Sediment source fingerprinting: testing hypotheses about contributions from potential sediment sources

JEAN P. G. MINELLA[1,2], ROBIN T. CLARKE[1], GUSTAVO H. MERTEN[1] & DES E. WALLING[3]

1 *Institute of Hydraulic Research, Federal University of Rio Grande do Sul, PO Box 15029, 91501-970 Porto Alegre, RS, Brazil*
jminella@ppgiph.ufrgs.br

2 *Department of Soils, Federal University of Santa Maria, 97105-900, Santa Maria, RS, Brazil*

3 *Department of Geography, Exeter University, Amory Building, Rennes Drive, Exeter EX4 4RJ, UK*

Abstract Techniques involving composite fingerprints and multivariate mixing models are being increasingly used in catchment studies to establish the relative contributions of potential sources to the suspended sediment output. Such information is important both for understanding the fine sediment dynamics of a catchment and for targeting remediation measures required to reduce sediment-related environmental problems. A multivariate mixing model algorithm, which compares the concentrations of a range of geochemical properties of the suspended sediment load with those of a number s of potential sources, is commonly used to provide estimates of the relative contributions (P_1, P_2, ..., P_s) of those sources to the suspended sediment load. However, such models do not provide measures of the uncertainty associated with the P-values. This paper describes how the usual mixing model can be modified, such that the optimization procedure used to estimate the sediment proportions P_1, P_2 … contributed by different sources also provides measures of their uncertainty. This approach allows hypotheses concerning the P values to be tested, such as: (i) whether the individual P-values differ significantly from zero, and (ii) whether the P-values change significantly between events. To calculate the uncertainty associated with a P-value, a statistical model which considers the correlation between the tracer variables is used. This approach has been tested using data from a small rural catchment in southern Brazil where a sediment source investigation is in progress. Sediment samples collected during 48 storm events were used to establish the source contributions.

Key words suspended sediment; source fingerprinting technique; multivariate mixing model; uncertainty analysis; optimization; soil management; Brazil

INTRODUCTION

Controlling soil erosion and sediment transport to watercourses is essential to reduce the inputs of nutrients and agricultural pollutants to fluvial, lacustrine and marine systems. In addition, the reduction of erosion and sediment yield is regarded as an important indicator of the effectiveness of soil and water management techniques in catchments (DEFRA, 2004; Brils, 2005; Owens & Collins, 2006). In Brazil, information on sediment yield from catchments has particular importance, especially where land use is changing, since more than 80% of the national energy requirements are supplied by hydropower. Sediment deposition in impounding reservoirs can severely reduce the productive life of hydropower installations. There is therefore an ever-increasing need to monitor catchment sediment yields, to identify the key sediment sources and to predict sediment mobilization and transport using modelling methods.

Recognizing the complexity of sediment mobilization and delivery, Foster (2000), Walling (2006) and Owens & Collins (2006) have identified the need to improve methods for evaluating catchment sediment budgets and for monitoring changes in soil condition (land use and management). Procedures for tracing the source of sediment within catchments are also seen as providing a valuable complement to traditional monitoring methods. Source fingerprinting techniques are being increasingly used to establish the relative importance of the main sediment sources in a catchment (Walling, 2005). However, existing procedures for establishing the magnitude of the contributions from different sediment sources involve a number of complications, which have limited their usefulness (Symader & Strunk, 1992; Small *et al.*, 2002). These complications include: (a) enrichment or depletion of tracer concentrations during sediment mobilization and transfer to the catchment outlet; (b) non-conservative behaviour of sediment properties within the fluvial system.

It is important that methods for identifying and quantifying the contributions of individual sediment sources should provide an indication of the uncertainty associated with the results obtained. This is particularly important where significant resources are to be invested in management strategies that reduce mobilization of sediment and its transport to water-courses. This paper describes how the existing mixing model commonly used in sediment source fingerprinting investigations can be modified, such that the optimization procedure used to estimate the sediment proportions P_1, P_2 … originating from different sources also yields measures of their uncertainty. As a consequence, the approach allows hypotheses concerning the P values to be tested, such as: (i) whether the individual P-values differ significantly from zero, and (ii) whether the P-values change significantly between periods or events.

MATERIAL AND METHODS

Summary of the traditional source apportionment model

The traditional source apportionment model compares the geochemical properties of potential sediment sources with those of the sediment, in order to trace the source of that sediment, To do this, it is necessary to characterize the potential sources of sediment within the catchment both chemically and physically, so that samples of suspended sediment can subsequently be compared with the potential sources using a set of tracer properties that can be measured in both.

Considering a tracer variable, the expression relating the concentration in suspended sediment to the concentrations in the potential source materials (Yu & Oldfield, 1989) is:

$$y_i = \sum_{s=1}^{g} a_{is} P_s \qquad i = 1 \ldots m \tag{1}$$

where for the ith tracer-variable, y_i is the value of the variable i obtained from suspended sediment, a_{is} is the value for the tracer variable i associated with source material s, P_s is the relative contribution of source s to the suspended sediment load represented by the sediment sample, and g is the number of possible sources.

The objective is then to determine P_1 , P_2 , P_g subject to the constraints $0 \le P_s \le 1$ and $P_1 + P_2 + ... + P_g = 1$. For example, if there are three sources (s = 3), and m tracer variables:

$$y_1 = a_{11}P_1 + a_{12}P_2 + a_{13}P_3 \tag{2}$$

$$y_2 = a_{21}P_1 + a_{22}P_2 + a_{23}P_3$$

$$y_m = a_{m1}P_1 + a_{m2}P_2 + a_{m3}P_3$$

The values of P_1, P_2, P_3 are estimated by minimizing the term (Walling & Woodward, 1995):

$$\sum [(y_i - a_{i1}P_1 - a_{i2}P_2 - a_{i3}P_3) / y_i]^2 \tag{3}$$

as a function of P_1 , P_2 , P_3.

It is not possible to estimate the uncertainty in the proportions (P_1, P_2, P_3) with this method. In addition, the method does not allow for any correlation amongst the tracer variables y_i, nor is it possible to test hypotheses about the proportions: for example whether the contribution from one of the possible sources is sufficiently close to zero as to be negligible, or whether the contribution from a particular source is so close to one that it could be concluded that it is the principal source of the sediment. To calculate the uncertainty associated with P_i, it is necessary to include in the calculation a statistical model that considers the correlation between the y_i tracer variables.

A modified method

The modified model can be applied to cases in which there are replicate samples of suspended sediment, and where it is possible to assume that replicates of each tracer variable y_i (i = 1...m) either have a Normal distribution, or can be transformed to a scale in which it is approximately Normal. Replicate samples of suspended sediment can be obtained either by using several sampling points within the same section, or by treating single samples collected from individual runoff events as replicates of events within a period regarded as homogeneous. In either case, the variance-covariance matrix for the tracer variables, denoted by $\boldsymbol{S}$, can be calculated. This need for Normality is an important additional restriction on existing fingerprinting methods, but its advantage is that statistical methods based on the likelihood function can then be used, and this in turn provides a tool for testing hypotheses about the magnitude of contributions from different sediment sources, and for calculating measures of their uncertainty.

Suppose we have K samples of suspended sediment from g possible sources, and that each of m tracer variables y_i ($i = 1...m$) can be expressed in the form (1) above, or putting the m tracer variables in a vector $\boldsymbol{y}$ of dimension $m \times 1$

$$\boldsymbol{y} = \boldsymbol{AP} \tag{4}$$

where $\boldsymbol{A}$ is a matrix with dimension $m \times g$ and $\boldsymbol{P}$ is a $g \times 1$ vector. Because of the Normality assumption, the (multivariate) probability distribution of the tracer elements $\boldsymbol{y}$ is:

$$f(\boldsymbol{y};\boldsymbol{AP},\boldsymbol{S}) = 1 / [(2\pi)^{m/2} \mid \boldsymbol{S} \mid^{½}] \exp [-1/2(\boldsymbol{y} - \boldsymbol{AP})^T \boldsymbol{S}^{-1}(\boldsymbol{y} - \boldsymbol{AP})] \tag{5}$$

where $\mid \boldsymbol{S} \mid$ is the determinant of the variance-covariance matrix $\boldsymbol{S}$ and the symbol T denotes a vector transpose. Given K samples of $\boldsymbol{y}$, denoted by $\boldsymbol{y}_s$ ($s = 1...K$), the log of the Likelihood function L is:

$$\log_e L = \text{constant} - K/2 \log_e \mid \boldsymbol{S} \mid - ½ \sum_{s=1}^{K} (\boldsymbol{y}_s - \boldsymbol{AP})^T \boldsymbol{S}^{-1}(\boldsymbol{y}_s - \boldsymbol{AP}) \tag{6}$$

The expression on the right-hand side of equation (6) involves the g values of the vector $\boldsymbol{P}$, representing the proportions P_i ($i = 1...g$) from the g possible sources of sediment. These may be estimated by maximizing the likelihood function L, or more conveniently by maximizing log_eL with respect to the unknowns, and subject to the usual equality and inequality constraints:

$$P_1 + P_2 + \ldots + P_g = 1 \tag{7}$$

$$0 \le P_i \le 1, \quad i = 1...g$$

Maximizing the expression for log_eL is made simpler if the K samples provide a good estimate of the variance-covariance matrix $\boldsymbol{S}$. With $\boldsymbol{S}$ assumed known, maximization of log_eL is then equivalent to minimizing the least-squares function:

$$\sum_{s=1}^{K} (\boldsymbol{y}_s - \boldsymbol{AP})^T \boldsymbol{S}^{-1}(\boldsymbol{y}_s - \boldsymbol{AP}) \tag{8}$$

with respect to the proportions P_i, a form that is a slightly more complicated form of equation (3) above. In practice, it is convenient to assume initially that $\boldsymbol{S}$ is known, leading to estimates of P_i, and then to follow an iterative calculation in which $\boldsymbol{S}$ is successively estimated from the "residuals" (i.e. the differences between the $\boldsymbol{y}$ and the estimates $\boldsymbol{AP}$).

It is by incorporating the likelihood function L, and the maximization of $\log_e L$, that hypotheses about the proportions P_i can be tested. Suppose, for example, that we wish to test the hypothesis that a particular P-value, say P_1, is zero, such that this "source" would not contribute any sediment. Denote by H_0 this hypothesis that $P_1 = 0$, and let H_1 be the alternative hypothesis, $P_1 \neq 0$. To test whether the data are more supportive of H_0 than H_1, the ratio of the two maximized likelihood functions $max\ L_0$ and $max\ L_1$, are calculated; $max\ L_1$ is the maximum of the expression in equation (5) above, and $max\ L_0$ is a similar expression in which the likelihood is maximized after setting $P_1 = 0$. The test is based on the ratio of the two maximized likelihoods (see equation (6); Johnson & Wichern, 1998) and is known as the Likelihood Ratio (LR) test. The test statistic is $\Lambda = maxL_0 / maxL_1$, and the approximate test of the hypothesis is made by comparing $-2 \log_e\Lambda$ with the distribution of χ_2 with (in the case of the hypothesis H_0: $P_1 = 0$) $K - 1$ degrees of freedom. In a more general case where, say, p of the proportions P_i are set to zero, the degrees of freedom are $K - p$. Thus if the statistic $-2 \log_e\Lambda$ exceeds the 95% quantile of the distribution of χ_2 with $K - 1$ degrees of freedom, the hypothesis H_0: $P_1 = 0$ would be rejected at the 5% level of significance.

A case study

The model was applied to data assembled by a study aimed at identifying sediment sources in a small agricultural catchment in southern Brazil. A soil conservation programme has been established in the catchment and the study objectives included assessment of the effectiveness of this programme in reducing soil loss. To do this, 48 storm events were monitored (rainfall, water discharge and suspended sediment concentrations) over a period of four years (2002–2006) and samples used for identifying the sediment sources were collected during these events (Minella *et al.*, 2008).

The study catchment, referred to as the Arvorezinha catchment, is located in the state of Rio Grande do Sul in southern Brazil and covers an area of 1.19 km^2. The area is characterized by steeply rolling terrain with average slopes in the range 4–84%. The mean annual precipitation is 1600 mm distributed fairly evenly throughout the year. Land use within the catchment is predominantly agriculture in fairly small holdings, with much of the land being used for growing tobacco. Available information suggest that the mean annual suspended sediment yield of the catchment prior to the implementation of improved land management was about 145 t km^{-2} $year^{-1}$, with suspended sediment concentrations during major runoff events reaching a maximum of approx. 11 000 mg L^{-1} (Merten & Minella, 2005).

The data and the application of the model

With the series of monitored events extending from May 2002 to March 2006, it was possible to separate the events into two different periods. The first period with 19 events, from May 2002 to July 2003, preceded the introduction of soil conservation measures and is referred to as the period of "traditional soil management", in which the soil was tilled in the traditional manner and there were no soil conservation measures. This system of preparing the soil for crops causes intense erosion and increases the sediment yield. The second period, which comprised 29 events and extended from August 2003 to March 2006, is referred to as the period of "conservation soil management", in which conservation practices – specifically minimum tillage where crop residues are left on the surface – were introduced. The use of these techniques for soil preparation reduces surface runoff, erosion and sediment yield.

When applying this model, the events within each of the subgroups were considered as replicates that characterize the two different periods of soil management (traditional and conservation). It was therefore assumed that each period was effectively homogeneous in terms of the hydro-sedimentological conditions and processes, and that the events can be considered as replicates, to demonstrate the application and functioning of the model.

It is commonly assumed that, in an agricultural region, crop fields will represent the dominant sediment source. However, existing studies and observations undertaken in the Arvorezinha catchment have demonstrated that the unpaved roads and the stream channels also provide a significant contribution to the sediment yield at the catchment outlet (Minella *et al.*, 2008). It is therefore important that the model should be able to test the following hypotheses:

– the relative contributions of each source are the same for both periods (conventional and conservation);
– crop fields supply 100% of the sediment in each period;
– the contributions from stream channels and unpaved roads are effectively zero in each period.

RESULTS

The modified model was used to estimate the proportions of the measured suspended sediment yield that originated from the three potential sources, namely, crop fields, unpaved roads, and stream channels, subject to the usual constraints that the proportions sum to one, and that each proportion must lie between zero and one. The results are presented in Table 1, and relate to the two periods of traditional and conservation soil management. As mentioned above, the results are based on analysis of data from 19 and 29 events, respectively, for these two periods.

Table 1 The mean contribution of the three potential sediment sources to the sediment yield from the Arvorezinha catchment.

Sediment source	Relative contribution (%) Traditional soil management	Conservation soil management
Crop fields	61.1	52.6
Unpaved roads	37.3	29.0
Stream channels	1.6	18.4

Table 2 Analysis of the uncertainty associated with the estimated sediment source contributions, based on the likelihood ratio test.

Hypothesis (H_0)	Likelihood Ratio Test	Conclusion
Comparison between the periods		
The source contributions for the period under traditional soil management are the same as those for the period of under conservation tillage.	<0.05	Reject H_0
Period under Traditional Soil Management		
The contribution of the stream channels is zero	0.23	Accept H_0
The contribution from unpaved roads is zero	<0.05	Reject H_0
The contribution of the crop fields is equal to 100%	<0.05	Reject H_0
Period under Conservation Soil Management		
The contribution of the stream channels is zero	<0.05	Reject H_0
The contribution from unpaved roads is zero	<0.05	Reject H_0
The contribution of the crop fields is equal to 100%	<0.05	Reject H_0

The results presented in Table 1 show that the introduction of improved soil management practices in the Arvorezinha catchment has resulted in considerable changes in the relative importance of the three sediment sources. A comparison between the periods of traditional and conservation soil management shows that the contributions of crop fields and unpaved roads have fallen from 61.1 and 37.3% to 52.6 and 29.0%, respectively, while the contribution of the stream channels has increased from 1.6 to 16.8% (Minella *et al.*, 2008).

These results show that there was a substantial reduction in the proportion of the sediment contributed by the crop fields due to the implementation of soil conservation measures, but that the proportion of sediment contributed by erosion of the streams channels increased. The value of the modified model is that the significance of this result can be tested in relation to the hypotheses defined above. Table 2 presents the results of comparing the statistic $-2\ \log_e\Lambda$ with the χ_2 distribution, using the Likelihood Ratio test.

The first analysis was used to establish whether the proportions estimated for the period under traditional soil management differed statistically from the values estimated for the period under conservation soil management. This information is important, because the goal of improving soil management was the reduction of erosion and sediment yield. The LR test showed that the probability of observing the calculated value of $-2\ \log_e\Lambda$, if the null hypothesis were true, is extremely small, so that this hypothesis is rejected. Thus, it can be concluded that the relative contribution of the sources during the period under traditional soil management is different from that during the period of conservation soil management.

The second analysis tested hypotheses relating to contribution of the individual sources during the period when the study catchment was under traditional soil management. The hypothesis that crop fields contributed the entire sediment load, with unpaved roads and stream channels contributing nothing, was rejected. The hypothesis that the contribution of channel sources (1.6%) does not differ significantly from zero was accepted using the LR test.

Finally, the statistical test assessed the source contributions during the period when the study catchment was under conservation soil management. When considering the hypotheses that the crop field contribution was equal to 100% and that the contribution of unpaved roads and stream channels was equal to 0%, all null hypotheses were rejected.

CONCLUSION

This paper describes the modification of an existing source fingerprinting model used to quantify the relative contributions of individual potential sources to the suspended sediment yield at the catchment outlet. The modification requires replicate sediment samples from which the variance-covariance matrix for the different tracer variables can be established and it must also be assumed that the tracer variables either have a multivariate Normal distribution, or can be transformed to scales in which they are at least approximately Normally-distributed. Where these additional assumptions can be shown to be met, it becomes possible to test hypotheses about the magnitudes of the contributions from different sediment sources. Careful study of whether the assumptions are justified is of course essential.

Based on the study carried out in the Arvorezinha catchment over two different periods of soil management, it was concluded that the relative contributions from the suspended sediment sources were statistically different between the two periods. Furthermore, during the period under conservation soil management there was a significant reduction in the contribution of the crop fields and unpaved roads to the sediment yield at the catchment outlet, and a corresponding significant increase in the contribution from the stream channels. Although the model had estimated a contribution of 1.6% from the stream channels during the period under traditional soil management, this proportion did not differ significantly from zero.

The improvement of the existing source ascription model by the inclusion of methods of uncertainty analysis should contribute to better understanding, and more rigorous assessment, of the impact of improved soil and land management on hydro-sedimentological processes, as well in reducing sediment mobilization and transfer to watercourses.

Acknowledgements Financial support was provided by CNPq (National Counsel of Technological and Scientific Development). Adair Giongo and Ademir Giongo assisted with field sampling and the field monitoring programme.

REFERENCES

Brils, J. (2005) Commission will continue its efforts to overcome the lack of knowledge on sediment quality in the EU. *J. Soils Sediment* **5**(1), 48–49.

Department of Environment, Food and Rural Affairs (2004) Developing measures to promote catchment-sensitive farming. *DEFRA and HM Treasury Consultation*. DEFRA, London, UK.

Foster, I. D. L. (2000) *Tracers in Geomorphology*. John Wiley & Sons Ltd, Chichester, UK.

Johnson, R. A. & Wichern, D. W. (1998) *Applied Multivariate Statistical Analysis*, fourth edn. Prentice-Hall International, Inc, New Jersey, USA.

Merten, G. H. & Minella, J. P. G. (2005) Impact on sediment yield caused by intensification of tobacco production in a catchment in southern Brazil. In: *Sediment Budgets 2* (ed. by D. E. Walling & A. J. Horowitz), 239–244. IAHS Publ 291. IAHS Press, Wallingford, UK.

Minella, J. P. G., Walling, D. E. & Merten, G. H. (2008) Combining traditional monitoring and sediment source tracing techniques to assess the impact of improved land management on catchment sediment yields. *J. Hydrol.* 3**48**, 546–563.

Owens, P. N. & Collins, A. J. (2006) *Soil Erosion and Sediment Redistribution in River Catchments: Measurement, Modeling and Management.* CAB International, Wallingford, UK**.**

Small, I. F., Rowan, J. S. & Franks, S. W. (2002) Quantitative sediment fingerprinting using a Bayesian uncertainty estimation framework. In: *The Structure Function and Management Implications of Fluvial Sedimentary Systems* (ed. by F. J. Dyer, M. C. Thoms & J. M. Olley), 443–450. IAHS Publ. 276. IAHS Press, Wallingford, UK.

Symader, W. & Strunk, N. (199_2) Determining the source of suspended particulate material. In: *Erosion, Debris Flow and Environment in Mountain Regions* (ed. by D. E. Walling, T. R. Davies & B. Hasholt), 177–185. IAHS Publ. 209. IAHS Press, Wallingford, UK.

Walling, D. E. (2005) Tracing suspended sediment sources in catchments and river systems. *Sci Total Environ.* **344**, 159–184.

Walling, D. E. (2006) Tracing versus monitoring: new challenges and opportunities in erosion and sediment delivery research. In: *Soil Erosion and Sediment Redistribution in River Catchments: Measurement, Modeling and Management* (ed. by P. N. Owens & A. J. Collins). CAB International, Wallingford, UK**.**

Walling, D. E. & Woodward, J. C. (1995) Tracing sources of suspend sediment in river basins: a case study of the River Culm, Devon, UK. *Mar. Freshwater Res*. **46**, 327–336.

Yu, L. & Oldfield, F. (1989) A multivariate mixing model for identifying sediment source from magnetic measurements. *Quatern. Res.* **32**, 168–181.