

Piecewise Regression Using Cubic Spline-A Case Study

Sarita Gajbhiye Meshram^{1,2*} and P.L. Powar²

¹*Department of Water Resources Development & Management
Indian Institute of Technology, Roorkee (Uttarakhand), India*

²*Department of Mathematics and Computer Science*

R.D. University, Jabalpur, M.P., India

**gajbhiyesarita@gmail.com*

Abstract

Watershed erosion reveals spatio-temporal variability and determines sediment yield which is a key to appropriate watershed management. Considering several Curve Numbers (SCS, 1956) as nodal point in the domain, in this paper, we have constructed the approximate curve using cubic splines corresponding to the Sediment Yield Index (SYI). The aim of this construction is to predict SYI corresponding to any arbitrary curve numbers lying in the domain of construction. It is observed that cubic spline show a satisfactory performance, with Nash efficiency of 61.36%. We have established the quality of approximation by comparing the given value with the approximate value of SYI at certain nodal point and noted that these values are quite closed and hence it is concluded that the approximation is satisfactory and may be quite useful in field applications.

Keywords: *Watershed, Runoff Curve Number, Sediment yield index (SYI), Cubic Spline*

1. Introduction

Soil erosion may be defined as the detachment and transportation of soil. Water is the major agent responsible for this erosion. In many locations winds, glaciers, *etc.* also cause soil erosion [1]. The watershed planning and management highlights the management techniques to control erosion in the catchment/watershed area [2,3]. Due to lack of data availability, estimates are made by means of prediction models that evaluate soil loss or sediment yield separately [4]. AISLUS proposed an empirical relationship between SYI and area & delivery ratio. A different approach proposed an empirical relationship between delivery ratio and morphological characteristics of catchment such as catchment area, average relief or slope [5,6]. These models are popularly employed because of their simplicity and easily available data. It is of common experience that the sediment yield largely depends on the rainfall-generated runoff, which, in turn, relies on the parameter Curve number of the Soil Conservation Service Curve Number (SCS-CN) methodology. The larger the CN value the larger will be the sediment yield in agricultural watersheds, and vice versa. Since CN represents the runoff producing potential of a watershed and SYI the potential of sediment yield, it is quite logical that these two parameters, viz., SYI and CN, should exhibit some relationship between them [7,8].

A multitude of studies have applied cubic splines at regional and national levels [9, 10, 11, 12]. Spline applications are increasing day by day in the fields of medicine, agriculture, engineering and other sciences; they are limited in soil erosion modelling. Therefore, the objective of this paper is to apply cubic splines in erosion modelling, as there have proved to be the best approximation. They are used to obtain the value of approximate curve number corresponding to an arbitrary sediment yield index.

2. Study Area and Data Source

Narmada is the largest west-flowing river of the Indian peninsula. It is one of the important rivers of India. The Narmada River rises in the Amarkantak Plateau of Maikala range in the Shahdol district of Madhya Pradesh at an elevation of 1,057 meters above mean sea level at a latitude $22^{\circ}40'N$, longitude $81^{\circ}45'E$. Bamhani watershed (2542 km²), Manot watershed (4884 km²), Mohgaon watershed (3978 km²) in Mandla district Madhya Pradesh, India. The three adjacent watersheds namely Bamhani, Manot and Mohgaon (Figure 1) conjoin together to form an important southern sub-basin designated as Combine watershed (Gajbhiye *et al.* 2014c).

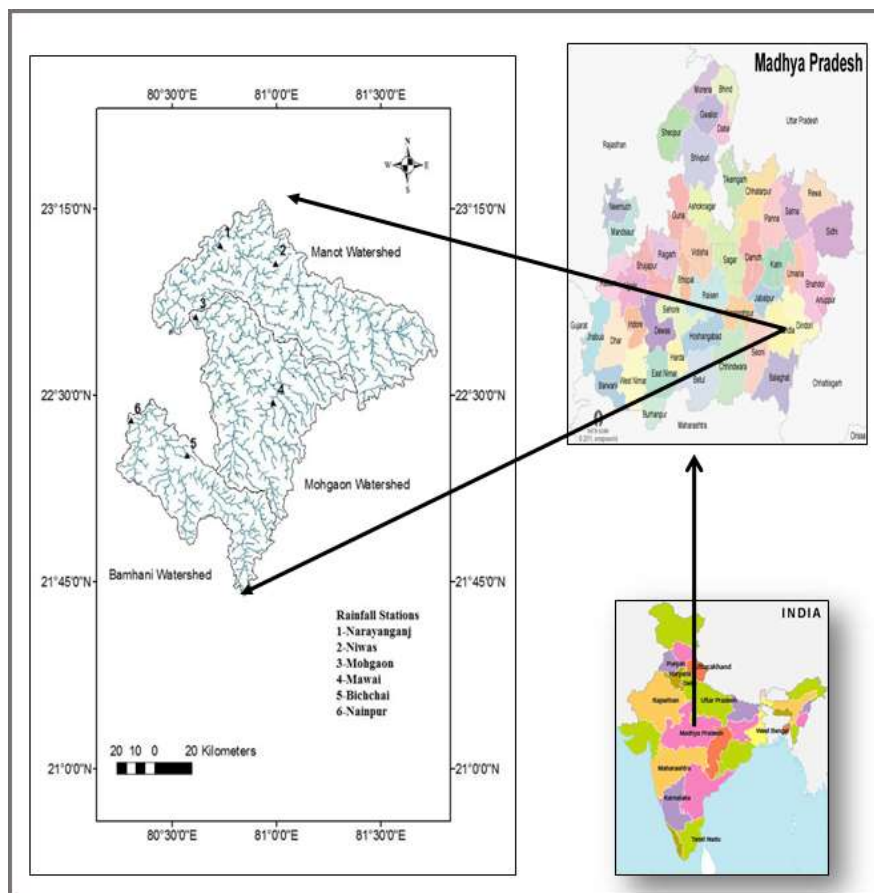


Figure 1. Location Map of the Study Area

The topographic map (scale of 1:50,000) was used for watershed delineation. The Landsat satellite data downloaded from Global Land Cover Facility (GLCF), Maryland website, were used to prepare the land use land cover map. The National Bureau of Soil Survey, Nagpur has prepared a soil map and Shuttle Radar Topography Mission (SRTM) data was used to prepare a slope map, which was also downloaded from the GLCF website.

3. Materials and Methods

3.1. Estimation of Runoff Curve Number

The SCS-CN method is based on the water balance Eq. (1) and two other Eqs. (2) and (3):

$$P = Ia + F + Q \quad (1)$$

$$\frac{Q}{P - Ia} = \frac{F}{S} \quad (2)$$

$$Ia = \lambda S \quad (3)$$

where, Ia is the initial abstraction; F is the cumulative retention; λ is the initial abstraction ratio (0.2); S is the maximum retention capacity (mm); P is the rainfall (mm), Q is the direct runoff (mm). For convenience in practical applications, S is defined in terms of a dimensionless parameter, CN , which varies in a more restricted range 0-100.

$$S = \frac{25400}{CN} - 254 \quad (4)$$

CN = the curve number (0-100), If $CN = 100$, that is an impermeable watershed. Conversely, $CN = 0$ corresponds to an infinitely abstracting watershed.

3.1.1. Land Use/Land Cover (LULC)

The supervised classification has been used in this study for classification of satellite data. The land use/cover classes were agricultural land, fallow/barren/open land, water body and forest. Areas occupied by each land use were determined and presented in Figure 2. Narmada watersheds have a good forest cover and agricultural land (Agricultural and Barren land). Maximum 65.45% area is covered with agriculture in Manot watershed and 59.03% forest in Bamhani watershed.

3.1.2. Soil Map

Soil map of the study area is generated with the aid of NBSSLUP map. The areas under different soils were identified and are presented in Figure 3. After analysis it was found that the area of all watersheds fall under the Hydrologic soil group 'B' (Loamy soil) and Hydrologic Soil Group 'D' (Clay soil).

To create a CN map (Figure 4), the hydrologic soil group and land use maps were uploaded to the Arc View platform. The hydrologic soil group field from the soil theme and the land use field from the land use map were selected for intersection. The weighted CN values for the study watersheds were computed using the formula:

$$CN_{LU} = \frac{\sum(CN_i \cdot A_i)}{A} \quad (5)$$

where, CN_{LU} = weighted curve number (dimensionless), CN_i = curve number of area i th assigned polygon, A_i = the area having CN_i (km²), A = The total area of watershed (km²)

3.2. Soil Erosion Estimation Using Sediment Yield Index Method

3.2.1. Slope Map

Digital Elevation Models (DEM) were generated from the SRTM data. From these maps, slope maps for the Narmada watersheds were generated. An investigation of reveals that about 95.10% (Bamhani), 96.88 % (Manot), 96.99% (Mohgaon) of the catchment area of the watershed fall under two slope categories: Gently and Moderately Sloping.

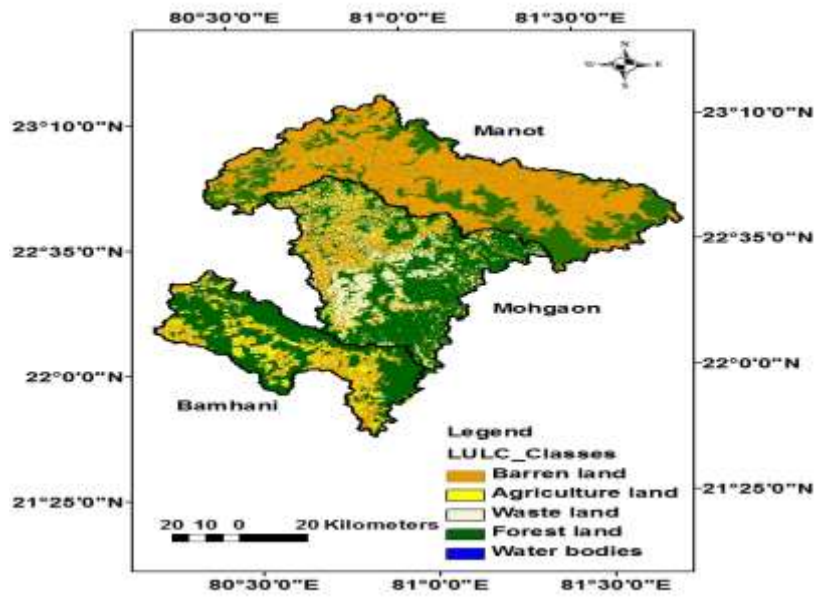


Figure 2. LULC Map

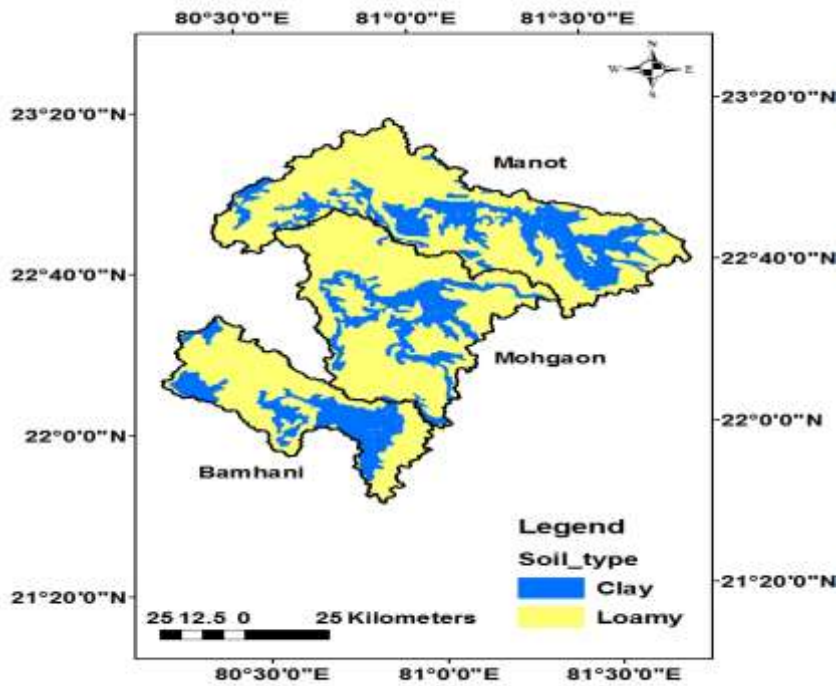


Figure 3. Soil Map

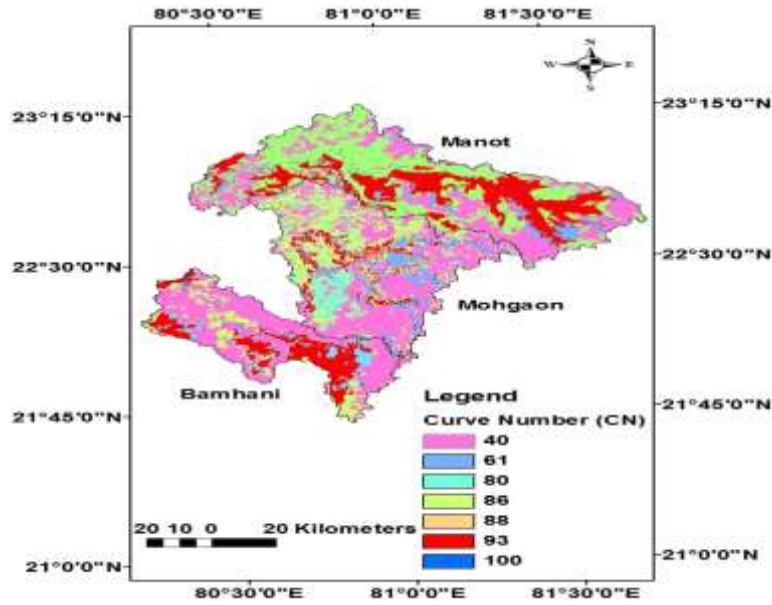


Figure 4. Spatial CN Map

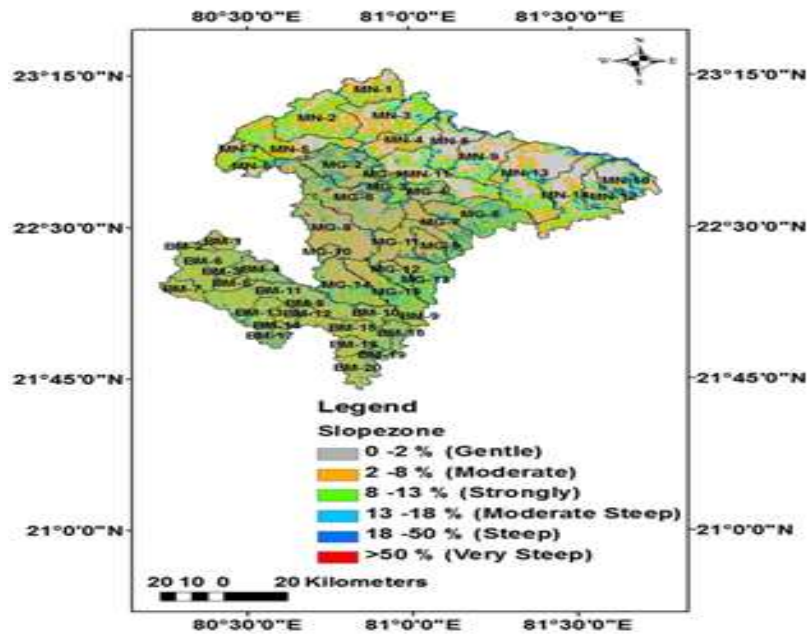


Figure 5. CIMU Map

3.2.2. Composite Intensity Mapping Unit (CIMU)

A Composite Intensity Mapping Unit (CIMU) map was prepared using the thematic map of slope, soil, and land Use (Figure 5). This map was then superimposed on the sub-watershed boundaries to obtain sub-watershed-wise CIMU maps.

3.2.3. Derivation of Sediment Yield Index (SYI)

To derive Sediment Yield Index (SYI), the methodology developed by AISLUS was followed, in which each erosion intensity unit is assigned a weightage value.

Sediment Yield Index (SYI) was calculated using the following empirical formula:

$$SYI = \frac{\sum_{i=1}^n (A_i * W_i * DR)}{A_w} * 100 \quad (6)$$

where, A_i = the area of the i th unit (CIMU) in km^2 , W_i = the weightage value of the i th unit (dimensionless), n = the number of mapping units, DR = delivery ratio (dimensionless) and A_w = the sub-watershed area in km^2 .

For developing a relationship between SYI and CN, the proposed procedure is outlined as follows: (1). Estimation of sub-watershed wise sediment yield index (SYI) and curve number (CN), (2). CN of different sub watersheds was arranged in increasing order of magnitude of CN, which will give the interval of CN for spline construction, (3). As in our study forty nine sub watersheds are there so eleven (BM 1, 7, 10, 13 14, 16, 19, MG 10, 15, MN 9, 12) were selected for construction of spline and the remaining were used for validation of the spline, (4). From the values of CN and spline coefficient compute the sediment yield index which corresponds to the interval of CN. Assume these are the computed SYI-values, (5). The computed SYI (SYI_c) was compared with the observed SYI (SYI_o) which was derived from the AISLUS method.

3.3. Cubic Splines (De Boor, 1978)

Consider a closed interval $[a, b]$ on a real line and define a partition of it as:

$$a = z_1 < z_2 < \dots < z_n = b$$

The z_i 's are called knots or the nodal points and at each z_i , the data $\delta(z_i)$ $i = 1, 2, \dots, n$ is given. Our aim is to construct a piecewise cubic interpolant f to δ . On each sub-interval, we define a piecewise polynomial $d_i(x)$ such that

$$f(z) = d_i(z) \quad \text{for some } d_i(z) \in \mathbb{P}_3 \quad (i = 1, 2, \dots, n - 1).$$

The following conditions have been imposed on the i^{th} polynomial piece d_i :

$$\left. \begin{aligned} d_i(x_i) &= \delta(z_i), & d_i(z_{i+1}) &= \delta(z_{i+1}) \\ d_i'(z_i) &= s_i, & d_i'(z_{i+1}) &= s_{i+1} \end{aligned} \right\} \quad (7)$$

$$i = 1, 2, \dots, n - 1$$

Here, s_1, \dots, s_n are free parameters. The approximate function f agrees with δ at z_1, \dots, z_n and it is continuous up-to the first order on $[a, b]$.

For calculating the coefficients of the i^{th} polynomial piece d_i , we used its Newton form:

$$d_i(z) = d_i(z_i) + (z - z_i)[z_i, z_i]d_i + (z - z_i)^2 [z_i, z_i, z_{i+1}]d_i + (z - z_i)^3 (z - z_{i+1})[z_i, z_i, z_{i+1}, z_{i+1}]d_i \quad (8)$$

Its coefficients were determined from the divided difference for d_i based on the input data:

This shows that, in terms of shifted powers $(z - z_i)^r$,

$$d_i(z) = c_{1,i} + c_{2,i}(z - z_i) + c_{3,i}(z - z_i)^2 + c_{4,i}(z - z_i)^3 \quad (9)$$

with, $c_{1,i} = d_i(z_i) = \delta(z_i)$, $c_{2,i} = d_i'(z_i) = s_i$,

$$c_{3,i} = \frac{d_i''(z_i)}{2} = [z_i, z_i, z_{i+1}]d_i - \Delta z_i([z_i, z_i, z_{i+1}, z_{i+1}]d_i) \\ = ([z_i, z_{i+1}]\delta - s_i) / \Delta z_i - c_{4,i} \Delta z_i$$

$$c_{4,i} = d_i'''(z_i)/6 = (s_{i+1} + s_i - 2[z_i, z_{i+1}]\delta) / (\Delta z_i)^2$$

Parameters s_2, s_3, \dots, s_{n-1} are determined from the condition that the approximation f should be twice continuously differentiable which is assured by the condition

$$d_{i-1}''(z_i) = d_i''(z_i), \quad i = 2, 3, \dots, n - 1 \quad (10)$$

$$\text{Or } 2 c_{3,i-1} + 6 c_{4,i-1} \Delta z_{i-1} = 2 c_{3,i} \quad (11)$$

$$\text{or } 2 \{([z_{i-1}, z_i]\delta - s_{i-1}) / \Delta z_{i-1} - c_{4,i-1} \Delta z_{i-1}\} + 6 c_{4,i-1} \Delta z_{i-1} \\ = 2 \{([z_i, z_{i+1}]\delta - s_i) / \Delta z_i - c_{4,i} \Delta z_i\} \quad (12)$$

$$\text{or } 2 ([z_{i-1}, z_i]\delta - s_{i-1}) / \Delta z_{i-1} + 4 c_{4,i-1} \Delta z_{i-1} = 2 ([z_i, z_{i+1}]\delta - s_i) / \Delta z_i - 2 c_{4,i} \Delta z_i \\ \text{or } 2 ([z_{i-1}, z_i]\delta - s_{i-1}) / \Delta z_{i-1} + 4 (s_i + s_{i-1} - 2[z_{i-1}, z_i]\delta) / \Delta z_{i-1} = 2 ([z_i, z_{i+1}]\delta - s_i) / \Delta z_i - 2 (s_{i+1} + s_i - 2[z_i, z_{i+1}]\delta) / \Delta z_i \quad (13)$$

$$\text{or } s_{i-1} \Delta z_i + s_i 2(\Delta z_{i-1} + \Delta z_i) + s_{i+1} \Delta z_{i-1} = b_i \quad (14)$$

$$\text{with } b_i = 3 (\Delta z_i [z_{i-1}, z_i]\delta + \Delta z_{i-1} [z_i, z_{i+1}]\delta), \quad i = 2, \dots, n - 1 \quad (15)$$

The remaining two free parameters s_1 and s_n are chosen such that $s_1 = 0$ and $s_n = 0$.

4. Construction of Cubic Splines

The SYI model, based on the spline was calibrated using the data of combine watershed (Table 1). The sub-watersheds (BM 1, 7, 10, 13 14, 16, 19, MG 10, 15, MN 9, 12) were used for spline construction and the remaining watersheds used for validation of the spline approximation.

Table 1. The Combine Watershed Data Set Utilized In the Present Study for Cubic Spline

watersheds	Curve Number	Sediment Yield Index	watersheds	Curve Number	Sediment Yield Index
MG 1	61.65	1252.84	MN 11	76.71	1347.6
MG 2	68.04	973.76	MN 12	67.84	1238.36
MG 3	56.65	1109.84	MN 13	73.76	1007.85
MG 4	65.64	965.63	MN 14	74.15	925.38
MG 5	73.51	943.09	BM 1	36.41	1070.47
MG 6	55.28	1054.65	BM 2	76.97	1159.31
MG 7	62.48	1322.99	BM 3	59.63	1107.11
MG 8	74.96	938.59	BM 4	43.95	1279.04
MG 9	56.17	1259.42	BM 5	63.96	1094.39
MG 10	78.28	1148.61	BM 6	54.08	914.97
MG 11	62.68	846.58	BM 7	86.73	796.42
MG 12	59.38	1337.36	BM 8	79.57	969.15
MG 13	54.49	956.43	BM 9	58.11	1062.07
MG 14	61.47	915.91	BM 10	61.22	1013.29
MG 15	54.35	1052.09	BM 11	58.09	923.82
MN 1	73.87	991.61	BM 12	69.97	971.02
MN 2	77.73	987.42	BM 13	64.21	1154.44
MN 3	72.03	986.82	BM 14	72.75	1177.68
MN 4	84.95	1032.24	BM 15	77.10	1032.71
MN 5	79.51	1173.33	BM 16	45.54	1115.03

MN 6	78.23	1088.42	BM 17	56.00	1264.42
MN 7	64.75	1031.08	BM 18	78.15	840.95
MN 8	59.36	1099.92	BM 19	57.55	1054.24
MN 9	76.71	940.97	BM 20	73.61	901.41
MN 10	66.42	1030.19			

We constructed the cubic splines as discussed in Section 3.3 for the data of combine watershed given in Table 1. Consider the curve number values as nodal points and sediment yield index values as data points. The piecewise cubic polynomials are, as shown in Table 2, Figure 6.

5. Validation of Approximate Sediment Yield Index (SYI)

For validation, the curve number values and the above coefficient (Table 2) were used in spline equation for the estimation of SYI. This computed SYI termed as SYIC was compared with the observed SYI (SYIo) which was derived from the AISLUS method. The computed SYI and observed SYI were compared through a line of perfect fit (Figure 7). The correlation between observed and computed SYI values exhibited a strong relationship with the coefficient of determination (R²) value of 0.62.

Table 2. Values of the Coefficient and Constructed Cubic Spline

CN	Coefficient				Cubic Splines
	C ₁	C ₂	C ₃	C ₄	
36.41-45.54	1070.47	0.00	1.89	-0.15	$p_1(x) = 1070.47 + 1.89(x - 36.41)^2 - 0.15(x - 36.41)^3$
45.54-54.35	1115.03	-2.65	-2.18	0.19	$p_2(x) = 1115.03 - 2.65(x - 45.54) - 2.18(x - 45.54)^2 + 0.19(x - 45.54)^3$
54.35-57.55	1052.09	3.10	2.84	-1.12	$p_3(x) = 1052.09 + 3.10(x - 54.35) + 2.84(x - 54.35)^2 - 1.12(x - 54.35)^3$
57.55-61.22	1054.24	-13.26	-7.95	2.32	$p_4(x) = 1054.24 - 13.26(x - 57.55) - 7.95(x - 57.55)^2 + 2.32(x - 57.55)^3$
61.22-64.21	1013.29	22.23	17.62	-3.10	$p_5(x) = 1013.29 + 22.23(x - 61.22) + 17.62(x - 61.22)^2 - 3.10(x - 61.22)^3$
64.21-67.84	1154.44	44.46	-10.19	1.19	$p_6(x) = 1154.44 + 44.46(x - 64.21) - 10.19(x - 64.21)^2 + 1.19(x - 64.21)^3$
67.84-72.75	1238.36	17.42	2.74	-1.79	$p_7(x) = 1238.36 + 17.42(x - 67.84) + 2.74(x - 67.84)^2 - 1.79(x - 67.84)^3$
72.75-76.71	1177.68	-85.38	-23.68	7.61	$p_8(x) = 1177.68 - 85.38(x - 72.75) - 23.68(x - 72.75)^2 + 7.61(x - 72.75)^3$
76.71-78.28	940.97	85.19	66.75	-23.4	$p_9(x) = 940.97 + 85.19(x - 76.71) + 66.75(x - 76.71)^2 - 23.42(x - 76.71)^3$
78.28-86.73	1148.61	121.58	-43.57	2.87	$p_{10}(x) = 1148.61 + 121.58(x - 78.28) - 43.57(x - 78.28)^2 + 2.87(x - 78.28)^3$

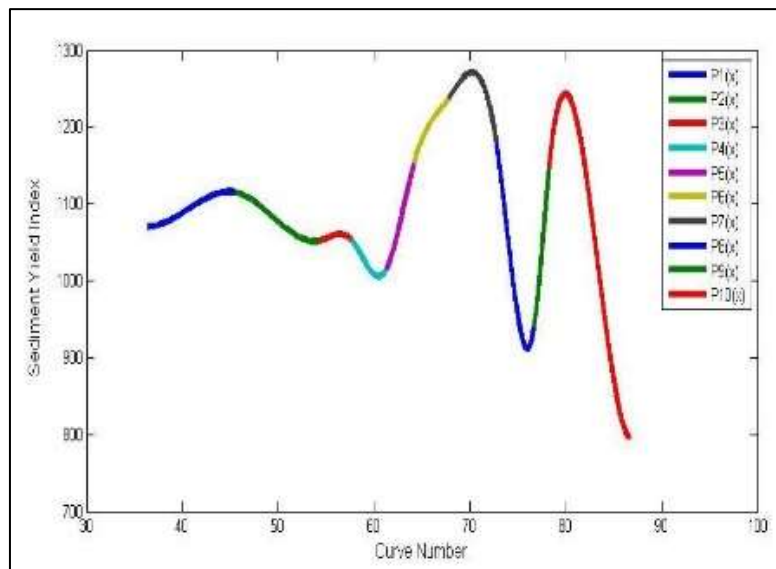


Figure 6. Cubic Splines Approximation of Given Dataset

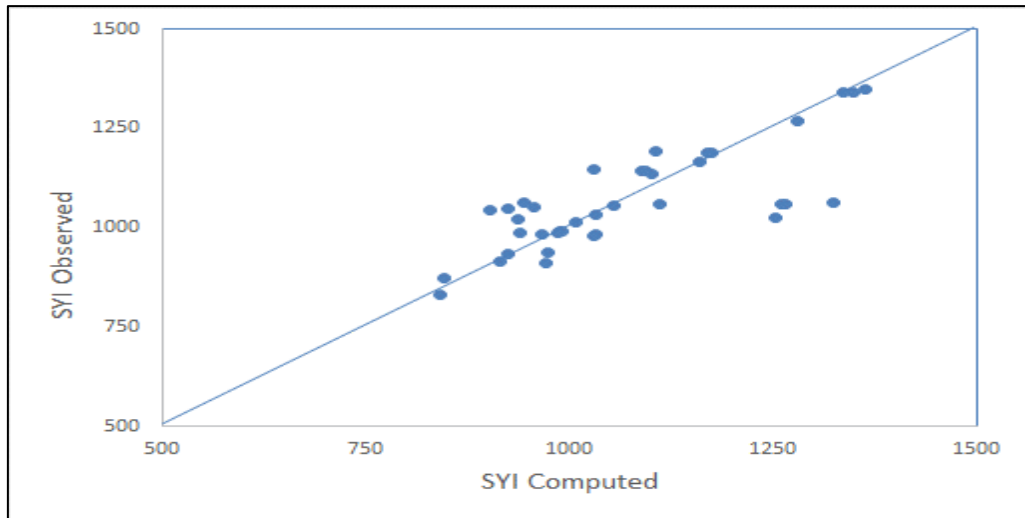


Figure 7. Scatter Plot between Predicted and Actual SYI

6. Conclusion

A cubic spline is developed between SYI and CN using eleven subwatershed data. This approximation is validated for sediment yield index for the remaining subwatersheds. Subsequently, the SYI (observed) data of combine watershed exhibit a strong correlation with SYI derived using the cubic spline. High R² values (0.62) support the general applicability of the proposed concept.

7. Advantages and Limitations

It has been noticed and established theoretically that the cubic spline are best approximants for most of the practical purposes. In this paper, the authors have verified that after computing the cubic approximation, at some particular nodal points, the actual values of SYI are very close to the approximation values. Hence, we conclude that the approximation is acceptable and the major advantage of this construction lies in getting an appropriate value of SYI at every arbitrary nodal point of the domain of curve numbers. In case when there is a more fluctuation in the data points, then we may not be able to compute appropriate cubic spline approximation.

Acknowledgments

The First author is thankful to UGC-New Delhi for providing financial support under the scheme of Dr. D.S. Kothari Postdoctoral Fellowship (DSKPDF).

References

- [1] S.Gajbhiye and S.K. Sharma, "Prioritization of Watershed through Morphometric Parameters: A PCA Based Approach", *Applied Water Science*, DOI: 10.1007/s13201-015-0332-9, (2015), pp. 1-15.
- [2] S. Gajbhiye, S.K. Sharma, S. Tignath and S.K. Mishra, "Development of a Geomorphological Erosion Index for Shakkar Watershed", *Geological Society of India*, vol. 86, no. 3, (2015a), pp. 361-370.
- [3] S. Gajbhiye, S.K. Mishra and A. Pandey, "Simplified sediment yield index model incorporating parameter CN. *Arabian Journal of Geosciences*, 8(4), pp 1993-2004 (2015b).
- [4] D. Nazzareno and G. Sergio, "An improved correlation model for sediment delivery ratio assessment", *Environment Earth Science*, vol. 59, no. 1, (2009), pp. 223-231.
- [5] S. Gajbhiye, S.K. Mishra and A. Pandey, "Hypsometric Analysis of Shakkar River Catchment through Geographical Information System", *Journal of the Geological Society of India*, vol. 84, no. 2, (2014b), pp. 192-196.

- [6] S. Gajbhiye, S.K. Sharma and C. Meshram, "Prioritization of Watershed through Sediment Yield Index Using RS and GIS Approach", *International Journal of u-and e-Service, Science and Technology*, vol. 7, no. 6, **(2014d)**, pp. 47-60.
- [7] S. Gajbhiye, S.K. Mishra and A. Pandey, "Prioritizing erosion-prone area through morphometric analysis: an RS and GIS perspective", *Applied Water Science*, vol. 4, no. 1, **(2014a)**, pp. 51-61.
- [8] S. Gajbhiye, S.K. Mishra and A. Pandey, "Relationship between SCS-CN and Sediment Yield", *Applied Water Science*, vol. 4, no. 4, **(2014c)**, pp. 363-370.
- [9] J.M. Chen, F. Deng and M. Chen, "Locally Adjusted Cubic-Spline Capping for Reconstructing Seasonal Trajectories of a Satellite-Derived Surface Parameter", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, **(2006)**, pp. 2230-2238.
- [10] A.S.V. Ravi Kant and Y.N. Reddy, "Cubic spline for a class of singular two-point boundary value problems", *Applied Mathematics and Computation*, vol. 170, no. 2, **(2005)**, pp. 733-740.
- [11] L. Desquilbet and F. Mariotti, "Dose-response analyses using restricted cubic spline functions in public health research", *Statistics in Medicine*, vol. 29, no. 9, **(2010)**, pp. 1037-1057.
- [12] M.S. Floater, "On the deviation of a parametric cubic spline interpolant from its data polygon", *Computer aided geometric design*, vol. 25, no. 3, **(2008)**, pp. 148-156.