A hybrid Monte Carlo–possibilistic method to evaluate soil erosion in an Alpine valley

F. Mazza, L. Longoni, M. Papini & D. Brambilla
Politecnico di Milano, Dipartimento di Ingegneria Idraulica, Ambientale, Infrastrutture Viarie e Rilevamento (D.I.I.A.R.), Italy

Abstract

The high number of complex processes involved in soil erosion and sediment delivery make their understanding and reproduction a difficult task. Alpine valleys, characterized by high slopes, are particularly susceptible to severe soil erosion. Two of the main consequences are silting of water reservoirs and fostering of shallow landslides. In the last decades several models for the evaluation of sediment production and delivery have been proposed. Different approaches can be split into two main categories: bottom-up and top-down models. Bottom-up models are designed to reproduce the main physical processes involved in soil erosion; these methods are really complicated from a computational point of view. Instead top-down models, like the Gavrilovic one, reproduce the phenomenon at the basin scale with a low number of parameters.

In this paper the authors present a hybrid Monte Carlo and possibilistic approach to the Gavrilovic method, designed to take into account uncertainties on input data. An example of application on a test basin situated in the Italian Alps is used to show the potential of the proposed method. The basin was split into sub-areas to reduce the subjectivity of the choice of empirical coefficients. A quantitative comparison between measures of average sediment yield and results obtained with the proposed method was performed.
Keywords: soil erosion, silting, Monte Carlo methods, epistemic uncertainty.

1 Introduction

Soil erosion and transport are two complex phenomena, acting on a wide range of scales, that are far from be fully understood and modeled, mainly because of the lack of knowledge about the physical mechanisms governing them. Many

research efforts have been done to assure a good understanding and reproducibility of soil erosion. These efforts are justified by the importance of the phenomenon from a technical point of view. As a matter of fact, sediment erosion and transport have many effects on human activities: removal of rich soil, promotion of shallow landslides and silting of water reservoirs. As a consequence, a number of methods have been proposed in the last fifty years for the computation of soil erosion, but none of them is able to reproduce its effects on a full range of scales in space and time. Thus, the choice of the best model must be done taking into account the scale of the study, the kind of desired results and data availability. According to De Vente and Poesen [1] it is possible to identify three families of models: empirical methods, physics-based models and conceptual models. Empirical methods allow one to compute average intensity of soil erosion at the basin scale by the use of empirical formulae. Physics-based models claim to reproduce soil erosion by simulating some of the most important physical processes. These models require a huge amount of data and are heavy from the computational point of view. Finally conceptual models, a middle-way between the previous two categories, are typically structured in conceptual compartments in which different physicals processes are clumped and reproduced by means of empiric relationships.

The aim of this paper is to modify a simple model to obtain a tool for reliable quantitative evaluation of sediment volumes produced and delivered to the closing section of a basin during a certain time period. According to previous considerations, a highly refined model is not necessary nor in space neither in time to assure this task. Therefore the Gavrilovic method [6, 7] has been adopted. In fact, although it is simplistic and has small number of required inputs, this method produces good results in alpine basins [4, 8]. The Gavrilovic model can also be considered the more quantitative method among semi-quantitative models [1]; it is featured by a very simple structure, then its integration with a hybrid Monte Carlo and possibilistic framework it is quite easy. This leads to overcome one of the main problems of soil erosion models: the lack of a degree of confidence on results. As a matter of fact usual approaches to soil erosion evaluation produce deterministic results. These results are based on subjective choice of parameters and on stochasticity of natural phenomena: it is thus better to look at them in a not-deterministic framework.

The paper is organized as follows. Section 2 recalls the main steps of the Gavrilovic method for soil erosion computation and sediment transport. In Section 3 the Gavrilovic method is combined with a hybrid Monte Carlo and possibilistic approach to consider uncertainty on input values and to propagate them to the results. In Section 4 the proposed method is applied to a test basin for the evaluation of sediments volumes during different time periods. In Section 5 obtained results are compared to measured values. Finally some conclusion about advantages and limitations of the proposed approach are given in Section 6.

2   The Gavrilovic model for soil erosion evaluation

The Gavrilovic model, in its original formulation, is a semi-quantitative empirical method for soil erosion and sediment yield estimation. It can be seen as the product of two components: an erosion evaluation procedure, used to compute W_g, and a sediment delivery one, that evaluates the percentage of sediments reaching the closing section of the basin through the routing coefficient R_u. Required inputs are the main topographic and hydrologic features of the basin and three descriptive coefficients (land use φ, type and extent of erosion X_a, soil resistance to erosion Y) used to describe land erosion susceptibility. The Gavrilovic method was originally developed for stream basins of Yugoslavia [2], but it has been successfully applied several times to alpine basins [1, 4, 5]. According to later improvements of the method, the basin can be split into subareas with homogeneous soil properties; then, values of coefficients can be assigned to each area [4, 6].

The following relationships allow one to compute the total mean annual discharge of eroded material G_g [m^3/year]:

$$z_i = Y_i \cdot Xa_i \cdot \left(\varphi_i + I_i^{0.5} \right) \tag{1}$$

$$\hat{z} = \sum_{i=1}^{N} z_i \cdot A_i \Big/ \sum_{i=1}^{N} A_i \tag{2}$$

$$T = \sqrt{\frac{t}{10} + 0.1} \tag{3}$$

$$W_{sp} = T \cdot H \cdot \pi \cdot \hat{z}^{3/2} \tag{4}$$

$$W_g = F \cdot W_{sp} \tag{5}$$

$$R_u = \frac{\left(L_p + L_a \right)}{F} \cdot \frac{\left(O \cdot D \right)^{1/2}}{\left(L + 10 \right)} \tag{6}$$

$$G_g = W_g \cdot R_u \tag{7}$$

where z_i is the potential erosion of the i-th area, I_i is the average slope steepness of the i-th area [%], A_i is the surface of the i-th area [km^2], N is the number of subareas, $\hat{z}$ is the global potential erosion coefficient, T is the temperature coefficient, t is the average annual temperature of the basin [°C], H is the average annual height of precipitations [mm/year], W_{sp} is the specific annual production of sediments [m^3/(km^2·year)], F is the surface of the catchment area [km^2], W_g is the average annual production of erosional sediments in the catchment [m^3/year], O is the perimeter of the basin [km], L is the length of the

basin [km], D is the mean level difference in the basin [m] and R_u is the sediment retention coefficient.

3 Development of a hybrid Monte Carlo approach to the Gavrilovic method

Despite of the chosen model, the prediction of soil erosion and sediment delivery is strongly dependent on the quality of input data. Typically input data are affected by uncertainty that may be caused by measurement errors, stochasticity of natural phenomena, or use of expert's opinions. From a conceptual, but also practical, point of view, uncertainty can be divided in two categories: random variability, which can be represented by probability distribution functions (aleatory uncertainty) and imprecision (epistemic uncertainty), that should be treated separately. In particular the possibility theory is well suited to represent epistemic uncertainty: possibility distributions, that are the core of possibility theory, are characterized by mathematical simplicity associated with a high representation power.

The distinction between aleatory and epistemic uncertainty is a critical point in soil erosion evaluation. While aleatory uncertainty affects variability of stochastic processes, like time evolution of temperatures and precipitations, epistemic uncertainty results from the subjective judgment and lack of exact correspondences between in-situ soil characteristics and tabled reference situations. Uncertainties have a strong effect on the results and cannot be discarded without reducing reliability of results. A hybrid model [7, 8] that combines Monte Carlo methods [9] and possibility theory [10, 11] allow one to represent and propagate uncertainty in a proper way.

3.1 Outline of the hybrid Monte Carlo and possibilistic method

Monte Carlo methods [9] have become popular tools for the study of systems with uncertain inputs. In this field, hybrid Monte Carlo and possibilistic method [7, 8] is able to merge stochastic variables and imprecise information. The latter can be represented in the framework of possibility theory. According to this theory, the only known information about an imprecise parameter is that its value belongs to certain intervals with a certain degree of possibility. A detailed description of the hybrid Monte Carlo and possibilistic method is out of the goals of this paper. Nevertheless, a simplified description of the approach is required to let the reader understand next sections.

A possibility distribution of an epistemic variable is a mapping function that assigns to each value of the variable x belonging to the range X (universe of discourse) a degree of possibility $\pi(x) \in [0,1]$ to be the correct value. The normalization condition for a possibility distribution is stated in eqn. (8): it states that there must be at least one value of x with total possibility in X. This statement is much weaker than the normalization condition of probability distribution functions, eqn. (9). An alternative numerical convenient way of representing unimodal possibility distributions is a set of nested confidence

intervals A_α, named the α-cuts of π(x). If π(x) is contiguous, the interval [$a_{\alpha,\inf}$, $a_{\alpha,\sup}$] has a possibility 1-α to contain the value of the imprecise variable x (Figure 2b).

$$\exists x \in X : \pi(x) = 1 \tag{8}$$

$$\int_X p(x) = 1 \tag{9}$$

$$\Pi(A) = \sup_{x \in A} \pi(x) \tag{10}$$

$$N(A) = 1 - \Pi(\overline{A}) = \inf_{x \notin A}(1 - \pi(x)) \tag{11}$$

A being a subset in the power set P(X) of X, possibility theory provides two evaluations of the likelihood of A: the possibility and necessity measures, that can be computed according to eqns. (10) and (11).

A hybrid approach allows propagate mixed uncertain information by combining a simple Monte Carlo method with the possibility theory. Named $\underline{Y}$ the vector of all input variables, $\underline{Y}_a$ the vector of aleatory input variables and $f(\underline{Y})$ the output variable, according to Baraldi et al [7] the numerical solution of the problem can be achieved with an algorithm composed by two nested loops:

1. outer loop: m_a repetitions of Monte Carlo sampling of vector $\underline{Y}_a$ (to process aleatory uncertainty)
2. inner loop: possibilistic interval analysis (to process epistemic uncertainty)

It is thus possible to compute the necessity and possibility measures for each i-th sample of the random vector. At the end of this process, results are clumped by the joint aggregation method [8]. Using probabilities of random extractions p_i as weights, the m_a realizations of possibility and necessity measures can be averaged by eqns. (12) and (13) to obtain respectively believe and plausibility measures. Considering a generic value u of $f(\underline{Y})$, a substitution of A with the interval (-∞,u] allows one to compute upper and lower limit cumulative distributions of u [7], see eqns. (14) and (15).

$$Bel(A) = \sum_i p_i \cdot N_i^f(A) \tag{12}$$

$$Pl(A) = \sum_i p_i \cdot N_i^f(A) \tag{13}$$

$$\underline{F}(u) = Bel(f(Y) \in (-\infty, u]) \tag{14}$$

$$\overline{F}(u) = Pl(f(Y) \in (-\infty, u]) \tag{15}$$

For a complete explanation of hybrid Monte Carlo and possibilistic methods, the reader is referred to Baraldi et al. [7] and to Baudrit et al. [8].

3.2 Hybrid Monte Carlo and possibilistic method applied to the Gavrilovic model

The computation of soil erosion with the Gavrilovic method is highly dependent on the choice of the three coefficients φ, X_a, Y. The approach here presented tries to overcome this problem by considering the level of uncertainty hidden in the choice, which is mainly due to the difficulty in finding a correspondence between tabled soil conditions and field situation. In addition, land use can change in a way that is difficult to predict: this can also be considered in the framework of possibility theory. The effect of stochasticity of natural phenomena is taken into account by sampling values of annual mean precipitation and temperature from their probability distributions. Subdivision of the basin in subareas allows one to reduce the inaccuracy on the choice of coefficients, being easier to find a correspondence between tabled situations and conditions of a little homogeneous area [4, 6]. The solution of the problem has been achieved with a Matlab script written for the purpose.

4 Application to the Tartano Valley

Tartano Valley is a medium basin (36.2 km^2) situated in the Italian Alps, approximately 100 km north from Milan. It extends in height from 1,148 meters a.s.l. to 2,504 meters a.s.l., with a mean altitude H_{sp} of 1861 meters a.s.l. The valley is interested by a variety of shallow landslides and accelerated soil erosion. The modified Gavrilovic method presented in this paper and the traditional one have been tested on the Tartano Valley. In Table 1 are reported the main topographic features of the basin. The analysis of rainfall heights and temperatures, recorded during the last century in a measurement station placed inside the basin area, allowed to define their respective probability distributions. Statistical inference was used to find suitable probabilistic distributions that represent randomness of annual average precipitations and temperatures. Independence between values of temperatures and rainfalls has been confirmed

Table 1:   Main topographic features of the basin.

Surface of the catchment area [km^2]	36.2
Perimeter of the catchment area [km]	27.0
Length of the principal waterways [km]	28.4
Length of the secondary waterways [km]	92.6
Minimum altitude [m a.s.l.]	1,148
Mean altitude [m a.s.l.]	1,861
Maximum altitude [m a.s.l.]	2,504

by a specific test. Pearson tests allowed one to affirm that both the data series can be fitted by Gaussian distributions. In accordance with these results the authors computed a mean rainfall height of 1,376 [mm/year], with an associated variance of 351 [mm/year] and a mean temperature of 3.0 [°C], with an associated variance of 0.4 [°C].

The area of Tartano valley has been zoned according to observed erosion process, land use and soil resistance to erosion. The area and the mean slope of each region have been extracted from a DEM (Digital Elevation Model) of the zone with square cells of 20x20 m^2. Possibility distributions of φ, X_a and Y coefficients have been defined for each region. In Table 2 the four key points of these distributions are reported according to the scheme shown in Figure 2.
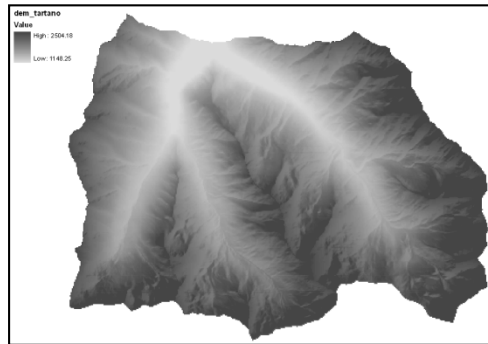

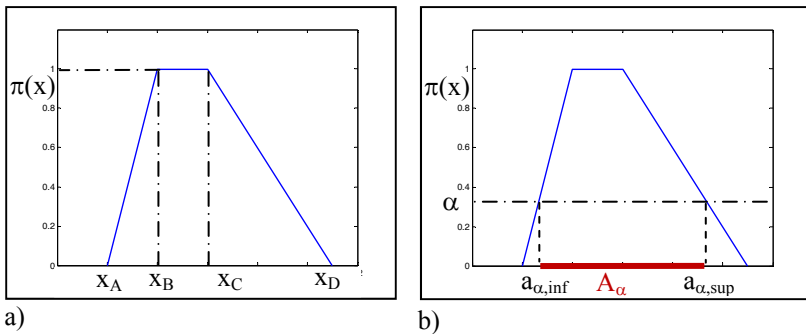Figure 1:   DEM of the Tartano Valley.

Figure 2:   Possibilistic distributions a) schematic representation of a unimodal possibilistic distribution; b) example of α-cut.

The proposed method has been applied to estimate sediment volumes delivered to the closing section of the basin during time periods of one year and sixteen years. For the latter case Gaussian distributions have been modified by reducing variances with a factor equal to 1/16. Finally, deterministic values of

parameters have been estimated (Table 2, column 8) and fed to the traditional Gavrilovic method to compute a reference deterministic result.

Table 2:   Homogeneous regions with respect to soil erosion: percentual surface (column 2), possibilistic distributions of coefficients for the Gavrilovic method (columns 4-7) and values used in the deterministic approach (column 8).

Area	Surface [%]	Land use	Parameter	x_a	x_b	x_c	x_d	$x_{deterministic}$
1	0.88	Medium/high density coniferous forest	φ	0.00	0.00	0.10	0.10	0.05
			X_a	0.05	0.10	0.20	0.22	0.15
			Y	0.75	0.80	0.85	0.90	0.83
2	57.61	Pastures	φ	0.25	0.30	0.40	0.50	0.35
			X_a	0.50	0.55	0.60	0.60	0.58
			Y	0.80	0.90	1.00	1.10	0.95
3	21.75	Scarce bushes	φ	0.15	0.20	0.30	0.30	0.25
			X_a	0.30	0.40	0.50	0.52	0.45
			Y	0.80	0.90	1.00	1.10	0.95
4	18.46	Moraines	φ	0.70	0.70	0.80	0.85	0.75
			X_a	1.00	1.00	1.00	1.00	1.00
			Y	2.00	2.00	2.00	2.00	2.00
5	0.32	Low density residential zone	φ	0.00	0.01	0.05	0.10	0.03
			X_a	0.00	0.00	0.03	0.05	0.01
			Y	0.80	0.90	1.00	1.10	0.95
6	0.39	Low density forest	φ	0.00	0.10	0.15	0.20	0.13
			X_a	0.20	0.30	0.35	0.37	0.33
			Y	0.80	0.90	1.00	1.10	0.95
7	0.58	Glaciers and reservoirs	φ	-	-	-	-	-
			X_a	-	-	-	-	-
			Y	-	-	-	-	-

5   Results

The method has been applied with the aim of defining the maximum and minimum volumes of sediments that are expected to reach the closing section of the basin. Sediment amounts have been computed for time periods of one and sixteen years. Data series of the period 1990-2006 have been used as reference values, Table 3.

Table 3:   Annual sediment yield (SY) into the reservoir at the Campo dam [4].

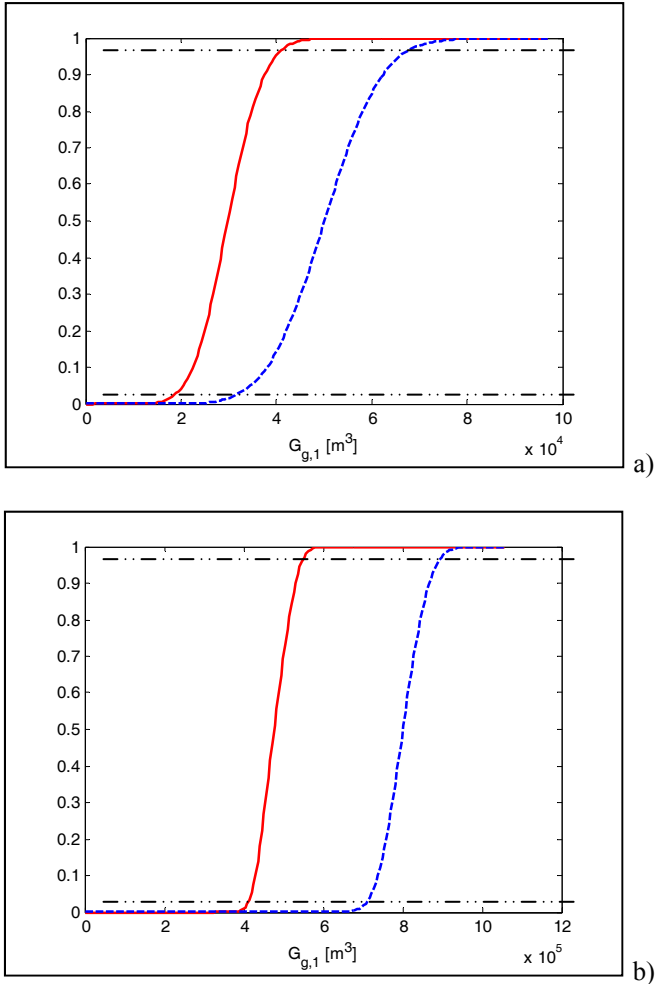| Year	1991	1992	1993	1994	1995	1996	1997	1998
SY (m^3)	34,073	43,504	53,605	36,737	26,264	39,749	35,314	32,800
Year	1,999	2,000	2,001	2,002	2,003	2,004	2,005	2,006
SY (m^3)	41,876	57,299	43,187	42,022	22,957	50,083	21,287	27,844
Mean SY value (m^3)		38,038						

Figure 4:   a) Upper and lower cumulative distribution functions of sediment amounts delivered to the closing section of the basin in one year; b) upper and lower cumulative distribution functions of sediment amounts delivered to the closing section of the basin in sixteen years. The dashed dotted lines show the boundaries of the two-sided 95% confidence intervals.

Results of the two simulations (one and sixteen years) are represented by upper and lower cumulative distribution functions shown in Figure 4. The difference variance of input random variables for the two considered time periods is reflected by a different dispersion of results: longer time period produce more concentrated results and steeper boundary cumulative distribution. Results obtained for a time period of one year show that, with a confidence level

of 95%, the amount of sediments delivered to the closing section of the basin should belong to the interval 18,724÷69,230 [m³], while the sediment amount computed with the deterministic application of the Gavrilovic method is 51,555 [m³]: obviously this point belongs to the previously forecasted interval, but it carries no information about the uncertainty of the result. Differently, the confidence interval found with the new method contains all the data points correspondent to measures of solid volumes (see Figure 5).
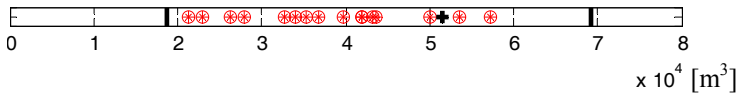

Figure 3:　　Results for the one-year period: dots corresponds to measured volumes of sediments, the cross correspond to the result of traditional application of the Gavrilovic model and vertical bars are the limits of the 95% two-sided confidence interval.

Similarly, the confidence interval for a time period of sixteen years is 410,632–900,977 [m³]. Comparison of the deterministic result, 824,887 [m³], and the measured one, 608,600 [m³], shows that the formed is biased. Traditional application of Gavrilovic provides no clues about this bias. This can be a problem when trying to forecast soil erosion in unstudied basins, where no measure of solid volumes is available. Differently, with the proposed method uncertainty appears in a manner that is understandable and useful to decision makers.

6 Conclusions

A new approach to the Gavrilovic method for reliable assessing of soil erosion was developed and tested on a medium-sized basin of Italian Alps. Hybrid Monte Carlo and possibilistic method was used to process, in a correct way, randomness and imprecision of input data. Unlike traditional deterministic approaches to soil erosion evaluation, the new method produces results in terms of confidence intervals, which provide the decision maker with the degree of uncertainty on the results.

Although further studies on different test sites are necessary to validate the performance of the proposed method, first results show that it can be a powerful tool for quantitative assessment of soil erosion and sediment routing. Since hybrid Monte Carlo and possibilistic method is very general, it is possible to apply it to different soil erosion models to study the phenomena at different scales and with different detail levels.

Acknowledgement

References

[1] De Vente, J. & Poesen J., Predicting soil erosion and sediment yield at the basin scale: scale issues and semi-quantitative models. *Earth-Science Reviews* **71**, pp. 95-125, 2005.

[2] Gavrilovic, S., Bujieni tokovi i erozija. *Gradevinski calendar*, Beograd, Serbia, 1976.

[3] Gavrilovic, Z., The use of an empirical method (erosion potential method) for calculating sediment production and transportation in unstudied or torrential streams. *International Conference on River Regime*, Hydraulics Research Limited, Wallingford, Oxon UK, pp. 411-422, 1988.

[4] Ballio, F., Brambilla, D., Giorgetti, E., Longoni, L., Papini, M. & Radice A., Evaluation of sediment yield from valley scope: a case study. *WIT Transactions of Engineering Sciences* **67**, pp. 149-160, 2010.

[5] Beyer Portner, N., Erosion des basins versant alpins suisses par ruissellement de surface. PhD Thesis, Laboratoire de Costructiones Hydrauliques-LCH, No. 1815, Lausanne, Switzerland.

[6] Globevnik, L., Holjević, D., Petkovšek, G. & Rubinić, J., Applicability of the Gavrilović method in erosion calculation using spatial data manipulation techniques. *IAHS-AISH Publication* 279, pp. 224-33, 2003.

[7] Baraldi, P., Popescu, I. C. & Zio E., Methods of uncertainty analysis in prognostics. *International Journal of Performability Engineering*, **6(4),** pp. 303-30, 2010.

[8] Baudrit, C., Dubois, D. & Guyonnet, D., Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Trans Fuzzy Syst*, **14(5),** pp. 593-608, 2006.

[9] Kalos, M. H. & Whitlock, P. A., *Monte Carlo methods, Volume I: Basics*. Wiley, 1986.

[10] Shafer G., *A mathematical theory of evidence*, Princeton University Press, 1976.

[11] Dubois, D., & Prade, H., Possibility Theory, Probability Theory and Multiple-valued Logics: A Clarification. *Annals of Mathematics and Artificial Intelligence* **32**, pp. 35-66, 2001.