

Identifying erosion areas at basin scale using remote sensing data and GIS: a case study in a geologically complex mountain basin in the Spanish Pyrenees

Santiago Beguería

Division Landscape Dynamics, GIS and Hydrology - Faculty of Geosciences - Utrecht University. PO Box 80.115, 3508TC Utrecht, The Netherlands

&

Instituto Pirenaico de Ecología - CISC, Campus de Aula Dei.
PO Box 202, 50080 Zaragoza, Spain

e-mail: s.begueria@geo.uu.nl

Inventory and monitoring of eroded areas at basin scale (Mm^2) can be very useful for environmental planning and can help to reduce land degradation and sediment yield to streams. Combined use of remote sensing images and auxiliary geocoded data has been widely used for mapping various environmental features, including surface erosion. Here an example is presented in the Yesa reservoir catchment in the Spanish Pyrenees. Several combinations of radiometric data (a sequence of images from different seasons of the year) and other geocoded information, including topographical (altitude and slope) and geological maps, were compared in their ability to predict previously identified erosive features. Multinomial logistic regression was used as classification method. The datasets were compared in terms of classification error statistics (sensitivity and specificity) using an independent random sample. The incorporation of lithological information improved the discrimination of eroded areas, but the same did not happen in the case of topographical information. Two final maps of eroded areas were obtained applying an equal predicted area rule and an equal error rate rule.

Keywords: erosion, sediment yield, mountain areas, data fusion, multinomial logistic regression, Spanish Pyrenees

1. Introduction

Sheet erosion, headcut advance and landsliding suppose major stress factors on vulnerable lands in mountain ranges and other areas in the world. Human activities like land clearing and deforestation, overgrazing, or in the other hand land abandonment can accelerate the natural rates of these processes, leading to land degradation. The immediate at site effect is the loss of soil. Off-site effects are related to the yield of sediment to the river network, which results in declining water quality and altered stream dynamics. Sediment yield is also the cause of reservoir sedimentation and damage to other hydraulic structures. For these reasons, there is great interest on producing accurate maps of active erosion areas and sediment sources, specially at a basin scale that better corresponds to the needs of environmental decision making.

The use of remote sensing data and GIS techniques have proved to have great potential in land use management and environmental planning at regional scales, offering good resolution and accuracy at low cost. Different methodologies have been applied for inventorying and monitoring the activity of different erosion processes, using band ratios (Pickup and Nelson 1984; Frazier and Cheng 1989), vegetation indices (Pickup and Chewings 1994; Tripathy et al. 1996), linear spectral unmixing (Koch 2000), combinations of reflective and microwave data sources (Koopmans and Forero 1993; García-Meléndez et al. 1998; Singhroy 1995; Singhroy et al. 1998; Metternicht and Zinck 1998), and combinations of remotely sensed data and other geocoded information (Floras and Sgouras 1999; Mati et al. 2000; Shrimalil et al. 2001; Zinck et al. 2001; Haboudane et al. 2002; Giannetti et al. 2001; Ma et al. 2003; Symeonakis and Drake 2004).

The purpose of this study was to develop a map of highly eroded areas in a mountain catchment. The area is characterized by high geologic heterogeneity, which hinders the discrimination of the areas affected by erosion. More specifically, erosion on bare soil or erodible rock outcrops had to be distinguished from resistant rock outcrops, which are very frequent in a mountain environment and tend to be spectrally similar to the former. To achieve this, a supervised classification procedure (multinomial logistic model) was used, and different datasets were checked. The ability of multi-temporal data (integration of images from different seasons) for discriminating erosion features was tested, and compared to the use of single images. The inclusion of auxiliary geocoded information, such as digital terrain model (DTM) or a lithological map was also tested. The various datasets were compared in their ability to predict observed eroded areas using an independent random sample, based on error rate (sensitivity and specificity) statistics.

2. Study area

The study area corresponds to the catchment of the Yesa reservoir, in the Central Spanish Pyrenees, with a total surface of 2191 km² (figure 1). Elevation ranges between 500 and 2800 m a.s.l., showing a general gradient towards the North. The geological setting is complex, and can be divided into four main units. From South to North these are: i) the Molasic basin, consisting mainly on lutites, sandstones and conglomerates, with a relatively high and steep relief; ii) the Inner depression, built on Eocene marls with a highly erodible behaviour; iii) the Eocene Flysch sector, consisting on thin alternating beds of calcareous sandstone and marls, generating moderate relief and medium erodibility; iv) the Inner Sierras, an overthrusting anticline composed of

Cretaceous and Paleocene limestone and sandstone; and v) the Axial area, consisting on mixed paleozoic rocks including highly erodible slates, greywackes and lutites. The main erosive features are related to gullying and head cutting in the subalpine belt due to forest elimination, diffuse erosion due to overgrazing and badland formation in specially favourable lithologies like marls.

Built in 1959 in the Aragón River, the Yesa reservoir has currently a storage capacity of 447 hm³. The reservoir experienced a process of siltation with sediments carried by the Aragón river that represented a loss of 21 hm³ by 1986, or 0.39 hm³ per year (López-Moreno et al., 2004). Determining the location and extension of severe erosive features within the reservoir's catchment can be of great value for environmental managers, allowing to plan specific actions to reduce land degradation and the yield of sediment to the river system that would ultimately affect the reservoir.

[Insert fig. 1 about here]

3. Data and methods

3.1. Data selection and preparation

I selected a set of three Landsat TM/ETM+ scenes corresponding to the months of June, August and October. The images were orthorectified using the algorithm of Palà and Pons (1995), as implemented in MiraMon GIS software. A coupled atmospheric and topographic correction was then performed following the method proposed by Gilabert et al. (1994) and a bi-directional reflectance model based on Minnaert (1961), to finally obtain ground reflectance data. Due to the existence of high correlation between different bands in the same image and between the same bands on correlative images, that highly affects the classification procedure, a factorial analysis was

performed on the images. The result was a reduction of the 18 spectral bands to 9 factorial bands, accounting for 96.8% of the original variance.

The auxiliary data set comprised a digital terrain model (DTM) consisting on elevation and slope maps, and a lithological map, having the same spatial resolution than the Landsat images. The elevation grid was interpolated from digitised isoline maps at 1:50000 scale, and the slope gradient was calculated by finite differences over a 3x3 window. The lithology map was elaborated from the 1:50000 geological maps of the Spanish Geological Survey, after rearranging the different legends in 12 common lithology classes. The lithology classes were classified into two groups: i) resistant rocks presenting a high resistance to erosion by runoff, and ii) erodible rocks with low resistance to erosion. The first group included active streambed and gravity deposits, sandstone, conglomerates, limestone and carbonatic sandstone. The erodible rocks included the moraine deposits, marls, marls with limestone, flysch, greywacke and slates.

Eight different data sets were prepared, in order to compare their ability for discriminating the erosion areas. The data sets were: unique satellite image (D1: spring image; D2: summer image; D3: autumn image), multi-temporal image (D4: factorial composition of spring, summer and autumn images), image composite and DTM (D5: incorporating slope gradient; D6: incorporating elevation), image composite and lithology (D7); and image composite, slope gradient and lithology (D8).

The use of multi-temporal radiometric information usually enhances the discrimination among vegetation types, benefiting from differences in the phenology of species. Bare soil and rock outcrops do not usually show large spectral differences along the year, although changes in the water content can be expected which can modify

the spectral response. Also, multi-temporal information can help in discriminating erosion areas from low coverage grasslands and alpine pastures with a short growing period, which can be confused with bare soils early in the season. Elevation and slope gradient were included due to their expected correlation with erosional features, specially in the case of the slope gradient. Finally, lithological information represents here a-priori knowledge about the substrate erodibility.

3.2. Selection of target land cover classes and sampling

The presence of bare soil and low resistant rock outcrops can be used in the study area as an indication of the existence of active erosion processes (severe sheet erosion, gully erosion and landsliding), with independence of the lithology group. This hypothesis was confirmed during several field campaigns, and defined the areas presenting bare soil or erodible outcrops as the target for the classification process.

A set of 20 land cover classes (Table 1) was defined in order to cover the natural variability on the study area and maximise the separability between bare soil / erodible outcrop classes and the rest of land cover classes. Vegetated areas were classified into deciduous forest, pine forest, shrubland, alpine pastures, meadows and crops. Special attention was given to the bare soil areas (less than 10% vegetation coverage, estimated from aerial photographs), which constituted the main objective of the classification process. They were classified into several kinds of resistant rocks and erodible rocks (Table 1).

[Insert table 1 about here]

Various training areas were selected at different spatial locations for each class, in order to capture all the possible radiometric and topographical intra-class variability and

to reduce self-correlation problems. A training sample was obtained extracting 200 individuals (pixels) for each class, for a total of 4000 cases. The limestone outcrops class was divided in two classes to reflect the different subpopulations that aroused from the previous analysis of the spectral signatures.

An independent random sample containing 510 points from the whole study area was also generated. To ensure an even spatial distribution of the points, the area was divided into 170 squares of 3x3 km, and three points were chosen randomly within each square with a minimum distance of 250 m between them. The points were then visually classified using high resolution aerial orthophotos. Each point was classified as either bare soil / erodible outcrop or vegetated / resistant outcrop. This sample was used to estimate a priori probabilities based on the proportions of each class in the random sample. Since this sample was not used for calibration, it was also employed at the validation stage.

3.3. Classification method

I chose multinomial logistic model as the classification method. The main advantage of logistic regression is its robustness against violation of normality and homocedasticity assumptions (Hosmer and Lemeshow, 2000), which is often the case in spectral imaging analysis. Logistic regression is a specific case of general linear model with a logit transfer function. A linear relationship is adjusted for the response category between the logarithm of the odds ratio and the independent variables:

$$\ln\left(\frac{p(y_1 | \mathbf{x})}{p(y_0 | \mathbf{x})}\right) = \mathbf{B} \cdot \mathbf{x} \quad (\text{eq. 1})$$

where $p(y_i | \mathbf{x})$ and $p(y_o | \mathbf{x})$ are the posterior probability of membership and non-membership to class Y , respectively, given the observation vector \mathbf{x} , and \mathbf{B} is a coefficient vector. Equation 1 can be developed for the posterior membership probability as:

$$p(y_i | \mathbf{x}) = \frac{\exp(\mathbf{B} \cdot \mathbf{x})}{1 + \exp(\mathbf{B} \cdot \mathbf{x})} \quad (\text{eq. 2})$$

The values of the coefficient vector \mathbf{B} were obtained using a maximum likelihood algorithm (SPSS 13.0).

The training sample obtained previously was used for calibrating the logistic model. This resulted in a classification model with 20 categories, which were later grouped into erodible and non erodible corresponding to the target of the study. The use of equal size sample groups is optimum for the calibrating algorithm, but the resulting posterior probabilities are biased with respect to the reality. This occurs because the model tends to reproduce in its output the frequency distribution of the classes as it is in the training sample. For this reason, the model needs to be corrected to yield unbiased results corresponding to the real proportions of each class. Following the Bayes' theorem this can be done using the following expression:

$$p_u(y_i | \mathbf{x}) = \frac{\frac{p_u(y_i)}{p_b(y_i)} p_b(y_i | \mathbf{x})}{\sum_{j=0}^1 \frac{p_u(y_j)}{p_b(y_j)} p_b(y_j | \mathbf{x})} \quad (\text{eq. 3})$$

where $p_b(y_i | \mathbf{x})$ represent the biased posterior probabilities as obtained using eq. 2, $p_b(y_i)$ are the biased prior probabilities, i.e. the proportion of each class in the training sample, and $p_u(y_i)$ are the unbiased prior probabilities, obtained from an independent

random sample. $p_b(y_I)$ in was 0.35 in this case, and $p_u(y_I)$ was estimated from the independent sample to be 0.175.

For comparison purposes the pixels where classified using the maximum probability rule. An individual was classified as bare soil / erodible outcrop if the unbiased posterior probability was higher than the prior probabilities in the calibration sample ($p_u(y_I) = 0.35$).

3.4. Validation and model comparison

Validation statistics were calculated by comparing the predicted erosion status with the observations in the independent sample. This allowed to compare the predictive power of each dataset. Various validation statistics can be formulated from the terms in the confusion matrix (table 2).

[Insert table 2 about here]

Values a and d in table 2 represent the proportion of correctly predicted positive and negative individuals, respectively. b represents the proportion of false positives, i.e. individuals incorrectly classified as belonging to class Y , commission error or error type I, and c represents the number of false positives, i.e. individuals incorrectly predicted as not belonging to class Y , omission error or error type II. A broadly used validation statistic is the success rate, which is the ratio of correct predictions to the total number of individuals, N :

$$\text{Success Rate} = \frac{a + d}{N} \quad (\text{eq. 4})$$

An important drawback of the success rate is that it is highly affected by the proportion of the positive and negative cohorts ($a+c$ and $b+d$, respectively) in the validation sample. This results in success rates biased in favour of the majority cohort, which may easily lead to wrong conclusions if this effect is ignored (Kubat et al., 1998; Beguería, in press). However, alternative metrics can be derived which overcome this problem, like the model's sensitivity and specificity (see, i.e., Forbes, 1995; Fielding and Bell, 1997):

$$S = \frac{a}{a+c} \quad (\text{eq. 5})$$

$$Sp = \frac{d}{b+d} \quad (\text{eq. 6})$$

The model's sensitivity expresses the proportion of positive individuals correctly predicted, or the probability that an individual belonging to class Y would be correctly predicted, $p(Y'_1|Y_1)$. The model's specificity is the proportion of negative individuals correctly predicted, or the probability that an individual not belonging to class Y would be correctly predicted, $p(Y'_0|Y_0)$. A highly sensitive model is one with great ability to correctly predict positive individuals. A highly specific model is one which do not yield many false positives, so it is very specific to the class which is being predicted. The values (1-sensitivity) and (1-specificity) represent the probabilities of committing error type II, $p(Y_0|Y_1)$, and error type I, $p(Y_1|Y_0)$, respectively. This statistics are not affected by the proportion of the two cohorts in the sample, since they only refer to either the positive or the negative cohorts.

The previous statistics are valid for comparing different predicted classifications, once a threshold value has been set for the posterior probabilities. However, different classifications might be obtained by raising or lowering the threshold value, resulting in different sensitivity/specificity (or error rate) values. A plot can be easily constructed

showing the sensitivity/specificity pairs yielded by classification models at all possible classification thresholds. The line resulting from joining this points is called the ROC (receiver operating characteristic) curve, and it has been used to determine the overall predictive power of a model, independently of the classification threshold set (Deleo, 1993). The area under ROC curve (AUR) statistic is often used as a global validation statistic, and it is included in the major statistical packages. It varies between 0.5 and 1, the former meaning no predictive power and the latter maximum predictive power.

Statistical criteria have been developed to determine whether the difference in AUR observed between two alternative algorithms or datasets are more likely to be random than real. In the case of a paired sampling, as is the case here (the same points are sampled for the different datasets), the test must consider the correlation between the areas under the ROC curve. Here I have used the test of Hanley and McNeil (1983).

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2r SE_1 SE_2}}, \quad (\text{eq. 7})$$

where A_1 and A_2 are the AUR of the two datasets, SE_1 and SE_2 are the AUR standard errors, and r is the estimated correlation between the two areas. The value of r depends on the average correlation between the posterior probabilities of the two cohorts and the average AUR (see table 1 of Hanley and McNeil, 1983).

The statistics used for comparing the different dataset were the success rate, sensitivity, specificity and the AUR. The proportion of the area that was predicted as eroded was also reported for comparison.

3.5. Error cost based classification and elaboration of the predictive map

As it was anticipated in the previous section, different classification models can be obtained by adopting alternative classification thresholds. The last step of the process, which is the elaboration of the predictive map, should not be based arbitrarily on the maximum probability classification. Different alternatives should be considered, in terms of the sensitivity/specificity pairs, or, equivalently, of the error rates. The final use of the predictive map should also be considered. In some applications (e.g., prediction of potential habitats for endangered species) the cost of false negatives can be very different than that of false positives. An error cost function can be defined by weighting the two types of error, and a classification threshold can be set to the value that minimises the total cost (Provost and Fawcett, 1998).

In many cases, however, the error costs are not clear, or they can be considered equivalent. The threshold could be set to the value that yields a total positive predicted area similar to the estimation obtained from the independent sample. Or else, a threshold could be found to balance the two error types.

In this case, a map was obtained by using the two above mentioned rules, i.e. equal predicted area and equal error rates.

4. Results

Table 3 shows the validation statistics obtained for the various datasets, based on the maximum probability classification rule. The results of the biased model are shown for comparison. The AUR statistic is the same for both biased and unbiased models, because the transformation on eq. 3 only has the effect of shifting the posterior probability distribution function, but it does not change its shape, what results in identical ROC-plot curves for both models (invariance property of the ROC plot).

All datasets tended to show high success rates, which were above 0.8 except in the case of dataset D3 (autumn image), which in general had the poorest performance of all. However, it is possible that this high results were influenced by the low proportion of eroded individuals in the validation sample, resulting in very optimistic values. This is confirmed to some point by the other validation statistics employed. Indeed, all the datasets showed low sensitivity values (around 0.4), indicating relatively poor ability to correctly predict positive cases. The specificity, however, was high for all the datasets (above 0.9 in most of the cases), meaning good ability to correctly predict negative cases.

The differences between the biased and unbiased models is clear comparing the values on table 3. The biased model tended to overestimate the occurrence of erosive features, as shown by the area predicted as eroded, which is in general higher than the estimated real value of 17.5%. The values of the unbiased model are much closer to this figure, indicating a more reliable picture of the extension of erosive features in the area.

The AUR statistic was good to very good. The AUR is the best statistic for comparing between datasets, since it is independent of the threshold value used. In general, adding more information resulted in better predictive ability, in agreement with the expectations. Among the three images, the spring (D1) and summer (D2) ones showed higher performance than the autumn one (D3), although the difference in AUR was not significant. The combination of all the images together (D4) was significantly better than the spring and autumn single images. The addition of slope gradient data (D5) further improved the predictive ability, although the change was very small and not significant with respect to the image combination alone (D4) or the summer image (D2). The addition of elevation data (D6) significantly decreased the predictive ability with respect to previous datasets. Finally, the incorporation of lithological data (D7, D8)

resulted in a significant increase in the AUR, specially without the concurrence of slope gradient (D7).

[Insert table 4 about here]

Considering this results, the dataset consisting on the image composite and the lithological data was selected for constructing the prediction map. Two alternative classification rules were set, equal predicted area and equal error rate rules. For the first case, a threshold value was found which yielded 17.5% positive predicted area, coinciding with the estimations from the independent sample. This resulted in slightly higher sensitivity and lower specificity than in the case of the maximum probability rule. For the second case the threshold was set to yield equal sensitivity and specificity values, which were 0.769. In this case the positive predicted area increased up to 32.4% of the total area. The resulting predictive maps are shown in figure 2.

[Insert figure 2 about here]

Large part of the eroded areas in the Yesa reservoir catchment appear related to marls outcrops in the Inner Depression, which are very close to the reservoir mouth and hence represent the highest risk of sediment delivery to it. Other large eroded areas are located in the headwaters of the Aragón River and its main tributaries in the Inner Sierras and the Axial zone. Erosional features appear related to various erodible rock types (upper part of flysch, greywackes, slates) and moraine deposits.

5. Discussion

5.1. Predictive ability for discriminating eroded areas

The good results of the validation support the hypothesis of the utility of combined radiometric and auxiliary data for mapping the spatial distribution of eroded areas. However, the predictive model appeared to be not too sensitive, meaning that it had difficulties in correctly discriminating positive (eroded) individuals, even with the most successful datasets. The specificity, on the contrary, was high in all cases, indicating a good ability to correctly predict the negative cases (non eroded). For obtaining a good sensitivity rate (above 0.75, model b in table 4) it was necessary to classify 32.4% of the study area as eroded, which almost doubled the value estimated from an independent sample. This facts couldn't be revealed by only considering the success rate, which is the only validation statistic used in many studies, so the use of the sensitivity/specificity pair is recommended for any further classification application.

5.2. Comparison between different datasets

It was found that, in general, the incorporation of additional information to the dataset improved the prediction ability, as was expected. The statement did not hold, however, for all kinds of data. The combination of three images from different moments in the year only allowed for very little improve of less than 0.02 AUR. This is most certainly due to very little variation in bare surface reflectance along the year, resulting in high redundancy between images. This result seriously questions the usefulness of multi-temporal image datasets for discriminating bare soil areas, and discourages undertaking the effort to process additional images if they are not available already to the researcher, or else there is evidence of temporal variation on the spectral response of bare ground in a specific study case.

The incorporation of DTM data (elevation and slope) did not improve the predictive ability much further, and even reduced it significantly in the case of the elevation. This was due to the high heterogeneity of the study area, which is composed of many different lithological formations and presents a complicated relief. As a consequence, eroded areas appear in many different configurations of elevation and slope, making difficult its discrimination using a linear regression model. It might be that the use of other regression techniques, like regression-trees, would allow to capture this information, although this would always require very careful selection of training areas to include all the possible configurations. It is possible, too, that in other, more homogeneous areas topographical information would improve the prediction of erosional features. Since topographical data in the form of digital terrain models is currently available for many parts of the world, it is suggested that this information should be systematically incorporated and its usefulness tested using techniques similar to the ones applied here.

The addition of lithological data, on the contrary, revealed to significantly improve the prediction ability. This result was expected, since one the key issues in the present application was to discriminate between resistant and erodible rock outcrops. Identical positive effect can reasonably be expected on similar applications in other study areas, and the effort of developing a geological map can be justified in many occasions.

5.3. Evaluation of the prediction map

The final product of the classification process was, of course, a map of heavily eroded areas. The approach followed here yielded two different results. The first one

represents a reliable picture of the total extent of eroded areas within the study area, as estimated using an independent random sample. This map, however, has a relatively poor sensitivity. Consequently, the exact location of the eroded areas may be imprecise, and high uncertainty remains due to false negatives (areas which were not predicted as eroded, but in fact they are). The second map corresponds to an equal error scenario, which means that the two types of errors (false negatives and false positives) are balanced. This map has higher sensitivity, so it is expected to include most of the actually eroded areas, but at the expense of having a high number of false positives too. Whereas the first map is a plausible picture of the extension of erosion in the study area, the second one represents a more conservative model, in which the predicted eroded area is almost double in size than the estimation from the independent sample. The option for a final solution should depend on the special needs of the end-users of the map.

6. Conclusions

The performance of different datasets including radiometric (Landsat) and other geo-referenced data in detecting eroded areas in a geologically complex mountain basin was tested. Of particular interest to this study was the discrimination of highly erodible rock outcrops like marls or fractured slates from other resistant rock outcrops, like limestone.

It is known that multi-temporal image datasets usually improves the predictive power of classification models, specially in the case of vegetation mapping. It failed, however, to substantially improve the discrimination of erodible bare rock outcrops.

The incorporation of DTM data (elevation and slope) did not improve the classification either, due to the complex configuration of the area, consisting in various kinds of rocks and a complex relief. The results in other areas can greatly vary, so the usefulness of DTM data should be assessed for every specific case.

The addition of geological information by means of a lithology map provided the best results of all the tested datasets, so it was selected for making the final map of eroded areas. Two classification rules based on the sensitivity/specificity (or error rates) pairs were used. An equal predicted area rule yielded a reliable estimation of the total area presenting erosional features, although it had some degree of uncertainty due to a high rate of false negatives. The equal error rates rule yielded a more balanced model with higher sensitivity, but resulted in a conservative map in which the extent of the predicted eroded area almost doubles the area estimated from an independent random sample.

Acknowledgments

This study has been supported by research projects PIRIHEROS (REN2003-08678/HID) and CANOA (CGL2004-04919-C02-01), funded by CICYT (Spanish Ministry of Science and Technology) and EU-FEDER, and "Programa de grupos de investigación consolidados" (BOA 48 of 20-04-2005), funded by Aragón region government. The author also wishes to acknowledge personal support by a post-doctoral grant funded by the Spanish Government Secretary for Education and Universities and the EU Social Fund.

References

BEGUERÍA, S., in press. Validation and evaluation of predictive models in hazard assessment and risk management. *Natural Hazards*.

DELEO, J. M., 1993. Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. In: *Proceedings of the Second International Symposium on Uncertainty Modelling and Analysis*, Computer Society Press, College Park, pp. 318–325.

FIELDING, A.H., BELL, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence / absence models. *Environmental Conservation*, 24(1): 38-49.

FLORAS, S.A. and SGOURAS, I.D., 1999, Use of geoinformation techniques in identifying and mapping areas of erosion in a hilly landscape of central Greece. *International Journal of Applied Earth Observation and Geoinformation*, 1:68–77.

FRAZIER, BE and CHENG, Y., 1989, Remote sensing of soils in the eastern Palouse region with Landsat Thematic Mapper. *Remote Sensing of Environment*, 28:317–325.

FORBES, A.D., 1995. Classification-algorithm evaluation: five performance measures based on confusion matrices. *Journal of Clinical Monitoring*, 11(3): 189-206.

GARCÍA-MELÉNDEZ, E., MOLINA, I, FERRÉ-JULIÁ, M and AGUIMET, J., 1998, Multisensor data integration and GIS analysis for natural hazard mapping in a semiarid area (Southeast Spain). *Advanced Space Research*, 21: 493–499.

GIANNETTI, F., MONTANARELLA, and L., SALANDIN, R., 2001, Integrated use of satellite images, DEMs, soil and substrate data in studying mountainous lands. *International Journal of Applied Earth Observation and Geoinformation*, 3:25–29.

GILABERT M.A., CONESE C. and MASELLI F., 1994, An atmospheric correction method for the automatic retrieval of surface reflectance from TM images. *International Journal of Remote Sensing* 15:2065–2086.

HABOUDANE, D., BONN, F., ROYER, A., SOMMER, S. and MEHL, W., 2002, Land degradation and erosion risk mapping by fusion of spectrally-based information and digital geomorphometric attributes. *International Journal of Remote Sensing*, 23: 3795–3820.

HANLEY, J.A., McNEIL, B.J., 1983, A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148: 839–843.

HOSMER, D.W. and LEMESHOW, S., 2000, *Applied logistic regression* (Wiley: New York), 373 pp.

KUBAT, M., HOLTE, R., MATWIN, S., 1998. Machine learning for the detection of oil spills in satellite images. *Machine Learning*, 30, 195–215.

MA, J. W., XUE, Y., MA, C.F., WANG, Z.G., 2003, A data fusion approach for soil erosion monitoring in the Upper Yangtze River Basin of China based on Universal Soil Loss Equation (USLE) model. *International Journal of Remote Sensing*, 24: 4777–4789.

KOCH, M., 2000, Geological controls of land degradation as detected by remote sensing: a case study in Los Monegros, north-east Spain. *International Journal of Remote Sensing*, 21:457– 473.

KOOPMANS, B.N. and FORERO, R.G., 1993, Airborne SAR and Landsat MSS as complementary information source for geological hazard mapping. *Photogrammetry & Remote Sensing*, 48: 28–37.

LÓPEZ-MORENO, J.I., BEGUERÍA, S., GARCÍA-RUIZ, J.M., 2004. Storage regimes of the Yesa Reservoir, upper Aragón river basin, Central Spanish Pyrenees. *Environmental Management*, 34(4): 508–515.

MATI, B.M., MORGAN, R.P.C., GICHUK, F.N., QUINTOR, J.N., BREWER, T.R. and LINIGER, H.P., 2000, Assessment of erosion hazard with the USLE and GIS: A case study of the Upper Ewaso Ng'iro North basin of Kenya. *International Journal of Applied Earth Observation and Geoinformation*, 2: 78–86.

METTERNICHT, G.I. and ZINCK, J.A., 1998, Evaluating the information content of JERS-1 SAR and Landsat TM data for discrimination of soil erosion features. *ISPRS Journal of Photogrammetry & Remote Sensing*, 53:143–153.

MINNAERT M., 1961, Photometry of the Moon. In G.P. Knipper and B.M. Middlehurst (eds.), *The Solar System. III: Planets and Satellites* (Chicago University Press), pp. 213–248.

PALÀ V. and PONS X., 1995, Incorporation of relief in polynomial-based geometric corrections. *Photogrammetric Engineering and Remote Sensing*, 61:935–944.

PICKUP, G. and CHEWINGS, V.H., 1988, Forecasting patterns of soil erosion in arid lands from Landsat MSS data. *International Journal of Remote Sensing*, 9:69–84.

PICKUP, G. and NELSON, D.J., 1984, Use of Landsat radiance parameters to distinguish soil erosion, stability and deposition in arid central Australia. *Remote Sensing of Environment*, 16:195–209.

PROVOST, F and FAWCETT, T., 1998. Robust classification systems for imprecise environments. In *Proceedings of the fifteenth national conference on artificial intelligence*. American Association for Artificial Intelligence, Maddson, WI.

SHRIMALIL, S.S, AGGARWALZ, S.P. and SAMRAL, J.S., 2001, Prioritizing erosion-prone areas in hills using remote sensing and GIS - a case study of the Sukhna Lake catchment, Northern India. *International Journal of Applied Earth Observation and Geoinformation*, 3:54–60.

SINGHROY, V., 1995, SAR integrated techniques for geohazard assessment. *Advanced Space Research*, 1:67–78.

SINGHROY, V., MATTAR, K.E. and GRAY, A.L., 1998, Landslide characterisation in Canada using interferometric SAR and combined SAR and TM images. *Advanced Space Research*, 21:465–476.

SYMEONAKIS, E. and DRAKE, N., 2004, Monitoring desertification and land degradation over sub-Saharan Africa. *International Journal of Remote Sensing*, 25:573–592.

TRIPATHY, G.K., GHOSH, T.K., and SHAH, S.D., 1996, Monitoring of desertification process in Karanataka state of India using multi-temporal remote sensing and ancillary information using GIS. *International Journal of Remote Sensing*, 17:2243–2257.

ZINCK, J.A., LÓPEZ, J., METTERNICHT, G.I., SHRESTHA, D.P., and VÁZQUEZ-SELEM, L., 2001, Mapping and modelling mass movements and gullies in mountainous areas using remote sensing and GIS techniques. *International Journal of Applied Earth Observation and Geoinformation*, 3:43—53.

Utrecht, 16 September 2006

Table 1. Land cover classes.

<u>Vegetated</u>	
1	Beech / Fir
2	Oak
3	Pine
4	Shrubland
5	Alpine pastures
6	Meadows
<u>Bare ground - non erosional</u>	
11	Riverbed deposits (gravel bars)
12	Gravity deposits
13	Sandstones
14	Conglomerates
15	Limestones 1
16	Limestones 2
17	Carbonatic sandstones
<u>Bare ground - erosional</u>	
21	Alluvial and moraine deposits
22	Marls
23	Marls / limestones
24	Flysch
25	Greywackes
26	Slates
27	Exposed soil

Table 2. Confusion matrix: Y_1 , belonging to class Y ; Y_0 , not belonging to class Y . Values a and d are the true positives and true negatives, and b and c the false positives and false negatives, or error type I and error type II, respectively.

		Observed	
		Y_1	Y_0
Predicted	Y'_1	a	b
	Y'_0	c	d

Table 3. Validation statistics for the eight data sets: D1, spring image; D2, summer image; D3, autumn image; D4, spring, summer and autumn images; D5, composite image + slope gradient; D6, composite image + elevation; D7, composite image + lithology; D8, composite image + slope gradient + lithology. Values are shown corresponding to a classification based on the maximum probability rule, for both unbiased (a) and biased (b) models. Equal AUR superscript letters indicate significant differences between datasets (confidence level $\alpha=0.05$).

General:	D1	D2	D3	D4	D5	D6	D7	D8
Threshold	0.350	0.350	0.350	0.350	0.350	0.350	0.350	0.350
AUR	0.745 ^a	0.769 ^b	0.714 ^c	0.786 ^{acd}	0.793 ^{ace}	0.645 ^{abdef}	0.889 ^{abcdef}	0.845 ^{acef}
(sd. error)	(0.029)	(0.031)	(0.034)	(0.026)	(0.027)	(0.039)	(0.016)	(0.023)
Unbiased model:								
Predict. area	17.7%	13.2%	17.7%	16.7%	13.4%	11.3%	14.2%	10.5%
Success rate	0.808	0.845	0.796	0.821	0.835	0.825	0.847	0.864
Sensitivity	0.456	0.433	0.422	0.467	0.411	0.322	0.467	0.411
Specificity	0.882	0.932	0.875	0.896	0.925	0.932	0.927	0.960
Biased model:								
Predict. area	30.3%	24.9%	39.6%	24.5%	9.4%	12.6%	20.2%	15.0%
Success rate	0.720	0.794	0.666	0.775	0.810	0.835	0.817	0.854
Sensitivity	0.567	0.622	0.678	0.556	0.511	0.389	0.556	0.511
Specificity	0.753	0.831	0.664	0.821	0.873	0.929	0.873	0.927

Table 4. Validation statistics for classification model based on dataset D7 (three images and lithology), based on equal predicted area rule (a) and equal error rate rule (b). Statistics for the maximum probability rule are also shown for comparison (c).

	a	b	c
Threshold	0.225	0.062	0.350
Predict. area	17.5%	32.4%	14.2%
Success rate	0.833	0.769	0.847
Sensitivity	0.522	0.767	0.467
Specificity	0.899	0.769	0.927

Figure captions

Figure 1. Location of the study area within the Iberian Peninsula. Elevation map (1) and geology (2). Main geo-structural units are: Molasic basin (a); Inner depression (b); Flysch sector (c); Inner Sierras (d); Axial zone (e).

Figure 2. Map of highly eroded areas, based on dataset D8 and two classification rules.

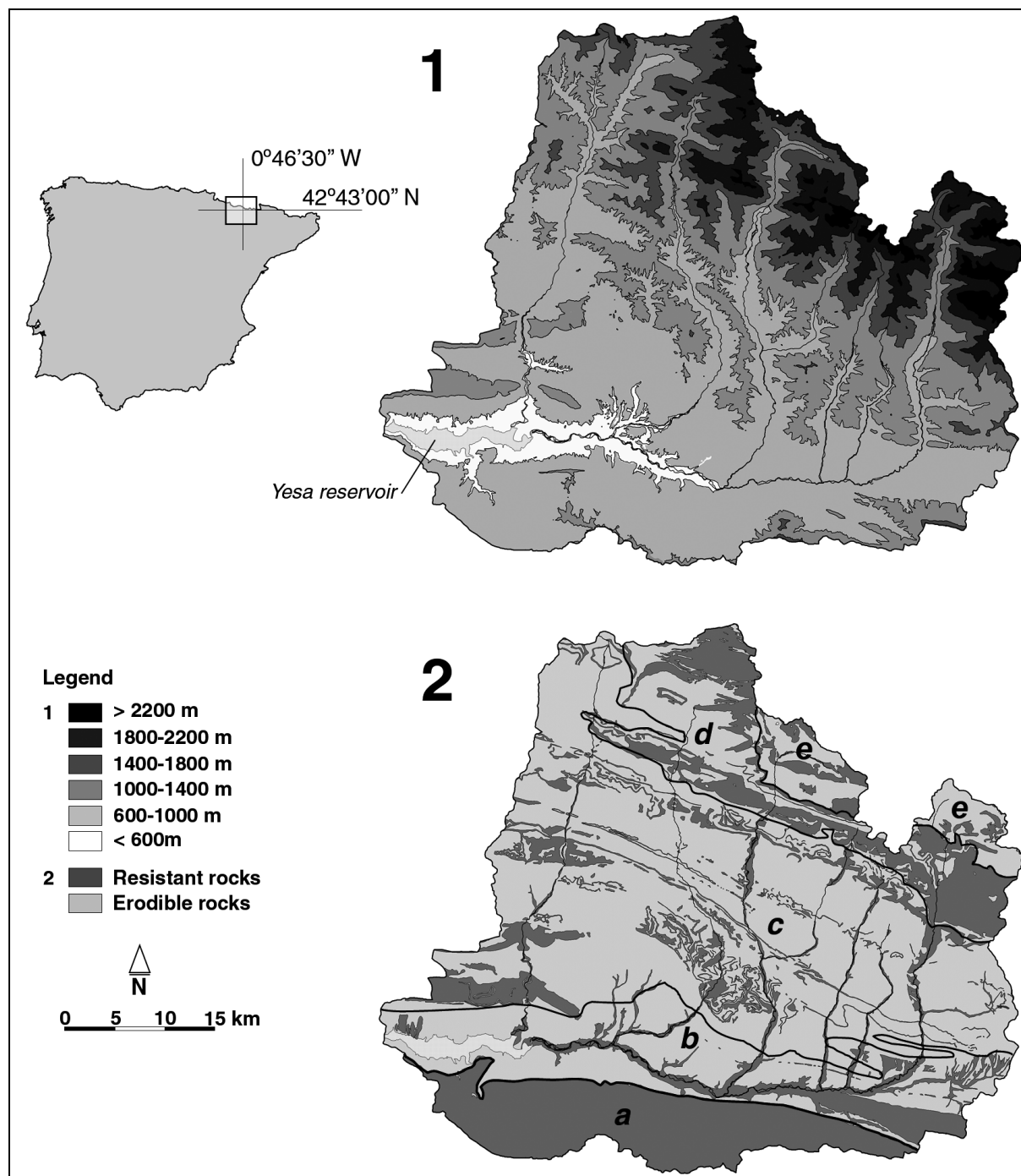


FIG. 1

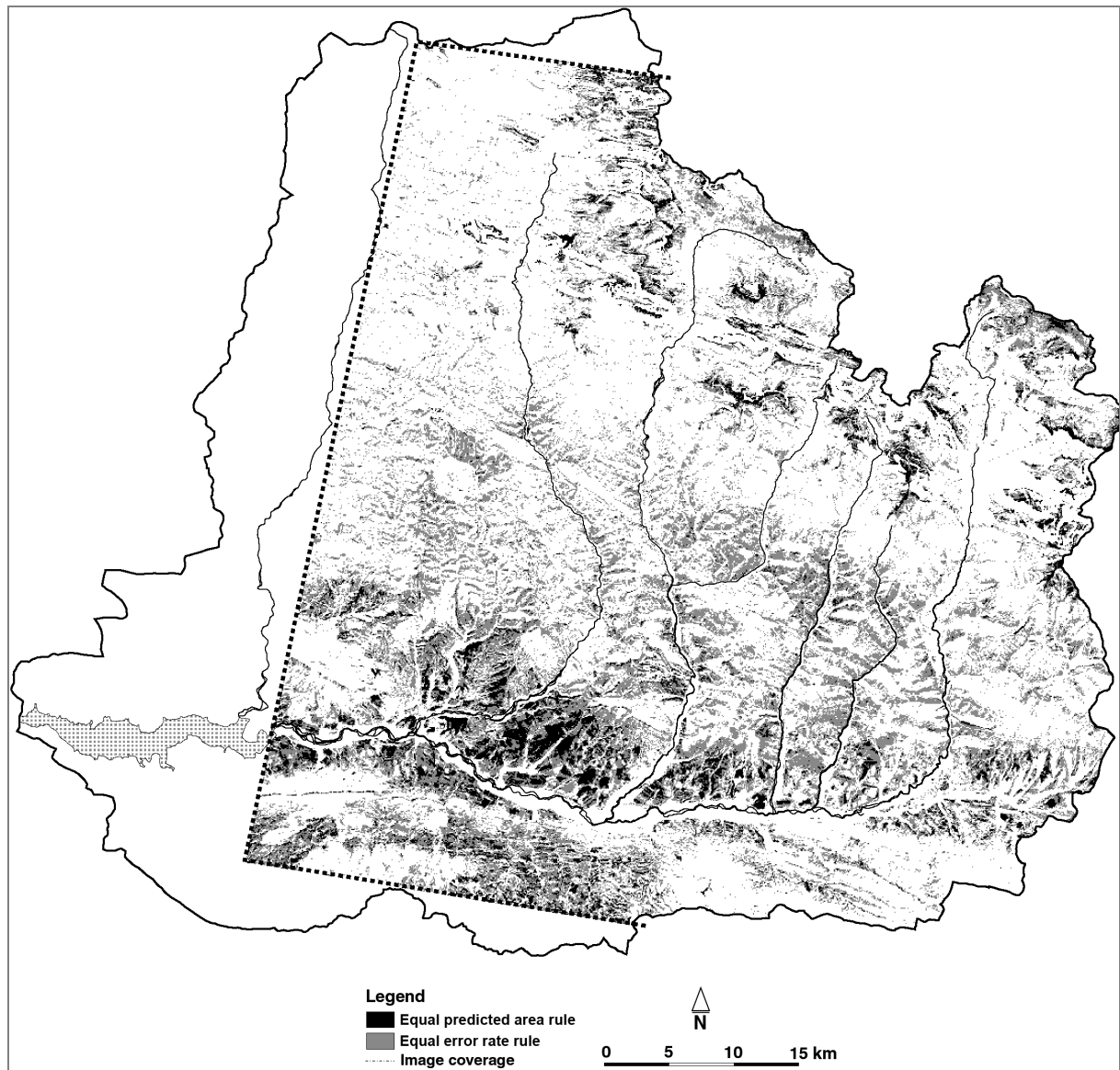


FIG. 2