ELSEVIER

Superficial and subterranean soil erosion in Tabasco, tropical Mexico: Development of a decision tree modeling approach

V. Geissen [a,*], C. Kampichler [b], J.J. López-de Llergo-Juárez [a], A. Galindo-Acántara [b]

[a] *El Colegio de la Frontera Sur, Administración Correos 2, Apartado Postal 1042, 86100 Villahermosa, Tabasco, Mexico*
[b] *Universidad Juárez Autónoma de Tabasco, División Académica de Ciencias Biológicas, Carretera a Cárdenas Km. 0.5 s/n, 86150 Villahermosa, Tabasco, Mexico*

Abstract

Tropical countries suffer from a lack of financial resources to monitor and model sources and outcomes of soil degradation and, therefore, a scarcity of data. In this study it was tried to model superficial and subterranean water erosion based on basic environmental variables such as geological formation, soil type or vegetation cover — data which are potentially available even in developing countries. Different forms of water erosion and karst or pseudokarst formations in Tabasco, Southeastern Mexico were estimated. The study region (3500 km^2) consists of plain and hilly areas with an annual precipitation ranging from 2000 to 4000 mm. Main land use form is rangeland. The area is geologically composed of sedimentary rocks, limestone, sandstone and conglomerates. Collecting data by field observations over the entire study area a cartography of different soil loss forms such as rill and gully erosion, mass movement, sinkholes and tunnels was created including field data from available maps of geological formation, soil type, precipitation and former land use 20 years ago. Additionally, data collected in the field such as actual land use, vegetation cover and inclination were added. In the entire study area 1039 sites were affected by soil erosion with 2435 single manifestations of soil erosion. 482 sites were found with one gully each, 57 sites with erosion rills with a total of 416, 392 sites with one mass movement each, 85 sites with sinkholes with a total of 1122, and 23 sites with one tunnel each. Rendzic Leptosols over Oligocene limestone are strongly affected by sinkholes and tunnels in regions with inclinations of less than 5°. This phenomenon is explained by karstic and pseudokarstic formation. Superficial erosion forms are mainly found on rendzic Leptosols over shale sandstone and on eutric or peli-eutric Vertisols on Andesit in areas with inclinations between 5 and 30°. The application of classification trees allowed to successfully predict the occurrence of sinkholes and tunnels (degree of agreement between prediction and observation: very good to excellent). Even in areas with scarce data-bases this could allow for easy identification of high-risk areas. Predictive success of the occurrence of the different forms of superficial soil losses, however, is low. Superficial erosion seems to be mainly caused by more specific pedological factors. Nevertheless, automated induction of classification trees can be a valuable tool for preliminary data analysis and hypothesis generation in areas with lack of local expertise and can guide erosion risk mapping and soil conservation planning.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Linear soil erosion; Mass movement; Qualitative modeling; Classification trees; Subterranean soil erosion; Tropical soils

1. Introduction

Land use changes in tropical countries have led to severe soil degradation and losses of important soil functions. One of the most destructive and insidious process, which is steadily increasing as a result of anthropogenic activity in these countries, is water erosion (Pla Sentís, 1997; Servenay and Prat, 2003). However, the tropics typically suffer from a lack of financial resource to research, monitor and model sources and outcomes of soil degradation (Santana et al., 1989; Martinez, 1997). Frequently there is only few information available on the parameters which are needed to apply empirical or process based erosion models like USLE (Wischmeier and Smith, 1978), RUSLE (Chakroun et al., 1993) or WEPP (Ascough and Livingston, 1995); if used at all they are applied in a simplified form (Lu et al., 2004). While these erosion models are data-

intensive and aim to yield quantitative predictions of soil losses, often for comparatively small areas and local conservation planning, the tropics need predictive instruments that help in identifying entire landscapes as prone or less prone to certain erosion types (Vrieling et al., 2002). Another weakness of the above mentioned modelling approaches if applied in tropical countries is that they do not allow for the prediction of subterranean soil losses caused by karst or pseudokarst formations (Botschek, 1999), erosion forms which are abundant and widely distributed in the tropics (Gerstenhauer, 1978; Keqiang et al., 2004). In summary, the tools used for land-use planning purposes in tropical countries should (i) be simple, (ii) predict incidence of all forms of water erosion and (iii) work with as few explaining variables as possible. It has already been pointed out that a simple qualitative approach can be more effective in erosion risk assessment than the use of models that were not developed for the region to which they were applied (Vrieling et al., 2002).

We used three municipalities of the state of Tabasco in tropical south-eastern Mexico as a model area for testing a qualitative modeling approach. Mexico is a country suffering heavily from land degradation due to anthropogenic pressure

(SEMARNAT, 2002). Increasing deforestation over the past 60 years has led to increasing soil erosion in mountainous areas of Mexico. In this period in Tabasco woodland was reduced from 49.1% of the land surface to 13.6% (Palma and Triano, 2002) which caused severe problems of soil erosion in the hilly and mountain areas of the state. However, there is very few information on the actual state of soil degradation. A national memory about the actual state of soil degradation in Mexico has been published recently (SEMARNAT, 2002). According to this study, only 2.3% of Tabasco is affected by water erosion. However, the scale is so large that field observations often do not coincide with data of published maps. There are a few other studies that indicate severe soil losses in the hilly and mountain areas of Tabasco (Sánchez, 1997).

The aim of this study is (1) to describe the actual extension of the different forms of superficial and subterranean soil erosion (rill and gully erosion, mass movement, tunnels, sinkholes) in the above-mentioned model area in SE Mexico, (2) to use basic environmental variables available from published maps or from direct observation to identify the main factors responsible for the susceptibility for water erosion, and (3) to design a simple
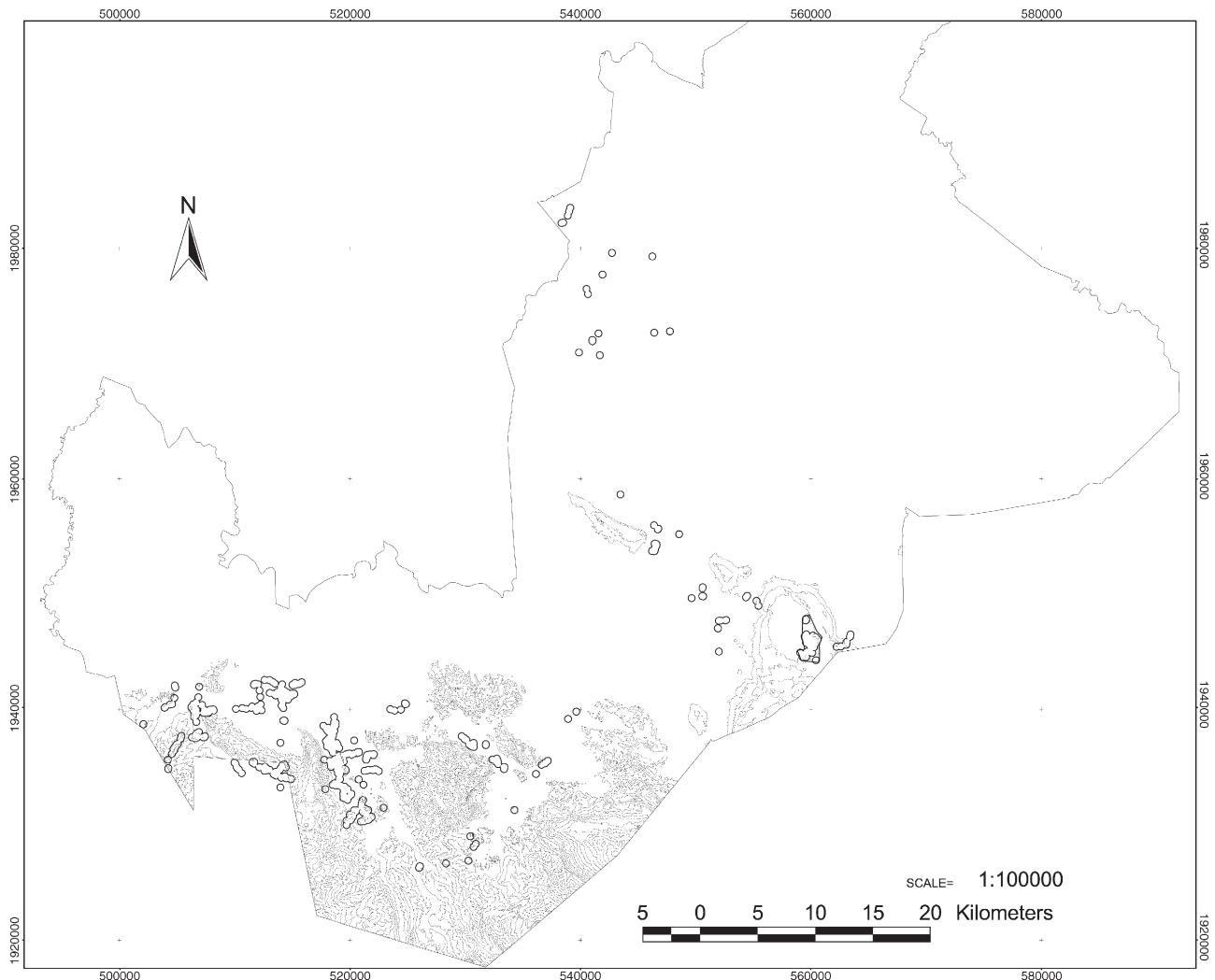

Fig. 1. Localization and topography of the study area in the municipalities Teapa, Tacotalpa and Macuspana, Southern Tabasco, SE Mexico.

model of occurrence and absence of erosion types by use of classification trees (Breiman et al., 1984; Ripley, 1996). Classification trees have the advantage of representing relationships between explanatory and dependent variables in a tree-like structure and natural language – most commonly as IF–THEN rules – and, thus, are more comprehensible than mathematical formulae or matching tables. They already have been used in erosion modeling (De la Rosa et al., 1999; Vrieling et al., 2002; Shrestha et al., 2004); in contrast to these studies, however, we do not use classification trees to represent local expert knowledge – which hardly does exist – but we induce them automatically by applying appropriate data-mining software on data available from maps and on field observations. We expect this qualitative approach to be robust to the scarcity of information available on geological, pedological, landscape and climatic conditions and expect it to be potentially useful for the identification of areas susceptible to water erosion and, thus, for tropical land use planning.

2. Study area

The study area is formed by the adjacent municipalities Teapa, Tacotalpa and Macuspana in the south of Tabasco, SE Mexico (Fig. 1). The state of Tabasco is a paradigmatic study area for deforestation and land degradation during the last 60 years as has been pointed out above. The area is characterized by a warm and humid tropical climate with high precipitation throughout the year. The annual average temperature is 25.4 °C and the mean annual precipitation amounts to 3862.5 mm. Maximum precipitation occurs in September with 600 mm, minimum precipitation in April with 150 mm (INEGI, 2000). The months from May to September are characterized by typical tropical rains with high intensities, the months between October and April are character-

ized by rainfalls with moderate intensities. The whole study area has a size of approximately 3500 km². The northern part of the study area is located in a plain at an elevation of less than 20 m a.s.l. (50% of the area), the central and southern part is hilly with elevations between 20 and 300 m a.s.l. (43%), and the southernmost part is mountainous with elevations from 300 to 800 m .a.s.l. (6%) and from 800 to 1600 m a.s.l. (1%), respectively. The dominant soils in the plain are Gleysols and Fluvisols over alluvial sediments. In the hilly and mountainous areas dominant soils are Luvisols, Vertisols and Leptosols over sandstone, shale sandstone and limestone. Dominant land use form is rangeland with 52%, 17% are used as primary or succession forest, 4.7% as agricultural land and 16% are natural reeds (Conafor, 2000).

3. Materials and methods

3.1. Field work and data origin

We mapped the whole study area according to the Manual of Evaluation of Actual Erosion by the German Association of Water Management (DVWK, 1996). We made field trips in the whole area to estimate the different soil erosion forms that occurred in the region. We distinguished between linear erosion forms – rill erosion (depth < 40 cm) and gully erosion (depth > 40 cm) – mass movement, and additionally soil losses caused by karst or pseudokarst processes such as sinkholes and tunnels. A single site was defined by homogeneous environmental conditions described by identical values of inclination, actual vegetation, percentage of vegetation cover and percentage of superficial rocks, which were determined in the field. We measured length, depth and width of each gully, length of each rill, width and depth of each mass movement and diameter of each sinkhole and tunnel. The location of each erosion site was determined in the most

Table 1
Variables included in the prediction model

Variable	Origin	Scale	Classes	Description and comments
Geological formation	Geological map	Nominal	Oligocene limestone Paleocene shale–sandstone etc.	
Soil type	Soil map	Nominal	Rendzic Leptosols Eutric Vertisols etc. (Table 3)	Soil units (FAO, 1998)
Annual precipitation	Precipitation map	Ordinal	2000–2500 mm, 2500–3000 mm, 3000–3500 mm, 3500–4000 mm, 4000–4500 mm	
Elevation	Topographical map	Ratio	Continuous values	Minimum 0 m, maximum 1600 m
Inclination	Field work	Nominal	0–5°, 6–10°, 11–20°, 21–30°, 31–40°, 41–50°	
Vegetation	Field work	Nominal	High primary forest Succession forest Pastures according to their dominant grass species such as *Brachiaria decumbens*, *Cynodon plectostachyus*, etc. Agricultural land Uncultivated land	Height >= 30 m — Fallows and abandoned pastures
Vegetation cover	Field work	Ordinal	0–5%, 5–10%, 10–20%…, 90–100%	
Former land use (in 1985)	Land use map	Nominal	High primary forest medium primary forest Succession forest Agricultural land Rangeland Reeds	

detailed maps available for the region which were topographical maps (1:50,000), soil, geological and precipitation maps (1:250,000) as well as a map on former land use as determined in 1985, all published by the Mexican National Institute of Statistics and Geographic Information (INEGI, 1983, 1984, 1985, 1987, 1990a,b,c) and the Mexican National Forest Commission (Conafor, 2000). Consequentially, for each site showing a manifestation of erosion we had a list of independent variable values available (Table 1).

3.2. Data analysis and modeling

We used the Chi-Square-tests to determine significant deviances of the observed frequencies of erosion forms on the different soil types, geological formations and inclinations from the hypothetical expected frequencies if erosion manifestations were distributed at random throughout the study area. Proportions of soil types, geological formations and inclination classes (0–5°, 6–10°, 11–20°, 21–30°, 31–40°, 41–50°) were obtained from the published maps (INEGI, 1985, 1987, 1990a,b,c). No

distribution data were available for detailed vegetation types and vegetation cover. Consequently, for these parameters we could not apply this type of analysis.

We modeled occurrence and absence of erosion types by means of automatic classification tree induction (Breiman et al., 1984; Ripley, 1996). Classification trees, also called decision trees, are graphical classification support tools developed within the subfield of Artificial Intelligence called Machine Learning (Finlay and Dix, 1996). Their mean advantage is that they are simple to understand and interpret and allow for a plain representation of relationships between explicative and response variables. For classification, each one of a set of cases (objects, situations, in our case: sites) is run down the tree; at each node a decision is made with regard to an attribute of the case (in our case: factors responsible for the susceptibility for water erosion) until it reaches a terminal node. Each terminal node contains a label of classification. For example, a classification tree for geometric shapes would ask for the number of vertices and kind of symmetry at the nodes and would yield a class label ("triangle", "circle", etc.) for every case run down the tree. In

Fig. 2. A. Soil map of the municipalities Teapa, Tacotalpa and Macuspana, Southern Tabasco, with the localization of the different soil erosion forms (sites affected by soil erosion are marked in the map). B. Geolog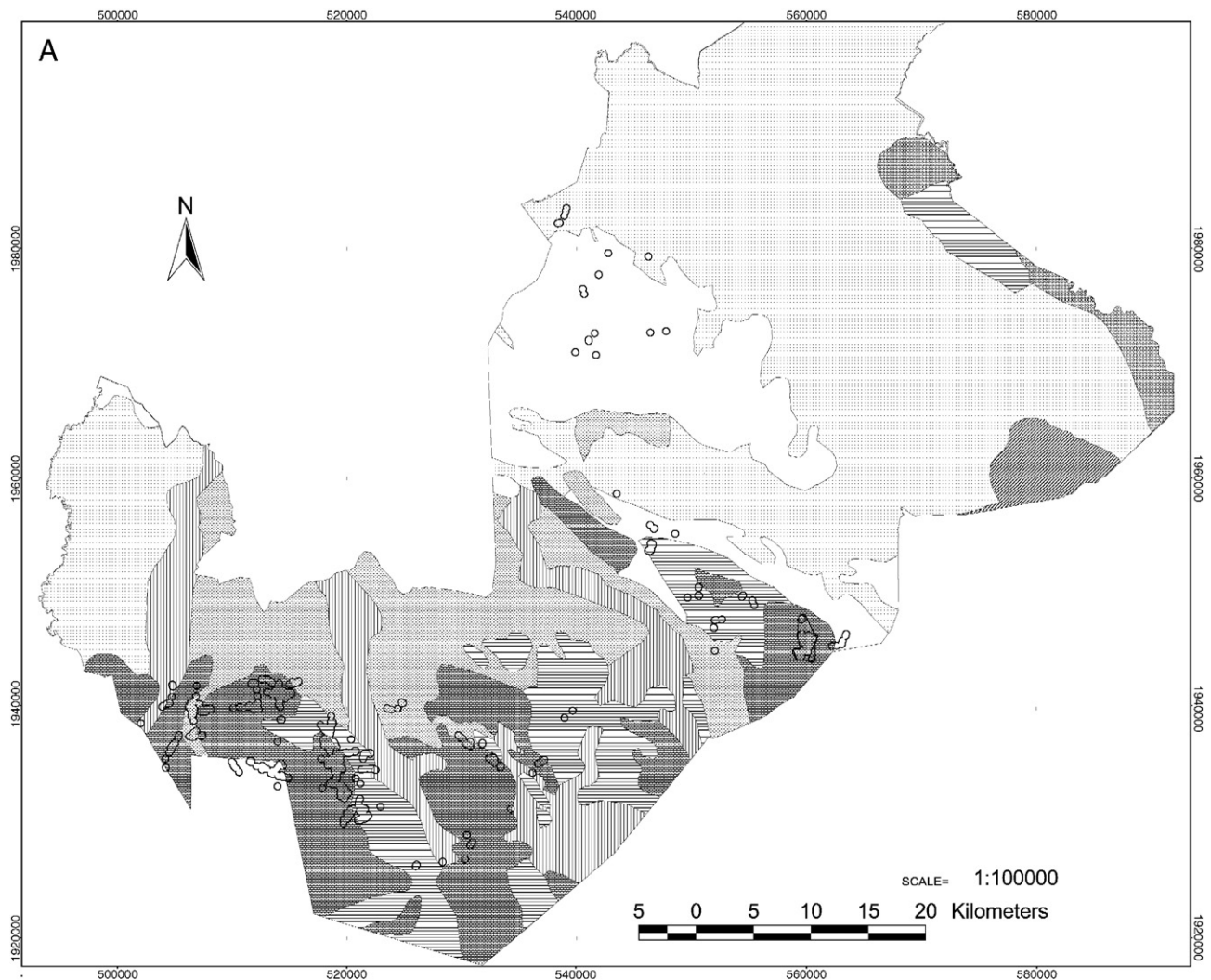ical map of the municipalities Teapa, Tacotalpa and Macuspana, Southern Tabasco, with the localization of the different soil erosion forms (sites affected by soil erosion are marked in the map).
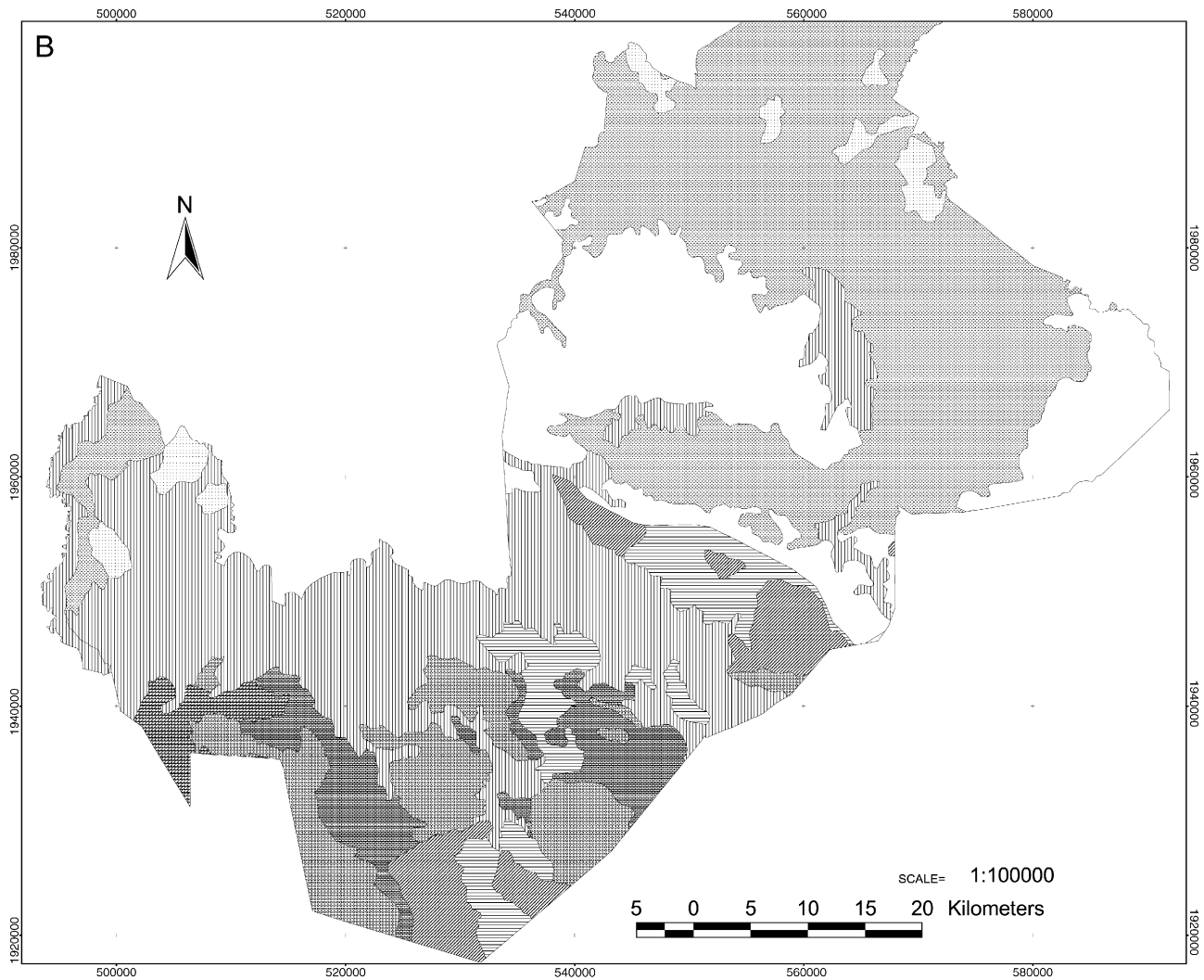
Fig. 2 (*continued*).

Machine Learning the main objective of classification trees is inductive learning, that is, learning by taking examples and generalizing. Algorithms developed for automatic tree induction split the original set of cases into smaller and smaller pieces in all possible ways applying the fundamental idea for subset selection that the data in each of the descendant subsets should be "purer" than the data in the parent subset, a task that was unthinkable before the computer age. Classification trees are increasingly used in environmental data analysis and modelling and have entered advanced statistical textbooks (for example, Venables and Ripley, 1999). We used geological formation, soil type, elevation, inclination, precipitation, actual vegetation, vegetation cover, and former land use as predictor variables (Table 1) for classification tree induction. Due to lacking data of some environmental variables on some of the 1039 observed sites we used only 967 sites for modeling. We developed a model for every erosion type, counting sites with manifestations of the corresponding erosion type and using the remainder of the sites as absences for the given erosion type. The resulting classification trees, thus, predict erosion types only for the area actually suffering erosion since no sites without erosion entered the data-base.

Classification tree induction was performed with the package tree (Ripley, 2005) of the statistic programming environment R (R Development Core Team, 2004). Agreement between model predictions and observations were measured by determining sensitivity (=proportion of correctly predicted occurrences) and

Table 2
Occurrence of different soil loss forms depending on the inclination

Inclination (°)	Area occupied (%)	Gullies	Rills	MM	Sinkholes	Tunnels	Total (n)
0–5	50	7*	15*	13*	100*	87*	199
6–10	35	16*	15*	8*	0*	0	117
11–20	8	38*	26*	19*	0*	9	276
21–30	3	31*	36*	36*	0	4	319
31–40	3	6	5	19*	0	0	105
41–50	1	0	0	6	0	0	23
Total sites (n)		482	57	392	85	23	1039

For ease of reading the number of sites showing a given type of erosion in a given inclination class are expressed as percentage of the total for the respective erosion type. MM, mass movement; *, significant differences ($p < 0.05$) between the distribution of the inclinations and the occurrence of soil erosion forms.

Table 3
Occurrence of different soil loss forms depending on soil type

Soil type/erosion form	Area occupied (%)	Gullies	Rills	MM	Sinkholes	Tunnel	Total (n)
Eutric Fluvisols	9.8	1.2*	1.7*	1.8*	0.0*	0.0*	14
Eutric Gleysols+Fibric Histosols+Mollic Gleysols	3.75	0.4*	0.0*	2.3*	0.0*	0.0*	11
Ferric Acrisols	1.6	0.0	0.0	0.0	0.0	0.0	0
Rendzic Leptosols	10.8	52.3*	54.4*	51.7*	100*	100*	593
Gleyic Luvisols	1.3	0.0	0.0	0.0	0.0	0.0	0
Cromic Luvisols	17.1	1.2*	1.7*	5.1*	0.0*	0.0*	27
Eutric Vertisols	13.9	7.5	15.8	8.9	0.0*	0.0*	80
Peli-eutric Vertisols	8.0	27.0*	22.8*	26.8*	0.0*	0.0*	248
n.d.		10.4	3.6	3.4	0.0*	0.0*	66
Total (n)		482	416	392	1122	23	1039

For ease of reading the number of sites showing a given type of erosion for a given soil type is expressed as percentage of the total for the respective erosion type. MM, mass movement; *, significant differences ($p < 0.05$) between the distribution of the soil types and the occurrence of soil erosion forms.

specificity (=proportion of correctly predicted absences) and by Cohen's Kappa (Cohen, 1960) applying the equation

$$\kappa = \frac{P_C - [P_D P_V + (1-P_D)(1-P_V)]}{1 - [P_D P_V + (1-P_D)(1-P_V)]}$$

where P_C is the proportion of correct predictions, P_D is the proportion of observed occurrences, and P_V is the proportion of predicted occurrences; Cohen's Kappa was evaluated according to Monserud and Leemans (1992): <0.05, no; 0.06–0.20, very

poor; 0.21–0.40, poor; 0.41–0.55, fair; 0.56–0.70, good; 0.71–0.85, very good; 0.85–0.99, excellent; 1.00, perfect agreement between model prediction and observation.

4. Results

In the entire study area we observed 1039 sites that had been affected by soil erosion with 2435 single manifestations of soil erosion. The cartography shows 482 sites with one gully each, 57 sites with erosion rills with a total of 416, 392 sites with one

Table 4
Occurrence of different soil loss forms depending on combination of soil type and geological formation

Soil type	Geol. form.	Area (%)	Gullies	Rills	MM	Sinkh.	Tun.	Total (n)
Rendzic Leptosol	Ks_limestone	33.8	2.0*	0*	5.4*	0*	13.0*	19
	To_limestone	16.9	4.3*	0*	2.0*	100*	82.6*	115
	Q_alluvial	14.8	4.8*	0*	3.4*	0*	0*	19
	Te_shale–sandstone	9.9	0.4*	3.2*	0.5*	0*	0*	3
	Tpal_shale–sandstone	15.7	79.8*	67.7*	66.0	0*	0*	360
	Tm_sandstone	0.2	0	0	0	0	0	0
	Ts_andesit	8.7	8.7	29.0*	22.7*	0*	4.4	77
Total (n)			252	31	203	84	23	593
Chromic Luvisols	To_limestone	3.5	33.3	0	20.0	0	0	6
	Te_shale–sandstone	6.2	50.0*	0	15.0	0	0	6
	Tm_sandstone	71.6	16.7	100	65.0	0	0	15
	Q_alluvial	24.9	0	0		0	0	
Total (n)			6	1	20			27
Eutric Vertisol	Ks_limestone	4.0	19.4*	0	8.6	0	0	10
	To_limestone	1.4	0	0	0	0	0	
	Te_shale–sandstone	2.3	0	0	0	0	0	
	Tm_sandstone	0.9	0	0	0	0	0	0
	Tpal_shale–sandstone	1.2	16.7*	22*	25.7*	0	0	17
	Ts_andesit	1.2	58.3*	78*	45.7*	0	0	44
	Q_alluvial	88.9	5.5*	0*	20.0*	0	0	9
Total (n)			36	9	35	0	0	80
Peli-eutric Vertisol	Ks_limestone	8.9	0	0	0	0	0	0
	To_limestone	17.5	3.0	0*	0.9*	0	0	5
	Te_shale–sandstone	29.5	10.7*	15.3*	20.0	0	0	37
	Tpal_shale_sandstone	30.1	75.3*	69.2*	67.6*	0	0	178
	Q_alluvial	13.9	10.7	15.3	11.4	0	0	28
Total (n)			130	13	105	0	0	248
n.d.	n.d.		50	2	13	1		66
Total (n)			482	57	392	85	23	1039

For ease of reading the number of sites showing a given type of erosion for a given combination of soil type and geological formation is expressed as percentage of the total for the respective erosion type. MM, mass movement; *, significant differences ($p < 0.05$) between the distribution of the combinations between soil type and geological formation and the occurrence of soil erosion forms.

Table 5
Occurrence of different soil loss forms depending on vegetation

Vegetation/ erosion form	Gullies (%)	Rills (%)	MM (%)	Sinkholes (%)	Tunnel (%)	Total (n)
Uncultivated land	17.8	14.0	23.7	2.3	65.2	204
Without vegetation	0.21		0.76			4
Secondary forest					13.0	3
Cynodon dactylon	0.42					2
Eleusine indica	1.4	1.7	1.8			15
Cynodon plectostachyus	7.1	5.3	4.8	88.2	21.7	136
Pennisetum ssp.	0.21	3.5	0.25			4
Brachiaria humidicola			0.76	2.3		5
Hyparrhenia rufa	1.4					7
Corn					7.0	6
Paspalum notatum	32.0	47.3	36.0			322
Brachiaria decumbens	29.0	24.5	27.5			262
Paspalum virgatum			1.02			4
n.d.	10.5	3.7	3.31			65
Total (n)	482	57	392	85	23	1039

For ease of reading the number of sites showing a given type of erosion for a given vegetation is expressed as percentage of the total for the respective erosion type. MM, mass movement; n.d., not determined.

mass movement each, 85 sites with sinkholes with a total of 1122, and 23 sites with one tunnel each (Fig. 2). That means that rills and sinkholes often were found aggregated on one field. The occurrence of soil erosion is concentrated in certain areas of the region (Figs. 1 and 2). The width of the gullies varied from 1 to 65 m, their length from 1 to 200 m and their depth from 1 to 12 m. The rills varied from a length of 6 to 228 m. The mass movements varied from 2 to 80 m in width and 1 to 12 m in depth. The diameter of the sinkholes varied from 4 to 21 m, the diameter of the tunnels from 4 to 10 m.

4.1. Inclination

The plains dominated by Fluvisols, Gleysols and Histosols (47% of the whole study area) are hardly affected by soil erosion (Tables 1, 2). The majority of the gullies, rills and mass

Table 6
Occurrence of different soil loss forms depending vegetation cover

% Veg. cover	Gullies	Rills	MM	Sinkholes	Tunnels	Total (n)
0–10	43.9	29.8	48.5	1.2	4.4	421
10–30	27.8	35.1	27.5	1.2	4.4	265
30–50	13.3	14.0	12.2	0.0	17.4	124
50–80	11.6	15.8	10.7	58.8	60.9	170
80–100	3.1	5.3	1.0	38.8	13.0	59
Total (n)	482	57	392	85	23	1039

For ease of reading the number of sites showing a given type of erosion for a given class of vegetation cover is expressed as percentage of the total for the respective erosion type. MM, mass movement.

Table 7
Classification matrices of the models predicting presence and absence of erosion types on 956 sites in Southern Tabasco, Mexico

Erosion type	Prediction	Observation		Cohen's Kappa
		Present	Absent	
Rills	Present	0	0	0.00 ("no")
	Absent	55	912	
Gullies	Present	291	207	0.36 ("poor")
	Absent	138	331	
Mass movement	Present	58	16	0.19 ("very poor")
	Absent	318	575	
Tunnels	Present	17	2	0.81 ("very good")
	Absent	6	942	
Tunnels [a]	Present	18	18	0.60 ("good")
	Absent	5	926	
Sinkholes	Present	82	21	0.86 ("excellent")
	Absent	2	862	
Sinkholes [a]	Present	71	17	0.81 ("very good")
	Absent	13	866	

[a] Not including explanatory variable "vegetation".

movements are concentrated in areas lower than 100 m a.s.l. with inclinations of 10 to 30° (Table 2). The high frequency of the superficial soil erosion forms on these inclinations is significantly different from the frequency distribution of the inclination-classes in the study area. All sinkholes and 87% of the tunnels are located in an area between 200 and 300 m a.s.l. with an inclination of less than 5° (Table 2). This occurrence is also significantly higher than expected from the frequency distribution of the inclinations in the area.

4.2. Soil type and geological formation

Although only 10.8% of the study area is occupied by rendzic Leptosols, approximately 50% of the gullies, rills and mass movements occur on this soil type (Table 3, Fig. 2A) which is significantly higher than expected from the soil type frequency distribution in the area. 66 to 80% of the superficial erosion forms that occur on rendzic Leptosols are located on shale sandstone of the Paleocene which is significantly higher than expected from the frequency distribution of the geological formations in the area (Table 4, Fig. 2A, B). Leptosols over limestone, Alluvial, and Oligocene shale sandstone are significantly less affected by superficial soil erosion as expected from their frequency distribution in the area (Table 4, Fig. 2A, B).

34 to 37% of the superficial erosion forms are located on eutric or peli-eutric Vertisols (Table 3, Fig. 2A). However, the occurrence of these erosion forms on eutric Vertisols (7.5–15.8%) corresponds to the percentage of this soil type in the region (13.9%). 46 to 78% of the superficial soil erosion forms are located on eutric Vertisols over Andesits, 16.7 to 25.7% of them occur on Vertisols over shale sandstone of the Paleocene (Table 4, Fig. 2A, B). Eutric Vertisols over Cretaceous limestone are also significantly more affected by gully erosion than expected by the geological formation frequency distribution in the region.

Peli-eutric Vertisols are strongly affected by superficial erosion forms. Gullies, rills and mass movements occur on this soil type significantly more often than expected from the percentage of this

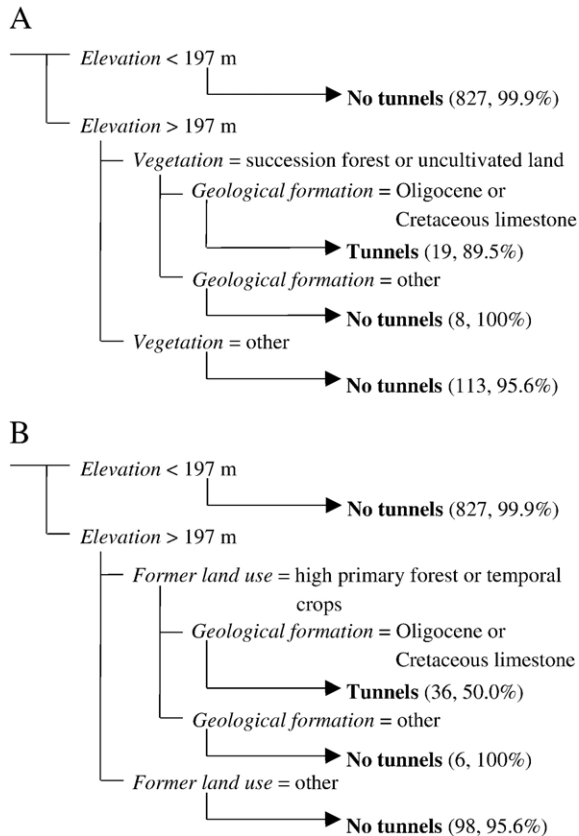284 V. Geissen et al. / Geoderma 139 (2007) 277–287

Fig. 3. Classification trees for occurrence and absence of tunnels. (A) Full model including all environmental variables, (B) alternative model excluding the variable "vegetation".

soil type in the area (8%). While peli-eutric Vertisols over limestone, Alluvial and Eocene shale sandstone are hardly affected, peli-eutric Vertisols over Paleocene shale sandstone are strongly affected by gullies, rills and mass movement.

Chromic Luvisols are significantly less effected by soil erosion in this region (Table 3, Fig. 2A). Although they occupy 17.1% of the area, only 1.2 to 5.1% are affected by superficial soil erosion.

All the sinkholes and 82.6% of the tunnels are located on rendzic Leptosols over limestone of the Oligocene (Tables 3, 4, Fig. 2A, B). No other soil types and hardly no other geological formations are affected by these subterranean soil loss forms.

4.3. Vegetation

Independent of the soil type, vegetation plays an important role for erosion processes. Under forest vegetation and agricultural use no superficial erosion occurs in the study area. However, 13% of the tunnels occur under secondary forest. The pastures mainly affected by superficial soil erosion are dominated by the grass species Paspalum notatum and Brachiaria decumbens, whereas 88% of the sinkholes are found under Cynodon plectostachyus and 65% of the tunnels under uncultivated land (Table 5). Not surprisingly, increasing vegetation cover led to decreasing superficial soil losses (Table 6). 65 to 76% of the superficial erosion forms occur on areas with vegetation cover of less than 30% (Table 5).

However, 59% of the sinkholes and 61% of the tunnels occur in areas with a vegetation cover of 50 to 80%.

4.4. Classification trees

The prediction success of the classification trees for the superficial erosion types rills, gullies and mass movement is poor with Cohen's Kappa never more than 0.29 (Table 7). Rill erosion cannot be modeled at all as the classification tree fails completely to predict rill occurrences. Sensitivity and specificity of the gully model are modest with values of 68% and 62%, respectively. Sites without mass movement occurrence are satisfactorily predicted by the respective classification tree (specificity: 97%); however, prediction success for sites with mass movement occurrence is low (sensitivity: 15%).

The classification trees for subterranean water erosion perform much better, characterized by a sensitivity of 74% (tunnels) and 98% (sinkholes), a specificity of almost 100% (tunnels) and 98% (sinkholes), and high values of Cohen's Kappa (Table 7). Both models include the explanatory variable "vegetation" (Figs. 3A and 4A). We induced additional classification trees leaving out vegetation to discover relationships between erosion and the variables independent of actual land use. The alternative tunnel model performs considerably worse with Kappa dropping to 0.60, the alternative sinkhole model is almost as good as the full model (Table 7); the most important variable next to geological formation in both trees is the type of former land use (Fig. 4A and B). Sinkholes appear on Oligocene limestone exclusively; areas covered by natural vegetation
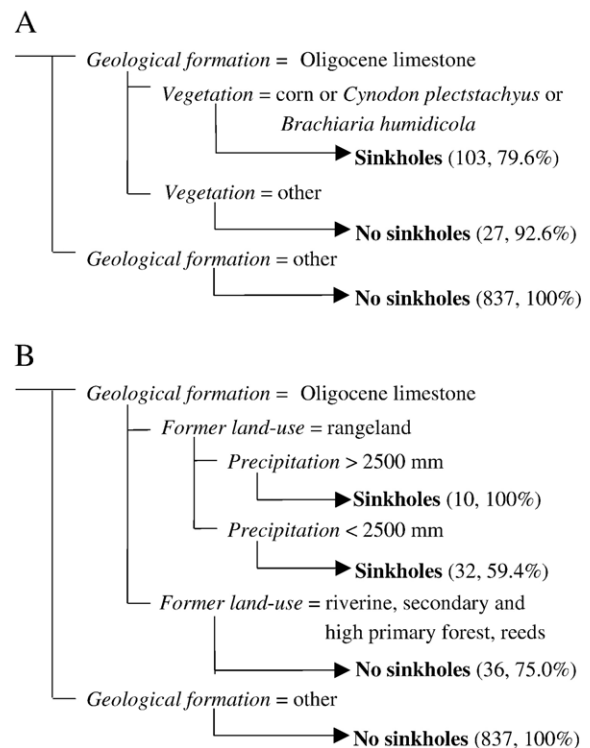

Fig. 4. Classification tree for the occurrence and absence of sinkholes. (A) Full model including all environmental variables, (B) alternative model excluding the variable "vegetation".

Table 8
Basal nodes of the classification trees for superficial erosion forms and the fitted probabilities for occurrence and absence

Erosion form	Basal node	Predicted class	Fitted probability (%)	Erosion risk (%)
Rills	Geological formation=Oligocene or Cretaceous limestone	No rills	100	0
	Geological formation=other	No rills	93.2	6.8
Gullies	Geological formation=sandstone or Oligocene limestone	No gullies	87.0–	13.0
	Geological formation=other	Gullies	50.3	50.3
Mass movement	Geological formation=Oligocene limestone	No mass movement	92.3	7.7
	Geological formation=other	No mass movement	56.3	43.7

20 years ago (mainly primary and secondary forests) do not show sinkholes, whereas the majority of sites that were already used in 1985 for rangeland suffered from this form of subterranean water erosion. Tunnels are predicted on elevations of more than 197 m a.s.l. on Oligocene or Cretaceous limestone if the actual vegetation is agricultural use or succession forest.

Classification trees not only predict occurrence or absence of a given erosion type on a given plot; they can also be used to evaluate erosion risk for a plot with a given combination of environmental variables. As shown in Figs. 3 and 4 at each terminal node ("leaf" in classification tree terminology) the trees provide a fitted probability P for the predicted class which simply can be interpreted as erosion risk: P in the case of assignment to class "occurrence" (e.g., tunnels), or $1-P$ in the case of assignment to class "absence" (e.g., no tunnels). Even when classification criteria (sensitivity, specificity, Cohen's Kappa) are low these probabilities – at least of the main bifurcation of the trees – give valuable information (Table 8). Probabilities of about 50% (as for example for gullies and the branch "geological formation=other" in Table 8) yield worthless predictive models with Cohen's Kappa close to zero, but still are useful results for erosion risk mapping.

5. Discussion

5.1. Superficial erosion

The analysis of single environmental factors by χ^2-tests showed strong relationships between the occurrence of superficial soil erosion forms and inclination, vegetation, soil type and geological formation. We mainly found inclinations of 10 to 30° to be affected by gullies, rills and mass movements. Morgan (1994) also described that areas with lower inclination are less affected by gully erosion and mass movement. Areas with inclinations of more than 30° are less affected in the study area due to the fact that these sites still are covered with forest vegetation. The soil types that are mainly affected by superficial

erosion are rendzic Leptosols and peli-eutric Vertisols. These soil types are characterized by a low infiltration rate which leads to a high runoff rate of rainwater (Palma and Cisneros, 2000). Rendzic Leptosols and peli-eutric Vertisols affected by superficial erosion are mainly located over Paleocene shale sandstone which show an intensive weathering. The formation of saprolit in these geological formations may favor earthflow in hilly areas and lead to the formation of linear erosion forms.

Due to lack of data on spatial proportions of actual vegetation types in the study area, we have no statistical proof; however, we suppose that the grasses Brachiaria decumbens and Paspalum notatum are not adequate to prevent superficial soil erosion as local livestock farmers in Tabasco believe. This fact was also confirmed by Souza and Seixa (2001) for soils in Brazil and by Semiday et al. (2002) for soils in Puerto Rico. Instead, Moreno (1998) proposed to establish silvopastoril systems for erosion control.

Not surprisingly and as described by many other authors (Wischmeier and Smith, 1978; Hutson, 1982; Morgan, 1994) we confirmed that superficial soil erosion decreases with increasing vegetation cover.

Although inclination, soil type, geological formation and vegetation cover had a strong influence on the occurrence of superficial soil erosion forms, it was not possible to create successful models to predict the appearance of superficial soil erosion forms. This is due to the fact that we could predict absences of certain soil erosion forms under certain conditions (high specificity), but the model did not predict their occurrence in a sufficient way (low sensitivity). If, for example, mass movement mainly appears on peli-eutric Vertisols on Paleocene shale sandstone this does not necessarily mean that the majority of these soils are affected by this erosion form. The models show, that we cannot predict precisely where superficial soil erosion forms will occur, however we (i) can identify combinations of environmental factors with high risk of being affected by certain superficial erosion forms and (ii) we can exclude several areas which most probably will not be affected. Nevertheless, additional information on more specific soil properties such as infiltration rate, organic matter, texture etc. seems to be indispensable to allow for better modeling success.

5.2. Subterranean erosion

Tunnel erosion or subterranean soil losses and sinkhole formation are not mentioned in the soil degradation maps of SEMARNAT (2002) and are not considered in any predictive model for soil erosion such as USLE, WEPP, SIDASS etc. (Wischmeier and Smith, 1978; Lane and Nearing, 1989; De la Rosa et al., 2005). These phenomena are typical for limestone regions in the tropics (Gerstenhauer, 1966, 1978).

40 years ago Gerstenhauer (1978) described the entire study area still to be covered with an undisturbed soil layer under tropical rainforest. Deforestation in this area brought about changes in the water balance and consequently a lowering of the water table. Li and Wang (1990) describe the following consequences for karst (sinkhole) formation: a lowered groundwater table leads to increased groundwater velocity and, therefore, it eats away and erodes the soil cover to form a soil cave at the

interface between the soluble bedrock and the cover. If the soil cover is thin and has a poor structure, the soil cave enlarges until a collapse occurs (Li and Wang, 1990). Furthermore rainwater infiltration plays an important role for karst collapse (Keqiang et al., 2004). The more water that enters into the bedrock zone, the higher is the dissolution of the limestone. In Leptosols rainwater infiltration into the bedrock layer is higher than in soils with a thick layer. Therefore, the tunnels which had been covered by a soil layer under forest collapsed after deforestation and sinkhole formation has taken place in the last 40 years in the study area. The above-mentioned theories also explain the fact that sinkholes only occur on rendzic Leptosols and none of the sinkholes occur on Luvisols or Vertisols which are also associated with Oligocene limestone. Li and Wang (1990) explain that if the soil cover is thick, a natural balanced arch will develop and the cave will not collapse.

In contrast to the superficial soil erosion forms, predictive potential of the models for subterranean erosion forms is very high. They show that the only important factors for the occurrence of sinkholes were geological formation (Oligocene limestone) and former or actual land use. The important factors for the prediction of tunnels were elevation, geological formation (Cretaceous or Oligocene limestone) and actual vegetation type. While sinkholes were only predicted on Oligocene limestone, tunnels were also predicted in Cretaceous limestone. This may be due to the fact that the latter type of limestone forms "kegelkarst" due to its chemical composition (Gerstenhauer, 1978). This formation leads to limestone hills and residual cones. The importance of the elevation can be easily explained by the appearance of Cretaceous and Oligocene limestone only in areas over 197 m a.s.l. The prediction of tunnels under succession forest and uncultivated land may be due to the fact that in these sites the vegetation cover is dense with deep roots, and therefore the formation of sinkholes is prevented. Tunnel formation is a natural process in limestone. However, the occurrence of sinkholes is strongly influenced by human activity as shown in this study.

Whereas in the areas previously used as forest no sinkholes are predicted, areas with former use as pasture are strongly affected. This most probably also is due to the change in the groundwater table after deforestation. The processes of groundwater table lowering described above, which led to sinkhole formation, do not occur instantaneously but require certain time for their development. Sinkholes occur with a lag time of several years. Our results suggest the need to count with a lag time of 20 years or more, as sites that still were forested in 1985 do not yet show sinkhole formation whereas sites where deforestation took place earlier are strongly affected. The identification of a time lag is extremely important in conservation planning and can help to determine susceptible areas where erosion will almost inevitably occur in the near future unless protection measures will be taken.

In Tabasco in many areas used as rangeland, the grass *Cynodon plectostachyus* is sown to protect soils against erosion processes (Geissen and Morales Guzman, 2005). This works well in areas with medium inclination to prevent them from superficial erosion. However, *Cynodon plectostachyus* does not prevent sinkhole formation, because in order to inhibit sinkhole formation

not superficial flow but groundwater properties must be changed. A pasture, such as *Cynodon plectostachyus*, does not influence soil water dynamics in a sufficient way. The only way to reduce sinkhole formation appears to be reforestation in order to decrease infiltration and to stabilize soil by tree roots.

6. Conclusions

Manifestations of water erosion do not occur randomly in a tropical landscape but show distinct relationships with the landscape's geological, pedological, topographic and biological features. Subsurface erosion seems to be dependent on rather basic environmental properties, in the case of Tabasco by geological formation, soil type, inclination and dominant grassland species. Its occurrence can be satisfactorily predicted by simple models. Even in areas with scarce data-bases – as it is the case for most tropical countries – this could allow for easy identification of high-risk areas and sustainable land-use planning. However, superficial erosion seems to be mainly caused by more sophisticated environmental factors (for example, soil texture, infiltration rate, etc.) which might be harder to identify in the tropics due to the lack of detailed small-scale cartography. The mapping of pedological factors is highly desirable and seems to be an indispensable prerequisite for the determination of water erosion risk in the tropics.

Qualitative modeling by automated classification tree induction can be a valuable tool for preliminary data analysis and hypothesis generation. As pointed out by Vrieling et al. (2002), qualitative modeling normally relies on the knowledge of experts that have worked in the region for a long time (e.g., Shrestha et al., 2004 in the Nepalese Himalaya). In case there is no local expertise available – as will be the case in the bulk of area in developing countries – the data must speak for themselves and data-mining methods may yield first insight into the environmental variables–erosion relationship. The resulting models may guide erosion risk mapping and soil conservation planning as well as identifying knowledge gaps and requirements for future data collection.

Acknowledgements

References

Ascough II, J.C., Livingston S.J., 1995. WEPP User Summary. *USDA-ARS NSERL Rep.* 11, USDAARS National Soil Erosion Research Laboratory, West Lafayette, IN.

Botschek, J., 1999. Zum Bodenerosionspotential von Oberflächen — und Zwischenabfluss. Bonner Bodenkundliche Abhandlungen 29 (Bonn.).

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Chapman and Hall, New York.

Chakroun, H., Bonn, F., Fortin, J.P., 1993. Combination of single storm erosion and hydrological models into a geographic information system. Farm Land Erosion: In Temperate Plains, Environment and Hills, 261–270.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Conafor, 2000. Inventarío forestal. Comisión nacional forestal. http://www.conafor.gob.mx/programas_nacionales_forestales/Geomatica.

De la Rosa, D., Mayol, F., Moreno, J.A., Bonsón, T., Lozano, S., 1999. An expert system/neutral network model (ImpelERO) for evaluating agricultural soil erosion in Andalucia region, southern Spain. Agriculture, Ecosystems and Environment 73, 211–226.

De la Rosa, D., Díaz-Pereira, E., Mayol, F., Czyz, E.A., Dexter, A.R., Dumitru, E., Enache, R., Fleige, H., Horn, R., Rajkay, K., Simota, C., 2005. SIDASS Project Part 2. Soil erosion as a function of soil type and agricultural management in a Sevilla olive area, southern Spain. Soil & Tillage Research 82, 19–28.

DVWK, 1996. Bodenerosion durch Wasser — Kartieranleitung zur Erfassung aktueller Erosionsformen. Merkblaetter zur Wasserwirtschaft 239/1996. Hamburg.

FAO, 1998. World Reference Base for Soil Resources. World Soil Resources Reports, vol. 84. Rome.

Finlay, J., Dix, A., 1996. An Introduction to Artificial Intelligence. UCL Press, London.

Geissen, V., Morales Guzman, G., 2005. Fertility of tropical soils under different land use systems — a case study of soils in Tabasco, Mexico. Applied Soil Ecology 31, 169–178.

Gerstenhauer, A., 1966. Beitraege zur Geomorphologie des mittleren und noerdlichen Chiapas (Mexico) unter besonderer Berücksichtigung des Karstformenschatzes. Frankfurter Geographische Hefte, vol. 41. Frankfurt.

Gerstenhauer, A., 1978. Der tropische Kegelkarst in Tabasco (Mexico). Zeitschrift für Geomorphologie, Supplementband 2, 22–48 (Karstmorphologie).

Keqiang, H., Bin, W., Dunyun, Z., 2004. Mechanism and mechanical model of karst collapse in an over-pumping area. Environmental Geology 46, 1102–1107.

INEGI, 1983. Carta Geológica, Teapa. 1:250.000. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

INEGI, 1984. Cartas de Efectos Climáticos. Villahermosa E15-8. 1:250.000. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

INEGI, 1985. Carta Edafológica. Villahermosa E15-8. 1:250.000. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

INEGI, 1987. Carta de uso de suelo y vegetación. Villahermosa E15-8. 1:250.000. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

INEGI, 1990a. Carta Topográfica. Teapa, E15D21, 1:50.000. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

INEGI, 1990b. Carta Topográfica. Tacotapla, E15D21, E13D31. 1:50.000. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

INEGI, 1990c. Carta Topográfica. Macuspana, E15D22, 1:50.000. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

INEGI, 2000. Anuario estadístico del Estado de Tabasco. INEGI — Gobierno del Estado. Mapas. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.

Lane, L.J., Nearing, M.A., 1989. USDA water erosion prediction project (WEPP). NSERL Report, vol. 10. National Soil Research Laboratory, USDA–ARS–MWA, 1196 SOIL Building, West Lafayette, IN.

Li, Z., Wang, Z., 1990. The Basis of Engineering Geology. Publishing Company of China Geoscience University, pp. 176–179.

Lu, D., Li, G., Valladares, G.S., Batistella, M., 2004. Mapping soil erosion risk in Rondonia, Brazilian Amazonia: using RUSLE, remote sensing and GIS. Land Degradation & Development 15, 499–512.

Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with Kappa statistic. Ecological Modelling 62, 275–293.

Moreno, L., C.E., 1998. Recuperation of soils of slopes degraded by moderate erosion and the use of biological-forestry methods. Suelos Ecuatoriales 28, 184–187.

Morgan, R.P.C., 1994. Soil Erosion and Conservation. Silsoe College, Cranefield University, Longman.

Palma, L.D.J., Cisneros, D.J., 2000. Plan de uso Sustentable de los Suelos de Tabasco, 2^a. edición. ISPROTAB — Fundación Produce Tabasco — Colegio de Posgraduados, Villahermosa, Tabasco, México.

Palma, L.D.J., Triano, S.A., 2002. Plan de uso Sustentable de los Suelos de Tabasco, vol. II. Colegio de Posgraduados — ISPROTAB, Villahermosa, Tabasco, México.

Pla Seníts, I., 1997. A soil water balance model for monitoring soil erosion processes and effects on steep lands in the tropics. Soil Technology 11, 17–30.

R Development Core Team, 2004. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org.

Ripley, B., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge.

Ripley, B., 2005. Classification and Regression Trees. R Package Version 1.0-19.

Sánchez, E.D.C., 1997. El riesgo de la pérdida del suelo por erosión hídrica en Tabasco. In: SEMARNAP (Ed.), Los suelos de Tabasco — Restauración, Conservación y uso, pp. 27–31.

SEMARNAT, 2002. Evaluación de la Degradación del Suelo Causado por el Hombre en la República Mexicana Escale 1:250.000. Memoria Nacional, México.

Semiday, N., Orengo-Santiago, E., Más, E.G., 2002. Weed suppression and soil erosion control by live mulches on upland coffee plantations. Journal of Agriculture of the University of Puerto Rico 86, 155–157.

Servenay, A., Prat, C., 2003. Erosion extension of indurated volcanic soils of Mexico by areal photographs and remote sensing analysis. Geoderma 117 (3–4), 367–375.

Shrestha, D.P., Zinck, J.A., van Ranst, E., 2004. Modelling land degradation in the Nepalese Himalaya. Catena 57, 135–156.

Souza, C.R., Seixa, F., 2001. Erosion control in forest road slopes by organic covers. Scientia Forestalis 60, 45–5.

Venables, W.N., Ripley, B.D., 1999. Modern Applied Statistics with S-PLUS. Springer, New York.

Vrieling, A., Sterk, G., Beaulieu, N., 2002. Erosion risk mapping: a methodological case study in the Colombian Eastern Plains. Journal of Soil and Water Conservation 57, 158–163.

Wischmeier, W.H., Smith, D.D., 1978. Predicting Rainfall Erosion Losses. Agriculture Handbook 537. United States Department of Agriculture. Science and Education Administration.