An improved method for evaluation of sediment loads for river management

PIERO DE BONIS, GRAZIA FATTORUSO, SERGIO GRAUSO, FILIPPO ONORI & ANDREA PAGANO

ENEA (National Agency for New Technology, Energy and Environment), Centro Ricerche Portici, via del Vecchio Macello, I-80055 Portici (Napoli), Italy

e-mail: debonis@portici.enea.it

Abstract When sediment production and river transport data are not available for planning purposes, simple and reliable evaluation models are necessary. A geomorphic model, based on the parameterization of drainage basins with the geomorphic quantitative analysis theory is proposed in the literature to estimate suspended sediment transport. This method uses a linear regression between geomorphic parameters and river sediment loads. Even though in many cases this approach provides reliable estimates, differences still occur between measured and estimated values. This paper presents the preliminary results of a study that applies a multivariate analysis technique (cluster analysis) to increase the reliability of the method. In addition we show how geographical information system (GIS) technology can support the analysis and the model application by obtaining the digital fluvial network directly from a digital elevation model and computing geomorphic parameters.

Key words soil erosion; fluvial suspended sediment transport; geomorphic modelling; multivariate analysis; Sicily

INTRODUCTION

Understanding physical processes, like soil erosion and fluvial sediment transport, is important in the management of river systems. This is because of the consequences these processes may have on human activities within the drainage basin and, at a different temporal and spatial scale, along the coastal zone influenced by the river. Suitable tools are therefore required to enable the risk associated with the processes of erosion and sediment transport to be predicted. In terms of land planning and management it is important to provide measures, or at least good estimates, of various environmental parameters such as river and basin morphology, land use, river sediment transport and discharge. If this sort of information is missing, one should be able to make up with reliable and easy-to-employ estimation models.

For many years, researchers have tried to address this issue using different approaches: physically-based hydraulic models (i.e. WEPP, Eurosem, SHE, Medrush and others) and agronomic (Universal Soil Loss Equation) or geomorphological empirical formulae have been set up to predict the transport and/or production rate of sediments. It is important to point out that going from single slope scale to drainage basin scale, different factors are responsible for erosion processes. Therefore, the approach and the more suitable estimation methodology must vary according to the case being studied (Kirkby & Cox, 1995).

Sediment production and fluvial sediment transport are two different processes, but they are closely related. In particular fluvial sediment load can approximately be considered as an *index* of the potential erosion rate on slopes. Consequently, relationships have been developed by some authors (e.g. Ciccacci *et al*., 1980, 1987; Cooke & Doornkamp, 1974) between the suspended load and the geomorphic parameters of a drainage basin defined by Horton (1945) and Strahler (1957). The formulae used in quantitative geomorphology theory give only a rough estimate of sediment transport and erosion and certainly they are not suitable at a detailed scale. Nonetheless these formulae are of some use at the basin scale and thus for land-management needs. Moreover, they are reliable and easy to employ. The aim of the present work is to make this methodology more reliable and easier to use.

To reduce the differences still existing in many cases between calculated and measured sediment load data, an attempt was made to establish the best correlation between geomorphic parameters and riverine sediment transport using a statistical analysis technique, which included both multiple linear regression and cluster analysis. To obtain the geometric and orographic parameters (and the associated geomorphic parameters) the drainage network was analysed using an automatic GIS procedure.

MATERIALS AND METHODS

Data sets

The present study consists of two parts, each relating to a different set of data. The first data set is the geomorphic parameters calculated by Ciccacci *et al.* (1987) for 20 Italian basins. These data were chosen to compare the results and verify the effectiveness of the statistical technique employed. The second data set is the geomorphic parameters obtained from a sample of drainage basins used to test both the statistical approach and the automatic network extraction technique and analysis. Channel networks were delineated for 17 basins with available sediment load (Tu) records located in Sicily (the island region of southern Italy chosen as a test-area). A grid digital elevation model (DEM) by the Italian Army Geographical Institute (which supplies cartography with national coverage) was used, with pixel size of 20 m × 20 m. Channel network extraction and basin delineation were performed by using ESRI-ArcView GIS 3.2a and extensions.

Multiple regression formulae

Multiple regression formulae were obtained starting with the four independent variables among the geomorphic parameters which were most strongly correlated with the measured sediment load of the drainage basins. A *t*-test was then used to eliminate those variables which are not necessary to estimate the dependent variable.

Cluster analysis

Hierarchical cluster analysis is a multivariate procedure for detecting natural groupings in data (Milligan, 1980; Todeschini, 1997; Wilkinson *et al.*, 2000). When a data set with cases (objects) and variables is given, hierarchical cluster analysis joins the two closest cases as a cluster and continues, in a stepwise manner, joining an object with another object, an object with a cluster, or a cluster with another cluster until all objects are combined into one cluster. Output from hierarchical clustering is usually represented as a dendrogram (Sokal & Sneath, 1963). The aim of using such a statistical approach is to identify those basins which show a better correlation between some geomorphic parameters (or groups of parameters) and suspended load. This allows one to determine (and therefore to apply) different *families* of regression formulae depending on the result of the clustering.

River network delineation using GIS

GIS-based methods are being increasingly used to delineate channels (flow paths) and basins, and automatically extract these parameters for use in hydrological models. This can be done using DEMs derived from elevation data sets obtained from maps or images (air-photo or satellite stereo-pairs). The standard method for working with grid DEMs in hydrology involves pits filling, computation of flow directions and support areas, and channel network extraction and basin delineation (O'Callaghan & Mark, 1984; Jenson & Domingue, 1988; Mark, 1988). A widely used method for specifying flow directions, designated D8 (eight directions method) is to assign flow from each grid cell to one of its eight neighbours, either adjacent or diagonally, in the direction of steepest downward slope (O'Callaghan & Mark, 1984). Once flow directions are assigned, the basin for a given outlet is delineated and support areas contributing to each grid cell are estimated by counting the number of grid cells that drain through each grid cell. For delineating drainage networks a *support area threshold* applied to the grid of drainage area is used. Channels and start points are mapped as those grid cells where the support threshold is exceeded.

Since our objective was to produce data for use in geomorphic analysis, the constant drop property (Tarboton *et al.*, 1991; 1992) was initially followed to set up the right support area threshold. Then the following geomorphic parameters were calculated for use in the statistical analysis, by implementing *ad hoc* algorithms and computing procedures written in Avenue language:

(a) Rbar: arithmetic average of bifurcation ratio,
(b) Rbdar: arithmetic average of direct bifurcation ratio,
(c) Rbpon: weighted average of bifurcation ratio,
(d) Rbdpon: weighted average of direct bifurcation ratio,
(e) Rpon: weighted index of bifurcation,
(f) Re: elongation ratio,
(g) Rc: circularity ratio,
(h) Rh: height ratio,
(i) Ga: hierarchical anomaly density,
(j) Da: hierarchical anomaly index,
(k) Dd: Drainage density,

(l) Sup: Surface area,
(m) Gp: average slope,

It was decided to set the support area threshold to 600 for all basins since such a value provides a better match between the river networks obtained from the DEM and those depicted in official cartography.

RESULTS AND DISCUSSION

The idea of using cluster analysis, to sort out basins which may be similar with respect to some of the geomorphic parameters, has been successfully employed to improve the sediment load estimates given by Ciccacci *et al.* (1987). The authors' 20 basins can be divided into two groups by cluster analysis (distance: Euclidean; linkage: average):

– cluster 1 containing basins 2, 4, 5, 6, 7, 17, 18;
– cluster 2 containing basins 9, 10, 11, 12, 13, 14, 15, 16, 19, 20.

Three basins (1, 3 and 8) were left out since they did not adequately fit in the cluster analysis (this is not a problem, since we intend to proceed increasing the number of samples in future work, so that even those basins would fit into some clusters). New linear regressions between measured Tu data and geomorphic parameters were then performed within the two groups. Cluster 1 basins were analysed using a simple regression containing only the Dd parameter, while cluster 2 basins were analysed using a multiple regression with Dd and the Fournier's index (*F*). For comments about the significance of the simple and multiple regressions, see the quoted work of Ciccacci *et al.* (1987). In Table 1, the percentage errors between the measured Tu and the values obtained using the regression formulae are shown.

It is important to point out that more accurate results may be obtained with a larger number of samples, but even with the limited number of basins and geomorphic parameters considered, the average percentage error between modelled and measured sediment load improved. In fact, it has changed from an average error value of about 14% (Ciccacci *et al.* (1987) to 11% considering all 20 basins. A more accurate result was also obtained considering the two clusters separately, producing average errors of about 9.5% and 12% for clusters 1 and 2 respectively. It is also important to point out the result obtained from the test cluster, shown in the last row of Table 1, which demonstrates that the cluster technique is independent of the small sample size considered. In fact, the average error does not improve when a further limited sample of six basins (randomly chosen) were considered. The actual regression formulae used to estimate the suspended load are not reported here since they strongly depend on the cluster. What this study demonstrates is a general method, not a formula which can be applied in any situation.

In the second part of this study, the data from the test-area basins were standardized to avoid problems resulting from different scales. Then, the distance between objects was computed using Pearson's distance and the linkage was performed using complete linkage. Such analysis divided the 17 basins into two clusters:

– cluster 1 containing basins 1, 2, 5, 6, 8, 12, 15;
– cluster 2 containing basins 3, 4, 7, 9, 10, 11, 13, 14, 16, 17.

Table 1 Percentage error and average error for basins in Cluster 1, Cluster 2, test Cluster.

Basin	2	3	4	5	6	7	9	10	11	12	13	14	15	16	17	18	19	20	Average
Cluster 1 with respect to Dd	7.9		3.13	16.94	6.59	19.76									10.86	1.31			9.48
Cluster 2 with respect to Dd and F							1.05	28.22	39.3	0.04	15.49	1.83	5.14	5.59			17.53	6.66	12.08
Cluster 1 and 2 combined	7.9		3.13	16.94	6.59	19.76	1.05	28.22	39.3	0.04	15.49	1.83	5.14	5.59	10.86	1.31	17.53	6.66	11.01
Test Cluster with respect to Dd	3.45	0.09				12.04		38	31.59				13.42						16.43

Table 2 Percentage error and average error for multiple regression relative to all basins with respect to RBDPON, SUP and Dummy.

Basin	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Average
Error	114.65	2.07	27.22	42.38	27.45	38.28	37.02	37.62	8.61	30.65	56.66	3.5	78.01	40.95	20	91.23	38.27	40.86

Table 3 Percentage error and average error for basins in Cluster 1, Cluster 2 with respect to RBDPON and SUP.

Basin	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Average
Cluster 1	17.23	6.75			35	17.55		4.18				31.2			9.09			17.29
Cluster 2			23.82	30.69			32.19		25.37	34.16	60.36		70.4	22.93		102.55	23.76	42.62
Cluster 1 and Cluster 2 combined	17.23	6.75	23.82	30.69	35	17.55	32.19	4.18	25.37	34.16	60.36	31.2	70.4	22.93	9.09	102.55	23.76	32.19

Two different approaches were then followed to perform multiple regressions. In the first, a dummy variable denoted by Du was introduced. This variable was set to 1 for the basins in cluster 1 and 0 for the others. The regression formula used in this case is:

$$\log(Tu) = 2.3284 - 0.3021 * Rbdpon + 0.2486 * Sup - 0.3419 * Du \quad (1)$$

The estimates of Tu calculated using equation (1) are listed in Table 2.

In the second case, different regression formulae were applied to each of the clusters. The results obtained are shown in Table 3. The regression formulae used in this case are not presented because they cannot be applied in other circumstances, but once again the method presented is general.

In both cases the use of cluster analysis sharpens the estimate one may obtain using standard quantitative analysis. The values reported in Tables 2 and 3 show that the dummy variable is not the best choice in this case. It is possible that some basins have a smaller percentage error by using a dummy variable, but the overall error is smaller when different regression formulae are used for different clusters. Also in this part of the study estimates can be improved when a greater number of basins is used.

CONCLUSIONS

The main result of this preliminary study is that the suspended load of a basin can be estimated by recognizing the cluster it belongs to and then applying a regression formula obtained using only basins in this cluster, and this may be achieved without using the basin to determine the cluster itself. This encourages us to proceed in this direction since adding new elements will only improve the estimates.

Regarding the data extracted from DEMs, the present work can be considered as a starting point for further analysis. If the support area threshold parameter is not to be chosen subjectively, the right method to delineate a drainage network consistent with the morphology must be set up, to have the related geomorphic parameters suitable for calculating fluvial suspended load.

Another aspect that may improve the accuracy of the prediction is to include new parameters, such as the basin lithology, in the regression formulae or, at least, in the cluster analysis.

REFERENCES

Ciccacci, S., Fredi, P., Lupia Palmieri, E. & Pugliese, F. (1980) Contributo dell'analisi geomorfica quantitativa alla valutazione dell'entità dell'erosione nei bacini fluviali (Contribution of quantitative geomorphic analysis to the evaluation of erosion rate in drainage basins). *Boll. Soc. Geol. It.* **99**, 455–516.

Ciccacci, S., Fredi, P., Lupia Calmieri, E. & Pugliese, F. (1987) Indirect evaluation of erosion entity in drainage basins through geomorphic, climatic and hydrological parameters. In: *International Geomorphology 1986* (ed. by V. Gardiner), part II, 33–48. John Wiley, New York.

Cooke, R. U. & Doornkamp, J. C. (1974) *Geomorphology in Environmental Management*. Clarendon Press, Oxford, UK.

Horton, R. E. (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Bull. Geol. Soc. Am.* **56**, 275–370.

Jenson, S. K. & Domingue, J. O. (1988) Extracting topographic structures from digital elevation data for geographic information system analysis. *Photogramm. Engng and Remote Sens.* **54**(11), 1593–1600.

Kirkby, M. J. & Cox, N. J. (1995) A climatic index for soil erosion potential (CSEP) including seasonal and vegetation factors. *Catena* **25**, 333–352.

Mark, D. M. (1988) Network models in geomorphology. In: *Modelling in Geomorphological Systems* (ed. by M. G. Anderson), 73–97. John Wiley, New York.

Milligan, G. W. (1980) An examination of the effects of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* **45**, 325–342.

O'Callaghan, J. F & Mark, D. M. (1984) The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics and Image Processing* **28**, 328–344.

Sokal, R. R. & Sneath, P. H. A. (1963) *Principles of Numerical Taxonomy*. W. H. Freeman, San Francisco.

Strahler, A. N. (1957) Quantitative analysis of watershed geomorphology. *Trans. Am. Geophys. Un.* **38**(6), 913–920.

Tarboton, D. G., Bras, R. L. & Rodriguez-Iturbe, I. (1991) On the extraction of channel networks from digital elevation data. *Hydrol. Processes* **5**, 81–100.

Tarboton, D. G., Bras, R. L. & Rodriguez-Iturbe, I. (1992) A physical basis for drainage density. *Geomorphology* **5**, 59–76.

Todeschini, R (1997) *Introduzione alla Chemiometria* (Introduction to chemometry). Università degli Studi di Milano, Milan.

Wilkinson, L., Engelman, L., Corter, J. & Coward, M. (2000) Cluster analysis. In: *Systat 10—Statistics I*, 53–85. SPSS Inc., Chicago.