE Sicily). *Geomorphology* **2008**, *95*, 260–273.

53. Della Seta, M.; del Monte, M.; Fredi, P.; Palmieri, E.L. Direct and indirect evaluation of denudation rates in Central Italy. *CATENA* **2007**, *71*, 21–30.

54. Carter, A.J.; Scholes, R.J. *Generating a Global Database of Soil Properties*; Council for Scientific and Industrial Research (CSIR) Environmentek: Pretoria, South Africa, 1999.

55. Batjes, N.H. *A Homogenized Soil Data File for Global Environmental Research: A Subset of FAO, ISRIC and NRCS Profiles (Version 1.0)*; Working paper and preprint 95/10b; International Soil Reference and Information Centre: Wageningen, The Netherlands, 1995.

56. *Digital Soil Map of the World, Version 3.5*; Food and Agriculture Organization of the United Nations, Rome, Italy, 1995.

57. Royal Irrigation Department. *Monthly and Annual Suspended Sediment in Main River Basins of Thailand*; Sediment and Water Quality Group, Hydrology Division, Royal Irrigation Department: Bangkok, Thailand, 2011. (In Thai)

58. De Luís, M.; Raventós, J.; González-Hidalgo, J.C.; Sánchez, J.R.; Cortina, J. Spatial analysis of rainfall trends in the region of Valencia (East Spain). *Int. J. Climatol.* **2000**, *20*, 1451–1469.

59. Petras, I. *Arcview Arealrain Extension*; Department of Water Affairs and Forestry: Pertoria, South Africa, 2001.

60. Landau, S.; Everitt, B.S. *A Handbook of Statistical Analyses Using SPSS*; CRC Press Company: London, UK, 2003.

61. Pallant, J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS*; Allen & Unwin: Crows Nest, NSW, Australia, 2005.

62. Haan, C.T. *Statistical Methods in Hydrology*, 2nd ed.; The Iowa State Press: Ames, IA, USA, 2002; p. 496.

63. Kaiser, H. An index of factorial simplicity. *Psychometrika* **1974**, *39*, 31–36.

64. Bartlett, M.S. A note on the multiplying factors for various chi square approximations. *J. R. Stat. Soc.* **1954**, *16*, 296–298.

65. Hair, J.F.; Black, B.; Babin, B.; Anderson, R.E.; Tatham, R.L. *Multivariate Data Analysis*, 6th ed.; Pearson Printice Hall: Upper Saddle River, NJ, USA, 2006.

66. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*; Harper Collins: New York, NY, USA, 2001.

67. Brejda, J.J.; Moorman, T.B.; Karlen, D.L.; Dao, T.H. Identification of regional soil quality factors and indicators: I. Central and southern high plains. *Soil Sci. Soc. Am. J.* **2000**, *64*, 2115–2124.

68. Andrews, S.S.; Mitchell, J.P.; Mancinelli, R.; Karlen, D.L.; Hartz, T.K.; Horwath, W.R.; Pettygrove, G.S.; Scow, K.M.; Munk, D.S. On-farm assessment of soil quality in california's central valley. *Agron. J.* **2002**, *94*, 12–23.

69. *National Engineering Handbook, Section 3: Sedimentation*, 2nd ed.; United States Department of Agriculture, Soil Conservation Service: Washington, DC, USA, 1983.

70. Noble, R.; Cowx, I. *Development of a River-Type Classification System (D1), Compilation and Harmonisation of Fish Species Classification (D2)*; Final Report of Development, Evaluation & Implementation of a Standardised Fish-Based Assessment Method for the Ecological Status of European Rivers—A Contribution to the Water Framework Directive (FAME); FAME Group, University of Hull: Hull, UK, 2002; p. 51. Available online: https://fame.boku.ac.at/downloads/ D1_2_typology_and%20species_classification.pdf (accessed on 10 August 2013).

71. Stevens, J. *Applied Multivariate Statistics for the Social Sciences*; Lawrence Erlbaum: Mahwah, NJ, USA, 1996.

72. Kazama, S.; Suzuki, K.; Sawamoto, M. Estimation of rating-curve parameters for sedimentation using a physical model. *Hydrol. Process.* **2005**, *19*, 3863–3871.