18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Tools for soil erosion mapping and hazard assessment: application to New Caledonia, SW Pacific

Rouet, I. [1], **D. Gay** [1,2], **M. Allenbach** [1], **N. Selmaoui** [1,2], **A.G Ausseil** [3], **M. Mangeas** [4], **J. Maura** [5], **P. Dumas** [4,6], **and D. Lille** [5]

[1] *Pôle Pluri-disciplinaire de la matière et de l'Environnement, EA 3325, Université de la Nouvelle-Calédonie*
[2] *Equipe de Recherche en Informatique et Mathématiques, EA 3791, Université de la Nouvelle-Calédonie*
[3] *Landcare Research, New Zealand*
[4] *Unité ESPACE, US 140, Institut de Recherche pour le Développement*
[5] *Bluecham SAS*
[6] *Centre des Nouvelles Etudes sur le Pacifique, EA 4242, Université de la Nouvelle-Calédonie*
Email: isabelle.rouet@univ-nc.nc

Abstract: Soil erosion is a major issue around the world. In the Pacific Islands, erosion has a strong impact on terrestrial and coastal ecosystems (mountains, alluvial plains, mangroves, coral reefs). Bush fire, deforestation and/or human activities accelerate erosion in mountainous areas especially on lateritic soils. Frequent and intense precipitation events (tropical depressions, cyclones) regularly strip particles from these sensitive soils. Any eroded sediment is quickly transported to coastal plains and to the sea along the main drainage lines, with immediate and recurrent impact on human activities such as open-cast mining, farming and fishing. There is therefore a need to identify key components of erosion processes for sustainable environmental management and planning. Particularly, we need to identify where the main erosion areas are, and what areas are impacted by severe sedimentation. If these areas are properly identified, management and mitigation can become effective. Hazard assessment is also required to plan new developments with limited erosion impacts. However, both mapping of active erosion and the assessment of hazard at a regional scale is time-consuming, costly and rarely updated. In that respect, two approaches have been tested in New-Caledonia:

- Fast mapping of active erosion-linked areas: a Geographical Information System (GIS) based method was developed to model expert analysis from satellite images and topographical data. Bare ground surfaces were first mapped by remote sensing. Automated expert rules were then used to label pixels into different classes in relation to erosion and sedimentation. This method is fast, using only a few thematic layers. It provides a baseline for mapping inventory and for future updates that are essential for monitoring and hazard modelling. The quality of the mapping results are reasonable at the regional scale: the map shows zones of sediment production or accumulation, and aggravated erosion zones such as trails or open pit mines. The few identified errors seem to be easy to correct.

- Optimisation of hazard modelling: a multidisciplinary approach based on data mining and geological knowledge was developed to efficiently assess erosion hazard. The method uses a readily available dataset used in an expert study area and a remote sensing product as training data. The results were consistent with general and local knowledge of erosion, suggesting that data mining is relevant for erosion hazard study. Furthermore, predictive models of erosion (i.e. hazard assessment) built with data mining methods seemed to give more realistic results than expert assessment.

These preliminary results suggest that erosion assessment could become more efficient in the near future in New Caledonia. The mapping method is rapid and efficient and identifies erosion-linked areas in the most sensitive geological regions. Improvements to the method have been identified and a more detailed typology will make classification even more attractive. Likewise, data mining analysis could also be a useful tool for hazard assessment at the regional scale, even with sparse data and a partial inventory. Processing time for basic mapping was greatly reduced and hazard assessment improved with these methods. These semi-automated approaches will lead to better general and local knowledge.

Keywords: *natural hazard, soil erosion, remote-sensing, GIS mapping, data mining*

1. INTRODUCTION

Lateritic soils in the inter-tropical zone are particularly prone to erosion problems (Laganier 1994, Dumas 2004). These soils originate from long processes of weathering, and their degradation by erosion strongly impacts on fragile coastal ecosystems. The runoff following high precipitations events like cyclones or tropical depressions, which are common in the tropical zone, is the main erosion factor. It can lead to a rapid removal of soil, depending on the loss of vegetation cover due to bush fires, deforestation or open-cast mining activities. Once the soil matter has flushed from the hillslopes through sheet or gully erosion (also due to past mining activities), sediments are rapidly transported to the coastal zone through the small catchments in the region. Flash floods following high precipitation events lead to major sediment inputs and periods of hypersedimentation which disturb the coastal area (Laganier 1994, Dumas 2004). Destruction of farming areas and marine habitats (such as coral death) are the main immediate impacts. In New Caledonia, the coastal lagoon has recently been added to the World Heritage site list of the UNESCO which is likely to increase public awareness on the impacts of erosion. This country is also one of the biggest lateritic nickel ore producers in the world. However, mining practices in the past consisting of flushing tailings straight onto the hillslopes, leading to uncontrolled discharges and intense erosion. Runoff in open-cast mines and on trails dramatically increases transfer of sediments to the drainage network and, even if runoff is controlled in newly implemented mines, old mines continue to be a major source of sediments. These issues are additional to a natural sensitivity of soils for erosion, especially in ultramafic zones which are well known for high level of erosion hazard (Maurizot & Lafoy 2003).

Assessment and management of impacts of erosion processes in coastal areas can only be done if data exist for inland catchment areas. However, there is no inventory map for sources and transitional zones of sediments available for the Grande Terre, the main island of New Caledonia. A necessary step consists of identifying these sensitive areas and of assessing their evolution by modelling the hazard. The mapping method needs to be fast in these rapidly evolving inland systems. We need to be able to compare maps to identify new erosion or hypersedimentation areas, especially in uninhabited zones. At present, mapping is very time-consuming because of the geographical extent and the high demand on expert interpretation (field and/or photo-interpretation mapping). Likewise, data collection in the field for hazard modelling is very difficult due to the terrain. Hazard is thus rarely assessed, and only made on the basis of expert knowledge. Two areas of improvements are needed: (1) minimise processing time for mapping and (2) optimise hazard assessment. This paper presents a compilation of recent methods for erosion/sedimentation areas and erosion hazard mapping at regional scales (1:50,000 and 1:100,000 respectively). It was tested in a catchment (~470 km²) located in the "Massif du Sud" ultramafic terrane of New Caledonia.

2. DATA AND METHODS

2.1. Automated mapping

As regional scale expert mapping is too time-consuming, we developped a computer-based method to model expert analysis. The geomorphological analysis has been automated with GIS rules to produce a basic map of active erosion-linked areas. The method is based on satellite image processing and an automated expert rule generation (Rouet et al. 2008) and is divided into three steps:

1. Bare ground mapping by remote sensing: a bare ground surface outside urban areas is likely to be an area subject to erosion or an area of sediment accumulation, which are both active erosion processes. The map was developed from a SPOT 5 image (1B orthorectified, with no pan-sharpening, 10 m resolution) using a method developed by the Bureau de Ressources Géologiques et Minières (BRGM). The detection is based on the brilliance index IB ($IB = \sqrt{XS1^2+XS2^2+XS3^2}$, with $XS1$, $XS2$ and $XS3$ the first three bands of SPOT 5), categorised into 20 equal interval classes, of which only relevant classes are kept. This method has been tested in the past on ultramafic terrane of New Caledonia and is considered to be reliable (Maurizot & Rouet 2006). The resulting map is a rapid inventory of bare ground surfaces. However, there is no distinction within the different bare ground areas, and an additional level of classification is needed.
2. Automated classification into erosion zones (erosion and accumulation areas): basic expert rules have been translated into decision rules in an objective way to model expert analysis using raster GIS. A triangular irregular network (TIN) Digital Elevation Model (DEM) was built using ArcGIS

from 1:10,000 topographical maps (Govt New Caledonia). This was then converted to a raster of 10 m resolution. Track and river vectors were converted into rasters at the same 10 m grid as the DEM. A slope layer was generated from the DEM for the morphological analysis (ArcGIS). The classification followed 3 steps: (i) Identification of bare ground areas over-imposed on tracks and reclassification of pixels into track pixels; (ii) Identification of remaining bare ground areas super-imposed or connected to rivers. These river context pixels are separated between accumulation areas (where slope < R, with R a threshold slope value) and channel erosion (slope > R); (iii) Classification of remaining pixels into hillslope erosion areas (Nickel mines: slope < M; erosion zone: slopes > M, with M a threshold slope value for open-cast mining zones). Local knowledge was used to identify the relevant thresholds for New Caledonia ultramafic terrane: R was set at 6 degrees (Maura 2008) and M was set at 25 degrees (Laganier 1994).

3. Validation.

2.2. Hazard assessment

It is often difficult to assess hazard at a regional scale (Saby 1998, Le Bissonnais et al. 2002, Luneau 2006). In many cases, there is either no data or no data available at a decent scale. The relative importance of erosion factors are usually empiricaly assessed. Hazard models are therefore qualitative, with three or more levels of intensity, and quantification cannot be done. However, hazard maps are very important for decision makers. A data mining approach has been tested for erosion hazard analysis (Gay et al. 2007), with the same spatial dataset used in an expert hazard analysis (Luneau 2006). The data mining approach followed three steps: (i) attribute ranking, (ii) association rule mining and (iii) predictive model building. A pre-processing step was necessary for the association rule mining as discrete attributes were needed.

Pre-processing

A simplified version of hazard assessment was used. The classes of bare ground obtained with the brilliance index were segmented into two larger classes “bare ground” or “not bare ground”. The class layer becomes binary and the various erosion types are not identified at this stage.

Eight descriptive spatial layers (used in Luneau 2006) were used in addition to the class layer: elevation, slope, planform and profile curvature, tracks, vegetation cover, soil type and rainfall. Each descriptive layer is processed at the same grid reference (30 m). Each pixel is therefore described by eight attributes and is labelled “bare ground” or “not bare ground”. Data layer values are then extracted and formatted as a data table for use in the data mining algorithms: each row is a pixel and each column is an attribute. We obtained $9*10^5$ rows. The data is composed of 30,000 bare ground rows (~3%) and 870,000 non-bare ground rows (~97%).

The association rule mining step needs discrete or binary attributes. We therefore applied a discretisation/binarisation method to convert the attribute values. We used an entropy-based supervised discretisation method (Fayyad & Irani 1993): considering a continuous attribute A with values in Dom(A), *n* split points are chosen in Dom(A) with regards to a class entropy-based measure in order to make (*n*+1) intervals of values in Dom(A). A is then divided into (*n*+1) binary attributes. In our tests, data discretisation produced 139 binary attributes. Discretisation/binarisation and the data mining steps have been processed using the WEKA platform (Witten & Frank 2005) – an open-source software providing a collection of machine learning and data mining tools.

Attribute ranking

The first step is a simple ranking of the set of attributes with respect to discriminative measures. We used Information Gain (IG), Gain Ratio (GR) and chi-squared statistics (Shannon 1948). The first ranked attributes are considered the most discriminative, i.e. relevant to characterise bare ground (soil erosion).

Association rule mining

An association rule π is of the form $\pi : X \rightarrow Y$, where X and Y are disjoint sets of attributes. When introduced in Agrawal & Srikant (1994), association rules were used in datasets made of transactions (rows) and items (columns). In our case, transactions are pixels and items are descriptive attributes. First, association rules using frequency and confidence were used. The frequency of a rule π in a dataset r (defined as $freq(\pi,r)$) is the relative number of pixels in data containing X and Y. The confidence is defined as $conf(\pi,r) = freq(\pi,r)/freq(X,r)$. This framework allows us to discover potentially relevant associations between sets of attributes reflecting trends within the database. The goal is to identify the most frequent configurations for existing erosion areas and to check if there are pixels that meet these criteria. A set of association rules are then extracted according to Agrawal & Srikant 1994. A 0.1 frequency means that 10% of the objects (pixels) follow the rule. A 0.7 confidence suggests that when the condition of the rule is met, its consequence is verified in 70% of cases.

Predictive model building

The attribute ranking and the association rules allow to search within the database similar configurations that do not necessarily correspond to bare ground but can lead to it. This method refers to evaluation of the hazard, which is equivalent to a probability of occurrence. Two common classification methods implemented in the WEKA platform were used to model the hazard: Naive Bayes (NB) (John & Langley 1995) and the C4.5 decision tree (Quinlan 1993).

3. RESULTS AND DISCUSSION

3.1. Active erosion area mapping by expert approach modelling

The resulting map (fig. 1A) was checked by flying over the study area by helicopter and by comparing with another map of erosion-linked surfaces that covered parts of the study area (Maurizot et al. 2005). The classification results reach 76% and 65% respectively for accumulation areas and for erosion areas. Classification errors can be attributed to shadows in the satellite image. A pre-processing step of satellite image to remove shadows effect could be added in the future to improve classification (flattening, Dymond & Shepherd 2004).

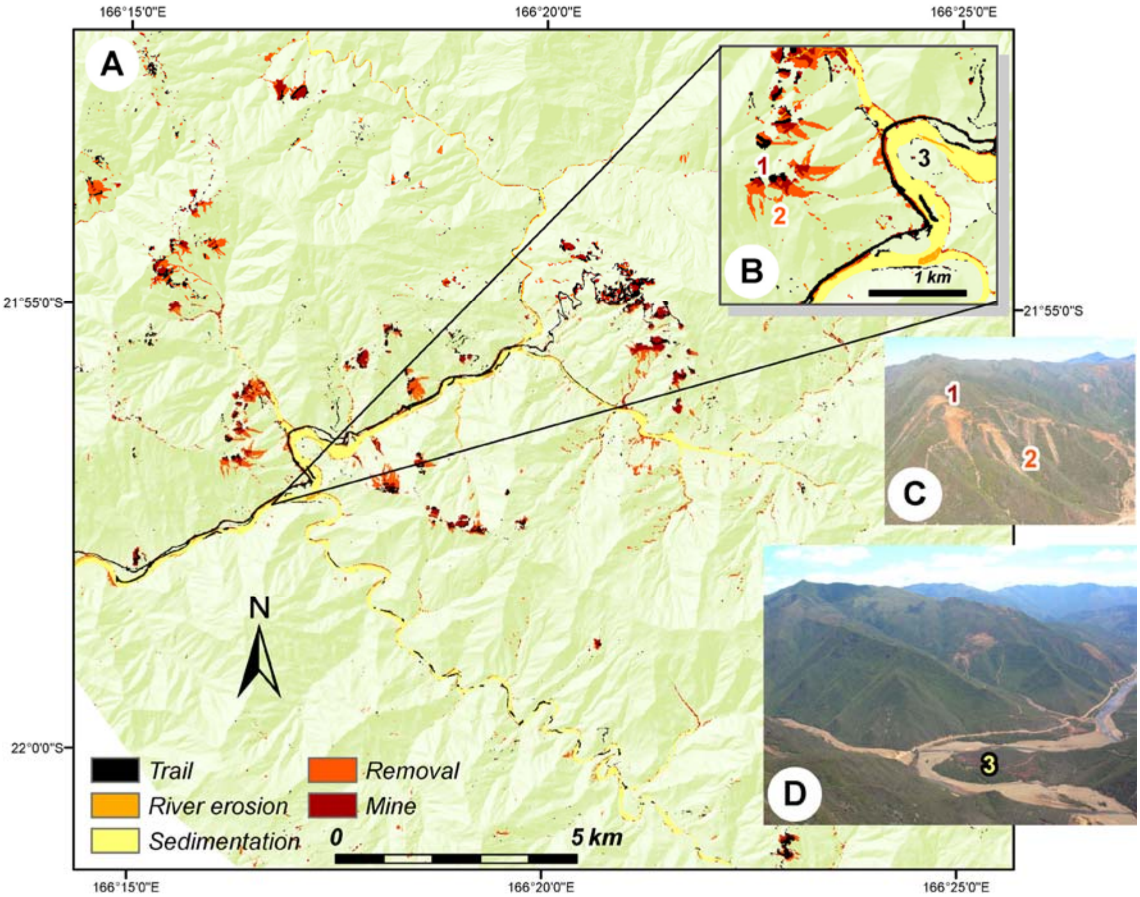

Figure 1: Spatial distribution of the main active areas highlighted by automated mapping (A, 1:100,000 mapping). B details 1:50,000 scale mapping. C & D pictures of landscape of B.

The mapping of mines seemed to be well defined after validation with aerial photographs and flight over the study area. The edges between old mining zones and their immediate impact on the hillslope were correctly identified. However, the M threshold was found to be irrelevant for very old mines, as mineral extraction could happen on a slope over 25 degrees (narrow stope, manual extraction). This does not affect the principle of the method used, as recent mining exploitations usually occur below the M threshold (large stopes and mechanisation). Integrating a map of old mines should resolve this issue. Some labelling errors were occasionally noted close to river systems, where the slope was between 6 and 25 degrees, due to small and narrow vegetation strips. The connectivity criterion could be insufficient, leading to a misclassification as a mine instead of erosion area. A simple adjustment of the connectivity criterion for the bare ground surfaces should reduce these minor errors, for example by choosing neighbouring pixels within a maximal distance of another river context pixel. A shape criterion could also be used.

3.2. Erosion hazard modelling

For hazard assessment, only hillslope erosion areas have been tested.

The attribute ranking results show that vegetation is the most relevant attribute for erosion areas (IG = 0.0398 & GR = 0.021). Depending on the measure used, other high-ranked attributes are soil type (IG = 0.016 & GR = 0.011), rainfall (IG = 0.008 & GR = 0.003), slope (IG = 0.008 & GR = 0.003) and trails (IG = 0.007 & GR = 0.029).

The analysis of association rules has been reduced to 10 samples of approximately 60,000 objects, respecting class distribution, to reduce computing processing time and one extraction has been performed per sample. An association rule was considered relevant if it appeared in extracted rules in each sample. Further investigation is needed to understand why the other landform descriptors seem less relevant. The association rules of attributes defining non-erosion areas identify forest or dense bush on outcropping bedrock. According to the results, erosion is more likely on areas without vegetation (with a 0.97 confidence), on open tracks (with 0.86 confidence) on lateritic soils (0.73 confidence). The results reflect expert knowledge and confirm that the methods tested are appropriate for the characterization of erosion hazard.

Results of classification from Naïve Bayes (NB) and C4.5 classifier methods are obtained by 10-fold stratified cross-validation (table 1). Prediction of non-erosion pixels appeared difficult (18,9 % of erosion pixels are misclassified) whereas prediction of erosion pixels was reasonable (86% and 88% correct classification respectively for NB and C4.5). The C4.5 classifier seemed to give slightly better classification results than the NB classifier, with an overall correct classification of 86% compared to 83%. The Kappa index values suggest that the results are better than would be expected through chance allocation.

		True non-erosion	True erosion	All	Kappa
NB	Predicted non-erosion	81%	19%	83%	0.734
	Predicted erosion	14%	86%		
C4.5	Predicted non-erosion	85%	15%	86%	0.673
	Predicted erosion	12%	88%		

Table 1: Confusion matrix for erosion hazard models for the NB and C45 classifiers.

Predictive models of erosion are shown in Fig. 2A. Fig.2B is an expert hazard map which has been built by a semi-empirical method (Luneau 2006). High hazard areas from the expert map are large (controlled by geology) whereas actual erosion areas are very small and dispersed. The data mining map provides a more realistic view of the spatial distribution of high hazard areas. It gives finer details of high hazard areas, at a level of detail that expert assessment through empirical approaches cannot give.

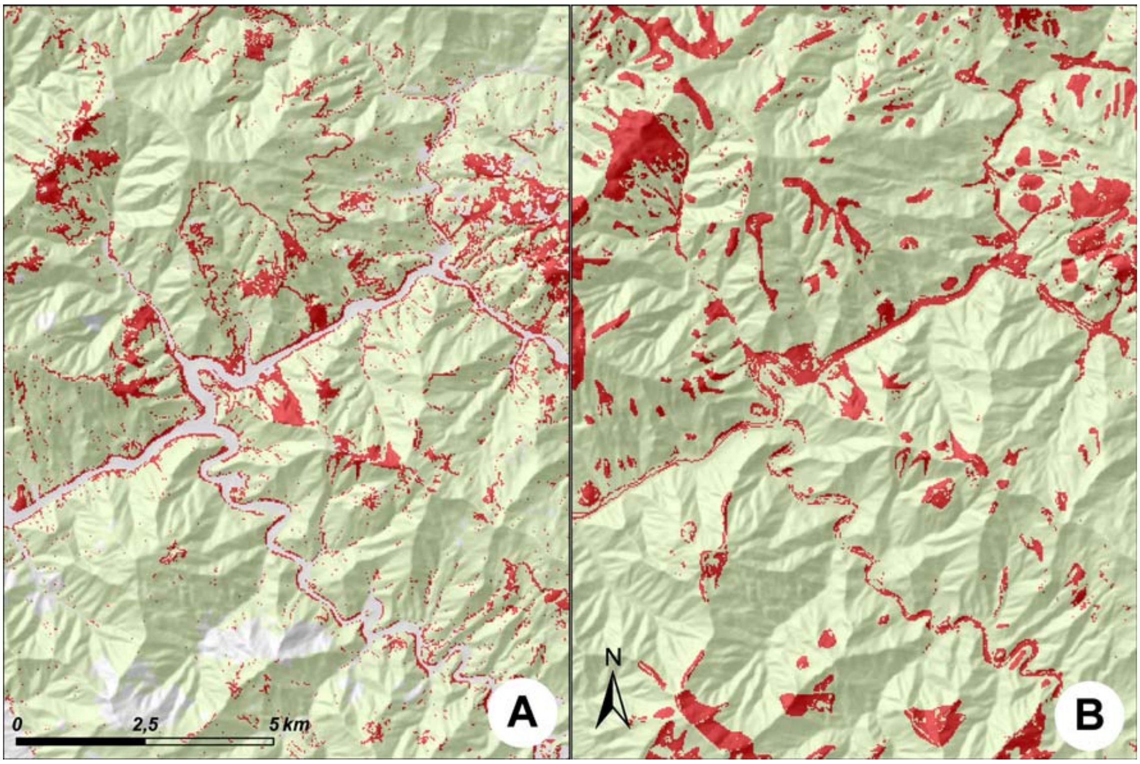

Figure 2: High hazard map performed by C4.5 classifier (A) and from Luneau (2006) (B).

4. CONCLUSION AND PERSPECTIVES

This paper shows methods that can be used as tools for land use and environmental management especially where data is sparse. The mapping method developed gives a rapid and regional view on the processes involved at a given date. This semi-automatic method is about to be optimised. We also plan to improve the method by correcting satellite image for the effects of sun angle, topography and atmospheric conditions. The resulting image (flattened image), without influence of shadows, will be used for automatic mapping of bare surfaces. Spatial resolution can also be improved with pan-sharpened images and very high resolution data will be tested. Applying this method to a temporal series of satellite images will allow monitoring the evolution of the morphogenic activity in the study area, under steady climate condition or after a storm event.

The first results from data mining indicate that the method can be useful for hazard assessment. Attribute ranking and association rules mining described the most contributing attributes for erosion occurrence. Predictive models from data mining algorithms show that hazard maps can be greatly improved, with more detailed maps. As the erosion mapping method improves, a better spatial resolution will be used to assess erosion hazard. The method will then become more attractive and faster to process, making it all the more cost-effective.

ACKNOWLEDGMENTS

The authors would like to thank P. Maurizot (BRGM) for his precious help in data preparation. We are grateful to D. Cluzel (University of New-Caledonia) for support and comments.

REFERENCES

Agrawal, R. and R. Srikant, (1994), Fast algorithms for mining association rules in large databases, Morgan Kaufmann, Actes 20th International Conference on Very Large Data Bases VLDB'94, 487-499.

Dumas, P. (2004), Caractérisation des littoraux insulaires: approche géographique par télédétection et SIG pour une gestion intégrée. Application en Nouvelle-Calédonie, PhD thesis, University of Orléans.

Dymond, J. R. and J. D. Shepherd (2004), The spatial distribution of indigenous forest and its composition in the Wellington region, New Zealand, from ETM+ satellite imagery. Remote Sensing of Environment 90, 116–125.

Fayyad, U. M. and K. B. Irani (1993), Multi-interval discretization of continous-valued attributes for classification learning", dans Actes 13th International Joint Conference on Artificial Intelligence, 1022-1027.

Gay, D., I. Rouet, M. Mangeas, P. Dumas, and N. Selmaoui (2007), Assessment of classification methods for soil erosion risks, dans MODSIM conference proceedings, 2659-2665.

John, G. H. and P. Langley (1995), Estimating continuous distributions in bayesian classifiers, Actes 11th Conference on Uncertainty in Artificial Intelligence, 338-345.

Laganier, R. (1994), Contribution à l'étude des processus d'érosion et des risques naturels dans les îles du Sud-ouest Pacifique (Nouvelle-Calédonie et Iles Salomon), PhD thesis, Science & Technics University of Lille-Flandres-Artois.

Le Bissonnais Y., J. Thorette, C. Bardet and J. Daroussin (2002), L'érosion hydrique des sols en France. INRA & IFEN (Ed.), 105 p.

Luneau, G. (2006), La spatialisation de l'aléa érosion en Nouvelle-Calédonie, Master thesis, ENSAT/SIGMA.

Maura, J. (2008), Méthodologie d'appréhension des phénomènes érosifs naturels et anthropiques dans les processus de gestion d'un patrimoine classé, sous influence industrielle minière, Master thesis, Sup Agro Montpellier.

Maurizot, P. and Y. Lafoy (2003), L'aléa naturel mouvement de terrain en Nouvelle-Calédonie. Synthèse des connaissances, BRGM technical report RP-52213-FR.

Maurizot P., Y. Lafoy, N. Mercier-des-Rochettes and M. Vendé-Leclerc (2005), Cartographie des formations superficielles et des aléas mouvements de terrain en Nouvelle-Calédonie - Zone de Bouloupari. BRGM, technical report RP-53926-FR, 38 p.

Maurizot, P. and I. Rouet (2006), Aide à la mise au point d'une méthodologie de traitement de la couverture territoriale SPOT5 pour le calcul des superficies dégradées par l'ancienne activité minière en Nouvelle-Calédonie, BRGM Rapport méthodologique RP-54787-FR.

Quinlan J. R. (1993), C4.5 : programs for machine learning, Morgan Kaufmann, San Francisco, USA.

Rouet I., D. Gay, M. Allenbach, N. Selmaoui, J. Maura and D. Lille (2008), Semi-automatic detection and labelling of erosion areas using domain knowledge. An appraisal from southeast New Caledonia (SW Pacific), dans Actes SPIE Asia-Pacific Symposium, Noumea, Vol. 7150.

Saby N. (1998), Etudes de méthodes d'évaluation de l'aléa érosif: application à la cartographie de l'aléa érosif sur une zone du sillon inter-andin (Bassin de Quito - Equateur). Master thesis of Ecole Nationale des Sciences Agronomiques de Rennes.

Shannon, C.E. (1948), A Mathematical Theory of Communication, *Bell System Technical Journal*, 27, 379–423 & 623–656.

Witten, I.H. and E. Frank (2005), Data Mining : Practical machine learning tools and techniques (2nd edition). Morgan Kaufmann Publishers Inc., San Francisco, USA.