Automated catena-based discretisation of landscapes for the derivation of hydrological modelling units

T. FRANCKE[*][†], A. GÜNTNER[‡], G. MAMEDE[†], E. N. MÜLLER[†], A. BRONSTERT[†]

† Institute for Geoecology, University of Potsdam, P.O. Box 60 15 53, 14415 Potsdam, Germany,
‡ GeoForschungsZentrum Potsdam (GFZ), Telegrafenberg, 14473 Potsdam, Germany

Correspondence   *Corresponding author. francke@uni-potsdam.de
http://brandenburg.geoecology.uni-potsdam.de/projekte/sesam

In hydrological and soil erosion modelling at large spatial scales, semi-distributed approaches may use representative hillslope profiles to reproduce landscape variability. Until now, the process of delineating landscape units as homogeneous parts of the landscape with regard to their terrain, vegetation and soil properties required expert knowledge and familiarity with the study area. In addition, the delineation procedure was often highly time-consuming and included a high degree of subjectivity. This paper presents a novel, semi-automated approach for the delineation of landscape units, the derivation of representative toposequences and their partitioning into terrain components. It incorporates an algorithm to retrieve representative catenas and their attributes for elementary hillslope areas based on elevation and other key spatial data frequently required as environmental model input, e.g. vegetation and soil data. An example application for the Ésera catchment in Spain illustrates that with the presented approach upscaling of hillslope properties becomes feasible for environmental modelling at large scales while ensuring reproducible results.

1. Introduction

In hydrological and erosion modelling, spatial discretisation of the landscape into modelling units is usually accomplished using a fully-distributed, semi-distributed or lumped scheme. While the fully distributed approach is relatively straightforward by often using raster-based input data as derived from, e.g., remote-sensing, the data volume and computational demand increase strongly with extent of the domain and finer grid resolution. This is a limiting factor for practical applications (e.g. Bathurst 2002, Garbrecht & Martz 2000). Moreover, raster cells of fixed resolution and non-adjustable shape impose an artificial discretisation of fluxes (e.g. overland flow) in the process representation.
Models that are not based on a raster representation require pre-processing steps in Geomorphometric Regionalisation (Schmidt & Dikau 1999) to describe the spatial domain in the model. At small scales, 'hillslope-based' models like WEPP (Flanagan & Nearing 1995) or KINEROS (Woolhiser *et al*. 1990) delineate the stream network and parameterise the surrounding runoff contributing areas (hillslopes) with a detailed and geo-referenced representation of their longitudinal profiles. Models designed for larger river basins like WASA (Güntner & Bronstert 2004), SWAT (Neitsch *et al*. 2002) and SWIM (Krysanova *et*

1

al. 2000) apply semi-distributed schemes. Conceptually, hillslopes undergo an object aggregation into functional units of a higher order (e.g. in Hydrological Response Units in SWIM, landscape units in WASA), which are represented by characteristic parameters. Georeference of the elementary object is only partially preserved by assigning areal fractions to higher-order objects (e.g. sub-basins) with known location.

Although GIS can greatly facilitate the retrieval of hillslope information, deriving the exact location of hillslope profiles is often performed manually (e.g. Maurer 1997) and introduces a certain degree of subjectivity (Cochrane & Flanagan 2003). Moreover, the process is labour-intensive or even unfeasible for larger catchments, because available GIS-tools allow further object-aggregation in crude ways only (Schmidt & Dikau 1999). Thus, for use in large-scale models like SWAT and SWIM, hydrological response units are usually derived by mere intersection of GIS-map layers such as land use, management and soil data. For each entity a mean value for slope is computed. This method naturally cannot preserve the intra-slope distribution of properties nor topological information.

Garbrecht & Martz (2000) presented object analysis methods ('Data Reduction methods') for deriving representative values for the length and slope of sub-catchments. However, depending on the definition of these parameters, significant differences may result from the different Data Reduction methods. For deriving basic morphometric hillslope parameters, Cochrane & Flanagan (2003) used a sophisticated weighted mean technique to compute the average slope values and the average length from all flowpaths for a given hillslope area. By this way, more details of the hillslope geometry are preserved and a complete profile instead of a single value for slope is produced. The results can be used in models that represent hillslopes as flowstrips of constant width (e.g. WEPP, KINEROS). Information on variable hillslope width as supported in models such as CATFLOW (Maurer 1997) is not generated, nor can additional attributes such as soil and vegetation be considered.

For semi-distributed models like WASA or SWAT, the derived hillslope properties must be upscaled by performing a further object aggregation. Hillslopes are grouped into classes that comprise profiles with similar distribution of topography, soil and vegetation properties along the hillslope. The resulting objects may be represented by a considerably reduced set of parameters, but the upscaling produces biased results when done manually, especially if multiple attributes are to be considered in the classification process.

Güntner & Bronstert (2004) used the SOTER concept (FAO 1993), transformed by Gaiser *et al*. (2003), to parameterise the sub-areas of river basins with similar hillslope characteristics. The SOTER concept was introduced to provide a consistent method for the worldwide delineation of the so-called SOil-TERrain-units, focussing on the properties mentioned above (topography, soils, etc.). However, this concept was not specifically designed with application to hydrological and erosion modelling in mind. Its implementation depends to a large degree on expert knowledge and thorough familiarity with the area of interest. Moreover, representing large spatially contiguous areas with a single hillslope profile is necessarily a rather strong simplification of reality, especially for relatively heterogeneous landscapes with different hillslopes types in close proximity.

To address the challenges in the parameterisation of 'hillslope-based' models described above, a novel, semi-automated algorithm for the delineation of landscape units is presented in this paper. Although tailored for use with the WASA model, the algorithm can potentially be applied for other models that require the derivation of representative hillslopes in the catchment. Nevertheless, the WASA terminology is used in this paper as illustrated in Fig. 1)1: landscape units (LU) are parts of the catchment that can be characterised by a typical toposequence. Toposequences are idealised hillslope profiles representing the upscaled properties of the hillslopes they represent. Toposequences always start at a local divide and end at the channel. They are composed of terrain components (3 TCs in Fig. 1). Each TC has a distinct slope gradient and soil- and vegetation-association. The term catena is used for a

hillslope profile of a particular hillslope area with a concrete spatial reference. The meaning of both toposequence and catena is not limited to morphometrical attributes, but is used more generally as a set of attributes (including soil properties, for example) along the length of a hillslope.

[Insert fig 1 about here]
Fig. 1: Explanation of terminology: landscape units (LUs) are homogenous parts of the catchment, represented by toposequences that consist of terrain components (TCs).
[A: catchment boundary; B: river; C: example of an EHA]

The presented algorithm called 'LUMP' (Landscape Unit Mapping Program) was designed to fulfil the following demands:
(1) The delineation of landscape units (LU) is to be automated. The algorithm should reduce subjective decisions on spatial discretisation to a minimum but at the same time allow for including prior knowledge. (2) For each LU, a representative toposequence must be computed. The toposequences have to be decomposed into terrain components (TCs). (3) The properties of the resulting LUs and TCs (area, slope, length, soil and vegetation properties, etc.) must be derived as input data for the hydrological model. (4) Besides a silent 'default' mode, a more sophisticated 'expert' mode is necessary to check and modify intermediate results, if desired. (5) For the purpose of visualisation and easy data import/export, the close interaction with a GIS is mandatory.
This paper describes how these functions are implemented in LUMP (Section 2). The quality of the algorithm and its limitations are discussed for an example application in Section 3.

2. Methods

LUMP assesses various representative properties of elementary hillslope areas, assigns similar areas to the same LU class and, finally, produces a spatially continuous map displaying the extension of the LUs within the area of interest. The partitioning of the representative toposequences of the LUs into TCs and computing the resulting TC parameters conclude the tasks of LUMP. The steps of the entire algorithm are illustrated in Fig. 2, detailed explanations are given below. LUMP consists of a set of interacting scripts for GRASS-GIS (GRASS Development Team, 2005) and Matlab™ (Mathworks, 2002). The scripts and a technical documentation are freely available (SESAM, 2006).

[Insert fig 2 about here]

Fig. 2: Steps performed in the LUMP process with reference to the respective section in this paper in brackets

2.1. Delineation of elementary hillslope areas (EHA)

2.1.1. Concept
An elementary hillslope area (EHA) is the basic unit that is used for the calculation of a representative catena. It comprises a contiguous slope area that can be characterised by a representative catena. All EHAs with a similar catena form a LU (see Fig. 1). This means that an EHA must be large enough to cover the range from the channel to the local divide but small enough to contain only one characteristic hillslope type. Thus, its size depends on the spatial scale of the hillslopes to be expected and the resolution of the digital elevation model (DEM) used. Consequently, the minimum size of the EHA also determines the resolution at which the spatial extent of the LUs will be generated.

3

2.1.2. Data used

The delineation of the EHAs requires a digital map of the stream network, provided directly or computed from a DEM. The DEM is also needed to derive flow accumulation with common GIS-operations that are used in later steps.

2.1.3. Algorithm

All flow-related GIS-operations described use a hydrologically corrected, i.e. filled DEM, according to common practice (Garbrecht & Martz 2000).

EHAs are equivalent to small subcatchments, which are hillslope areas that drain into the first adjacent downslope channel (Garbrecht & Martz 2000). Thus, their delineation can be done with standard GIS-algorithms (e.g. r.watershed with GRASS). By applying an appropriate threshold value for flow accumulation area, the catchment can be subdivided into areas that cover the extent of EHAs (Fig. 4). The resulting delineation can be modified manually by the user, if required.

[Insert fig 3 about here]

Fig. 3: Example of a delineation of a catchment into 660 elementary hillslope areas (Ésera basin, North-Eastern Spain, basin area: 1231 km²)

2.1.4. Limitations

Especially with ephemeral rivers, the delineation of the river network – and thus the separation between hillslope and river cells – is not always straightforward. The user has to ensure that the distinction made corresponds to distinct dominating flow regimes (i.e. flow on hillslopes and river flow) and their representation in the model.

Each EHA must have a minimum number of cells to be able to be processed in the next step. When using lower resolution DEMs (cell-size > 50 m), this may lead to EHAs that cover larger areas, which increases the probability of lumping different hillslope types within this EHA. Thus, low resolution DEMs will finally result in coarser and more generalised maps of LUs.

2.2. Derivation of representative catena for each EHA

2.2.1. Concept

The calculation of representative catenas is based on the created EHAs. Preliminary studies indicated that the concept of calculating representative catenas based on a 2D-domain (areal data) perform far better than methods based on randomly sampled hillslope profiles (linear data). The results of the latter are very sensitive to DEM-noise and to variations in the algorithm (Francke 2005, Francke et al. 2006). Instead, by calculating representative catenas from the EHAs, each single cell is included in the computation. The calculations are based on the approach given by Cochrane & Flanagan (2003), extended by additional attributes (e.g. soil, vegetation; further termed 'supplemental data'). Moreover, hillslope width as a function of the distance to the river is also computed.

2.2.2. Data used

This step requires two more grid maps that are generated from the DEM and the river network. The raster map 'relative elevation' contains difference between the elevation of a cell and the elevation of the rivercell it drains to. The map 'flowpath length' is generated from the travel distance of the waterflow from a cell to the river.

Additional spatial data relevant for hydrological and sediment modelling can be included as maps containing either quantitative data (e.g. aspect, LAI, erodibility, groundwater levels, etc) or categorical, i.e. nominal or classified data (soils, land use, aggregated prior knowledge like

hydrological response units or connectivity classes, etc). All data are used at the same resolution as in the DEM.

2.2.3. Algorithm

Cochrane & Flanagan (2003) proposed methods for the computation of a representative catena profile for a given hillslope, which comprise equations for calculating the representative catena length and the representative catena profile. The representative catena length is the length to be used when representing a hillslope as a series of rectangles with constant width (as in the WEPP model). The representative catena profile describes the gradient along the respective hillslope.

Representative catena length

Cochrane & Flanagan (2003) present the 'Chanleng''- and 'Calcleng''-method for the computation of the representative catena length. The Chanleng-method (calculation of catena length based on area and length of adjacent channel) is restricted to hillslopes draining to the sides of channels (i.e. no headwater slopes). It also requires the determination of the length of the adjacent channel reach, is consequently very sensitive to the resolution of the used raster map. Therefore, the Calcleng-method (calculation of catena length based on flowpath lengths) was chosen, which is independent of the calculation of the channel length. The original Calcleng-method is based on the processing of all flowpaths in the hillslope. For the easier-to-perform cell-based calculation, this translates to:

$$L = \frac{\sum_{c=1}^{m} (l_c \times a_c)}{\sum_{c=1}^{m} a_c} \qquad \text{eq. 1}$$

where
L : representative catena length [m]
m : number of cells which have no upslope contributing area [-]
l_c : flowpath length from current cell to river as contained in map 'flowpath length'' [m]
a_c : area of flowpath [m^2]
with
$$a_c \cong l_c \qquad \text{eq. 2}$$

which allows the calculation of L for each EHA. L determines the distance from the bottom of the hillslope at which the representative profile (computed below) is truncated. Fig. 4 illustrates the results of the calculation of the representative length for an example EHA.

Representative catena profile

For the calculation of the representative morphometric profile, the 'Linear Average Representative Slope Profile'' method is used. According to Cochrane & Flanagan (2003), this method produces results which are not significantly different from more elaborate methods such as the 'Exponentially Transformed Average Representative Slope'' and 'Weighted Average Representative Slope Profile''.

When generalising this concept, a representative value not only for slope but for any other attribute can be computed for each point along the representative catena. Again, the original calculation is based on the flowpaths within the hillslope. It implies that each single cell within the hillslope is considered as many times as it is a member of a flowpath. This is equal to weighting a cell's value by the flowpath density at this cell:

$$A_i = \frac{\sum\limits_{c=1}^{n} a_c * fd_c}{\sum\limits_{c=1}^{n} fd_c} \qquad \text{eq. 3}$$

where

A_i : value of attribute of mean catena at the distance i from the channel
n : number of cells in hillslope with distance i from the channel [-]
a : value of attribute at given cell
fd_c : flowpath density at given cell [-]

Flowpath density at each cell c is approximated as:

$$fd_c \cong \sqrt{fa_c} \qquad \text{eq. 4}$$

where fa is the flow accumulation (upslope contributing area) derived before.
The above calculation of the longitudinal profile uses relative elevation rather than slope as used by Cochrane & Flanagan (2003), because the former is independent of the choice of slope calculation methods and thus more robust than the derivative slope (Evans 1990).
For quantitative attributes, eq. 3 can be applied directly. For each categorical attribute with r classes, eq. 3 is processed r times for each class separately. Thus, a mean probability or fraction for each class is computed for every distance i from the channel (see section 'Classification of catenas" for details).
Fig. 4 depicts the results of the calculation of the mean catena profile for an EHA.

Additionally derived hillslope properties
For each EHA, the cumulative density of the number of cells $dens_{cum}$ is calculated as:

$$dens_{cum}(i) = k = |\{c_1, c_2, ...c_k\}| \qquad \text{eq. 5}$$

where c_{index} are all cells with

$$flowpath\_length(c_{index}) \le i \qquad \text{eq. 6}$$

$dens_{cum}$ is a measure of the distribution of hillslope area as a function of its distance from the river. The gradient of this function describes hillslope width. Thus, the areal convergence of EHAs, i.e. the change in width along the length, can be captured (see Fig. 4). This attribute can be used in the classification process and for hillslope parameterisation if required by the model.

[Insert fig 4 about here]

Fig. 4: Example of a representative catena computed from an EHA. Black dots mark individual cells within the EHA. The coarsely dashed vertical line denotes the representative length, which determines the top end of the mean catena profile (bold line).

For each EHA, the representative catena and its attributes is finally resampled to the resolution of the DEM and passed to the classification process.

2.2.4. Limitations
The concept of describing a three-dimensional landscape by using two-dimensional catenas is necessarily a simplification. Although the calculation of hillslope width along the length of the catena preserves some characteristics of the three-dimensional reality, this will hardly reflect the true process of flow concentration downhill, especially because the calculated hillslope width generally increases downslope. This 'convergence paradox" (Boogart & Troch, 2006) can explicitly only be dealt with if the resolution of the DEM allows the

identification of individual flow concentrating features at the hillslope. Alternatively, flow concentration may be treated implicitly within the model (e.g. Güntner & Bronstert 2004).

The number of cells within an EHA decreases with its size. Therefore, the calculation of the representative catena becomes more sensitive to the value of a single (possible erroneous) cell. On the other hand, very large EHAs may average over distinct hillslope types as described in section 'Delineation of EHAs'', which will lead to averaged and probably not very representative catenas.

2.3. Classification of catenas, generation of toposequences

2.3.1. Concept

The previous step produced a representative catena for each EHA. The length and relative elevation gain (difference in elevation between top and foot) of these catenas varies, as does the number of their respective catena points from discrete sampling, depending on the output resolution of the previous step. Beside the morphometrical data, a set of various supplemental attributes (quantitative and/or categorical) can optionally be associated to each point of a catena (as with 'LAI' and 'soils' in Tab. 1).

Tab. 1: Example properties of three (hypothetical) catenas, as returned by the derivation of representative catenas

[Insert tab 1 about here]

LUMP classifies all catenas into a given number of classes using cluster analysis. The attributes used in the clustering process are:

- horizontal and vertical length (elevation gain relative to foot of catena, expressed as single values for each catena)
- shape of hillslope profile, and
- sets of supplemental attributes, stored for each point along the catena, that further characterise hillslope properties.

The classification is not limited to a single value for each catena but regards attribute characteristics along the hillslope. LUMP enables the classification considering multiple attributes with different physical units or categorical data: The successive classification runs perform the classification for each single attribute separately, with the final class assignment resulting from the intersection of the single classification steps.

2.3.2. Algorithm

In cluster analysis, 'similarity'' of two objects is measured with the help of their distance in a multi-dimensional vector-space. Therefore, as a first step all catenas are resampled to a unit-resolution by converting all catenas to the same number of catena-points using linear interpolation which allows their representation as vectors with the same number of elements. The median of the number of sampling points of the catenas is used for determining the number of points u_res in the unit-resolution.

For categorical (i.e. classified) supplemental data, the class-ID merely reflects the membership of a certain class but is numerically meaningless as a quantitative measure. Therefore, any categorical attribute with n classes is internally converted to a vector v of the length n. The relatedness to class m is expressed as a fraction stored at the m'th component of v. This concept allows for incorporating the occurrence of multiple classes at one point (e.g. at catena point 1 soil classes D and E were encountered). It also enables the resampling described above by interpolation fractions and allows including the supplemental data in the clustering process. Tab. 1 and Tab. 2 illustrate this concept.

[Insert fig 5 about here]

Fig. 5: 60 catenas, resampled to a unit resolution of 10 catena points and normalized (example from Ésera catchment)

The elevation profiles of the resampled catenas is then normalised to a vertical extension of unity. This conversion results in an attribute vector which holds the normalised 'shape'' of the hillslope profile and will further be referred to as the shape-attribute. The true horizontal length and the elevation gain are stored as separate attributes of the catena. Tab. 2 gives an example of the internal representation of all attributes of the resampled catena 1.

Tab. 2: Internal representation of catena 1 (see Tab. 1) after resampling.

[insert tab 2 about here]

For each attribute included, LUMP classifies the set of catenas into the specified number of classes. Increasing the number of classes for an attribute considered predominant allows the user to force the algorithm to produce a more detailed classification with regard to that attribute. On the other hand, attributes which are, e.g., set to produce one class only, will not contribute to a further partitioning of the dataset (e.g. attribute 'LAI' in the example given in Tab. 3 is not used to further partition the dataset). The catena attributes 'Horizontal length' and 'Elevation gain' are treated together as a composite attribute (further referred to as xy-extent) as a measure of catena extent and mean slope. An adjustable weighting factor fac_y multiplying 'Elevation gain' allows emphasizing this element of the two-element vector c_{xy}. All other attributes are represented in c_a containing the attribute values along the entire catena:

$$c_{xy} = \{ L_x; \quad fac_y \cdot L_y \} \qquad \text{eq. 7}$$

$$c_a = \{a(1,1); \quad a(1,2) \quad ... \quad a(u\_res, nclasses[na])\}$$

where *nclasses(a)* is the number of classes of the attribute, which is 1 for all quantitative data. The final class membership of a catena results from the unique combination of the successive classification assignment according to each attribute (Tab. 3). Thus, all catenas with an identical classification assignment throughout all attributes are finally to the same class.

Tab. 3: Example of successive classification of 4 catenas with 4 attributes

[insert tab 3 about here]

LUMP uses either an unsupervised K-means clustering algorithm to produce the number of classes specified by the user or a supervised cluster algorithm based on pre-defined end members. Both options use squared Euclidean distances.

A dendrogram, the silhouette coefficient and a silhouette plot (Kaufman & Rousseeuw 1990) can give a visualisation of the quality of the separation and the distinctiveness of the classes. Since each node in the dendrogram represents a split of the respective subgroup, this figure can be a guide in selecting an appropriate number of classes for the given task and attribute. The vertical distance of the nodes usually decrease, indicating that increasing the number of classes yields progressively less improvements in the classification of the dataset.

A representative toposequence for each resulting class is computed by averaging the catena attributes of the members of the respective class:

$$a\_ts(k)_j = \text{mean}[a(h_k)_j] \qquad \text{eq. 8}$$

where $a\_ts(k)_j$ refers to the *j*th attribute of the toposequence representing class *k*. h_k is an index to all catenas belonging to class *k*.

The toposequences are then passed for further processing to the partitioning module (see following section) and stored for inclusion in input files of the model.

[Insert fig 6 about here]

8

Fig. 6: 518 catenas classified into 9 classes (3 classes in attribute shape, 3 classes in attribute xy-extent)
The classification results are re-imported into the GIS by re-classifying each EHA according to the membership of its representative catena (Fig. 7).

[Insert fig 7 about here]

Fig. 7: Example of delineated LUs as a result of a classification

2.3.3. Limitations
Resampling the catenas implies a loss of information for catenas of which the number of catena points is reduced. On the other hand, catenas with few points are internally resampled to a higher resolution. Thus, their number of catena points is increased, although the actual information is not that detailed.

Currently, the user has to specify the number of classes to be produced for each attribute. A silhouette plot and the dendrogram can help with choosing this number appropriately. An automatic selection of the number of classes will be added to future versions of LUMP.

An advantage of the applied method is the optional inclusion of multiple layers of supplemental information into the classification process. This option, however, also requires a certain amount of expert knowledge to adjust the respective number of classes accordingly. It is the responsibility of the user to choose appropriate numbers to produce LUs that are meaningful with regard to the intended modelling purpose.

2.4. Partitioning of the toposequences of the LUs into terrain components (TCs)

2.4.1. Concept
The classification step produces one representative toposequence for each LU. To describe different segments within the hillslope, the toposequences can further be partitioned into terrain components (TCs). A TC is an idealised representation of a continuous part of the toposequence having uniform slope and distinct characteristics of supplemental attributes. The algorithm sub-divides each toposequence into a user-specified number of TCs by delineating parts according to the definition above. Besides slope gradients, each available supplemental attribute can be included into the partitioning process. For each attribute a the respective weighting factor $facTC_a$ has to be specified. This weighting scheme allows including multiple attribute of different physical units.

2.4.2. Algorithm
Each toposequence is converted into a matrix **M** (u_res x $nrows$) so that each column contains the entire set of attributes for one point of the toposequence. $nrows$ is a function of the number of attributes na' at each point of the toposequence:

$$nrows = \sum_{a=1}^{na'} nclasses_a \qquad \text{eq. 9}$$

The user-specified weighting factor fac_a of each attribute a is adjusted according to eq. 10 to ensure consistent weighting independent of the number of classes used in categorical attributes:

$$facTC_a* = facTC_a \frac{1}{ncomp_a \cdot nclasses_a} \qquad \text{eq. 10}$$

where $ncomp_a$ denotes the number of components the attribute uses and $nclasses_a$ is the number of classes of the attribute, which is 1 for all quantitative data. Attribute weighting is performed by multiplying each row of **M** with the resulting weighting factor $facTC_a*$.

Tab. 4: Internal representation of a toposequence for TC-partitioning with all weighting factors set to 1.

[Insert tab 4 about here]

LUMP employs an optimisation towards minimum variance for delineating similar parts within a toposequence. This method partitions the toposequence in a way that the overall variance $v_{o,p}$ within the n_{TC} TCs is minimised throughout all possible permutations of partitionings p:

$$v_{o,p} = \sum_{j=1}^{n_{TC}} l_j \cdot \mathrm{var}(TC_j)$$

eq. 11

$$|v_{o,p}| \to Min$$

The factor l_i is a weighting term that equals the length of the respective TC.

The overall variance v_o of a given partition p is a vector of *nrows* elements. Each element contains a single variance value, computed from weighted sum of variances of one attribute. To compare the $v_{o,p}$ of different partitionings p the vector-norm of $v_{o,p}$ is used.

[Insert fig 8 about here]

Fig. 8: An example toposequence of a LU, partitioned to 5 TCs based on minimisation of variance

2.4.3. Limitations

The choice of an adequate number of TCs has to be made by the user, pondering the contradicting demands of an appropriate representation and a reasonable generalisation. Thus, it is a function of the landscape characteristics and the requirements and capabilities of the target model and has to be adjusted accordingly.

For sub-dividing the toposequence into TCs, the slope data are incorporated into the algorithm. The slope at each point of the toposequence is calculated from the horizontal spacing and the elevation gain to the next uphill point. At the most uphill point of the toposequence, however, the previous downhill point must be used, because no uphill point is available.

The TC-concept allows the occurrence of several soil and vegetation classes within a TC. In theory, a hillslope segment with constant slope and soils A and B alternating exactly at each point of the toposequence is conceptually a TC with uniform soil characteristics. This hypothetical example, however, has a rather high variance component for the soil attribute which might spuriously force the described algorithm to split this hillslope segment into several TCs.

3. Example application and discussion

3.1. Study area

An example run of LUMP was performed for the Ésera catchment (Central Spanish Pyrenees), which is located at about 42°20'N and 0°30'E. The catchment, with an area of 1231 km², is part of the Ebro Basin and is characterised by heterogeneous relief, vegetation and soil characteristics. Elevation increases from 430 m in the southern and central parts of the catchment (*Intermediate Depression* and *Internal Ranges*) to up to 3000 m asl in the northern parts (*Axial Pyrenees*, Valero-Garcés *et al.* 1999, see Fig. 9, left). The climate is a typical Mediterranean mountainous type with mean annual precipitation rates of 600 to 1200 mm and an average potential evaporation rate of 550 to 750 mm, both rates showing a strong south-north gradient due to topography. The vegetation includes deciduous oaks, agriculture, pastures and mattoral in the valley bottoms, evergreen oaks, pines and mattoral in the higher areas (see Fig. 9, center). While the northern parts are composed of Paloeozoic rocks, Paloegene and Cretaceous sediments, the lower parts are mainly dominated by Miocene continental sediments. These areas consist of easily erodible materials (marls,

sandstones, carbonates), leading to the formation of badlands (Fig. 9, right) and making them the major source of sediment within the catchment (Fargas *et al.*, 1996).

[Insert fig 9 about here]

Fig. 9: relief (left), land use (center) and occurrence of badlands (right) in the Ésera catchment, NE-Spain

3.2. Application of LUMP, results

Aiming at the parameterisation of a hydrological and sediment transport model, the geospatial data of elevation (30-m-DEM derived from ASTER imagery), the mean LAI (derived from the land use map, C.H.E 1998) and the occurrence of badlands (derived from orthophotos) were assumed important proxies for runoff and sediment dynamics and processed with the LUMP algorithm. Based on the delineation of the catchment into 930 EHAs, representative catenas were derived as described in the methods section. LUMP was configured to classify the catenas into three classes regarding the attributes hillslope-shape, xy-extent, LAI and into two classes with regard to badland-occurrence.
Of the 54 (3x3x3x2) possible LU-classes, 42 classes resulted. The final LU-delineation is the intersection of the four maps below (Fig. 10a-d), but for legibility, the map is displayed for each of the attributes with classes of similar properties grouped within the same shade.

[Insert fig 10 about here]

Fig. 10: LUs delineated by LUMP, grouped by similar classes within the attributes shape (a), xy-extension (b), LAI (c) and badland occurrence (d)
The LU-delineation viewed according to the attributes xy-extent, LAI and badland-occurrence (Fig. 10b-d) show clear correlations with the input maps of the respective attribute (Fig. 9). The central to southern parts of the catchment are covered by LUs with flat hillslope profiles, steeper and longer catenas are only found in the northern parts. This distribution matches the actual properties of the catchment. The LAI-aggregated map of the LUs (Fig. 10c) resembles the land use map with LUs of high LAI mainly located in woodland areas and low-LAI LUs to be found in the valley bottoms where agriculture and pastures prevail. The spatial resolution of the LU-map is considerably coarser than the land-use map as an effect of the minimum size of the EHAs, the low number of three classes used for this attribute and the resulting averaging effects. This also explains the characteristics of the LU-map with regard to badland occurrence (Fig. 10d). The general location of the badland areas and their distribution is reproduced adequately by the LU-map, but the level of detail is reduced during the upscaling process.
The quality of the representation of the shape-attribute is difficult to assess from Fig. 10a. Therefore, the LUs were aggregated by shape and xy-extension class, resulting in nine classes. The respective areas were re-analysed as described in the section 2.2. This procedure allows a visual validation of the distribution of hillslope properties within the delineated LUs, which are supposed to be similar in shape and xy-extent within a class. Furthermore, the consistency of the algorithm can be assessed by comparing the properties of the toposequences (generated by LUMP) and the representative catenas (derived directly from the assigned area of each LU using eq. 1 and eq. 3).

[Insert fig 11 about here]

Fig. 11: Mean toposequences of aggregated LUs (LUMP output) and representative catenas derived directly from the LU-areas
In Fig. 11, the 3 classes of the shape-attribute (straight, concave, convex) are arranged column-wise, the xy-extent-attribute (short-flat, long-flat, long-steep) is ordered in rows.

11

Comparing the different scatter-characteristics of all the cells in each of the nine LU-aggregations, it can be concluded that LUMP partitioned the catchment into distinctive classes. There however, remains a large variation in the hillslope morphometry within each LU, especially in the case of the combination concave/long-steep. This fact indicates that the low number of 3-by-3 classes for hillslope-morphometry (chosen for illustrative purposes) results in LU-classes that still comprise a considerable variety of morphometrical hillslope types. In order to represent the wide range of hillslope types in the catchment, more LU-classes are recommended to decrease the variance within the LUs.

Fig. 11 shows that the toposequences of the LUs closely resemble the representative catenas derived from the re-analysis of the respective areas. Slight deviations are only evident at the upper parts of some LUs. Thus, the algorithm proves to be consistent because it delineates LUs and produces respective toposequences that are equivalent to representative catenas that are derived directly from these areas.

3.3. Discussion

Although the delineation of LUs and the derivation of toposequences produced plausible and consistent results, their appropriateness is ultimately to be judged by the performance of model applications. The performance of the LUMP results as model input will strongly depend on the selected landscape attributes and the number of classes into which each attribute is classified. Test runs with the hydrological model WASA for the Ésera catchment suggest that simulation results in terms of river runoff are sensitive to shifts in the classification focus between attributes and to the resulting modelling units and parameters at the sub-basin scale (in average 75 km² in size), while this was not the case at the basin scale (1231 km²) (Francke et al. 2006).

Little experience yet exists for selecting the appropriate number of classes into which a specific landscape attribute should be classified. For hillslope erosion, some authors tried to quantify the relative importance of multiple process factors (e.g. Curtis *et al*. 2005, Schoorl *et al*. 2004). Scherrer & Naef (2003) classified various soil and terrain attributes according to their role for runoff generation processes. These findings may give an indication as to which attributes are to be included in the classification in great detail, i.e., in many classes. The transferability to other catchments, however, remains uncertain.

Moreover, the optimal set of class numbers will depend on the model used and its particular process representation. A model that does not consider a certain attribute in its process parameterisation is unlikely to improve performance when this attribute is resolved in great detail in the delineation process. Furthermore, the target variable (e.g. runoff coefficient, sediment yield) for which the simulation is to be optimised will likely affect the choice of attributes to be included and their number of classes in the classification. The investigation of the relations between model performance and the number of classes for different attributes for a given model and target variable is a future task and is beyond the scope of this study.

An unresolved problem remains the definition of the hillslope extent and the length of its representative catena. Saying that the hillslope should reach from the watershed divide down to the river, the determination of its length encompasses a scale problem, i.e., it depends on the resolution of the available river network information. This, in turn, depends either on the scale of an existing river network map or on the user-defined threshold to indicate at which flow accumulation value the river network starts when generating by means of GIS analysis. The lack of an unbiased definition for the initiation point of a river is an inherent problem when deriving parameters like hillslope length (Schmidt & Dikau 1999). Conceptually, the threshold should aim at separating the 'hillslope" and 'river" domain according to the prevailing transport processes and how they are most suitably represented by the respective model equations.

4. Conclusions

LUMP is a tool for the semi-automated delineation of landscape units and their partitioning into terrain components. It facilitates the preparation of spatial data for the application of semi-distributed, hillslope-based models and ensures reproducible results. LUMP allows for including expert knowledge by incorporating various landscape attributes and by 'weighing' them by a large number of classes according to the perceptual understanding of their impact on catchment processes. Thus, it overcomes several shortcomings of the discretisation strategies currently used in semi-distributed modelling: the hillslope-based parameterisation becomes feasible for larger spatial domains due to the automated algorithm while ensuring reproducible results by reducing subjective decisions.

The LUMP algorithm derives representative hillslope parameters and upscales the hillslope properties with the help of landscape units. In this way, the delineation of modelling units and their parameterisation can be performed automatically for meso- to large scale catchments where a manual procedure is unfeasible and upscaling is mandatory. In contrast with methods based on mere intersection of multiple layers, LUMP preserves information on the distribution of landscape parameters in relation to the river network and their topographic position and thus allows for addressing connectivity issues in model applications.

In the presented application example, LUMP showed a satisfying capability of delineating LUs. Depending on the chosen attributes and respective number of classes, different spatial discretisation schemes of the same study area may result. The optimum number of classes and the selection of attributes depend on the choice of the model used and the target variable for which the calculation is to be optimised. LUMP provides new opportunities for further research on this subject because it allows the efficient and reproducible investigation of the effects of spatial discretisation in semi-distributed modelling. The implications of applying the LUMP-derived results in meso-scale hydrological and sediment modelling are currently being investigated.

5. Acknowledgement

6. References

BOOGART, P.W. AND TROCH, P., 2006, The convergence paradox: are catchment divergent or convergent. *HESS-D,* forthcoming

BRONSTERT, A. AND PLATE, E. J., 1997, Modelling of runoff generation and soil moisture dynamics for hillslopes and micro-catchments. *Journal of Hydrology,* **198,** 177-195.

C.H.E, 1998, Usos de Suelos (1984/1991/1995) de la cuenca hidrográfica del Ebro; 1:100.000, Consultora de M. Angel Fernández-Ruffete y Cereyo, Oficina de Planificación Hidrológica, C. H. E., http://oph.chebro.es/

CSIRO, 1992, *TOPOG User Guide, Version 4.0.* , CSIRO, Canberra

COCHRANE, T. A. AND FLANAGAN, D. C., 2003, Representative Hillslope Methods for Applying the WEPP model with DEMs and GIS. *Transactions of the ASAE,* **4,** (46), 1041–1049.

CURTIS, J. A., FLINT, L. E., ALPERS, C. N. AND YARNELL, S. M., 2005, Conceptual model of sediment processes in the upper Yuba River watershed, Sierra Nevada, CA. *Geomorphology,* **68,** 149–166.

EVANS, I., 1990, General geomorphometry. In *Geomorphological techniques,* A. Goudie (Ed.) (London: Unwin Hyman) pp. 44-56.

FAO, 1993, Global and national soils and terrain digital databases (SOTER). *Procedures Manual, World Soil Resources Reports No.74.* Rome, Italy

FARGAS, D., MARTÍNEZ-CASASNOVAS, J. AND POCH, R., 1997, Identification of Critical Sedmient Source Areas at Regional Level. *Phys. Chem. Earth.,* **22,** (3-4), 355-359.

FLANAGAN, D. AND NEARING, M., 1995, *USDA Water Erosion Prediction Project - Hillslope profile and*

watershed model documentation. NSERL Report 10

FRANCKE, T., 2005, Consistency tests of the Landscape Unit Mapping Program (LUMP) for WASA. SESAM Project Working Paper, University of Potsdam, download from: http://brandenburg.geoecology.uni-potsdam.de/projekte/sesam/reports.php#workingpapers

FRANCKE, T., CREUTZFELDT, B., GÜNTNER , A., MAERKER , M., MAMEDE , G. , MUELLER , E.N. , 2006 Spatial discretisation in semi-distributed hydrological modelling using the Landscape Unit Mapping Program (LUMP). Poster presented at EGU assembly 2006, Vienna, download from: http://brandenburg.geoecology.uni-potsdam.de/projekte/sesam/reports.php#presentations

GAISER, T., GRAEF, F., HILGER, T., FERREIRA, L. AND STAHR, K., 2003, An information system for land resources in Piauí and Ceará. In *Global change and regional impacts: water availability and vulnerability of ecosystems and society in the semi-arid Northeast of Brazil,* T. Gaiser, M. Krol, H. Frischkorn and J. Araú jo (Ed.) (Berlin: Springer), pp. 267–278.

GARBRECHT, J. AND MARTZ, L. W., 2000, Digital Elevation Model Issues in Water Resources Modeling. In *Hydrologic and Hydraulic Modeling Support with Geographic Information Systems,* D. Maidment and D. Djokic (Ed.) (ESRI Press), pp. 1-27.

GERLINGER, K. AND SCHERER, U., 1998, Simulating Soil Erosion and Phosphorus Transport on Loess Soils Using Advanced Hydrological and Erosional Models. In *Modelling Soil Erosion, Sediment Transport and Closely Related Hydrological Processes,* W. Summer, E. Klaghofer and W. Zhang (Ed.) (Wallingford: IAHS Press), pp. 119-127.

GRASS DEVELOPMENT TEAM. 2005. Geographic Resources Analysis Support System (GRASS). Ver. 6.1, GNU General Public License. ITC-irst, Trento, Italy, http://grass.itc.it

GÜNTNER, A. AND BRONSTERT, A., 2004, Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas. *Journal of Hydrology ,* **297,** (1-4), 136-161.

KAUFMAN, L. AND ROUSSEEUW, P., 1990, *Finding Groups in Data: An Introduction to Cluster Analysis* Wiley)

KRYSANOVA, V., WECHSUNG, F., ARNOLD, J., SRINIVASAN, R. AND WILLIAMS, J., 2000, *SWIM (Soil and Water Integrated Model), User Manual.*, PIK-Report, Vol. 77, Potsdam Institute for Climate Research, Potsdam, Germany

MATHWORKS, 2002, MatLab – The language of technical computing. Version 6.5. The MathWorks, Natick, Massachusetts

MAURER, T., 1997, *Physikalisch begründete, zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten.* PhD-thesis, Universität Karlsruhe, Mitteilungen des Instituts für Hydrologie und Wasserwirtschaft, H. 61

NEITSCH, S., ARNOLD, J., KINIRY, J., WILLIAMS, J. AND KING, K., 2002, *Soil and water assessment tool-theoretical documentation, version 2000.* , Texas Water Resources Institute

SCHERRER, S. AND NAEF, F., 2003, A decision scheme to indicate dominant hydrological flow processes on temperate grassland. *Hydrological Processes*, **17,** (2), 391-401

SCHMIDT, J. AND DIKAU, R., 1999, Extracting geomorphometric attributes and objects from digital elevation models - semantics, methods, future needs. In *GIS for earth surface systems : analysis and modelling of the natural environment*, R. Dikau & H. Saurer, (Ed.) (Gebrüder Bornträger), pp. 153-173

SCHOORL, J. M., BOIX FAYOS, C., DE MEIJER, R. J., VAN DER GRAAF, E. R. AND VELDKAMP, A., 2004, The 137Cs technique applied to steep Mediterranean slopes (Part I): the effects of lithology, slope morphology and land use. *Catena,* **57,** (1), 15-34.

SESAM, 2006, Sediment Export from large Semi-Arid Catchments: Measurement and Modelling, project website, Potsdam, Germany, http://brandenburg.geoecology.uni-potsdam.de/projekte/sesam

VALERO-GARCÉS, B. L., NAVAS, A., MACHÍN, J. AND WALLING, D., 1999, Sediment sources and siltation in mountain reservoirs: a case study from the Central Spanish Pyrenees. *Geomorphology,* **28,** 23–41.

WOOLHISER, D., SMITH, R. AND GOODRICH, D., 1990, *KINEROS, A Kinematic Runoff and Erosion Model: Documentation and User Manual.* , Vol. ARS-77, A. R. S. U S. Department of Agriculture

Figure Caption

Fig. 1: Explanation of terminology: landscape units (LUs) are homogenous parts of the catchment, represented by toposequences that consist of terrain components (TCs).
[A: catchment boundary; B: river; C: example of an elementary hillslope area (EHA)]

Fig. 2: Steps performed in the LUMP process with reference to the respective section in this paper in brackets

Fig. 3: Example of a delineation of a catchment into 660 elementary hillslope areas (Ésera basin, North-Eastern Spain, basin area: 1231 km²)

Fig. 4: Example of a representative catena computed from an EHA. Black dots mark individual cells within the EHA. The coarsely dashed vertical line denotes the representative length, which determines the top end of the mean catena profile (bold line).

Fig. 5: 60 catenas, resampled to a unit resolution of 10 catena points and normalized (example from Ésera catchment)

Fig. 6: 518 catenas classified into 9 classes (3 classes in attribute shape, 3 classes in attribute xy-extent)

Fig. 7: Example of delineated LUs as a result of a classification

Fig. 8: An example toposequence of a LU, partitioned to 5 TCs based on minimisation of variance

Fig. 9: relief (left), land use (center) and occurrence of badlands (right) in the Ésera catchment, NE-Spain

Fig. 10 : LUs delineated by LUMP, grouped by similar classes within the attributes shape (a), xy-extension (b), LAI (c) and badland occurrence (d)

Fig. 11: Mean toposequences of aggregated LUs (LUMP output) and representative catenas derived directly from the LU-areas
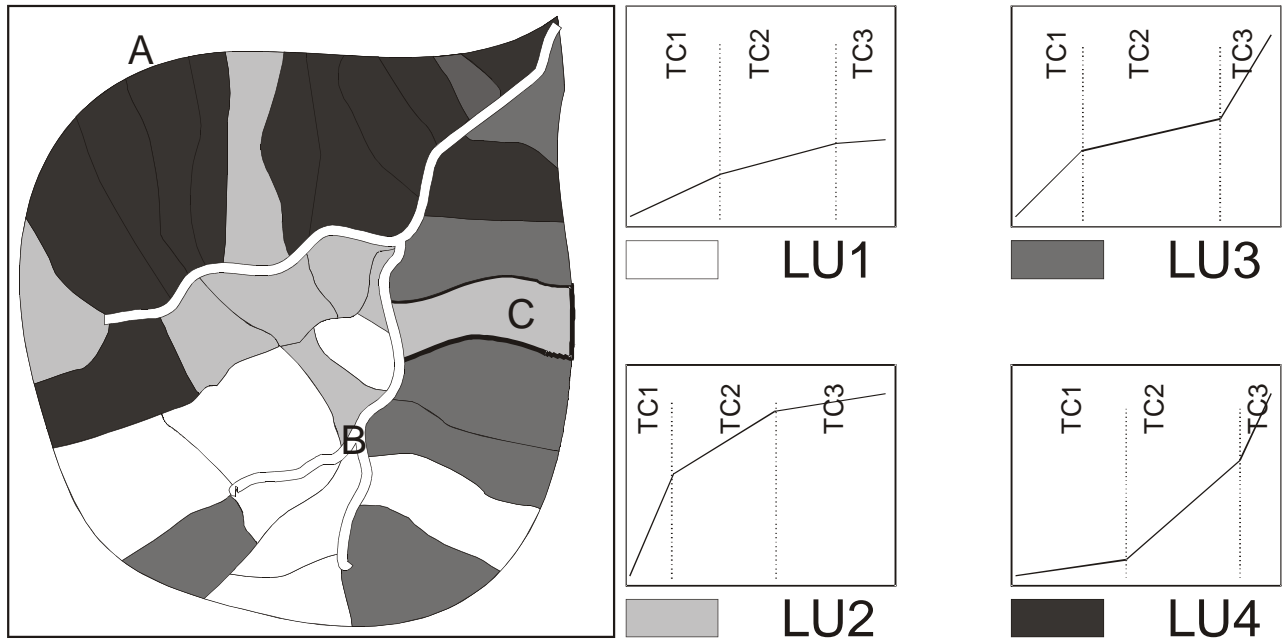
Table Caption

Tab. 1: Example properties of three (hypothetical) catenas, as returned by the derivation of representative catenas
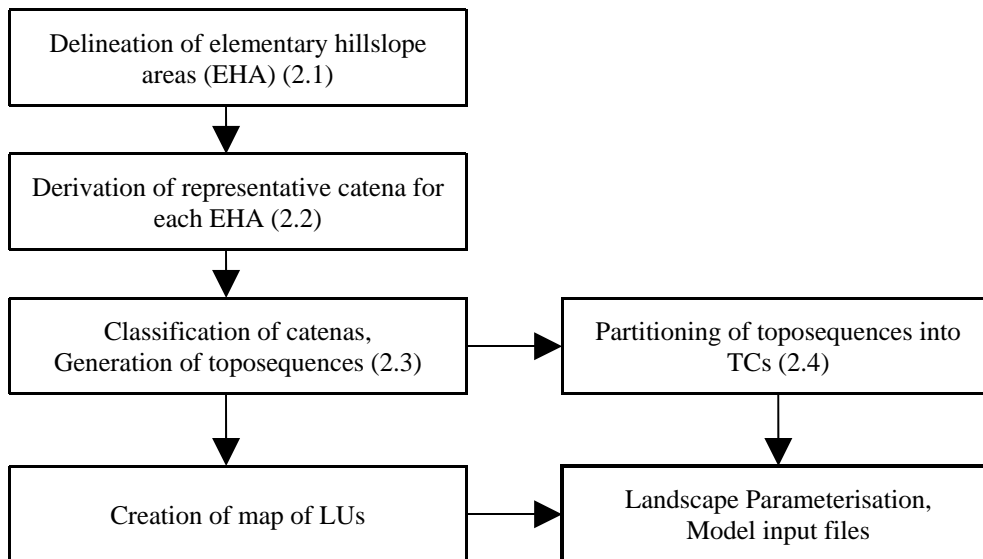
Tab. 2: Internal representation of catena 1 (see Tab. 1) after resampling.

Tab. 3: Example of successive classification of 4 catenas with 4 attributes

Tab. 4: Internal representation of a toposequence for TC-partitioning with all weighting factors set to 1.
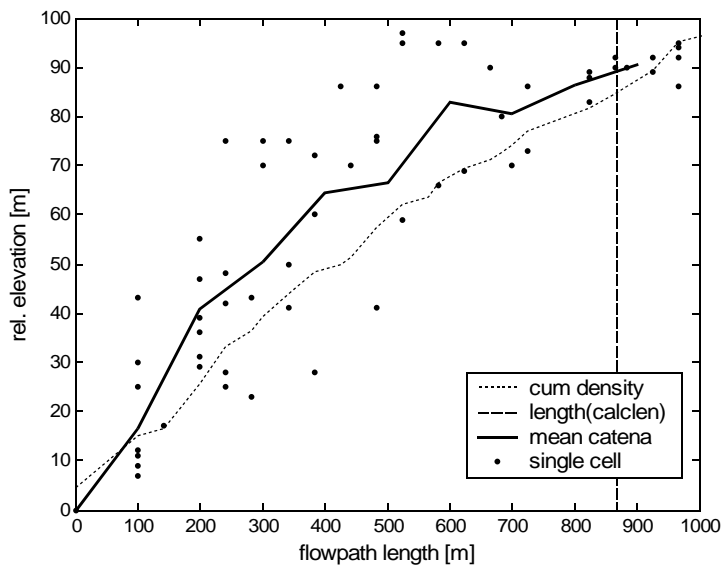
Fig. 1: Explanation of terminology: landscape units (LUs) are homogenous parts of the catchment, represented by toposequences that consist of terrain components (TCs).
[A: catchment boundary; B: river; C: example of an elementary hillslope area (EHA)]

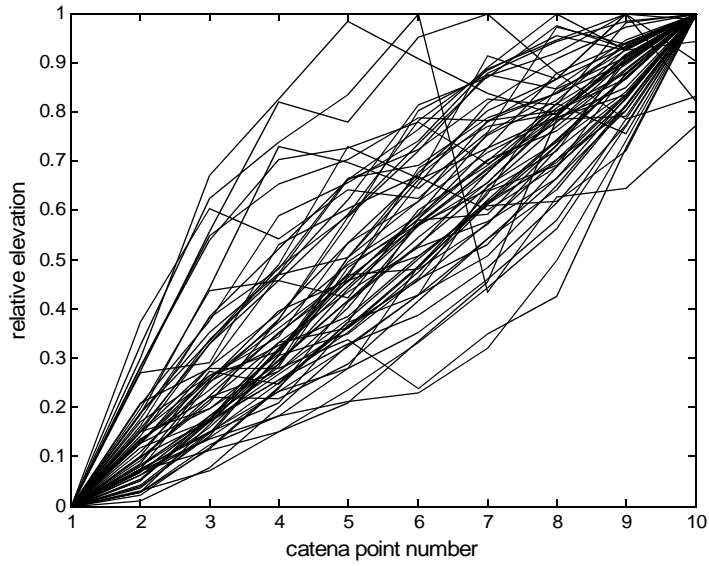

Fig. 2: Steps performed in the LUMP process with reference to the respective section in this paper in brackets
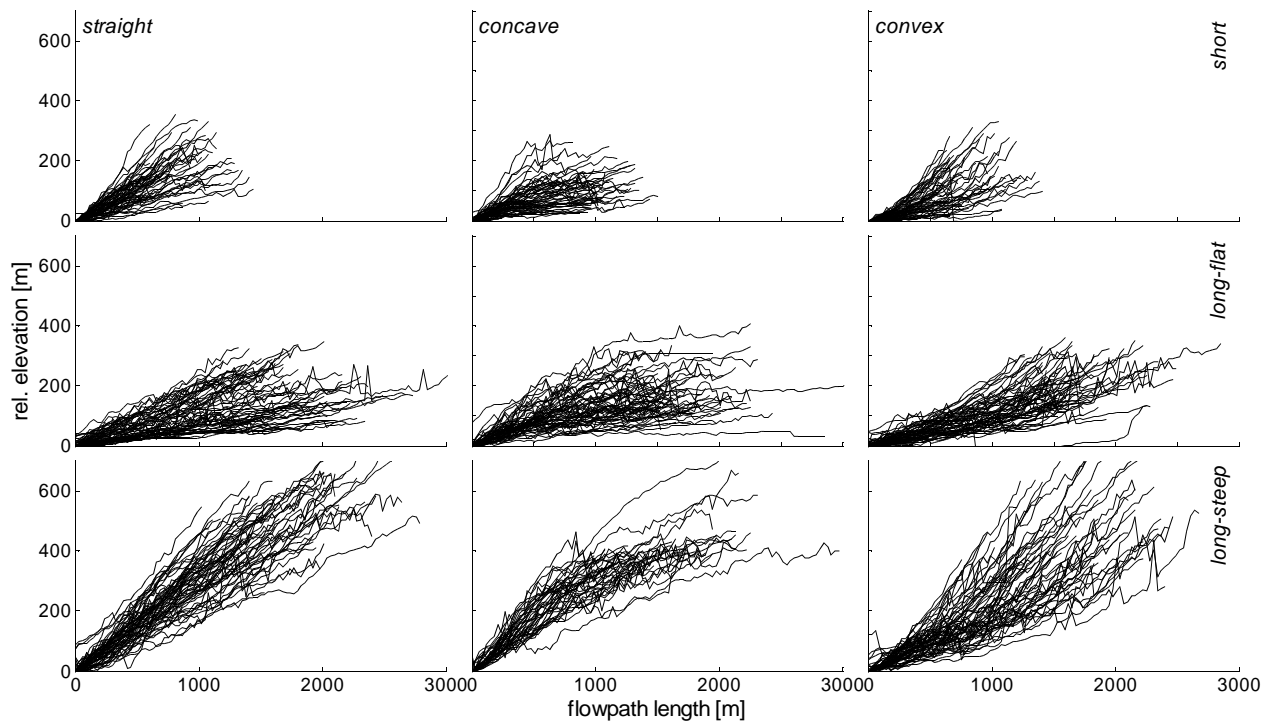
Fig. 3: Example of a delineation of a catchment into 660 elementary hillslope areas (Ésera basin, North-Eastern Spain, basin area: 1231 km²)
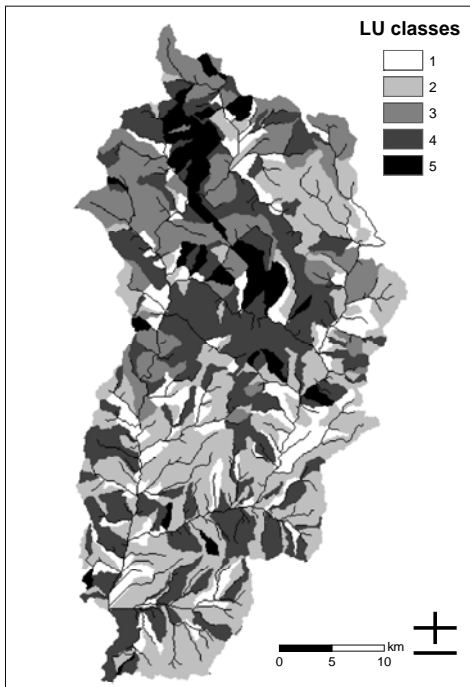

Fig. 4: Example of a representative catena computed from an EHA. Black dots mark individual cells within the EHA. The coarsely dashed vertical line denotes the representative length, which determines the top end of the mean catena profile (bold line).

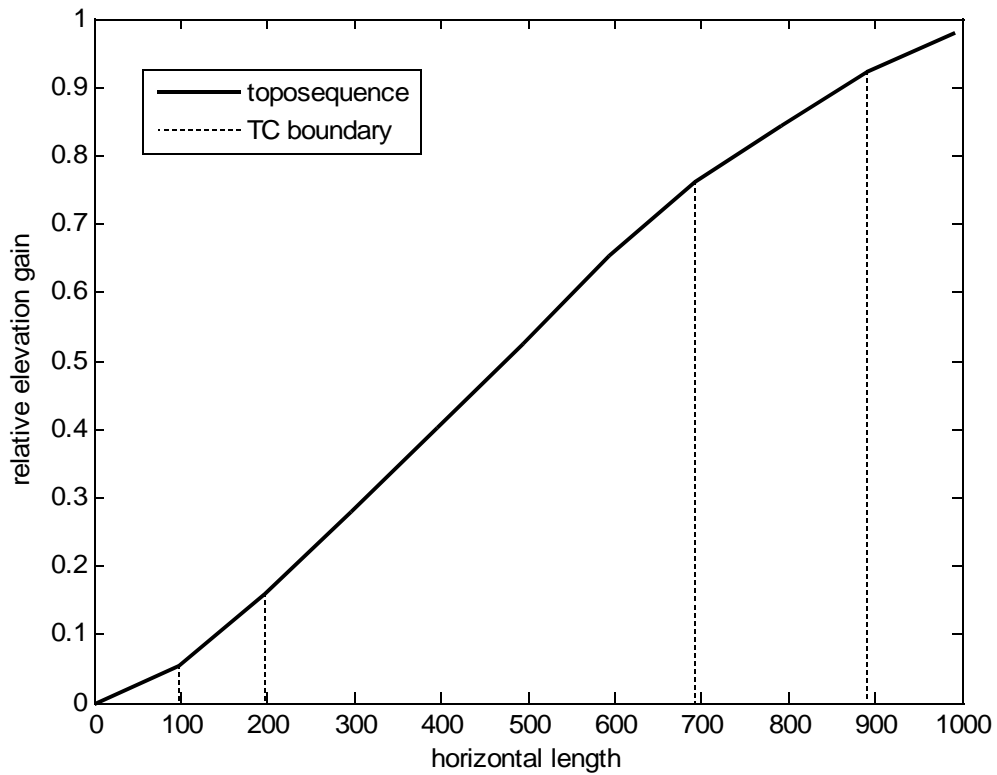**Fig. 5: 60 catenas, resampled to a unit resolution of 10 catena points and normalized (example from Ésera catchment)**

Fig. 6: 518 catenas classified into 9 classes (3 classes in attribute shape, 3 classes in attribute xy-extent)

Fig. 7: Example of delineated LUs as a result of a classification

Fig. 8: An example toposequence of a LU, partitioned to 5 TCs based on minimisation of variance
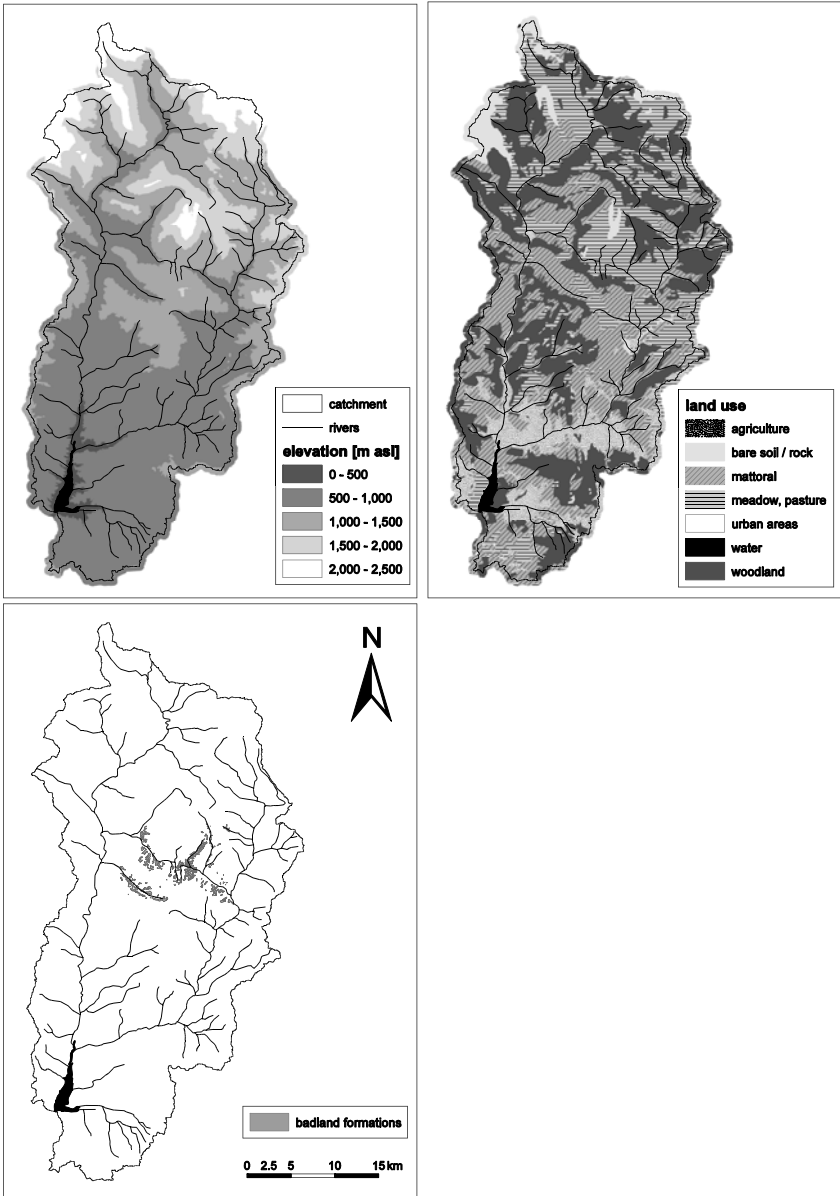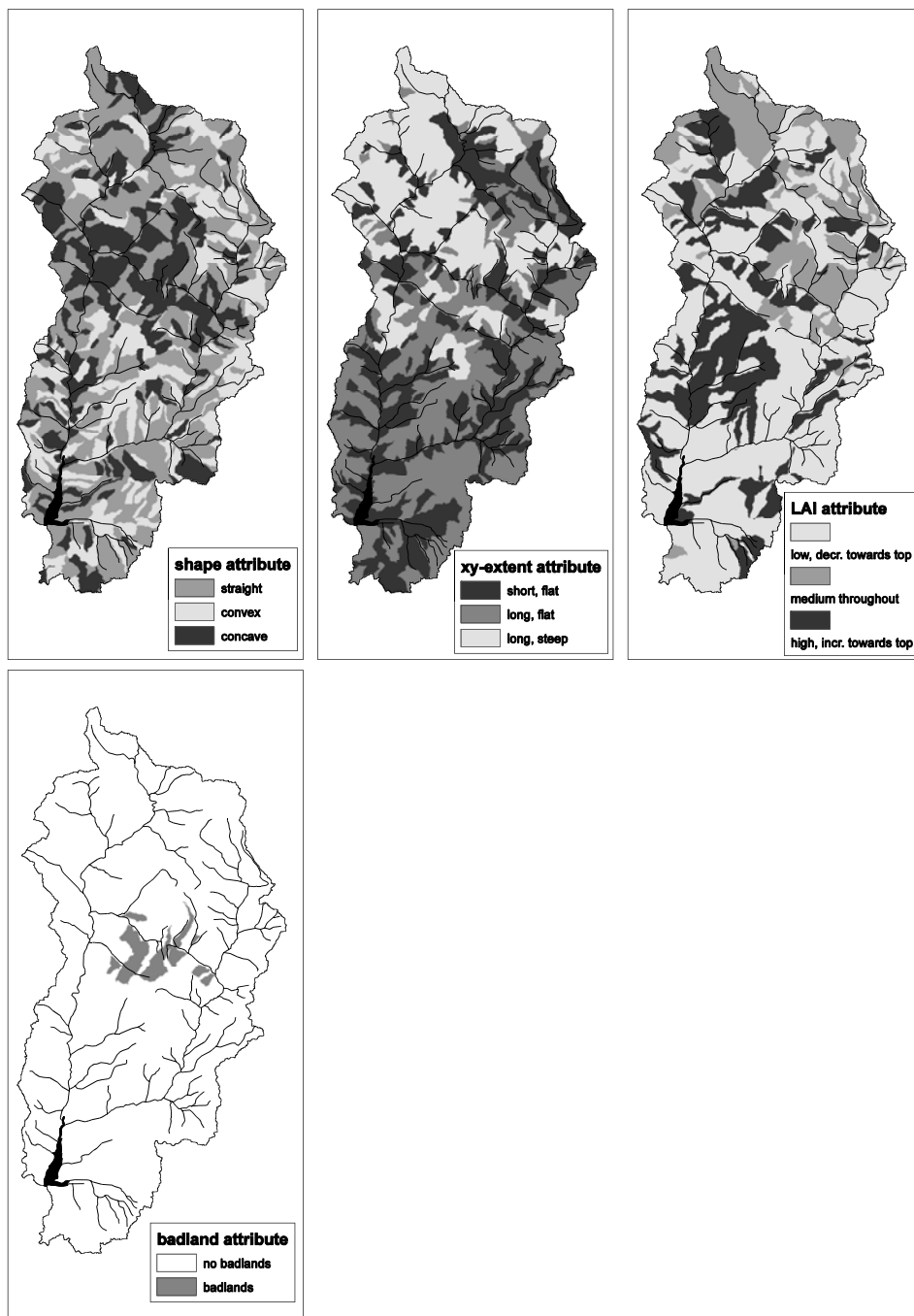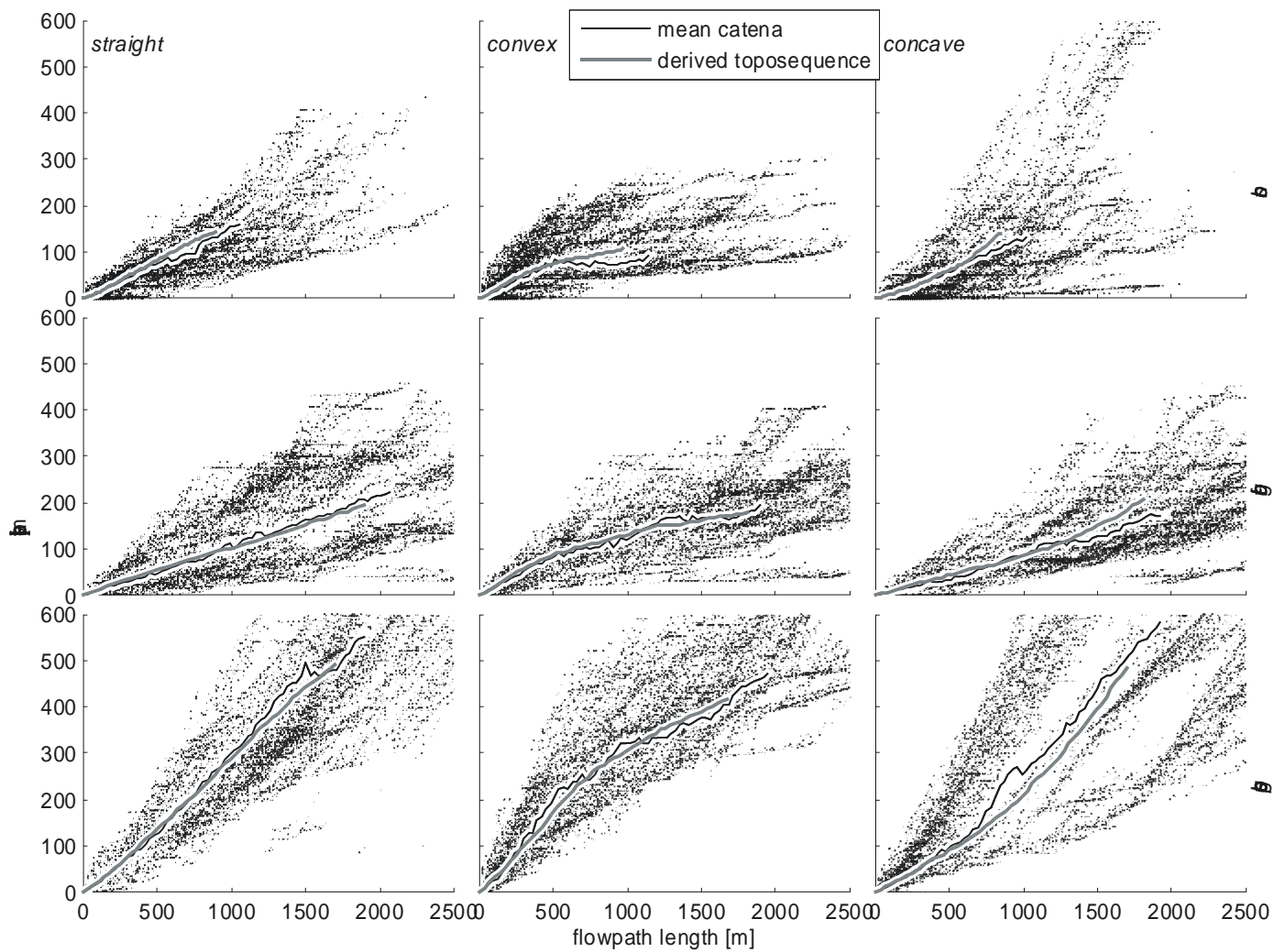
Fig. 9: relief (left), land use (center) and occurrence of badlands (right) in the Ésera catchment, NE-Spain

Fig. 10 : LUs delineated by LUMP, grouped by similar classes within the attributes shape (a), xy-extension (b), LAI (c) and badland occurrence (d)

Fig. 11: Mean toposequences of aggregated LUs (LUMP output) and representative catenas derived directly from the LU-areas

Tab. 1: Example properties of three (hypothetical) catenas, as returned by the derivation of representative catenas								
Catena-ID	point-ID	Elevation	LAI	Soils – areal fractions				
				Soil A	Soil B	Soil C	Soil D	Soil E
1	1	2025	9.1	0	0	0	0.5	0.5
1	2	2026	6.2	0.4	0	0	0.2	0.4
1	3	2108	3.0	0.4	0	0	0.2	0.4
1	4	2145	2.0	0	0.2	0	0.2	0.6
1	5	2211	3.0	0	0	0	0.6	0.4
2	1	2163	4.6	0.5	0	0.25	0	0.25
2	2	2229	2.5	0.25	0	0.5	0	0.25
2	3	2281	6.5	0.6	0	0.2	0	0.2
2	4	2352	2.7	0.2	0	0	0.6	0.2
2	5	2394	4.5	0.2	0	0	0.2	0.6
2	6	2470	3.3	0.4	0	0	0.4	0.2
2	7	2556	1.5	0.2	0	0	0.6	0.2
2	8	2637	1.5	0.2	0	0	0.2	0.6
3	1	1528	10.2	0	0	0	0.2	0.8
3	2	1592	6.5	0	0	0	0.4	0.6
3	3	1816	5.5	0	0	0.2	0.6	0.2
...								

Tab. 2: Internal representation of catena 1 (see Tab. 1) after resampling.									
Catena point, resampled	Horizontal length	Shape	LAI	Fraction					
				Soil A	Soil B	Soil C	Soil D	Soil E	
1	482.80	0.00	9.10	0.00	0.00	0.00	0.50	0.50	
2		0.00	7.10	0.28	0.00	0.00	0.32	0.40	
3		0.12	5.34	0.40	0.00	0.00	0.20	0.40	
4		0.34	3.78	0.40	0.00	0.00	0.20	0.40	
5		0.38	3.55	0.35	0.00	0.00	0.15	0.50	
6		0.45	3.38	0.35	0.00	0.00	0.15	0.50	
7		0.49	2.76	0.30	0.05	0.00	0.20	0.45	
8	Elev. gain	0.59	2.27	0.11	0.15	0.00	0.20	0.55	
9		0.76	2.31	0.00	0.14	0.00	0.32	0.54	
10	186.00	1.00	3	0.00	0.00	0.00	0.60	0.40	

Tab. 3: Example of successive classification of 4 catenas with 4 attributes					
	Attribute				
Catena ID	horizontal length, elevation gain (2 classes)	shape (3 classes)	LAI (1 class)	soil (2 classes)	final classification (<= 2x3x1x2 classes)
1	1	1	1	1	**1**

2	2	2	1	2	**2**
3	1	3	1	2	**3**
4	1	1	1	1	**1**
…					

Tab. 4: Internal representation of a toposequence for TC-partitioning with all weighting factors set to 1.										
toposequence point	1	2	3	4	5	6	7	8	9	10
Slope	0.00	0.10	0.18	0.13	0.08	0.14	0.20	0.20	0.18	0.13
LAI	9.1	7.1	5.3	3.8	3.6	3.4	2.8	2.3	2.3	3.0
Soil A	0	0.28	0.4	0.4	0.35	0.35	0.3	0.11	0	0
Soil B	0	0	0	0	0	0	0.05	0.15	0.14	0
Soil C	0	0	0	0	0	0	0	0	0	0
Soil D	0.5	0.32	0.2	0.2	0.15	0.15	0.2	0.2	0.32	0.6
Soil E	0.5	0.4	0.4	0.4	0.5	0.5	0.45	0.55	0.54	0.4