Automated geomorphological classification of the buëch catchment (france) using multiple point geostatistics

Lucie Babel

MSc Thesis

Supervisors : Dr. Derek Karssenberg
Ekkamol Vannametee

May 2012

Utrecht University

Department of Physical Geography, Faculty of Geosciences

Acknowledgments

It is certainly a priviledge to work on a topic which interested one so much that putting an end to it felt difficult. For giving me this chance, and for involving me in research projects in which I always felt challenged and motivated under his supervision, I would like to thank very sincerely Dr. Derek Karssenberg.

I would like to thank Ekkamol Vannametee too for his support throughout the entire research and his patience on days of informatics despair. My thanks go also to Dr. Martin Hendriks for all informations about the Buëch region he provided me with during my field trip, as well as to Oliver Schmidt and Judith Verstegen for computer assistance.

Finally, I would like to thank Niels for lending me a laptop for almost a year, my father for his feedbacks and interest, my mother for her unconditional support, and the family Grimberg for their warm hospitality.

Abstract

The delineation of geomorphological units through field mapping is particularly time consuming in large catchments and involves a high degree of subjectivity. In the context of large-scale catchment modelling, an alternative to the draw-backs of field mapping is to resort to automated landscape classification. Traditional variogram-based simulation methods have been shown to be unappropriate when dealing with curvilinear features, as it is the case with landscape patterns. In order to overcome this limitation, multiple-point geostatistical methods have been developed in the last two decades, primarily in the field of geological reservoir simulation. In the present research, a multiple-point geostatistical technique based on the single normal simulation equation (snesim) was applied to the geomorphological mapping of the Buëch catchment, Southern France.

The multiple-point geostatistical technique used in this study relies on deriving the conditional relationship between the properties of a cell and its geomorphological unit from a training image. This approach follows two main steps: the storage of multiple point geostatistics (MPG) in a dynamical data structure (search tree), and the simulation. In the first stage, the training image is scanned; the number of occurrences of geomorphological units, associated to properties (attributes) discretized into classes, is stored in the search tree. In the second stage, all cells of the catchment are visited in a random order. For each cell, the conditional probability is retrieved from the search tree, on the basis of which a geomorphological unit is assigned to the cell.

The main research objective of this study was to assess the capability of the MPG automated classification method to reconstruct the geomorphology of the area. To this end, three different training areas of identical size were used. Six topographical attributes and two neighbourhood attributes were defined, along with six different number of classes. An optimization design aiming at identifying the best combination of topographical and neighbourhood attributes, as well as the most performing number of classes, was then followed for each of the training areas.

The algorithm reached a mapping accuracy of 47% of correctly classified cells (Kappa coefficient: 0.33). It was shown that the attributes "distance to higher elevations", "relative elevation" and "slope" were leading to the highest increase of mapping accuracy. Adding neighbourhood attributes to the best topographical combinations lead to contradictory results between the training areas and the types of evaluation. For two of the training areas, increasing the number of classes lead to an increase of the mapping accuracy. Large differences were observed in the classification of geomorphological units, with percentages of correctly classified cells ranging from 75% (debris slope) to 3% (active badlands). The units were mostly misclassified in relatively near units in terms of topography and geomorphology.

The geomorphology of the training area was shown to impact directly on the correct classification of geomorphological units. The more present a geomorphological unit was in the training area, the better its classification became, irrespective of its actual proportion in the Buëch catchment. The results were also shown to be largely influenced by the correct classification of the unit "debris slope", which covers about 40% of the total area of the Buëch catchment. Future research is required to derive new topographical attributes likely to improve the classification of units other than "debris slope". The possibility to include neighbourhood information should be explored into further extent. The number of classes and the class boundaries could be adapted too for better discrimination possibilities.

Contents

1 Introduction

Modelling discharge and water balance in a large scale catchment is data demanding and involves a high level of computational cost. One of the solutions is to use a relatively large model unit (10^4- 10^6 m^2) to reduce the model runtime. The model units should be distinctive in terms of physiographical characteristics and can be taken to be approximately identical to geomorphological units. The delineation of the geomorphological units can be done through field mapping, however this task is particularly time consuming in large-scale catchments and involves a high degree of subjectivity. An alternative to the drawbacks of field mapping in the context of large-scale catchment modelling is to develop an automated landscape classification.

Several automated methods can be found in the literature and will be briefly presented hereafter. A first approach consists of defining the membership of the cell on the sole basis of the properties of the cells itself, following either classification or cluster techniques. Classification techniques amount to define conditional statements linking the properties of the cell (typically derived from a digital elevation model) to a membership, i.e. a geomorphological unit. Although these techniques are highly flexible (Burrough and McDonnell, 1998), the difficulty lies in the identification of discrimination rules allowing for both effective differentiation and as low data loss as possible. The order in which the conditional statements are applied, as well as the boundaries values, which all need to be defined by the user on the basis of data analysis and/or expert knowledge, can have a large impact on the accuracy of the automatic mapping. Cluster techniques, on the other hand, are almost entirely automated. A well-known example is the k-means clustering method, where the values of the properties derived from the digital elevation model (DEM) are clustered by minimizing the sum of square errors inside each cluster (Burrough and McDonnell, 1998). The cells belonging to a same cluster have hence almost identical properties and can be considered as belonging to the same geomorphological unit. In comparison to classification techniques, clustering methods require no conditional rules to be set. Class boundaries cannot be defined a priori, but emerge from the data set. However, the user needs to determine which properties, allowing for the best discrimination, are considered.

Another strategy consists of approaching automated geomorphological mapping from the perspective of stochastic simulation. This approach is fundamentally different, in that the automated map is considered as a realization drawn from a conditional cumulative probability function. Traditionally, simulation methods have relied on two-points (variogram-based) statistics, of which a typical example is sequential Gaussian simulation. However, two-point geostatistics are often unable to reproduce the mathematical complexity of curvilinear landform patterns, which would require to take into account the correlations between a larger number of spatial locations (Liu, 2006; Caers and Zhang, 2004; Strebelle, 2002). In order to overcome this limitation, multiple-point geostatistical methods have been developed in the last two decades, primarily in the field of geological reservoir simulation (Strebelle, 2002; Caers and Zhang, 2004). The multiple-point geostatistical (MPG) approach makes use of a training image, which corresponds to an area which is already mapped. The training image is scanned; multiple-point statistics are derived and stored in a frequency database. For each non-mapped cell, a realization is then drawn from the frequency database on the basis of conditioning data (properties derived from the DEM and neighbouring cells), and a geomorphological unit is assigned. The use of the training image is a core difference between the MPG approach and the cluster and classification techniques. While cluster and classification techniques determine the membership of the target cell on the basis of its own properties only, the MPG approach infers the conditional relationships between the cell properties and its membership from a pre-existing dataset. The units mapped with the MPG approach are hence pre-defined, since already existing in the training area; with cluster and classification techniques, on the other hand, the units need to be specified on the basis of data analysis. Additionally, the structure of the MPG technique allows for a wider range of conditioning data: beside the properties of the target cell, neighbourhood information can also be included to determine the membership of the cell.

This research is part of a larger project on nested hydrological modelling of large catchments, which uses geomorphological units as input. A research on automated geomorphological classification using k-means clustering and classification techniques was conducted earlier on the same study area (Schuur, 2009). This study aims at developing and applying a multiple-point geostatistical approach for automated geomorphological classification, with using DEM as only input. The main research objective is to assess the capability of the MPG automated classification method to reconstruct the geomorphology of the area.

In order to reach this objective, three further research questions are defined:

- How does the number and the type of DEM-derivatives and neighbourhood attributes impact the quality of the automatic mapping?
- How does the number of classes of topographical attributes impact the quality of the automatic mapping?
- How does the choice of the training area impact on the quality of the automatic mapping?

In Chapter 2, the study area and its geomorphological units are presented. Chapter 3 describes the methodology used in the different stages of the research. The results obtained with the multiple-point geostatistical approach are presented in Chapter 4 and discussed in Chapter 5, in relationship with the research questions defined above. A general conclusion is provided in Chapter 6.

2 Study area

The study area is located in the southern French Prealps (Hautes-Alpes department) and extends approximately between Veynes and Serres. It is part of the Buëch catchment, which is itself a confluent of the Durance river and flows in a north-west to south-east axis through the study area. The elevation ranges between 700m along the tributaries to a maximum of 1700m at the mountain crest; most of the area is however situated under 1300m above sea level.

The study area underwent two orogenies: the Alpine orogeny, which deformed Mesozoic deposits along north-north-east to south-south-west foldings, and the formation of the Pyrenees, which resulted in east to west foldings (Asch et al, 2010). Most important for the current geomorphology of the area is the lithology, which is characterized by a succession of massive limestones and dark marls and presents hence large differences in terms of resistance to erosion (Asch et al, 2010). The climate of the Quartenary, with its alternation of glacial and periglacial periods, is a second important factor in the formation of the current landscape.

Figure 1: Location of the study area, from Asch et al (2010).

2.1 Main geomorphological features

Two main types of geomorphological features are found in the study area: denudational forms, resulting from erosional processes, and accumulation forms.

2.1.1 Denudational forms

The massive limestones from the upper Malm and the lower Cretaceous are very resistant to erosion. They are forming steep ridges of bare rock, of 10 to 60m height, generally extending above debris slopes (Asch et al, 2010). Referred to as *hogbacks*, these cliffs are usually only some meters wide and hence almost absent on two-dimensional geomorphological maps, albeit they are frequently encountered in the study area. Next to massive limestones, the lithology of the area consists of several layers of black marls, which show very low resistance to erosional processes. Water incises marls in narrow gullies, connected in a hierarchical way and separated by higher ridges; these complexes are known as *badlands*. Badlands are considered as active when erosional processes are still ongoing. In this case bare marls are almost impermeable to water, which leads to overland flow (Schuur, 2009). Vegetation can start developing on badlands when the slope of the ridges separating gullies decreases, hence allowing for soil formation. Inactive badlands are typically covered by more than 10% by vegetation, which enables water infiltration (Asch et al, 2010).

A further denudational form characteristic of the region is the *glacis*. Glacis are concave hill slopes resulting from solifluction and overland flow, mostly formed during glacial periods in the absence of vegetation. As such, glacis often end some meters above the bottom of the valley, hence representing the old elevation of the valley. They are covered by a layer of sediment allowing for subsurface flow and ranging from some decimeters to some meters (Asch et al, 2010). Glacis are in principle connected at their highest part to hogbacks; if this connection is not present anymore, they are often referred to as *glacis remnants*.

2.1.2 Accumulation forms

Debris slopes (also known as scree slopes) are steep slopes covered by loose material eroded from the hogbacks. Debris slopes are hence found at a relatively high position in the landscape, under hill crests. The size of the deposited material can highly vary and the slope can be covered by vegetation, depending on the frequency of rock falls (Asch et al, 2010). *Mass movements* (landslides) are the least present geomorphological unit in the region and confined to a limited area in the eastern part of the catchment ; they are recognizable by the scarp under which they develop, as well as their lobate shape. Whereas gravity is the main geomorphological factor in the formation of debris slopes and mass movements, alluvial fans and colluvium are resulting from water-related depositional processes. *Alluvial fans* are characterized as fan-shaped features, generally convex perpendicularly to the steepest gradient. The conic surface of this unit originates from the transport of material from a source point by water in the downstream direction (Asch et al, 2010). In the study area, alluvial fans are divided into two categories on basis of the type of deposited material. Fine alluvial fans consist of marl deposits; their slope angle is generally not exceeding 10 degrees and the infiltration capacity is relatively low (Schuur, 2009). Coarse alluvial fans, which originate from calcareous scree eroded from hogbacks and transported over debris slopes, reach both higher slope angles and infiltration capacities. *Colluvium* are located at the foot of slopes and result from the deposition of mostly calcareous material eroded from upstream accumulation units. Their form is less easily recognizable than that of alluvial fans. They are often covered by a thick layer of soil enabling for a relatively high infiltration capacity (Schuur, 2009).

Two further accumulation forms relative to fluvial processes can be identified. Sediments deposited along the river at its current level are referred to as *river plains*. River plains are flooded regularly on a yearly basis, which implies that sedimentation processes are relatively active (Asch et al, 2010). *River terraces*, on the other hand, are old river plains situated at the ancient, higher level of the river. Their form is distinctive, since consisting of a highly flat surface with steep sides.

2.2 General position of geomorphological units

The geomorphology of the Buëch catchment is typical in that several sequences of units can be identified on a repetitive basis. On a top to bottom axis, hogbacks occupy the highest position in the landscape, on hill crests; under them, debris slopes extend on steep slopes, which cover a large percentage of the study area. Glacis are generally found among debris slope, or just underneath. Both inactive and active badlands, as well as colluvium, extend below debris slopes too. On a bottom-top axis, river plains are found at the lowest elevations inside subcatchments. They are generally surrounded at a slightly higher elevation by river terraces, but also by coarse and fine alluvial fines, which are connected to debris slopes at their upper extremity. These relatively clear sequences of units suggest that the relative position of cells towards both hill crests and rivers, as well as the neighbouring units, might be of importance in order to classify the geomorphology in an automated way.

3 Methodology

3.1 Multiple point geostatistics

The multiple point geostatistical method used in this research was developed by Strebelle (2002) as an extension of the algorithm of Guardiano and Srivastava (1993), and is commonly referred to as single normal equation simulation (snesim). Albeit snesim was first intended for the modelling of geological reservoirs, the non-specific character of its method renders it applicable to any type of landscape heterogeneity.

The snesim approach consists of using one or several training images, from which all multiple-point statistics necessary to the simulation of areas to be mapped are read. The training image can be of synthetic nature or correspond to a part of the landscape which properties are known. The snesim simulation algorithm follows two main steps: the storage of multiple point geostatistics in a dynamical data structure (search tree), and the simulation. In the first stage, the training image is scanned; the number of occurrences of the states of interest, associated to properties discretized into classes, is stored in the search tree (Strebelle, 2002). In the second stage, the cells that are not belonging to the training image are visited in a random order. For each cell, the conditional probability is retrieved from the search tree, on the basis of which a state is assigned to the cell. This method will be described into further details in the next section, in relationship with its actual implementation in this research.

3.1.1 Application of single normal equation simulation (snesim) to geomorphological classification

The main assumption underlying the use of the snesim algorithm for geomorphological classification is that the occurrence of geomorphological units is conditioned by topographical attributes; the nature of this conditional relationship is inferred from the training image. The training image corresponds to a part of the Buëch catchment, which geomorphology is known on the basis of a field map. For the purpose of the research, it is considered that the only data available about the rest of the Buëch catchment is a digital elevation map (DEM).

Six topographical attributes are defined and derived directly from the DEM (section 4.3). Their continuous values are discretized into a given number of classes. The number and type of topographical attributes, as well as the number of classes, is set prior to the application of the snesim algorithm and varies from one experimentation to another according to the optimization design (section 4.7). Beside topographical attributes, attributes relative to the neighbourhood of the cell are also used in the present research in the last step of the optimization design (section 3.4). Those neighbourhood attributes have categorical values (geomorphological units) and hence do not require to be discretized into classes.

The first stage of the snesim algorithm can be referred to as the "training" phase; only the training area is considered. For each cell of the training area, the algorithm scans the class of values of topographical attributes the cell belongs to (derived from the DEM), the class of neighbourhood attributes (if used), as well as the geomorphological unit of the cell (derived from the field map). The classes of attributes assigned to each cell are referred to as patterns. For each pattern (e.g. slope [2]; planform curvature [5]; distance to river [1]), the number of occurrences of geomorphological units are counted. It is expected that a given pattern may be more easily associated to one type of geomorphological unit than to another, which allows for

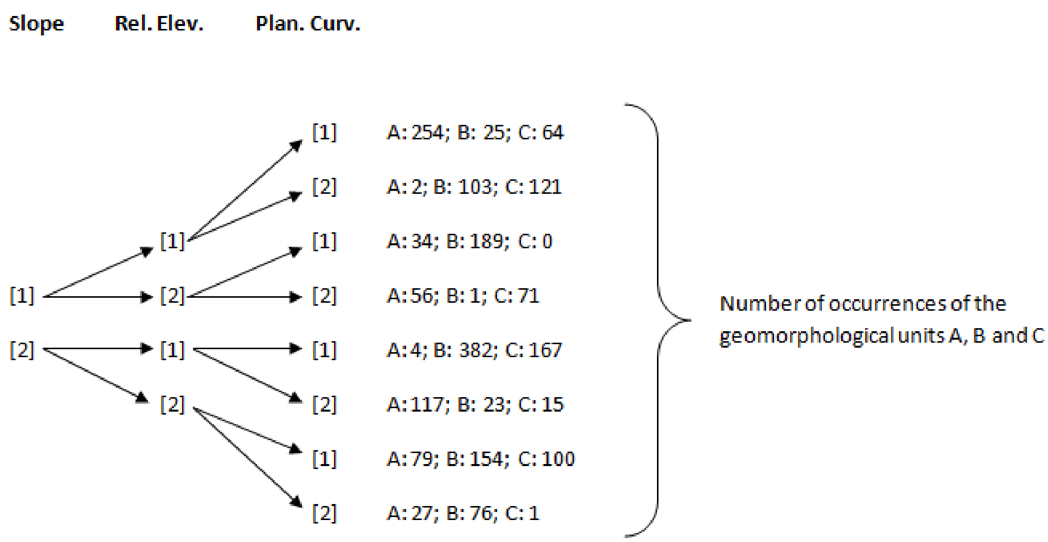

Figure 2: Hypothetical search tree resulting from the use of three attributes (slope, relative elevation and planar curvature) and two classes.

a characterization of geomorphological units by means of topographical attributes. By dividing for each pattern the number of cells belonging to a given geomorphological units by the number of cells showing the same pattern, the conditional probability can be obtained and described by the following equation:

$$Prob\,\{G \mid A\} = \frac{N(G)}{N(A)} \tag{1}$$

Where G is a given geomorphological unit, A a given pattern and N the number of cells. Pursuing on the former example, the hypothetical pattern "slope[2], planform curvature [5], distance to rivers [1]" could correspond in 60% of the cases to alluvial fans, and in 40% of the cases to colluvium. These conditional probabilities are stored in a search tree (Figure 2).

The next stage of the snesim algorithm consists of the automatic mapping. All cells of the Buëch catchment are visited in a random order and the the pattern of each cell is derived from the DEM (topographical attributes) or from neighbouring cells that are already mapped (neighbourhood attributes). At the beginning of the mapping, the number of mapped cells is too low to enable the calculation of neighbourhood attributes. Since neighbourhood attributes are generally added after topographical attributes, they are disregarded when they do not possess mapped neighbours; in these cases, only topographical attributes are used, until the next cell to be visited is surrounded by enough mapped neighbours for neighbourhood attributes to be used. The fact that neighbourhood attributes cannot always be computed is of importance for the order in which they are considered. If a neighbourhood attribute that can rarely be used is placed before another, both will be disregarded if the first one cannot be computed. If the attribute that is more easily computed is placed in front of the other, neighbourhood attributes will altogether be used more frequently. Note that the order of topographical attributes is not relevant, since all topographical attributes can be directly derived from the DEM.

On the basis of the search tree, a realization of the conditional probability associated to the pattern is drawn and geomorphological units get assigned to the cells. The fact that realizations are drawn from conditional probabilities implies that the type of geomorphological units assigned to given cells might vary from one model run to another. It should be noted that the entire catchment is mapped automatically, including the training area.

3.2 Data

The main inputs for the automatic classification of geomorphological units in the Buëch catchment are a digital elevation map and a geomorphological field map. The digital elevation map has a resolution of 37.5m. The geomorphological field map is available from an earlier study about automatic hydromorphological classification based on cluster analysis and rule based methods, carried on at Utrecht University in 2008 (Schuur, 2009). A field trip of four weeks was conducted by two students and resulted in a field map of 1:10'000 scale; data from earlier field works were used to preclassify the area prior to field mapping. The legend of the field map contains 12 geomorphological units: hogback, river terrace, river plain, coarse alluvial fan, fine alluvial fan, colluvium, active badlands, inactive badlands, glacis, glacis remnant, mass movement and debris slope.

The criteria used for the drawing of the geomorphological map are particularly important for two reasons. First, the snesim algorithm derives all conditional relationships between topographical attributes and geomorphological units from the field map; the results are hence fundamentally dependent on the field map. Second, the geomorphological map is used as only reference to assess the accuracy of the automatic mapping. This implies that in order to be accurate, the automatic map should resemble as much as possible to the geomorphological map, irrespective of whether the geomorphological map depicts the reality or not. A geomorphological map being necessarily subjective, it can be expected that differences will be obtained with any geomorphologist mapping the area. The topographical and neighbourhood attributes defined for the application of the snesim algorithm should thus reconstruct the "geomorphological eye" of the geomorphologists.

The check list used for the geomorphological mapping during the 2008 field work is retrieved in table 1 and furnishes an example of the criteria used by the geomorphologists. Long distance observations were carried from about 150 observation points; at one third of those, short distance observations were completed. Given that the DEM is the only input for the automatic mapping, it can be noted that the short distance observations of table 1 cannot be derived; the same holds for the vegetation and the colour of the units.

Long distance observations	Short distance observations
Slope	Kind of material
Aspect	Grain size
Shape of the unit	Sorting
Size of the unit	Color of the material
Colour	Thickness of the unit
Vegetated	Presence of soil layer
Position in the landscape	Thickness of soil layer
Position in the landscape compared to rivers and non permanent streams	Type of vegetation

Table 1: Checklist for field classification used during the geomorphological mapping, from Schuur (2009).

3.3 Topographical attributes

The attributes were chosen in order to correspond to the criteria used by geomorphologists in the field, and to meet the restriction to be derivable from a DEM. Three topographical attributes were retained from the checklist used during the geomorphological mapping of the Buëch catchment (Table 1): slope, position in the landscape (relative elevation), and position in the landscape compared to rivers and non permanent streams (distance to rivers). Three additional topographical criteria were assumed to be relevant during field mapping: planar curvature, profile curvature, and distance to highest elevations. All topographical attributes are presented hereafter. Furthermore, it was considered that the type of geomorphological units found at the proximity of the cell to be mapped plays an important role in the classification and should thus be included in the algorithm. The two neighbourhood attributes "unit of window majority" and "attribute of downstream neighbour" have been defined accordingly (section 3.4).

3.3.1 Slope

The slope attribute is computed by applying the PCRaster function `slope` on the DEM map. The function `slope` takes the elevation of the eight nearest cell neighbours into account (3x3 window) and uses a third-order finite difference method (PCRaster Team, 2011).

3.3.2 Planform curvature

Planform curvature corresponds to the second derivative of elevation per horizontal distance, perpendicularly to the steepest gradient. This attribute is calculated by applying the PCRaster function `plancurv`, which uses the eight nearest neighbours of the cell in a 3x3 window.

3.3.3 Profile curvature

Profile curvature is the second derivative of elevation in direction of the steepest gradient, and is calculated with the PCRaster function `profcurv`. Similarly to the function `plancurv`, `profcurv` is computed on the basis of the eight nearest cell neighbours. Positive planform or profile curvatures indicate a convex slope, while negative values indicate a concave slope.

3.3.4 Distance to river

The distance from each cell to the nearest river is computed following four steps. First, a local drain direction (LDD) map of the catchment is created by determining the flow direction of each cell on the basis of the steepest gradient from the cell to one of its eight contiguous neighbours, by making use of the PCRaster function `lddcreate` applied to the digital elevation model (DEM). Second, the material flowing out of each cell (which corresponds to the sum of the material of both the cell and its upstream cells) is computed by using the function `accuflux`, applied to the LDD map. Third, a Boolean map is created by assigning a "true" value to cells with an accuflux value larger than 1000, in order to identify rivers in the catchment. This threshold was set in order for the Boolean map to resemble the actual hydrological map of the catchment, which was available from earlier field works in the area (Schuur, 2009). Finally, the distance from each cell to the "true" value towards which the path over the local drain direction map is the shortest is computed with the function `ldddist` (Figure 3).

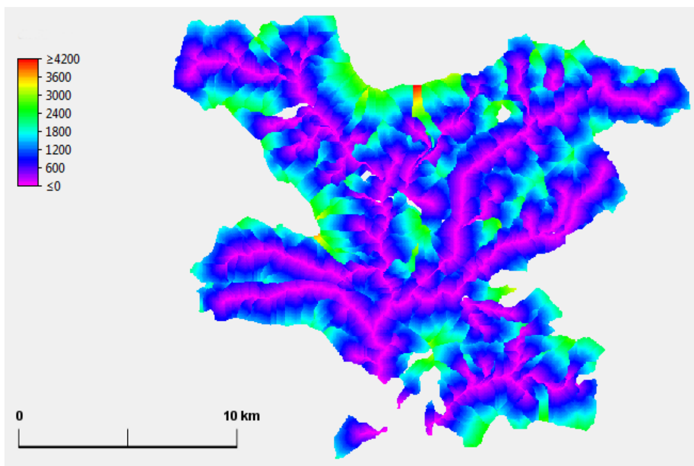

Figure 3: Map of the distance to rivers in the study area. Rivers are assigned a 0 value.

3.3.5 Relative elevation

The relative elevation of each cell corresponds to its normalized elevation (a value between 0 and 1) in regard to the lowest and highest elevation of the Buëch catchment (Figure 4), and is calculated as:

$$e_r = \frac{e - e_{min}}{e_{max} - e_{min}} \tag{2}$$

with e_r, the relative elevation of the cell, e, the elevation of the cell, e_{max}, the maximum elevation in the catchment, and e_{min}, the minimum elevation in the catchment.

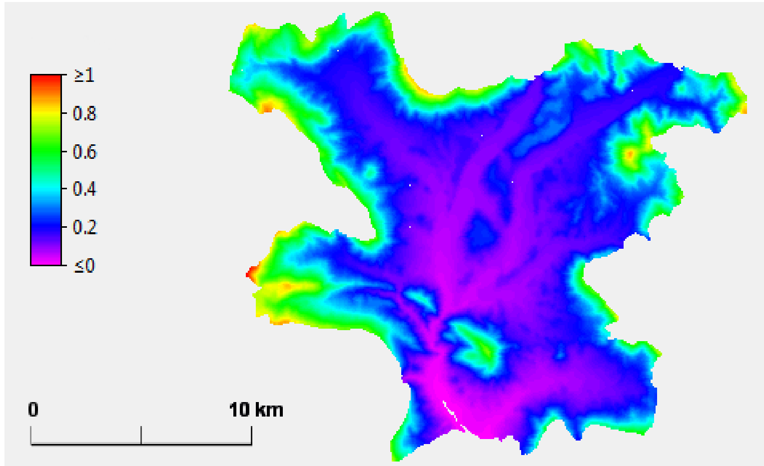

Figure 4: Map of the relative elevation in the study area

3.3.6 Distance to highest elevations

The distance to higher elevations is defined as the distance to the nearest cell with a relative elevation higher than 0.45 (Figure 5). This threshold corresponds approximately to the lowest limit of the unit "debris slope" in the catchment and was set accordingly. This provides a means to calculate the distance from a cell to debris slopes, which were shown in section 2.3 to extend above a relatively fixed sequence of units. A Boolean map is first created by assigning "true" values to each cell located above this threshold. The distance from each cell to "true" cells is then computed by using the `spread` PCRaster function, which calculates the shortest distance available for each cell.

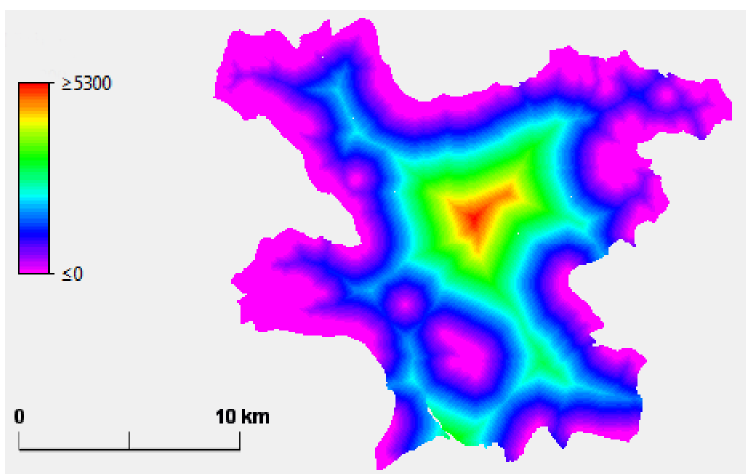

Figure 5: Map of the distance to highest elevations in the study area. Cells with a relative elevation higher than 0.45 are assigned a 0 value.

3.4 Neighbourhood attributes

3.4.1 Unit of downstream neighbour

The "unit of downstream neighbour" attribute returns the geomorphological unit of the downstream neighbour of each cell, as identified by the local drain direction map. This attribute is computed with the PCRaster function `downstream`. If the cell is a pit and does not have any downstream neighbour according to the LDD map, its own value is returned.

3.4.2 Unit of window majority

The "unit of window majority" attribute corresponds to the most occurring geomorphological unit found in a 3x3 window around the cell, and is computed with the PCRaster function `windowmajority.` If no unit is more represented than another in the window, the size of the window is iteratively enlarged by twice its initial length, until a dominating unit is found. This implies that the size of the window might differ between the training and the mapping phase. If some cells of the window have not been mapped yet (and are hence assigned a missing value), only the classified cells will be taken into account for the computation of the most occuring cell value.

3.5 Classes

Six different numbers of classes were chosen: 2, 3, 5, 7, 10 and 12 classes. The class boundaries are set such that each class includes an equal number of cells, in order to have equal probabilities to associate cells to each class. This is attained by dividing for each topographical attribute the total number of cells mapped by this attribute by the number of classes, and retaining as class boundaries the values of the cells which ranking order corresponds to:

$$R = \frac{N_{tot}}{C_{tot}} B \tag{3}$$

with R, the ranking of the cell, N_{tot}, the total number of cells, C_{tot}, the total number of cells, and B, the class boundary number.

3.6 Training areas

Three training areas were defined and will be referred to under the following names: the Northern subcatchment, the Savournon subcatchment, and the dotted training area. Both the Northern subcatchment and Savournon subcatchment (Figure 6) were delimited by selecting manually their outlet and applying the PCRaster function `subcatchment`, which identifies on the basis of the local drain direction map all cells belonging to the subcatchment. In order to define the dotted training area (Figure 7), nine cells were randomly selected over the entire catchment by making use of the PCRaster random generator function `uniform`. Circular patches were then created with the function `spread` (see paragraph 3.2.6), for which the radius was fixed at 980 cells, so that each patch has the same size.

All training areas have almost identical sizes, with the largest training area (Savournon subcatchment) being 1.3% larger than the smallest, dotted training area. Each of them represents approximately 8.3% of the total Buëch catchment.

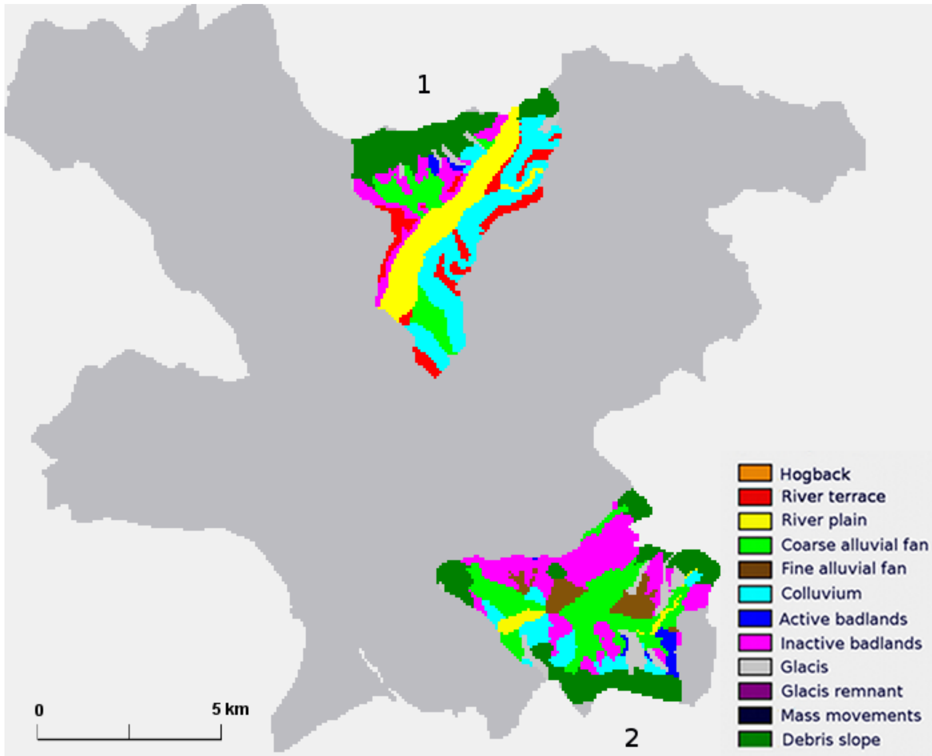

Figure 6: Location and geomorphology of the Northern subcatchment (1) and of the Savournon subcatchment (2)

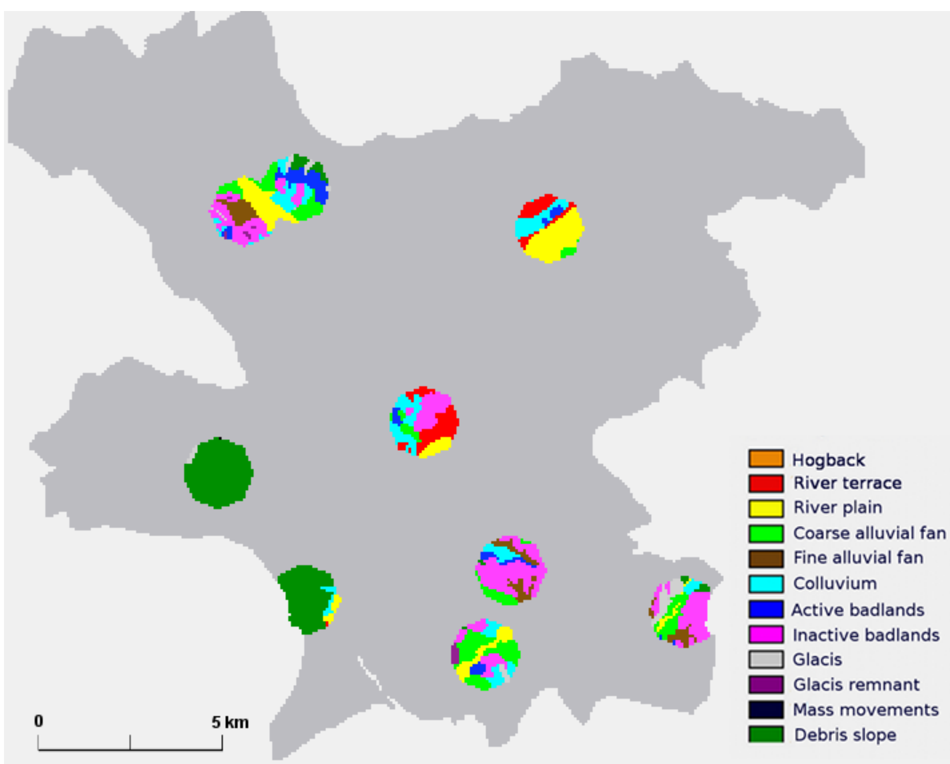

Figure 7: Location and geomorphology of the dotted training area

3.7 Optimization design

An optimization design was conducted in a systematic way for the three training areas (Figure 8) due to the important time limitations associated with model runs[1]. Three main stages can be distinguished: automated classification with single topographical attributes, automated classification with combined topographical attributes, and automated classification with combined topographical and neighbourhood attributes.

In the first stage, the geomorphology of the catchment is mapped automatically by solely one topographical attribute at a time. This is repeated for each number of classes, so as to obtain 36 automated geomorphological maps of the catchment. These maps are then evaluated on the basis of one point-to-point criterion (Kappa statistics) and one lumped criterion (sum of square errors of unit proportions), which allows to rank the topographical attributes in terms of the accuracy of their resulting automated map.

In the second stage, the ranking of topographical attributes is then used to determine five combinations. The first combination consists of the two best performing topographical attributes; the second combination consists of the three best performing ones; and so forth. The fifth combination hence contains all topographical attributes. The geomorphology of the catchment in then mapped with each of the five combinations of topographical attributes. The mapping is repeated for each number of classes; it should be noted that the number of classes is kept identical for each attribute inside of the combination. As a result, 30 geomorphological maps are obtained and evaluated with the same methods used in the first stage.

In the third stage, the combinations are ranked and the best performing combination (with its associated number of classes) is selected. The neighbourhood attributes are then added to obtain four combinations: two combinations with only one neighbourhood attribute added at a time, and two combinations with both neighbourhood attributes, but in a different order. The geomorphology of the catchment is mapped with the combinations of topographic and neighbourhood attributes. This step is executed with three number of classes: the best performing number, one number of classes lower, and one number of classes higher (the available number of classes being 2, 3, 5, 7, 10 and 12)[2]. Twelve maps are hence obtained and evaluated on the basis of Kappa statistics and sume of square errors of unit proportions, similarly to the first and second stage.

Finally, the automated geomorphological maps of topographical combinations with or without neighbourhood attributes are evaluated on the basis of one additional lumped criterion (patch size) and two hydrological criteria (saturated hydraulic conductivity and runoff) which are detailed in section 3.8.

[1]The running time of the automated classification program containing the "unit of window majority" attribute exceeds by about 35 times the one of non-neighbourhood combinations.

[2]If for instance the best number of classes is 5, the two other numbers of classes to be selected would be 3 and 7

First stage	Second stage	Third stage
Distance to higher elevations (H)		
Relative elevation (R)	HR	
Slope (S)	HRS	HRS + Unit of downstream neighbour
Distance to River (D)	HRSD	HRS + Unit of window majority
Planar curvature (Pl)	HRSDPl	HRS + Unit of downstream neighbour + Unit of window majority
Profile curvature (Pr)	HRSDPlPr	HRS + Unit of window majority + Unit of downstream neighbour

Table 2: Example of optimization design for the Northern subcatchment. The ranking of the topographical attributes of the first stage is used to determine the five combinations of the second stage. The first combination consists of the two best performing attributes, the second of the three best performing attributes, and so forth. The best performing combination of the second stage (HRS) is then selected for the third stage, during which two neighbourhood attributes are added.

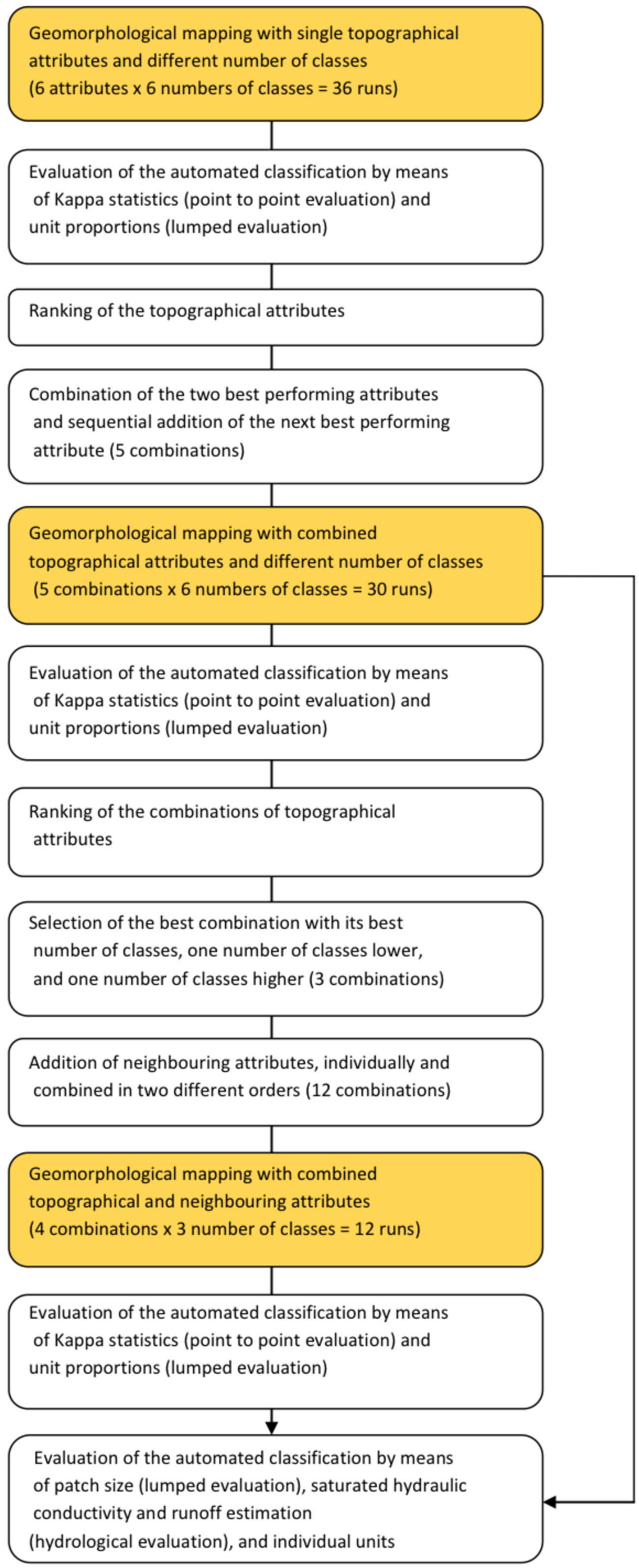

Figure 8: Optimization design followed in the present research

3.8 Evaluation criteria

One point-to-point criterion and two lumped criteria are chosen to assess the accuracy of the automated maps. Since the purpose of the automated classification technique is the application for large-scale hydrological modelling, the maps are additionnaly evaluated on the basis of two hydrological criteria. Finally, the percentage of correctly classified cells ss also calculated to obtain further insights into the mapping accuracy of individual units.

3.8.1 Point-to-point evaluation

The automated geomorphological maps are evaluated on a point-to-point basis by means of Kappa statistics. The Kappa coefficient is a measure of cell-by-cell agreement between two categorical maps, which is corrected for the expected probability of random agreement (Hagen, 2002). The geomorphological category of each cell on the automated map and on the field map are compared, and the fraction of correctly classified cells *P(a)* is calculated. To calculate the expected probability of random agreement *P(e)*, the probability of a cell of the automated map to be categorized as belonging to a certain geomorphological unit is multiplied by the probability of a cell of the field map to be categorized as belonging to the same unit; this operation is repeated for each category and summed. The Kappa coefficient κ can then be computed according to the following equation:

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)} \tag{4}$$

The Kappa coefficient ranges from -1 to 1 (Viera and Garrett, 2005). A value of 1 represents perfect agreement, 0 corresponds to an agreement that would only be obtained by chance, and negative values represent disagreement.

Kappa statistics are computed with the Map Comparison Kit, a software tool for map comparison of the Netherlands Environmental Assessment Agency (MNP) and developed by RIKS BV.

3.8.2 Lumped evaluation

Sum of square errors of unit proportions The sum of square errors of unit proportions is defined by the following equation:

$$SSE = \sum_{i=1}^{n} (\hat{P} - P_i)^2 \tag{5}$$

where $\hat{P}$ is the proportion of unit i in the automated map, P is the proportion of unit i in the geomorphological map, and n is the number of units.

Patch size The patch size criterion is computed with the Map Comparison Kit (see 3.7.1), which makes use of a "moving windows based structure" method. Patches - contiguous cells of identical category - and their size are first derived from the automated map and the field map. The patch sizes are then assigned to each cell on both the automated map and the field map. In the next step, moving averages are computed by calculating the weighted average of the neighbourhood of each cell (Hagen-Zanker, 2006). The difference between the moving average maps yields the global patch size difference.

3.8.3 Hydrological evaluation

Saturated hydraulic conductivity Saturated hydraulic conductivity values (Appendix A) are assigned to each geomorphological unit on the basis of their soil properties and the corresponding estimates of Rawls and Brakensiek (1983). The sum of square errors of hydraulic conductivity can then be computed by comparing the values of saturated hydraulic conductivity on a cell-to-cell basis on both the automated map and the field map.

Runoff estimates The runoff evaluation makes use of the saturated hydraulic conductivity maps to construct a static hydrological model, representing one rainstorm event with Hortonian runoff. Runoff is only considered to occur when the sum of the rainfall intensity and the inflow from the upstream neighbours of the cell are exceeding the saturated hydraulic conductivity. Three rainfall intensities are considered: 7mm, 10mm and 15mm. Runoff is computed with the PCRaster function `accuthresholdflux`, which return the amounts of material transported out of the cell.

Four locations were chosen in the catchment (Figure 9). Location 1, 2 and 3 drain from subcatchments. Location 4 corresponds to the outlet of the Buëch catchment; however, due to interruptions in the local drain direction map, it does not drain from the entire catchment. The subcatchment of location 2 is included in its drainage area, unlike the subcatchments of location 1 and 3.

For each automated map, runoff is computed at each of the four locations with three different intensities. The same operation is conducted once with the field map. The absolute difference between runoff generated by the automated map and runoff generated by the field map is then used as evaluation criteria.

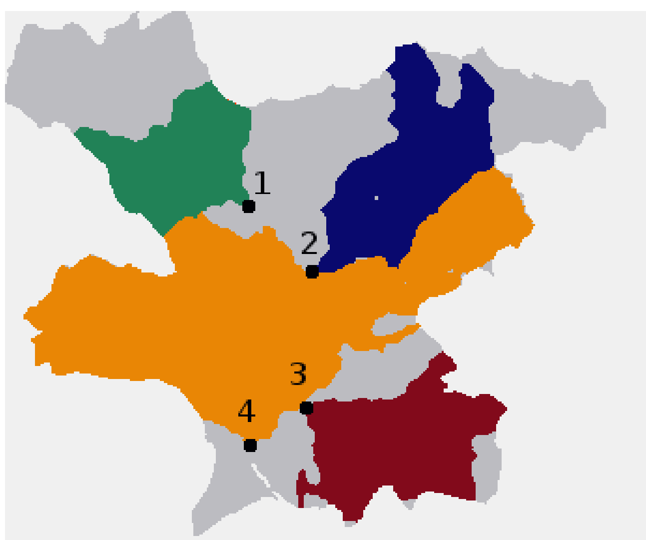

Figure 9: Location of the four outlets used for the runoff estimation and of their drainage areas. The drainage area of outlet 4 (orange) includes the drainage area of outlet 2 (blue).

3.8.4 Percentage of correctly classified cells

The percentage of correctly classified cells C was computed in PCRaster by comparing the automated map and the geomorphological map with a Boolean operation. For each unit i on the geomorphological map, the following equation was applied:

$$C = \frac{N_{i_{true}}}{N_i} 100 \tag{6}$$

where $N_{i_{true}}$is the number of cells of unit i on the automated map which locations correspond to cells of unit i on the geomorphological field map, and $N_i$is the total number of cells of unit i on the geomorphological field map.

4 Results

The best resulting maps are presented hereafter, along with some general remarks about the differences in terms of the geomorphology of the training areas. The detailed results are then presented in the order of the optimization design. When the evaluations of the Northern subcatchment and of the Savournon subcatchment are similar, these are treated together.

4.1 Best resulting maps

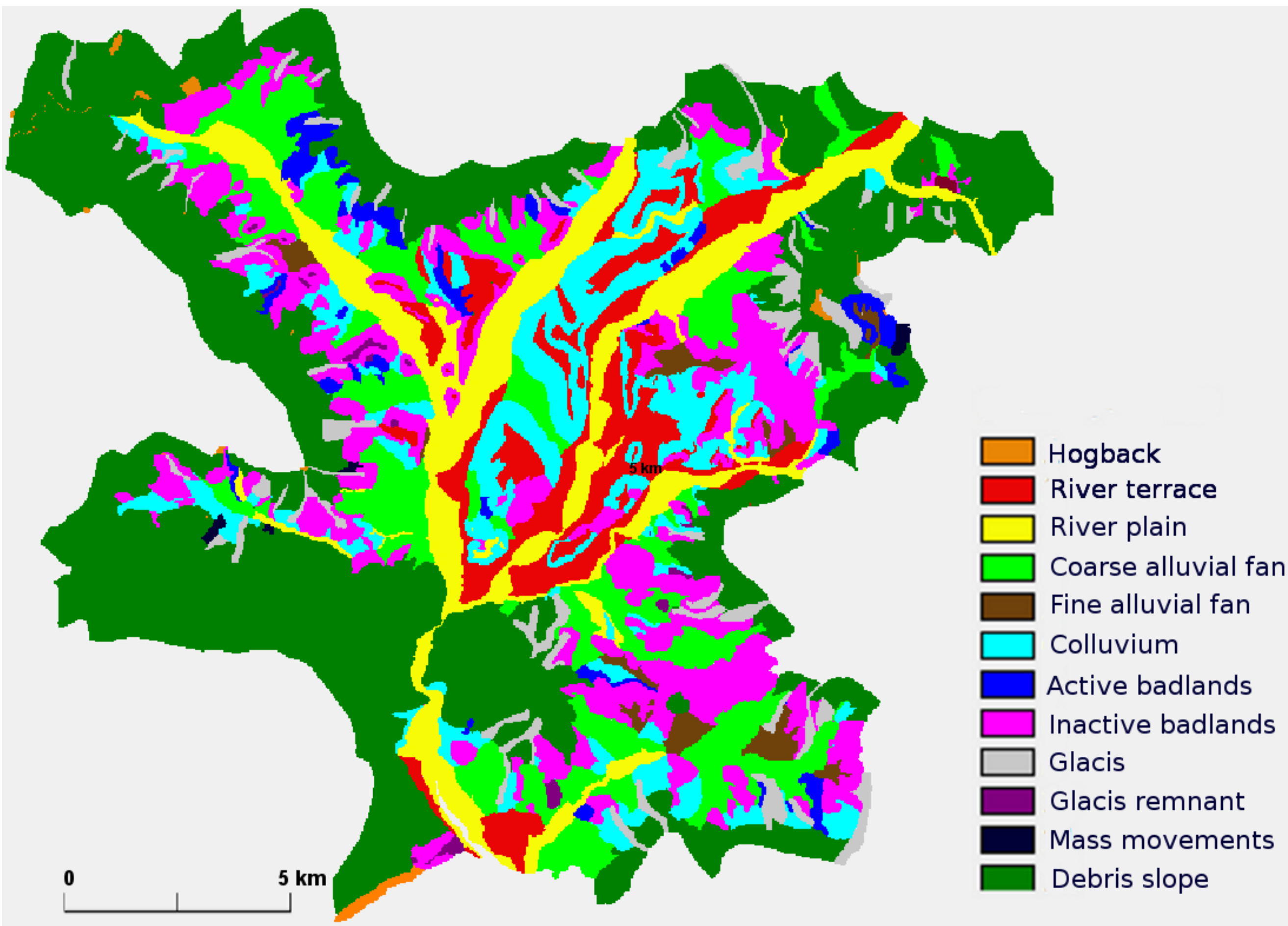

Figure 10: Geomorphological field map of the study area

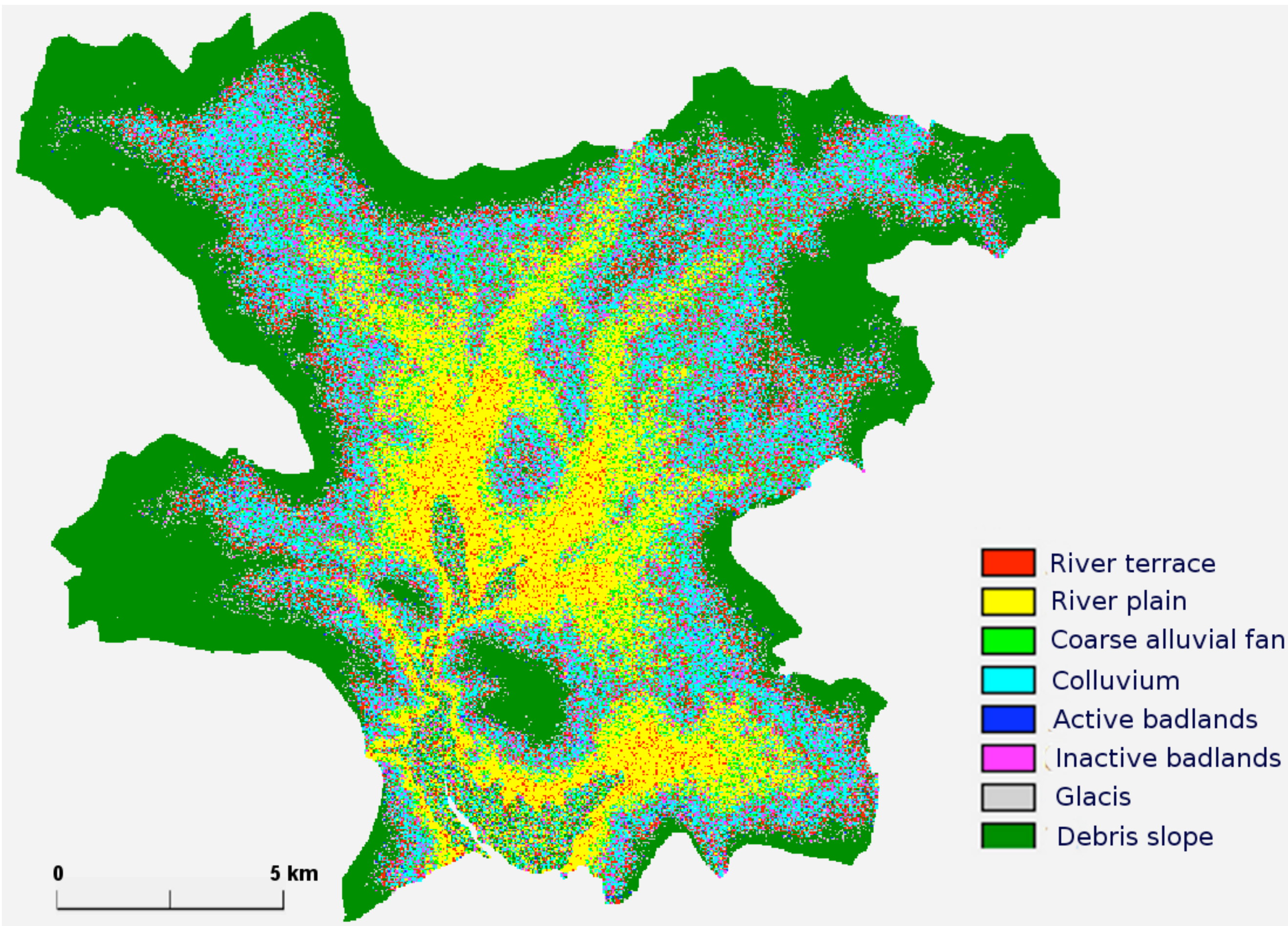

Figure 11: Automated map obtained when using the attribute "relative elevation" at 12 classes, with the Northern subcatchment as training area. On a point-to-point basis, this map lead to the most accurate results among individual topographical attributes. Note that the number of mapped units is inferior to the actual number of units of the geomorphological field map (Figure 10). Since the multiple-point geostatistical approach derives all conditional relationships from the training area, only the units present in the training area (in this case the Northern subcatchment) are used for the classification.

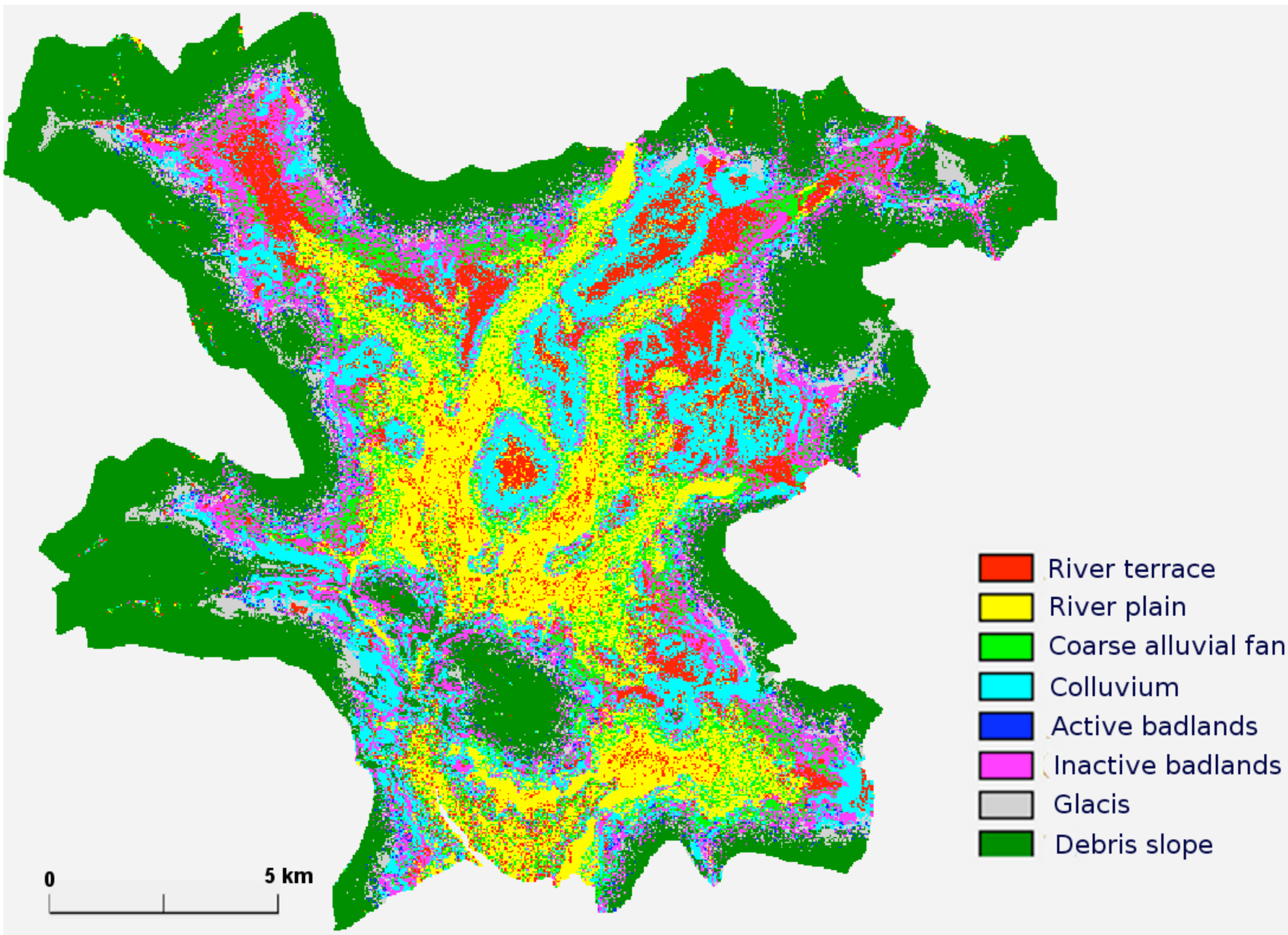

Figure 12: Automated map obtained when using the combination of attributes HRS (distance to highest elevations, relative elevation, slope) at 12 classes, with the Northern subcatchment as training area. Among combinations of topographical attributes, this map was the most accurate in terms of point-to-point accuracy.

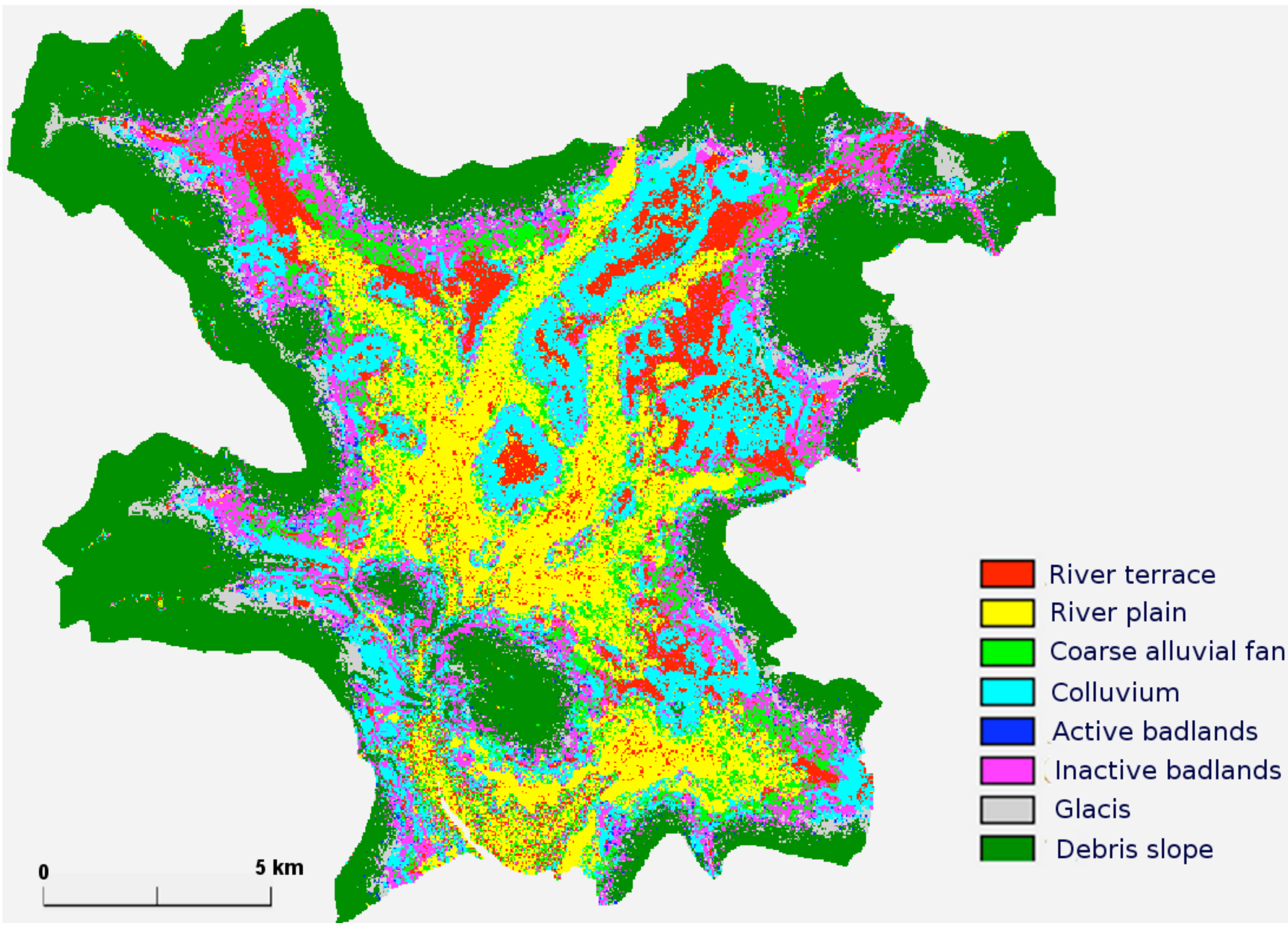

Figure 13: Automated map obtained when using the combination of topographical and neighbourhood attributes HRSW (distance to highest elevations, relative elevation, slope, unit of window majority) at 12 classes, with the Northern subcatchment as training area. Of all automated maps, this map achieved the highest point-to-point accuracy.

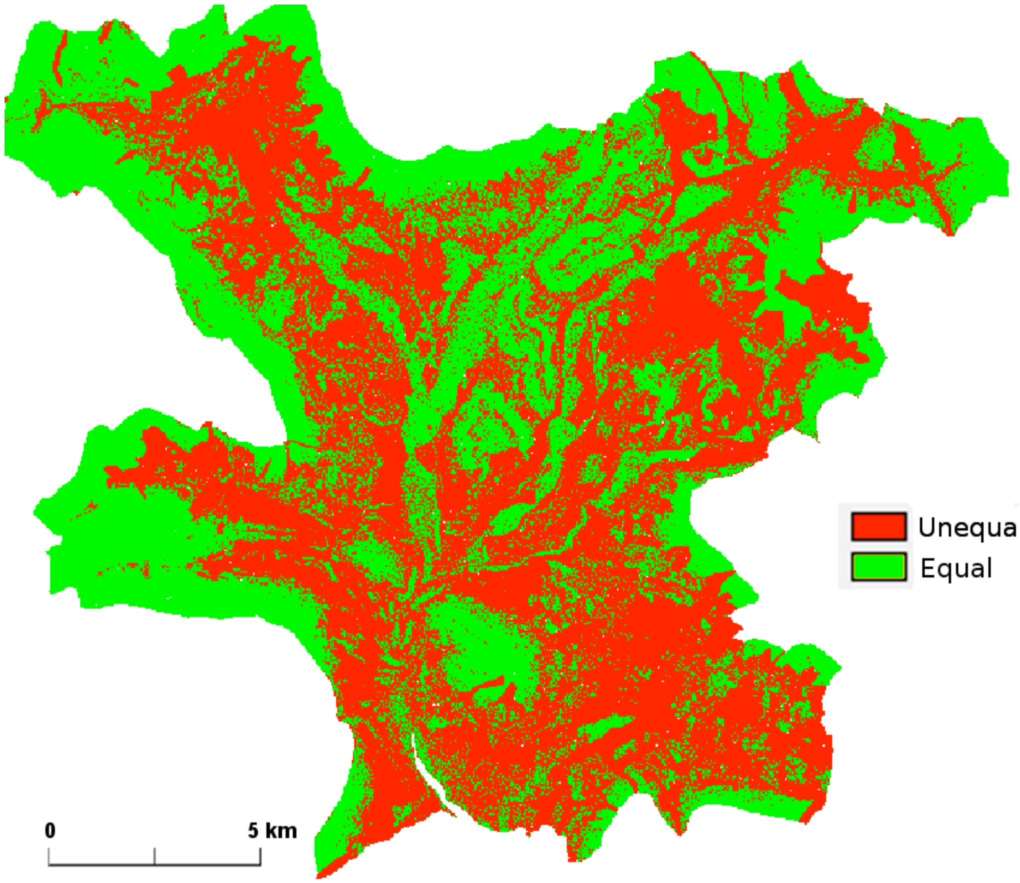

Figure 14: Boolean map representing correctly and uncorrectly mapped cells for the most accurate map obtained in this research (HRSW at 12 classes with the Northern subcatchment as training area; see figure 15).

4.2 Training areas

The training areas are different regarding geomorphology. River plains and colluvium are largely over-represented in the Northern subcatchment in comparison to the proportions found in the entire Buëch catchment (Figure 15). In the Savournon subcatchment, on the other hand, alluvial fans and inactive badlands are over-represented while the subcatchment is entirely deprived of river terraces, and shows a very limited number of river plains. The dotted training area is less different from the Buëch catchment on average, except for inactive badlands, which are over-represented too. Finally, all training areas share one characteristic: they do all include significantly less (between 17 and 19 percentage points) debris slope cells than is to be found in the entire catchment.

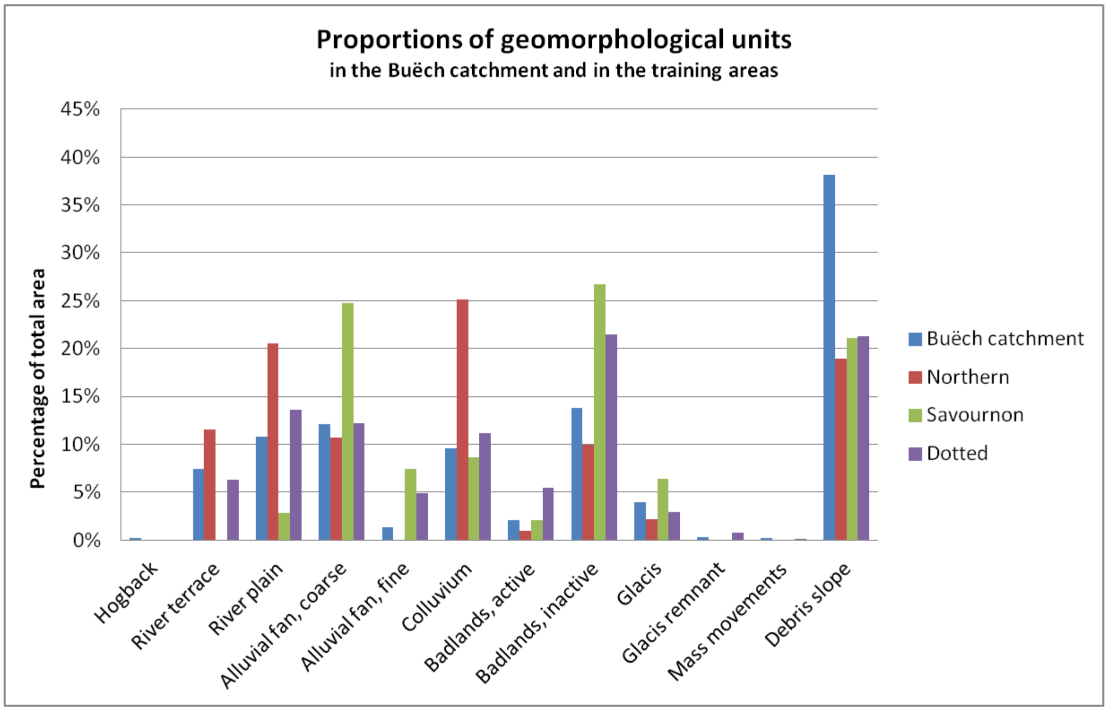

Figure 15: Proportions of geomorphological units in the Buëch catchment and in the training areas.

4.3 Individual topographical attributes

4.3.1 Northern Subcatchment and Savournon Subcatchment

The evaluation of individual topographical attributes yields similar results for the Northern subcatchment and the Savournon subcatchment. In both cases the type of evaluation (point-to-point based or lumped) has no effect on the ranking of the attributes.[3]

In terms of Kappa statistics, two groups of attributes showing similar performances can be distinguished (Figure 16). The first group, which performs significantly better, consists of the attributes "distance to higher elevations", "relative elevation" and "slope". The second group consists of "distance to river", "planar curvature" and "profile curvature". For all attributes and the two subcatchments, the Kappa accuracy appears to increase as the number of classes rises. When evaluating the attributes with the sum of square errors (SSE) of unit proportions, however, the distinction between the two groups is less clear, and the performance stays relatively stable over the different number of classes.

[3]The attribute "distance to the rivers", which ranks in both cases higher with the Kappa evaluation than it does with the SSE of unit proportions, is the only exception.

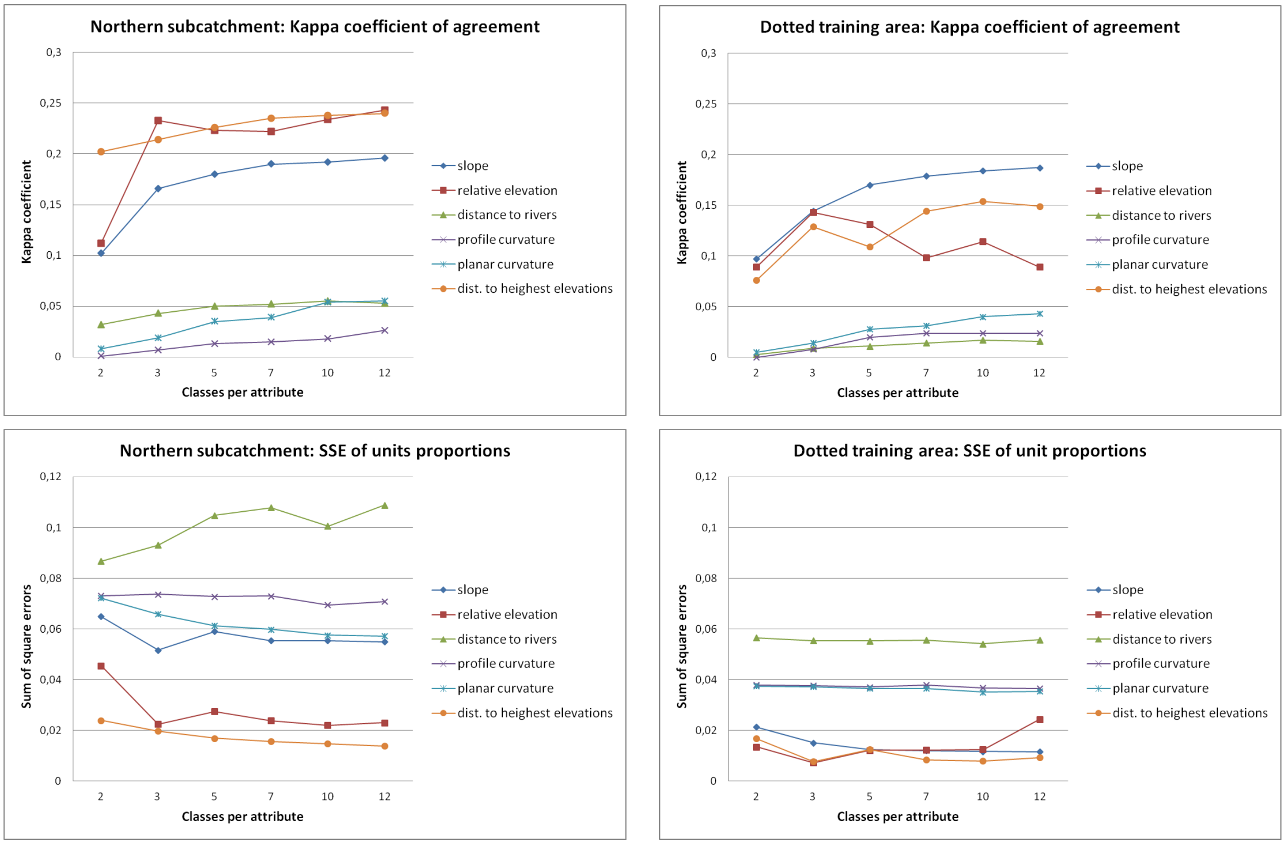

Figure 16: Evaluation of the automatic classification with individual topographical attributes, in terms of Kappa accuracy and sum of square errors of unit proportions. The results obtained with the Savournon subcatchment are comparable to those of the Northern subcatchment.

4.3.2 Dotted training area

When using the dotted training area, the attributes "relative elevation" and "distance to rivers" rank lower in terms of Kappa statistics than with the Northern and Savournon subcatchments (Figure 16). In all other cases, the general pattern resulting from the evaluations is in many points identical to the one obtained with the two subcatchments. First, the type of evaluation appears not to impact the ranking of the attributes, with the exception of the attribute "slope". Second, the same two groups of attributes can be observed; this pattern is even clearer with the SSE of unit proportions than it is with the two subcatchments. Third, the second group of attributes responds to an increase of the number of classes in a way similar to what can be observed with the two subcatchments.

The main difference between the results obtained with the dotted training area and those obtained with the two subcatchments lies however in the large variability of the evaluation of the attributes "distance to higher elevations" and "relative elevation" when modifying the number of classes. Both attributes show a peak of accuracy at 3 classes; above it, the performance of "relative elevation" is decreasing, while the performance of "distance to higher elevations" is increasing again above 5 classes.

Of all training areas, the highest Kappa values are reached with the Northern subcatchment, which shows the largest standard deviation among attributes. The dotted training area, with which the standard deviation is the lowest, is leading to the best results in terms of unit proportions.

4.4 Combined topographical attributes: point-to-point and lumped evaluation

To simplify the reading of the next sections, the following abbreviations are used in the text.

H	Distance to higher elevations	Pl	Planar curvature
R	Relative elevation	Pr	Profile curvature
S	Slope	A	Unit of downstream neighbour
D	Distance to Rivers	W	Unit of window majority

4.4.1 Selection of the combinations of topographical attributes

The five combinations of topographical attributes were selected on the basis of the Kappa accuracy and the sum of square errors of unit proportions of the automated map obtained with individual topographical attributes (section 4.2). In the cases where the two types of evaluation lead to different rankings of attributes (a situation observed with the attribute "distance to rivers"), the ranking of the Kappa accuracy was retained.

4.4.2 Northern Subcatchment and Savournon Subcatchment

Kappa statistics With both the Northern subcatchment and the Savournon subcatchment, the HRS combination is performing best and its accuracy follows a logarithmic curve, increasing as the number of classes rises (Figure 17). The difference in accuracy between the HRS combination and the other combinations is however smaller in the case of the Savournon subcatchment.

With the Northern subcatchment, the combinations of attributes other than HRS and HR show a peak of accuracy at 3 classes; their performance is then steadily decreasing as the number of classes increases. This pattern is not found when using the Savournon subcatchment, where all combinations follow a trend similar to the HRS combination.

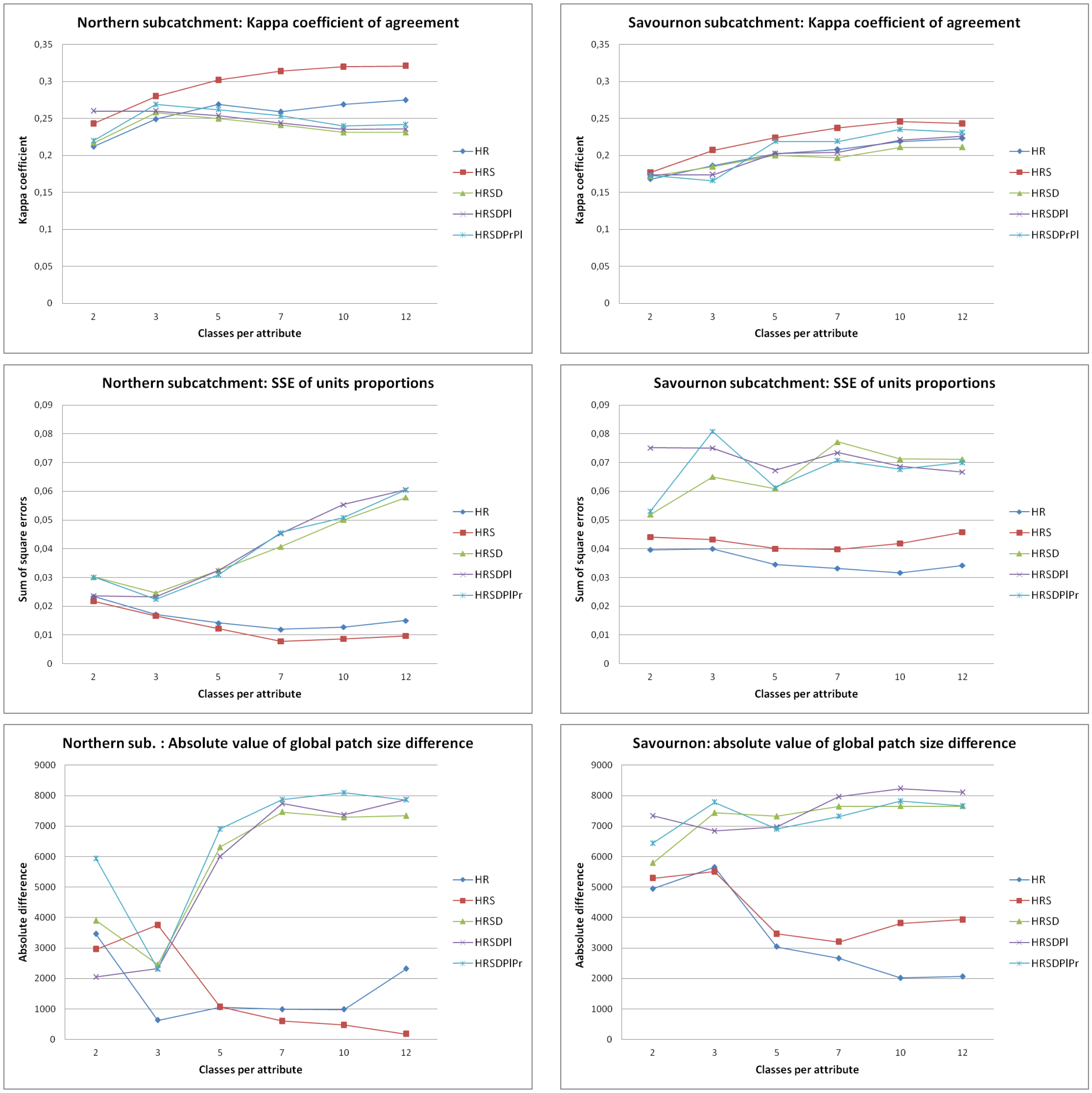

Figure 17: Evaluation of the automatic classification with combined topographical attributes, with the Northern subcatchment and the Savournon subcatchment.

Unit proportions The evaluation of all combinations in terms of unit proportions reveals two groups of combinations, present when using any of the two subcatchments as training area. The first group is made of HR and HRS and shows a clearly higher accuracy. The second group is made of all other combinations, i.e. HRSD, HRSDPl and HRSDPlPr. It is worth noticing that the first and best performing group consists of combinations of the very same attributes that were already outclassing others when used individually.

Albeit both subcatchment share these characteristics, the responses of the combinations to an increase of the number of classes are very different. In the case of the Northern subcatchment and identically to the Kappa evaluation, the second group of combinations reaches a peak of accuracy at 3 classes, after which its accuracy decreases linearly as the number of classes rises. The accuracy of the first group is improving with the number of classes, until reaching a maximum at 7 classes. In the case of the Savournon subcatchment, however, the values of SSE of unit proportions stay relatively stable, and unlike observed by means of Kappa statistics, the HR combination performs better than HRS.

Patch size The global patch size of the geomorphological field map is 10288 cells. This implies that a patch size difference of 8000 cells represents a difference of 78% towards the actual patch size found on the field map. The lowest difference obtained with the Northern subcatchment corresponds to a difference of approximately 2%. The global patch size difference follows a pattern rather similar to the SSE of unit proportions for both subcatchments, with the same two distinct groups of combinations. The minima and maxima observed with the evaluation of unit proportions are also found in terms of patch size. The only difference lies in the behaviour of HR and HRS at 3 classes with the Northern subcatchment. The accuracy peak at 3 classes, which was observed on the second group of combinations with both the Kappa evaluation and the SSE of unit proportions, is now found for HR too; at the opposite, HRS reaches its highest error at 3 classes.

4.4.3 Dotted training area

Kappa statistics Combining attributes with the dotted training area leads to results significantly different from those observed with the two subcatchments, albeit the evaluation was shown to be relatively similar in terms of individual attributes (section 4.1.2). The performance of all combinations follows a positive trend and peaks at 3 classes (Figure 18). However, whereas the combinations of two and three attributes are performing best with the two subcatchments, the ranking of combinations with the dotted training area follows the number of attributes in decreasing order: the more attributes, the better the combination performs. The combination containing all attributes (HSRPlPrD) is therefore the most accurate, while the combination with solely two attributes (HS) is the least performing one.

Unit proportions The SSE of unit proportions of the combinations follows the same structure as the attribute "distance to higher elevations" taken individually: two minima (i.e. accuracy peaks) are reached at 3 and 7 classes, while the error is the highest at 5 classes. Above 7 classes, the error increases again. The ranking of the combinations is generally alike that of the Kappa evaluation, apart from the HS combination, which ranks steadily first above 3 classes.

Patch size General trends are difficult to identify in terms of patch size with the dotted training area. However, some characteristics common to the two above mentioned evaluations can be observed. First, HSRPlPrD shows a large peak of performance at 3 classes, which is found on a lesser extent with HS and HSRPl too. The low performance at 5 classes, which was also found with the SSE of unit proportions, can be identified on all combinations but one.

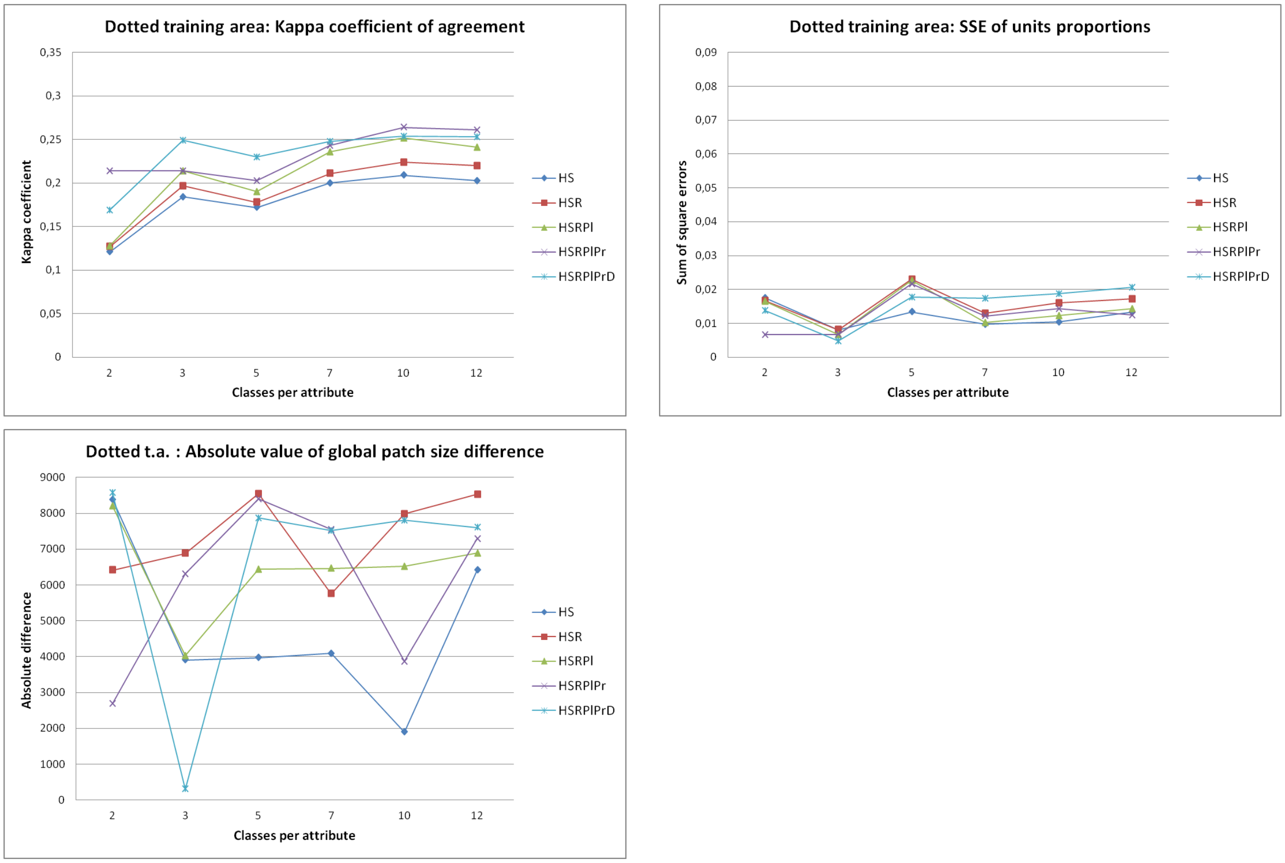

Figure 18: Evaluation of the automatic classification with combined topographical attributes, with the dotted training area.

4.4.4 Summary

With the Northern subcatchment and the Savournon subcatchment, the Kappa accuracy of the topographical combinations is increasing as the number of classes rises. The highest Kappa values are obtained with HRS. With the dotted training area, the combination containing all attributes (HSRPlPrD) is performing best and an accuracy peak is observed at 3 classes. In terms of unit proportions, two groups of combinations with similar accuracy can be observed with the Northern and the Savournon subcatchment; the first group, which consists of HR and HRS, is made of the attributes which were already performing best individually. With the dotted training area, the evaluation of unit proportions follows the same structure as the attribute H taken individually. With all training areas, the evaluation of patch size difference shares similarities with the evaluation of unit proportions.

The highest Kappa accuracy is obtained with the Northern subcatchment, similarly to the evaluation of individual attributes. The SSE of unit proportions is 2 to 3 times lower with the dotted training area than with the two subcatchments.

4.5 Combined topographical attributes: hydrological evaluation

4.5.1 Saturated Hydraulic Conductivity

For all training areas, the evaluation of combinations in terms of SSE of saturated hydraulic conductivity and SSE of unit proportions give identical rankings, peaks and responses to an increase of the number of classes. The only noticeable difference is found with the dotted training area, where the combinations yield less accurate results with 2 classes than with 5 classes in terms of SSE of saturated hydraulic conductivity; this situation was inverted with the evaluation of the SSE of unit proportions.

4.5.2 Runoff

Northern Subcatchment From the three training areas, the Northern subcatchment leads to the most heterogeneous runoff evaluation: the ranking of combinations is varying significantly depending on the outlet and the rainfall intensity (Appendix B). Although general conclusions are difficult to draw, three general patterns can be identified.

- In the first pattern, (outlet Nr. 1, 7mm and 10mm), the HRS combination performs clearly worse than the other combinations and reaches its highest error at 7 classes.
- In the second pattern (outlet Nr. 2, all rainfall intensities), the accuracy of all combinations increases with the number of classes. The HRSD combination yields the best results, while the worst results are obtained with the HRS combination.
- In the third pattern (outlet Nr. 1, 15mm, and outlet Nr. 4, 10mm and 15mm), the accuracy of HR and HRS is significantly higher than with other combinations, and improves as the number of classes increases. A maximum is reached at 7 classes. In an opposite way, the accuracy of all other combinations is decreasing as the number of classes increases.

The third pattern resembles in a striking way the results obtained when evaluating the combinations in terms of the SSE of unit proportions. It could hence be inferred that when the drainage area and/or the rainfall intensities are large, the proportions of units is decisive for the hydrological response. On smaller drainage areas such as those belonging to the outlets Nr. 1, 2 and 3 with rainfall intensities below 15mm, the influence of unit proportions is difficult to identify: combinations which are yielding particularly accurate results in terms of SSE of unit proportions show the highest square error of runoff.

Savournon Subcatchment The patterns of square error of runoff are more constant with the Savournon subcatchment than with the Northern subcatchment (Appendix B). Although some differences are observed between outlets, the response stays approximately the same for each given outlet when modifying the rainfall intensity. Two general observations can be made. First, the error is decreasing when the number of classes increases. Second, HR and HRS often show the highest errors of all combinations until reaching 10 or 12 classes, exception made of outlet Nr. 3, where they steadily rank last.

Just as with the Northern subcatchment, the results of outlet Nr. 4 with 15mm resemble strongly the results obtained with the evaluation of the SSE of unit proportions (Figure 19). The same conclusion appears thus to hold for the Savournon subcatchment.

Dotted training area The dotted training area shows the most constant results of the three training areas in terms of runoff evaluation (Appendix B). The ranking of combinations is generally following that of the Kappa evaluation, SSE of unit proportions and SSE of saturated hydraulic conductivity. The HSRPlPrD combination tends hence to be the best performing combination up to 7 classes. Three exceptions are found at outlet Nr. 1 (15mm) and outlet Nr. 3 (10mm and 15mm), where the square error of HSRPlPrD is increasing from 3 classes on and clearly exceeding that of the other combinations.

Whereas the runoff evaluation at outlet Nr. 4 with 15mm showed results very similar to those observed with the SSE of unit proportions for both subcatchments, it is remarkably different with the dotted training area. A large error peak can be identified at 3 classes, which is for all other types of evaluation the class number leading to the best results. Additionally, the largest error of all combinations is obtained with HSRPlPrD for that class number only. In other words, the results of the square error of runoff and the results of the SSE of unit proportions are exactly symmetrical at 3 classes.

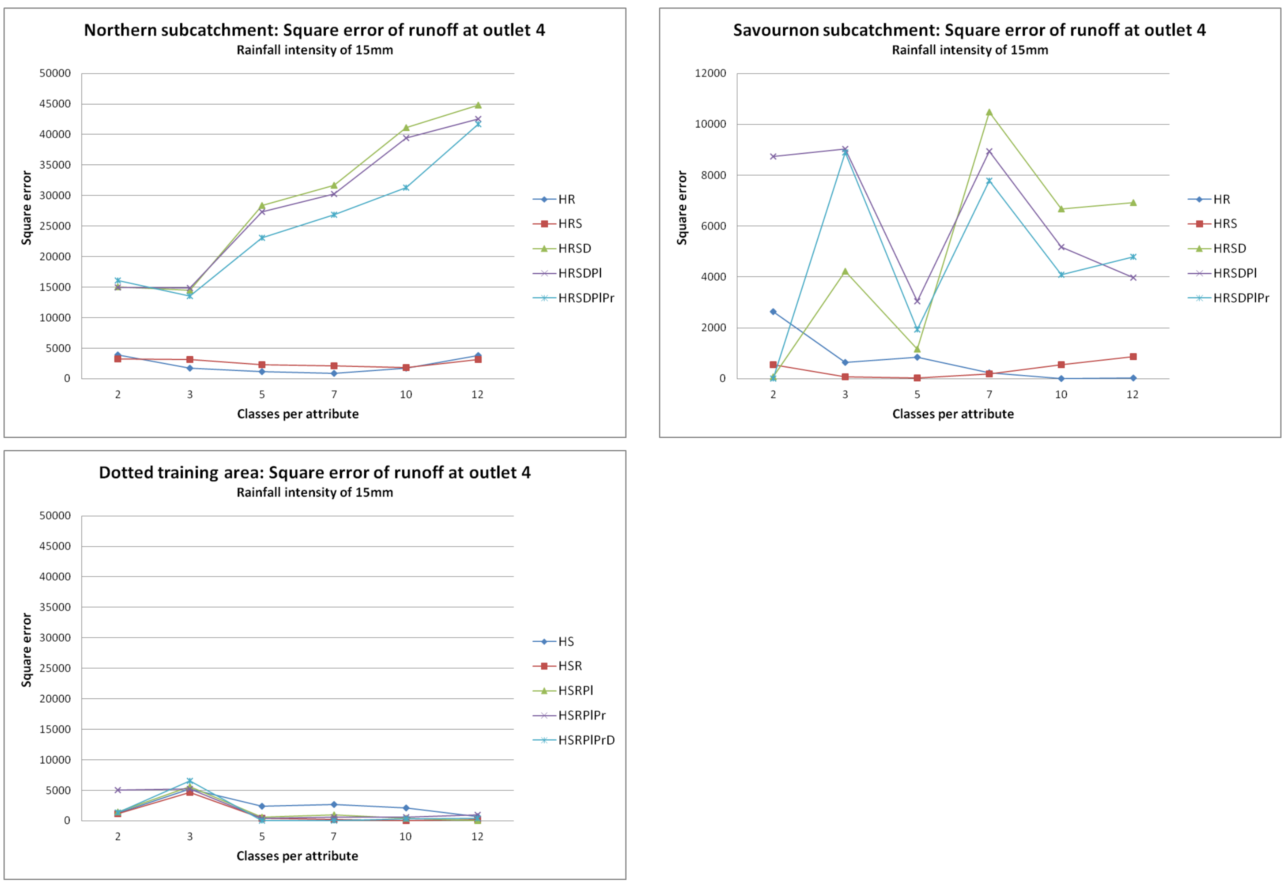

Figure 19: Square error of runoff estimation, as obtained with combinations of topographical attributes at outlet 4, with a rainfall intensity of 15mm.

4.5.3 Summary

With all training areas, the evaluation in terms of saturated hydraulic conductivity is similar to the evaluation of unit proportions. The evaluation of runoff estimation, on the other hand, gives highly variable results, both regarding the ranking of combination and the magnitude of the square error. No conclusions could be drawn from the runoff estimations obtained with the Northern subcatchment. With the Savournon subcatchment, the ranking of the combinations stays identical per outlet, irrespective of the rainfall intensity. The results are more constant with the dotted training area, where the ranking of the combinations is in most cases similar to all other types of evaluations.

It can be observed with the two subcatchments that for outlet 4, which drains from the largest area, the results resemble the evaluation of unit proportions when the highest rainfall intensity (12mm) is used. With the dotted training area, however, the results are then opposite to all other types of evaluations at 3 classes. This anomaly could not be explained.

4.6 Impact of neighbourhood attributes

4.6.1 Point-to-point and lumped evaluation

Adding neighbourhood attributes does not improve nor worsen univocally the performance of the best topographical combinations. With the Northern subcatchment and the Savournon subcatchment, neighbourhood attributes generally improve point-to-point accuracy (Kappa coefficient). However, this accuracy is worsened by neighbourhood attributes when using the dotted training area, except at 3 classes. In terms of SSE of unit proportions, neighbourhood attributes worsen the results with both the Northern subcatchment and the Savournon subcatchment, while they improve the results with the dotted training area. Finally, the accuracy in terms of patch size is improved by neighbourhood attributes with the dotted training area; general trends are difficult to identify with the Northern subcatchment and the Savournon subcatchment.

The ranking of combinations with neighbourhood attributes is strongly dependent on the type of evaluation, the training area and, to a lesser extent, on the number of classes. The performance of the neighbourhood attributes is presented in further detail in the following paragraphs.

Kappa statistics With all training areas and irrespective of the number of classes, adding the neighbourhood attribute "attribute of window majority" (W) with or without "attribute of downstream neighbour" (A) leads to the best results in terms of Kappa statistics among neighbourhood combinations (Figure 20). In this type of evaluation and for all training areas, W and WA show almost identical results, and A does not appear to contribute to a further improvement of Kappa accuracy. This fact is supported by the observation that adding solely A to the best topographical combination with the Northern and Savournon subcatchments does not lead to any clearly noticeable change; adding AW, however, does increase the accuracy, albeit to a lesser extent than W and WA. It could hence be concluded that the attribute of window majority W is mainly responsible for increasing Kappa accuracy. Its role is lessened when it is added after the attribute of downstream neighbour.

With the dotted training area at 2 and 3 classes, where the neighbourhood attributes are worsening the results, these conclusions are valid too in that W is the attribute decreasing the accuracy the less; adding A to it does not contribute significantly to improve the accuracy obtained when adding W only.

SSE of unit proportions With the Northern and the Savournon subcatchments, the results of the evaluation by means of the SSE of unit proportions are opposite to the Kappa evaluation. Where neighbourhood attributes contribute to improve the point-to-point accuracy, they increase the error in terms of unit proportions. W and WA, which lead to the biggest improvement for Kappa statistics, give the worst results of SSE of unit proportions. An exception to this symmetrical structure is found with A and AW: while AW is always more performing than A for Kappa statistics, it stays more performing in terms of SSE of unit proportions too.

The proportional increase of error of unit proportions, relative to the best performing topographical combination, is larger than the proportional increase of point-to-point accuracy with the Northern Subcatchment, and smaller with the Savournon subcatchment. Adding W to HRS at 12 classes with the Northern subcatchment decreases the accuracy of the proportions of units by around 14% and increases the Kappa accuracy by 4%. With the Savournon Subcatchment, the decrease of accuracy of unit proportions is of 4%, while the increase of Kappa accuracy reaches almost 9%.

When using the dotted training area, the ranking of neighbourhood attributes in terms of SSE of unit proportions is identical to the Kappa evaluation. Notwithstanding, neighbourhood attributes are always improving the accuracy of unit proportions, irrespective of their positive of negative impact on Kappa accuracy. The rankings are identical in that attributes which are increasing the Kappa accuracy the most (or decreasing it the less) are also decreasing the error in unit proportions the most.

For class 2 and class 5, the proportional decrease of error of unit proportions when adding W is always larger than the proportional decrease of Kappa accuracy. For class 3, which is the only class number for which neighbourhood attributes improve Kappa accuracy, both types of evaluation lead to a proportional increase of accuracy, of 6% for the Kappa coefficient and about 5% in terms of unit proportions.

Patch size For the dotted training area, neighbourhood combinations rank similarly when evaluated with either patch size of SSE of unit proportions. At 3 classes however, AW, W and WA follow opposite trends to the SSE of unit proportions by decreasing the patch size accuracy.

When using the Northern and Savournon subcatchments, the ranking of neighbourhood combinations is modified with each class number. The type of attributes leading to an increase or a decrease of the patch size difference differs with each number of classes, rendering general conclusions particularly difficult to draw. It was shown in section 4.2 that the evaluation of patch size difference without neighbourhood attributes tended to follow exactly the same pattern as the SSE of unit proportions, for all training areas; it appears that when neighbourhood attributes are added, this relationship can only be observed with the dotted training area at 2 and 5 classes.

4.6.2 Hydrological evaluation

SSE of hydraulic conductivity Adding the neighbourhood attributes AW, W and WA increases the accuracy in terms of hydraulic conductivity in all training areas and with all class numbers. The attribute A worsens the results with the Northern subcatchment at 7 classes, and with the Savournon subcatchment at 7 and 12 classes; it increases the accuracy in all other cases.

With every training area, the ranking of neighbourhood attributes stays generally identical when modifying the number of classes. The ranking is identical in the Northern and Savournon subcatchment: W leads to the largest increase of accuracy, followed by WA, AW and A. Similarly to the Kappa evaluation, the attribute W appears hence to have the largest positive impact on the accuracy. In this evaluation however, adding A after W leads to worse results than those obtained when adding W alone. This fact was not observed in terms of Kappa statistics and does not hold for the dotted training area, where WA is performing better than W for two of the three class numbers.

Runoff evaluation Three general observations can be made about the influence of neighbourhood attributes on runoff estimation.

First, for all training areas and outlets, the proportion of neighbourhood combinations improving runoff estimation tends to decrease as the rainfall intensity increases. This can best be observed with the Northern subcatchment, where at a rainfall intensity of 7mm, all neighbourhood combinations improve runoff estimation, irrespective of the number of classes. With a rainfall intensity of 12mm, however, at least one neighbourhood combination per number of classes is increasing the square error of runoff; at outlet 4, all neighbourhood combinations worsen the results.

Second, for each given number of classes, the ranking of neighbourhood attributes stays in most cases identical when the rainfall intensity increases. Exceptions are found with the Northern subcatchment, which results are more heterogeneous than those obtained with the two other training areas.

Third, with the Savournon subcatchment and the dotted training area, the impact of neighbourhood combinations can be observed pair-wise. A and AW show similar influence on runoff estimation, just as W and WA. With both training areas, A and AW lead generally to the highest increase of accuracy in the majority of the evaluations.

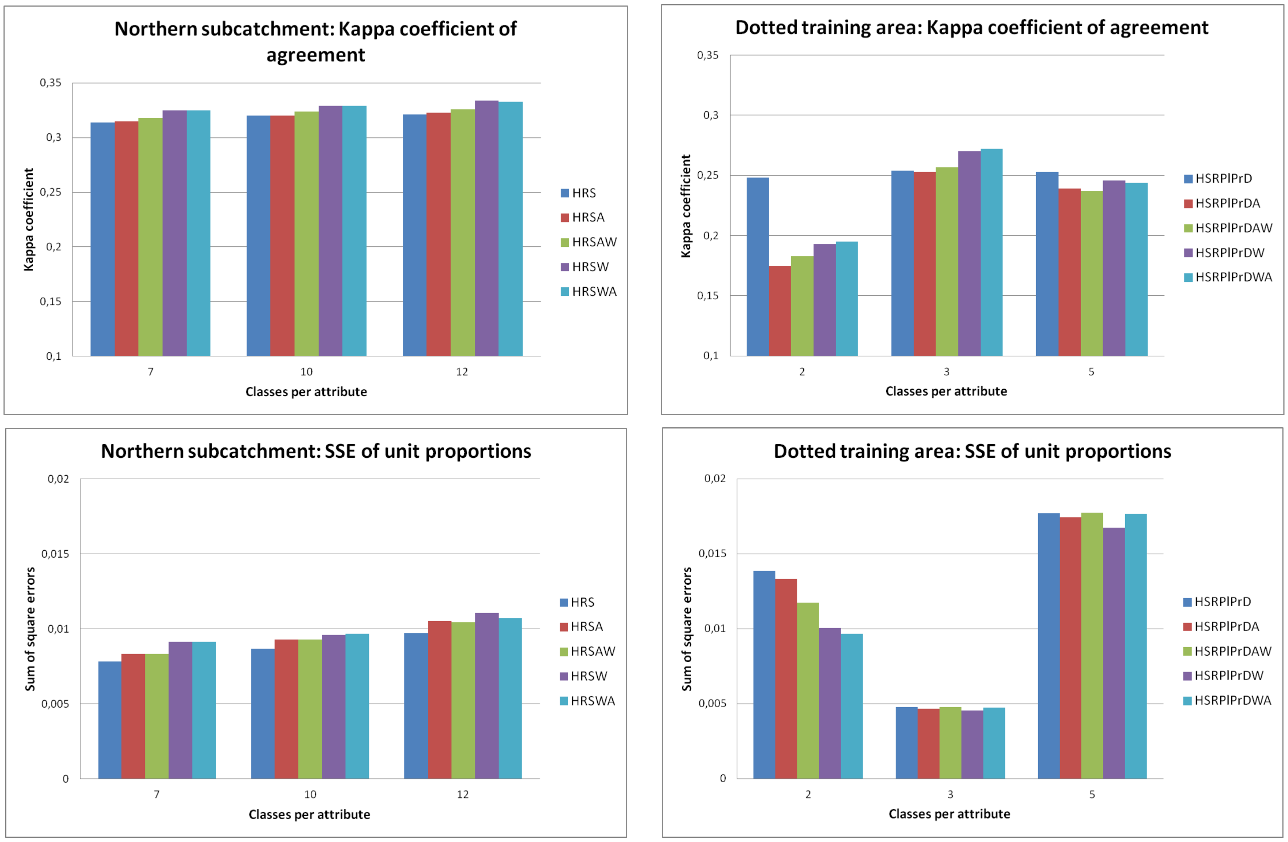

Figure 20: Impact of neighbourhood attributes on Kappa accuracy and sum of square error of unit proportions, with the Northern subatchment and the dotted training area.

4.6.3 Summary

With the Northern and the Savournon subcatchments, adding neighbourhood attributes to the best topographical combination increases Kappa accuracy, but also increases the sum of square errors of unit proportions. Inversely, neighbourhood attributes decrease Kappa accuracy with the dotted training area at 2 and 5 classes, and increase the accuracy in terms of unit proportions. The attribute "unit of window majority" has the largest impact on Kappa statistics. Its contribution to improving the accuracy is lessened when it is added after "unit of downstream neighbour".

The impact of neighbourhood attributes on patch size difference is difficult to analyze, as it varies with each number of classes and with each training area. The sum of square error of hydraulic conductivity, on the other hand, is decreased by the neighbourhood attributes AW, W and WA with all training areas and all class numbers. In terms of runoff estimation, less neighbourhood attributes are increasing runoff estimation as the rainfall intensity increases. The ranking of the attributes stays however relatively stable inside of each class number, irrespective of the rainfall intensity. The attributes A and AW, which have the lowest impact on Kappa statistics and SSE of hydraulic conductivity, show the highest improvement with both the Savournon subcatchment and the dotted training area.

4.7 Classification of individual units

4.7.1 Correctly classified cells

Influence of the combinations of attributes Irrespective of the training area, most units are best classified by the combination consisting of all topographical attributes, HRSPlPrD. The more topographical attributes, the higher the percentage of correctly classified cells. Consequently, HR (or HS with the dotted training area), the combination with the lowest number of attributes, is the combination yielding the lowest percentage of correctly classified cells. With the dotted training area, this observation holds for all units.

When the Northern subcatchment and the Savournon subcatchment are used as training areas, the unit "debris slope" is however best classified by the two combinations HR and HRS (Figure 21). The difference towards the combination HRSPlPrD, which is performing best otherwise, is particularly important. Debris slope are better classified by HRS than by HRSPlPrD by 29.7 percentage points with the Northern subcatchment, and 13.3 percentage points with the Savournon subcatchment. As a mean of comparison, the highest difference reached by HRSPlPrD over HRS with the Northern subcatchment, which concerns coarse alluvial fans, is of 7.7 percentage points.

Next to classifying "debris slope" cells more accurately, HR and HRS do also classify "glacis" cells better than other combinations with the Northern subcatchment; the difference in classification of HRS towards HSRPlPrD (7 percentage points) is however smaller than for the unit "debris slope".

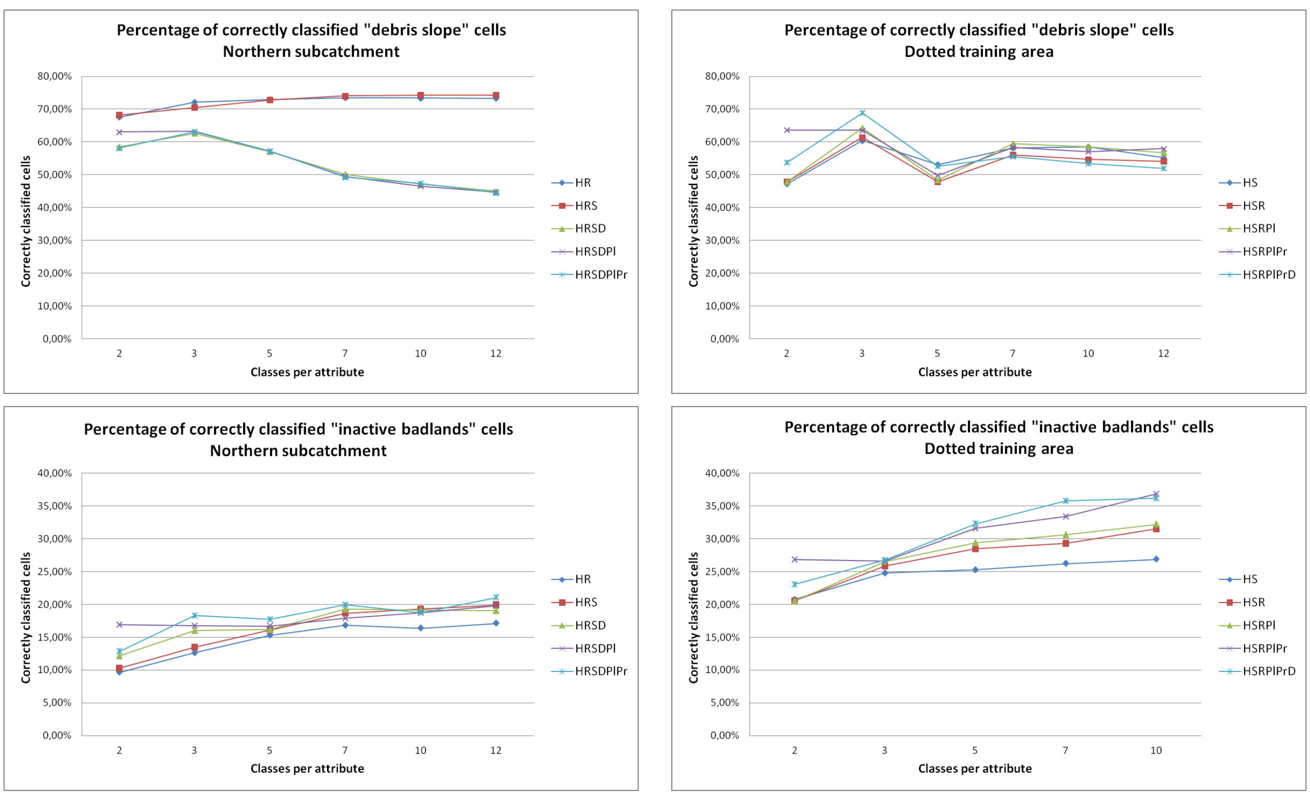

Figure 21: Percentage of correctly classified cells of the two most present units in the Buëch catchment, "debris slope" and "inactive badlands", with the Northern subcatchment and the dotted training area.

Influence of the number of classes In all training areas, the percentage of correctly classified cells tends to increase as the number of classes rises, irrespective of the combination. Two exceptions can be found with the unit "debris slope". With the Northern subcatchment and the Savournon subcatchment, the accuracy of the mapping of the "debris slope"cells is only increasing for HR and HRS. These observations can be put in relationship with the positive response of Kappa accuracy over an increase of the number of classes for both training areas. With the dotted training area, the percentage of correctly classified "debris slope" cells reaches a maximum at 3 classes and a minimum at 5 classes. These peaks can also be found in terms of SSE of unit proportions and Kappa statistics, albeit to a lesser extent.

Influence of the training area Large variations are found between the training areas in terms of best classified units. The unit "debris slope", which achieves the highest proportion of correctly classified cells with any of three training area, is the only constant.

The percentage of correctly classified cells per unit is on average larger with the Northern subcatchment than with the two other training areas. , For the best resulting maps, the average percentage of correctly classified cells reaches 31% with the Northern subcatchment, 26% with the Savournon subcatchment and 24% with the dotted training area.

Influence of neighbourhood attributes The influence of neighbourhood attributes on the percentage of correctly classified cells reached by the best topographical combination is relatively low. For none of the units does adding neighbourhood combinations lead to modifying the ranking of the topographical combination, the values with neighbourhood attributes being very close to those without neighbourhood attributes. The highest difference reached is an improvement of around 7 percentage points of the classification of river plain cells with the dotted training area.

Whether neighbourhood attributes improve or worsen the classification depends on the training area, the units and the number of classes. It can however be observed that all of them tend to improve the classification of the units "coarse alluvial fans", "inactive badlands" and "debris slope".

4.7.2 Misclassified cells

The confusion matrix, which is constructed to calculate the Kappa coefficient, can be used to analyze into what other units cells have been misclassified. It can be observed that units get misclassified into other units which location or geomorphological properties are relatively similar. The worst classified unit, "glacis", is often classified as "debris slope" or "inactive badlands". This is also the case with the unit "colluvium", which is reaching low percentages of correctly classified cells with the Savournon catchment. Large confusions in terms of topography or/and geomorphology, such as river plains being classified as debris slopes, are rare and of the order of a maximum of 15%.

The type of units into which cells get misclassified is highly influenced by the proportions of units in the training area. When using the Northern subcatchment as training area, where "colluvium" is the most present unit, all units get misclassified into colluvium by 6% to 31% (Figure **??**), while this percentage does not exceed 8% with the Savournon subcatchment, where it is only the 4th most present unit. Inversely, with the Savournon subcatchment, all units get misclassified into "badlands inactive" (the most present unit) by up to 56%, against 20% with

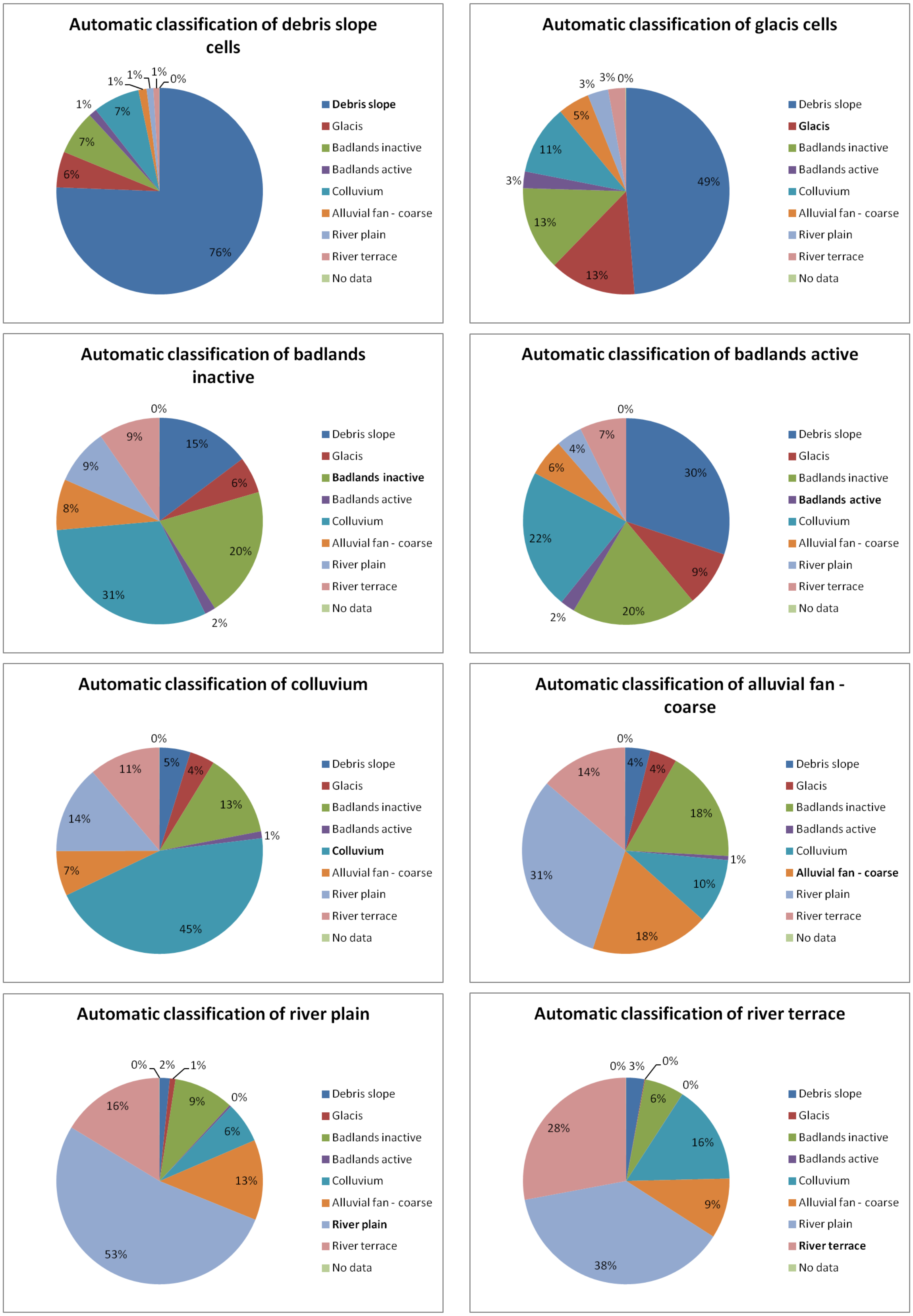

Figure 22: Classification of individual units with the most accurate automated map on a point-to-point basis (HRSW at 12 classes, Northern subcatchment as training area)

the Northern subcatchment, where it is the 6th most present unit.

Further remarks can be made about active/inactive badlands, and coarse/fine alluvial fans. Active badlands are misclassified as inactive badlands, but not the inverse: inactive badlands become rather misclassified as colluvium and debris slope. Similarly, fine alluvial fans are misclassified as coarse alluvial fans, while coarse alluvial fans become river plains and inactive badlands. These observations could be explained by the large differences between these similar units in terms of unit proportions, as inactive badlands and coarse alluvial fans are significantly more present in the training areas and in the Buëch catchment.

5 Discussion

A multiple-point geostatistical technique was used for an automated geomorphological classification of the Buëch catchment. An optimization design aiming at identifying the best combination of topographical and neighbourhood attributes, as well as the most performing number of classes, was followed for three different training areas of identical size. It was shown that for the two training areas which corresponded to subcatchments, the best combination of topographical attributes contained the attributes "distance to higher elevations", "relative elevation" and "slope", at 10 and 12 classes. For the dotted training area, the differences between the combinations are significantly lower than for the two subcatchment. The highest accuracy reached with the dotted training area was obtained with the combination made of all topographical attributes, at 3 classes. Adding neighbouring attributes to the best topographical combinations leads to contradictory results between the training areas and the types of evaluation. Overall, on a point-to-point basis, the most accurate automated map was obtained with the Northern subcatchment as training area, when using the attributes "distance to higher elevations", "relative elevation", "slope" and "unit of window majority" at 12 classes. A Kappa value of 0.33 was reached hereby; the average percentage of correctly classified cells per unit was of 31%. It was shown that the unit "debris slope" was particularly well mapped by the algorithm, at about 75% of accuracy. Poorly reproduced units, such as "glacis", were typically misclassified into relatively close units in terms of topographical or geomorphological properties.

5.1 Influence of the number and types of attributes

One of the research questions formulated for this study amounted to assess the impact of the number and types of attributes on the accuracy of the automated map. Irrespective of the training area under consideration, it was observed that three attributes were performing considerably better than others when taken individually. Those were "distance to higher elevations", "relative elevation" and "slope". When joined, these attributes lead to the best combination of topographical attributes for the two subcatchments; adding further attributes decreased the accuracy.

These results can be revisited in light of the ability of attributes to correctly classify individual units. It was shown that a large difference exists between combinations of attributes regarding the classification of the unit "debris slope"; when the Northern subcatchment was used as training area, the combination HRS classified "debris slope" up to 30 percentage points better than the combination HRSDPlPr. Given that "debris slope" is the most present geomorphological unit in the Buëch catchment, this difference plays a considerable role in the overall accuracy of the combinations.

In an opposite way to debris slopes, most geomorphological units are in fact better classified with the highest number of attributes, i.e. the combination HRSDPlPr. It can be inferred that if the proportion of "debris slope" cells over the catchment was significantly lower, the combination HRSDPlPr would outclass HRS in terms of overall accuracy. It appears thus that the difference in terms of correct classification of "debris slope" cells could provide on its own an explanation to the superiority of HRS regarding point-to-point accuracy (Kappa statistics) and unit proportions in the Northern and in the Savournon subcatchments[4]. More striking even

[4] With the dotted training area, the unit "debris slope" is best classified by the combination HSRPlPrD, similarly to other units; consequently, HSRPlPrD leads to the best overall accuracy.

is the similarity between the classification of "debris slope" cells by the combinations and the evaluation of the SSE of unit proportions: the ranking of the combinations and their response to an increase of the number of classes are identical in both evaluations. The classification of "debris slope" cells appears hence to have an even larger influence on the proportions of units than on point-to-point accuracy.

Two conclusions can be drawn from these observations. First, the unit "debris slope" seems to have been represented in a very performing way by the three attributes "distance to higher elevations", "relative elevation" and "slope". Since this unit, found on particularly steep slopes, occupies the highest position in the landscape (which was used in order to define the attribute "distance to higher elevations" too), those three attributes offer powerful discrimination possibilities. Second, the overall accuracy of the automated map might clearly differ from the results obtained in this research if the catchment to be mapped contains a lower proportion of "debris slope". As all other units reach much lower proportions of correctly classified cells, the overall accuracy can be expected to decrease when the proportion of "debris slope" decreases too.

5.1.1 Neighbourhood attributes

The geomorphological map of the region consists of relatively large patches of units. By integrating neighbourhood information in the algorithm, it was expected that the resulting automated map would be rendered smoother, hence more similar to the field map. It was shown that when the Northern and the Savournon subcatchments were used as training areas, neighbourhood attributes did indeed improve the point-to-point accuracy of the automated map in comparison to the best combination of topographical attributes; notwithstanding, the error in terms of unit proportions increased. With the dotted training area, the results were the opposite.

The opposite response of point-to-point and lumped evaluations to the addition of neighbourhood attributes can seem surprising, since the two types of evaluation lead to similar results in all other stages of the optimization design. The results show that the units which classification improved in all cases with neighbourhood attributes were the three most present units, i.e. debris slope, inactive badlands and coarse alluvial fans. It is possible that the opposite effect of neighbouring attributes on highly represented and rare units impacts on the evaluation of the mapping accuracy, since these units do not weight equally in the SSE of unit proportions. For example, "glacis" cells being about ten times less frequent than "debris slope" cells, the impact of one "glacis" cell on the proportions of units will be proportionnally larger than the impact of a "debris slope" cell.

5.2 Influence of the number of classes

A second research question concerned the influence of the number of classes on the mapping accuracy. The higher the number of classes, the higher the number of possible patterns; and hence the lower the amount of cells associated to each pattern during the training phase. As a consequence, it could be expected that an optimum of accuracy be found at a relatively low number of classes, allowing for a larger representativeness of each pattern. In an introductory article to the snesim algorithm, Strebelle (2004) typically advised to reduce the number of classes to 4 or beneath when using one attribute.

The results of this research show however that in terms of point-to-point accuracy, the performance of the best combinations of attributes increases as the number of classes rises. The most accurate map is obtained with 12 classes and 4 attributes, which represents $12^4 = 20736$ patterns, while the Northern training area is 16657 cells large. Even in the case where only one cell would be associated to each pattern, there are thus about 4000 patterns more than available cells.

An analysis of the search tree from which the best automated map was drawn (HRSW at 12 classes, with the Northern subcatchment as training area), reveals that only 882 of the 20736 possible patterns are used. In the large majority of the cases, only one type of unit is associated to each pattern. The number of cells belonging to a given pattern can be particularly high: approximately one third of all "river plain" cells, hence over 1000 cells, are for example associated to the pattern "distance to higher elevations [12], relative elevation [3], slope [1], unit of window majority [river plain]". The specificity of patterns towards given geomorphological units suggest that a high number of classes allows for an increased discrimination. As a second mean of illustration, with 12 classes only "debris slope" cells are associated to patterns including the highest class of "slope", one of the third highest classes of "relative elevation" and one of the four lower classes of "distance to higher elevations". When the number of classes is decreased, other units share the same characteristics.

An first important consequence of the high number of classes, and more precisely of the specificity of patterns which characterizes it, regards one of the properties of the multiple-point geostatistical approach: its stochasticity. The automated map resulting from the MPG approach is a realization of a conditional cumulative probability function, inferred at each cell from the search tree. However, if each pattern is only attributed to one geomorphological unit, as it is the case with 12 classes, there is for each pattern a 100% probability to correspond to that unit. In other words, the approach is almost deterministic and drawing multiple realizations from the search tree would result in near to identical maps. This would not be the case with a lower number of classes, where multiple geomorphological units share the same patterns.

A second consequence regards the accuracy of the automated map. If only one unit is associated with each pattern, then it implies that the errors of the automated mapping originate from heterogeneity in the patterns-units relationships between the training area and the mapping area. In other words, the criteria (attributes and classes) used to discriminate a geomorphological unit in the training area do not lead to the same unit being drawn in the rest of the catchment, because of topographical variability. It might hence be desirable to identify attributes that are less sensitive to spatial heterogeneity throughout the entire catchment.

It should be added that the number of classes was shown not to impact all combinations in an identical way. With the Northern and the Savournon subcatchment, increasing the number of classes did only improve the accuracy of the two best performing combinations (HR and HRS), while it decreased the accuracy of all others above 3 classes. These opposite responses could find a possible explanation in the different structure of the search tree in these situations. Adding attributes increases the number of possible patterns, just as does an increase of the number of classes. As presented above, using the highest number of classes with a small number of attributes still leads to a relatively high number of counts per pattern, given that only a small fraction of all possible patterns are used. Comparatively, raising the number of attributes appears to have a different effect: the number of patterns used in the search tree is extremely large, and most patterns only correspond to a few cells. This situation, combined with a high number of classes, leads to a form of over-specificity where each cell of the training area is stored individually; if slightly different topographical situations are encountered in the catchment, errors are easily made.

5.3 Influence of the training area

Similarly to the influence of the types of attributes, the effect of the training area on the mapping accuracy can best be discussed by analysing the classification of individual units. In fact, the ability of combinations of topographical attributes to classify geomorphological units correctly seems to be dependent on the proportions of units in the training area used. With all training areas, the three best classified units correspond to the three most frequent units in the corresponding training area, irrespective of their actual proportion in the Buëch catchment. This correlation between the correct classification of an unit and its proportion in the training area can be extended to nearly all units in the Northern subcatchment. Some exceptions are found with the Savournon subcatchment and the dotted training area.
The percentage of units in the training areas appears to impact not only on the type of units that are best classified, but also on the percentage of correct classification of these units. The average percentage of correctly classified cells was shown to be higher with the Northern subcatchment than with the Savournon subcatchment and the dotted training area. A possible explanation could be found in the more even distribution of the proportions of units with the Northern subcatchment. Six of the eight units present in the Northern subcatchment cover each 10% of the area or more; in the two other training areas, more than half of the units do not 10% each. Since the proportion of an unit in the training area appears to impact directly on its classification, training areas with overall well represented units could be expected to attain a higher percentage of correctly classified cells than those showing a limited number of highly present units and many low represented units. The ideal training area would hence be an area where the main units of the catchment would be present in relatively large proportions; in terms of overall accuracy, the presence of more rare units would not necessarily improve the results, albeit the training area would seem more representative of the regional geomorphology.

Two reasons could explain the better classification of well represented units. First, a larger number of cells per unit means a higher number of counts per unit in the search tree; as a consequence, the unit will have a higher probability to be drawn from the search tree. This fact is of importance when the number of patterns is low, i.e. when many units are associated to the same pattern. In the case where the number of patterns is very high and where only one unit is generally associated to each pattern, the number of counts per pattern is insignificant, given that the probability to draw that unit is in any case 100%. Rather, the number of cells will have an impact when it signifies that one unit get associated to various patterns: the second reason would hence relate to the higher representativeness achieved with highly present units. An unit with a large number of cells will necessarily cover a larger range of topographical situations: in the case of debris slopes, the area might cover more or less steep slopes, at various relative elevations, which will furnish the algorithm with a general representation of the locations where debris slopes might be expected. However, when only a very limited number of cells belong to an unit in the training area, the description of the unit by topographical attributes is extremely specific, while the unit might be encountered in the entire catchment in a much wider range of topographical situations.

5.4 Overall performance of the multiple-point geostatistical approach

On the basis of the research questions answered in the previous subsections, the general objective of this study was to assess the capability of the multiple-point geostatistical method to reconstruct the geomorphology of the area. It can be observed that this capability is highly dependent on the type of geomorphological units: the percentage of correctly classified cells ranges from 2% for "active badlands" to over 75% for "debris slope", hence from very low to very high agreement. Given that the most present units of the catchment are well mapped on average, the general pattern of the geomorphology can be considered to be well captured by the algorithm. On a more detailed basis, the patches of units are however barely recognizable and the map remains relatively noisy.

The study area was already mapped with a rule-based classification technique and a k-means clustering technique in an earlier research (Schuur, 2009), which allows to put the present results into perspective. It can be noted that the difference between units in terms of percentage of correctly classified cells is considerably larger with the MPG approach than with the two previous classification techniques, for which the majority of units reached between 20% and 30% of correct classification. With the rule-based and clustering techniques, all units are always represented; this is not the case with the MPG approach, where only the units present in the training area are mapped over the catchment. Notwithstanding, the average percentage of correctly classified cells was about 4 percentage points higher with the MPG approach, which classified "debris slope" significantly better. The Kappa coefficient of accuracy was not used in the evaluation of the other techniques. However, the overall accuracy of the different approaches can be informally compared through the total percentage of correctly classified cells. The confusion matrix reveals 47% of correct cells with the MPG approach, while the highest accuracy reached by the rule-based and clustering techniques was 32%. Although these results do not provide a direct means of comparison between the three automated classification techniques, they suggest that the MPG approach could be promising in the search for more accurate automated geomorphological mapping.

Two aspects should be taken into consideration when analysing the overall accuracy of the resulting automated maps. First, the geomorphological field map which was used as a basis for this study contains a large number of units, which do not always differentiate themselves from each other univoquely. The difference between active and inactive badlands, which relies on the percentage of vegetation cover, is for example rather subjective: the evolution from active to inactive badlands follows a continuum, which renders clear boundaries difficult to draw. From a hydrological point of view, the difference between badlands covered at 10% and 15% by vegetation might also be relatively insignificant. It can hence be wondered whether a such detailed distinction is appropriate for the ultimate purpose of the development of this technique, i.e. the application in large-scale hydrological modelling. In this regard, the misclassification of cells provides a meaningful insight of the performance of the MPG approach. The fact that units gets generally misclassified into relatively close units in terms of geomorphology and topography can be considered as very positive. With the Savournon subcatchment, active badlands reach barely 3% of correctly classified cells; however, they are misclassified at 51% as inactive badlands. This suggests that merging similar units together would lead to a different assessment of the accuracy of the MPG approach, which would be likely to perform considerably better.

Second, the evaluation is entirely based on the comparison between the automated map and the field map. The geomorphological field map is necessarily subjective, so that differences can be expected if other persons were to map the same area again. Additionally, some parts of the catchment are less easily reachable than others, which impacts the quality of the field mapping. An example can be found in the middle-west part of the catchment, of which a large area was solely mapped as debris slope, while the automated mapping assigned a larger diversity of units to it. This steep and forested area was both difficult to access and to map from long distance observation points. It is hence conceivable that the automated map was in some cases more close to reality than the field map. In conclusion, the evaluation of the performance of the MPG approach stays a relative evaluation, carried in comparison to another subjective, man-made map, and should not be considered as an absolute assessment of the automated classification to depict reality correctly.

5.5 Suggestions for future research

The application of the multiple-point geostatistical approach to automatic geomorphological mapping reveals several aspects in which technical and conceptual improvements are desirable. These will be treated hereafter in the order of the original research questions.

A first domain concerns topographical and neighbourhood attributes. It was shown that the topographical attributes used in this research, and in particular "distance to highest elevations", "relative elevation" and "slope", lead to a very accurate classification of "debris slope" cells. With all training areas, the correct classification of "debris slope" cells was approximately 20 percentage points higher than the classification of any other units. This large difference should be taken into account in order to define other attributes allowing for a better discrimination of the remaining units. Half of the attributes of this research were related to the relative position of the cell towards other elements, such as rivers and crests. Additional attributes taking a more detailed account of the DEM properties of the cell and of its neighbours could be defined. On a technical aspect, the attribute "relative elevation", which is calculated in this research on the basis of the minimum and maximum elevations of the entire catchment, would probably gain to be redefined on a subcatchment basis, as proposed by Schuur (2009).

The possibility to integrate neighbourhood information in the algorithm is a strength of the MPG technique over more traditional approaches, such as the k-means clustering and rule-based classification techniques. In this regard, exploring on additional neighbourhood attributes is of particular importance in order to assess more precisely the advantages and the drawbacks of this method in comparison to other automated mapping techniques. Since a large part of the regional geomorphology is highly influenced by downhill processes, further neighbourhood attributes could take the properties of the upstream area of the cell into larger account.

A second domain relates to the classes into which the values of the attributes are discretized. Three aspects deserve additional attention. First, since the highest point-to-point accuracy was obtained with the highest number of classes used in this research, additional runs should be realized in order to investigate whether the accuracy can still be improved by an increase of the number of classes. It is expected that above a certain threshold, the mapping accuracy will decrease again as the number of classes rises, as a consequence of the over-specificity of the search tree. Second, the class boundaries, which were defined so that each class contains an identical number of cells, could be defined instead on the basis of the values generally associated to given geomorphological units, in order to isolate units in particular classes. Third, the number of classes was kept constant in this research for each attribute inside combinations. It could however be set separately for each attribute, since discrimination possibilities are probably related in a different way to the number of classes depending on the topographical property under consideration.

A third domain concerns the training areas. The influence of both the geomorphology of the training areas and their location (subcatchments versus multiple spread areas) was assessed in this research. The issue of the impact of the size of the training area on mapping accuracy stays however unanswered. Since the size of the training area has a direct impact on the number of cells associated to patterns of the search tree, it can be expected that the response of the mapping accuracy to an increase of the number of classes and attributes will lead to different results when the size of the training area is modified.

Finally, In terms of hydrological evaluation, a next logical step would be to insert the automated map in the hydrological model for which it is meant to serve as input. Albeit using a full hydrological model implies that the exact effects of the automated map might be complex to differentiate from errors of other model parameters, it will provide a general idea of the suitability of the automated classification technique used in this research for the application in large scale catchment modelling.

6 Conclusion

Multiple-point geostatistical approaches to automated classification rely on deriving the conditional relationship between the properties of a cell and its membership from a training image. A multiple-point geostatistical technique based on the single normal simulation equation (snesim) was applied to the geomorphological mapping of the Buëch catchment, Southern France. The algorithm correctly reconstructed the general pattern of the geomorphology through an accurate mapping of the most present units; however, it lacked to distinguish unit patches. Large differences were observed in the classification of geomorphological units, with percentages of correctly classified cells ranging from 75% (debris slope) to 3% (active badlands). The units were mostly misclassified in relatively near units in terms of topography and geomorphology.

The results obtained in the present research were shown to be largely influenced by the correct classification of the unit "debris slope", which covers about 40% of the total area of the Buëch catchment. The high accuracy of its classification lead the combination which allowed for its best discrimination, "distance to highest elevations", "relative elevation" and "slope", to reach the highest overall accuracy over the Buëch catchment among topographical combinations, although most units were better classified by the combination containing all topographical attributes. The positive response of overall mapping accuracy to an increase of the number of classes appeared to be related to the high specificity of patterns it implied. A high number of classes allowed to reach a clearer discrimination between units, while still linking a representative number of cells to each pattern; this structure was improving the classification of the unit "debris slope". In contrast, increasing the number of attributes affected negatively the classification of "debris slope" cells by rendering the algorithm probably too sensitive to the topographical situation encountered in the training area.

The geomorphology of the training area was shown to impact directly on the correct classification of geomorphological units. The more present a geomorphological unit in the training area, the better its classification, irrespective of its actual proportion in the Buëch catchment. This relationship suggests that training areas with overall well represented units lead to a higher overall mapping accuracy than training areas with large differences in terms of unit proportions. This property could be explained by the higher number of cells associated to patterns in the dynamical data structure (search tree) when the area covered by an unit is enlarged.

In contrary to other classification techniques, the multiple-point geostatistical approach enables to incorporate neighbourhood information in the mapping algorithm. Neighbourhood attributes appeared to improve the accuracy of the automated map on a point-to-point basis, but to decrease it in terms of proportions of units. A possible explanation to this difference could be found in the opposite effect of neighbourhood attributes on well represented units and rare units.

Future research is required in order to derive new topographical attributes likely to improve the classification of units other than "debris slope". The possibility to include neighbourhood information should be explored into further extent, especially when considering the domination of downhill processes in the study area. The number of classes and the class boundaries could be adapted too in order to differentiate between units and to take the discrimination possibilities of topographical attributes into a more detailed account. Finally, the hydrological evaluation of the automated classification technique would gain to be carried on the hydrological model for which the automated map is meant to serve as input.

7 Appendices

7.1 Appendix A: Values of saturated hydraulic conductivity

Unit	Soil type	Ksat (m/h)
Hogback	outcorps	0
River plain	sandy clay loam	0.0043
Alluvial fan (coarse)	loam	0.0132
Alluvial fan (fine)	clay loam	0.0023
Colluvium	clay loam	0.0023
Active badlands	clay	0.0006
Inactive badlands	weathered rock	0.00001
Glacis	silt loam	0.0068
Glacis remnant	silt loam	0.0068
Mass movement	loamy sand	0.0611
Debris slope	sandy loam	0.0259

Table 3: Values of saturated hydraulic conductivity for given soil types, from Rawls and Brakensiek (1983). The correspondence between soil type and geomorphological units was obtained from Vannametee (personal conversation).

7.2 Appendix B: Runoff evaluation

Given the large differences of magnitude of the square error of runoff between outlets, the unit of the axis of the following graphs is set individually per outlet.

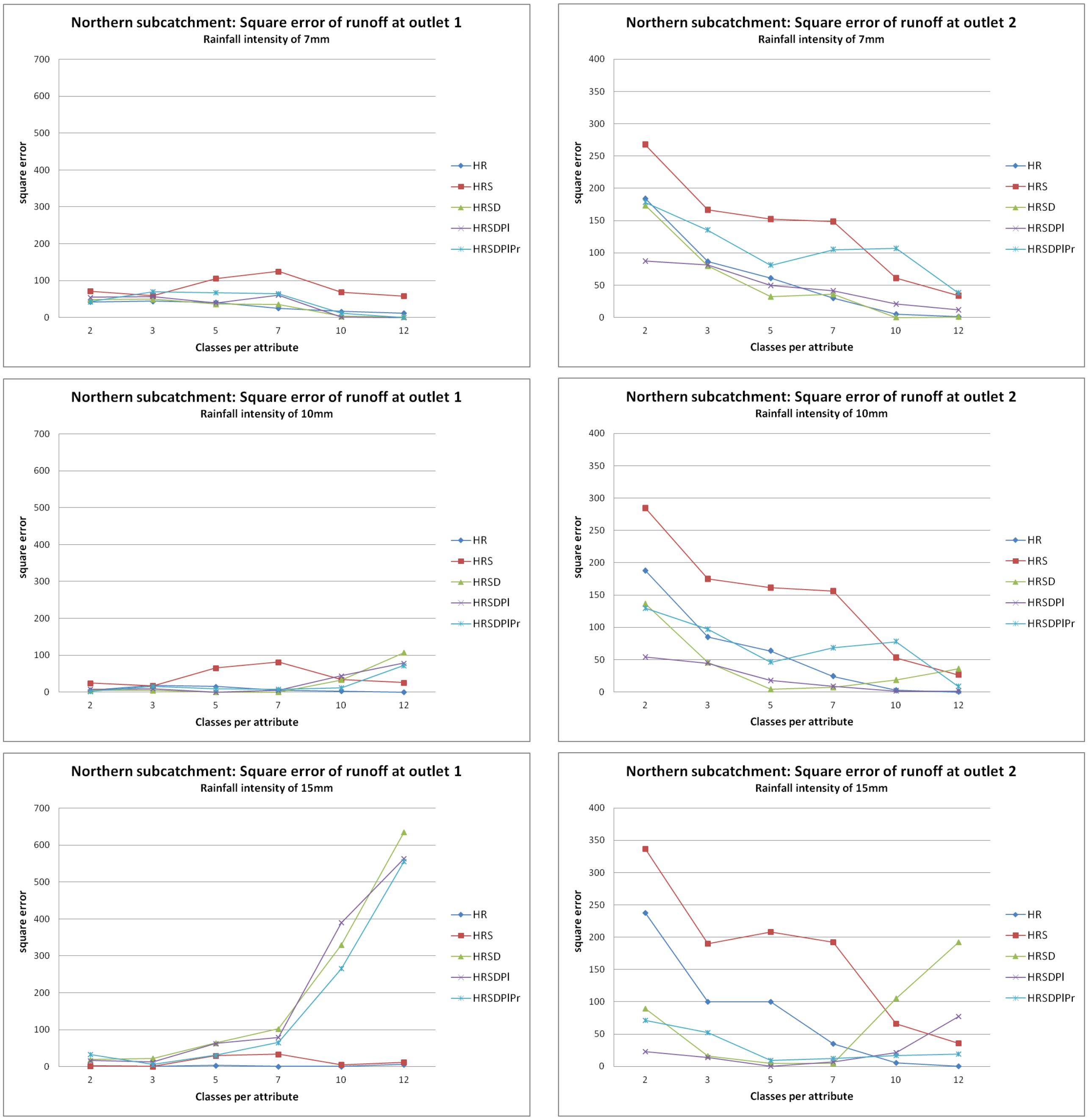

Figure 23: Square error of runoff estimation at outlet 1 and 2, obtained when using the Northern subcatchment as training area

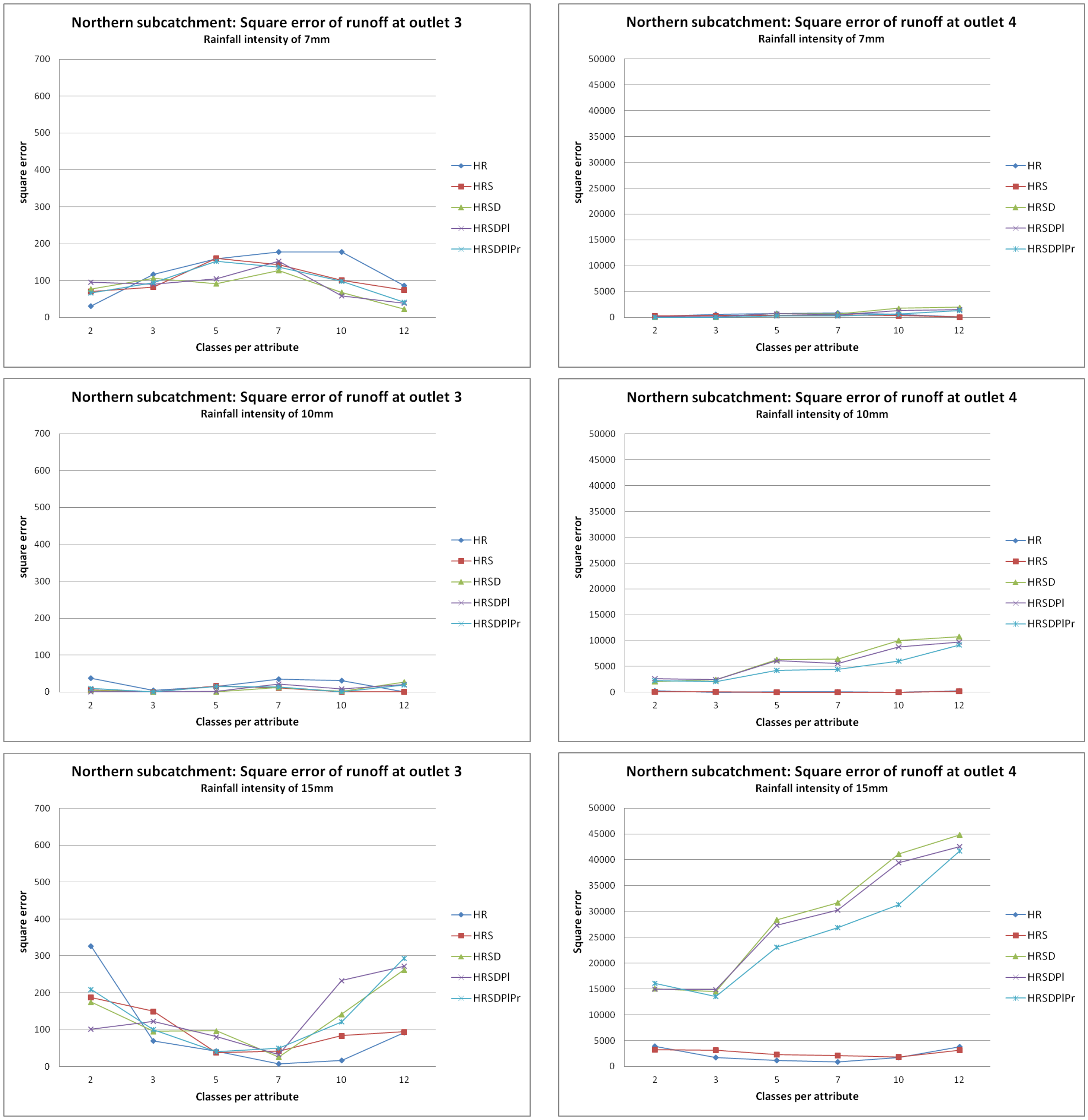

Figure 24: Square error of runoff estimation at outlet 3 and 4, obtained when using the Northern subcatchment as training area

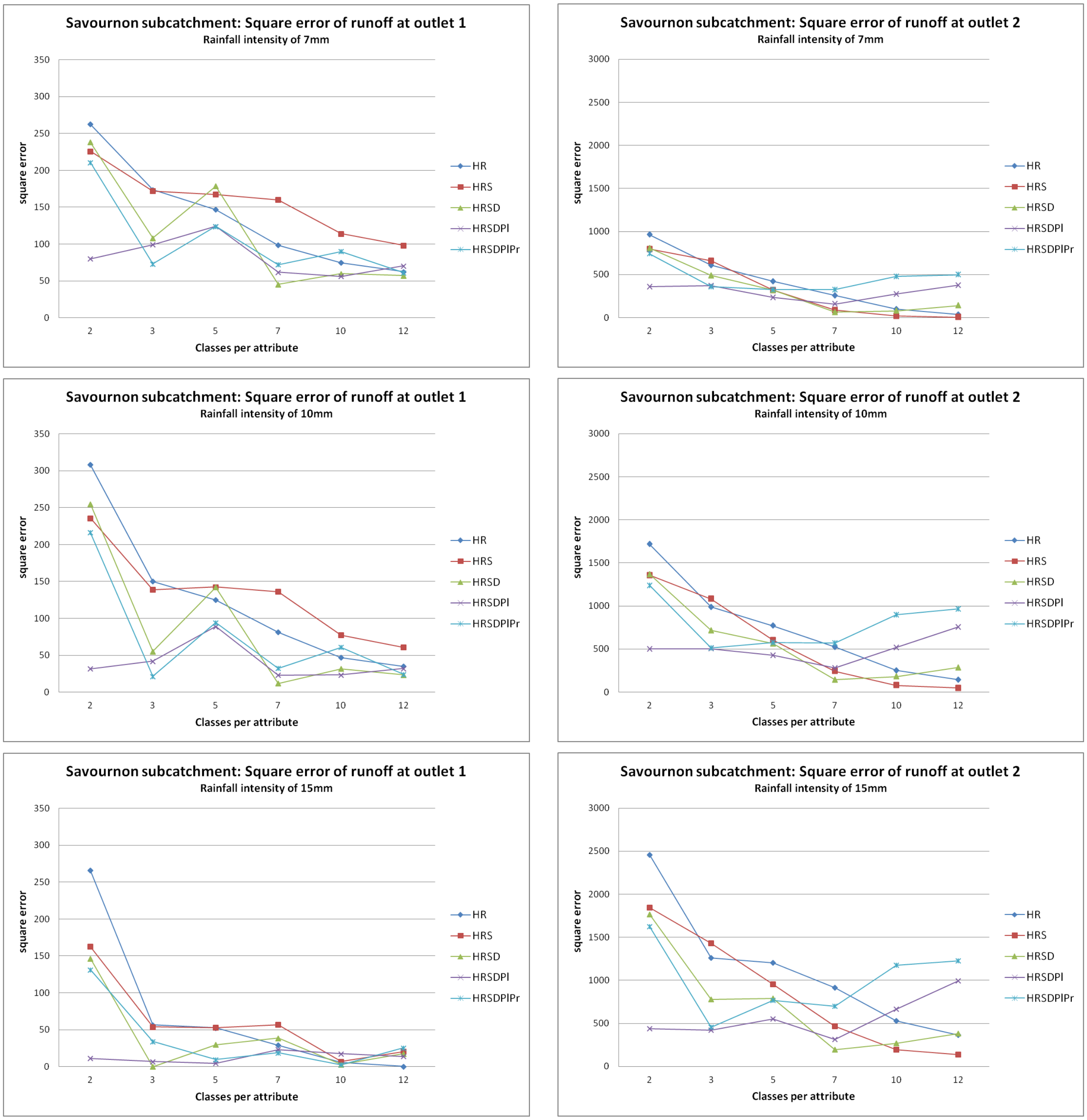

Figure 25: Square error of runoff estimation at outlet 1 and 2, obtained when using the Savournon subcatchment as training area

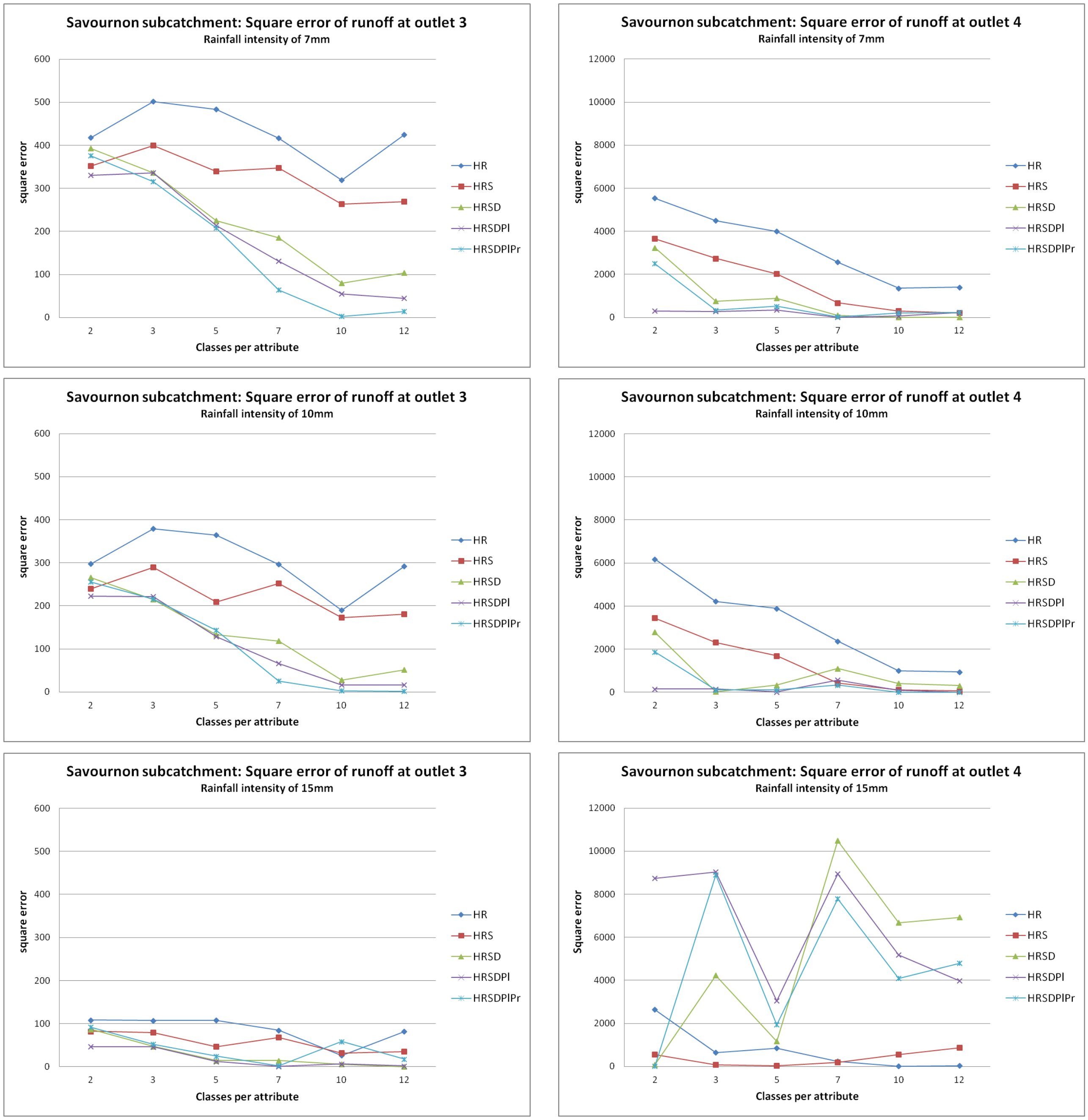

Figure 26: Square error of runoff estimation at outlet 3 and 4, obtained when using the Savournon subcatchment as training area

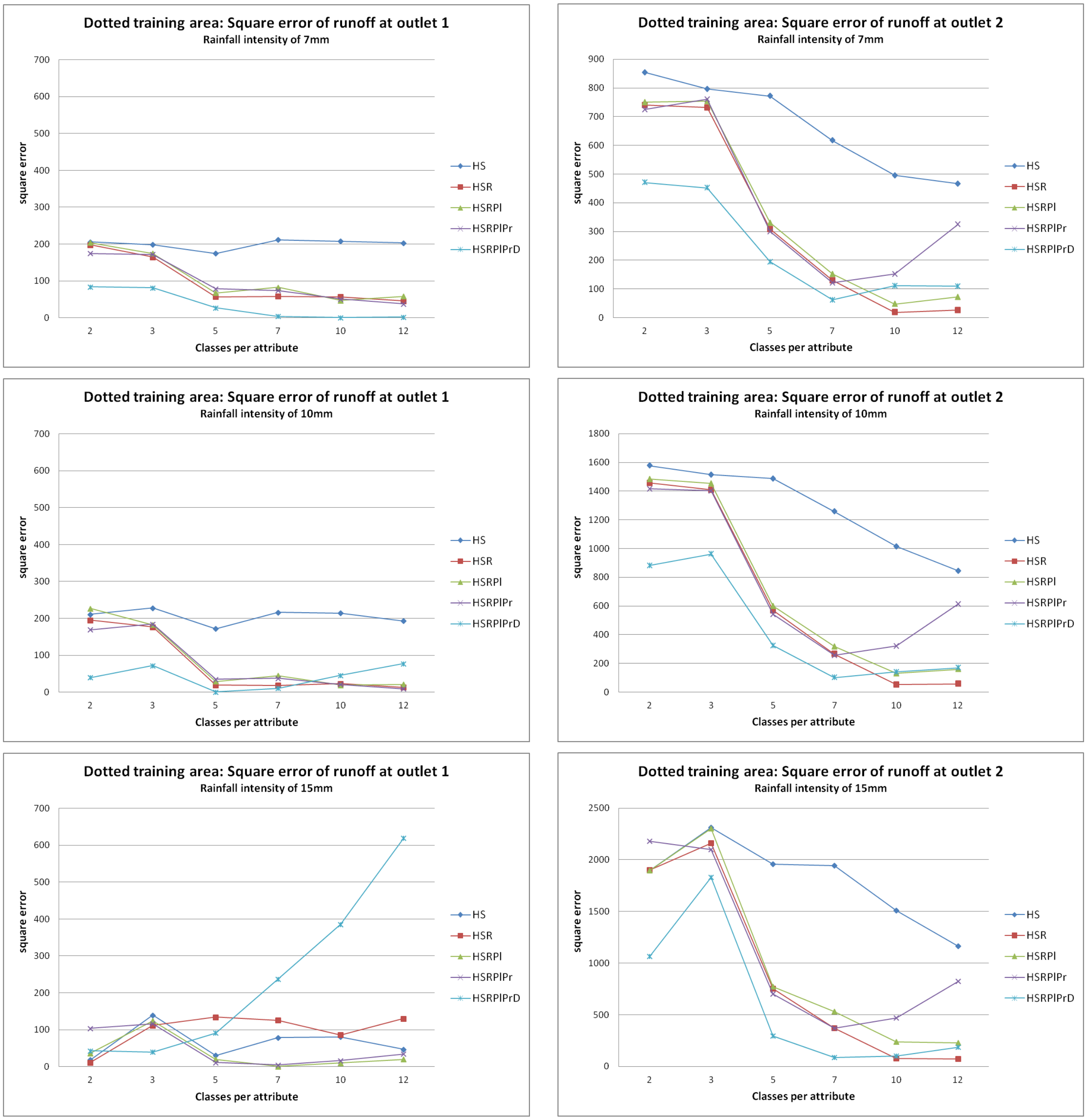

Figure 27: Square error of runoff estimation at outlet 1 and 2, obtained when using the dotted training area

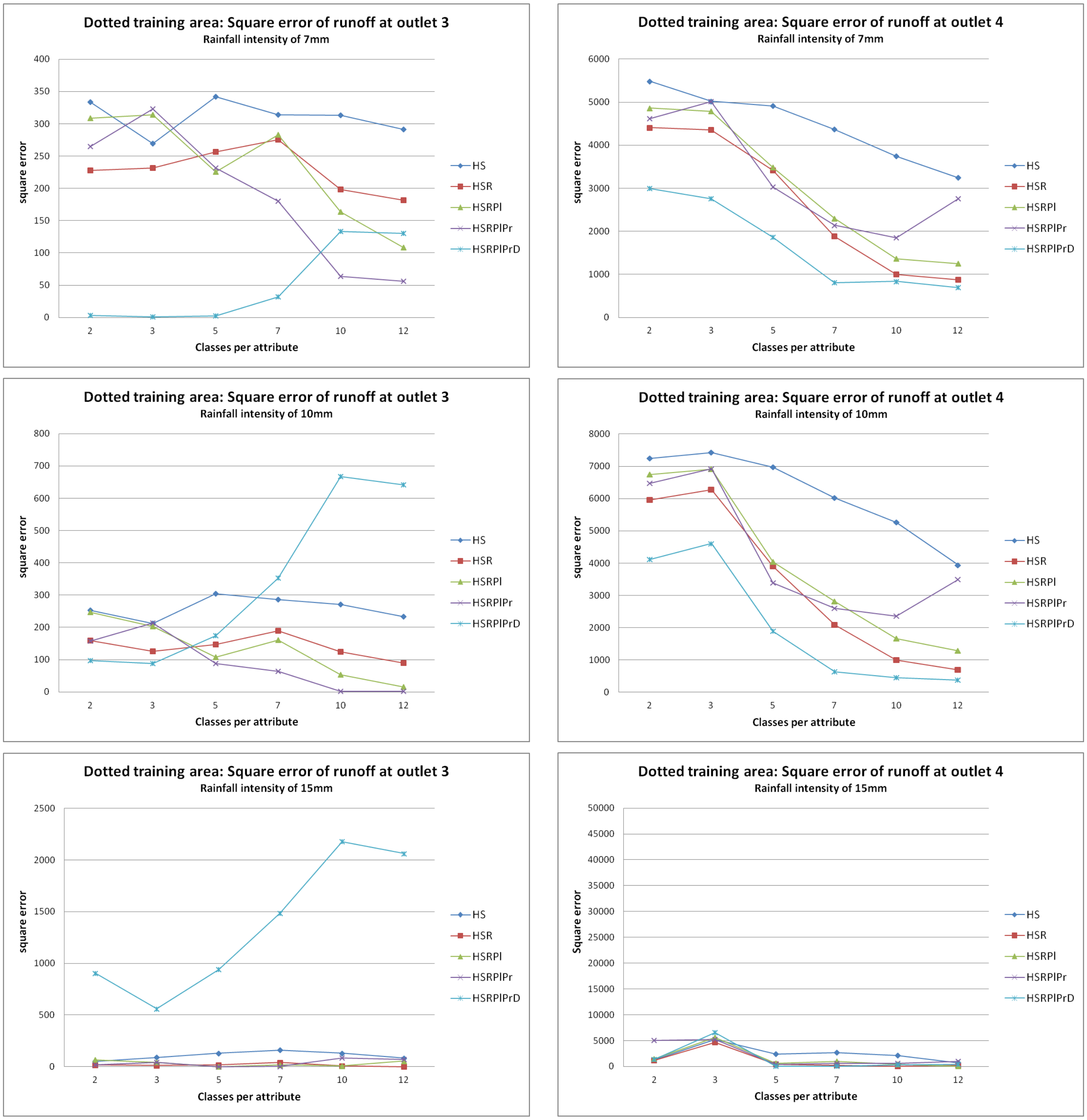

Figure 28: Square error of runoff estimation at outlet 3 and 4, obtained when using the dotted training area

References

[1] Asch, Th.W.J. van, J.H. van den Berg, P.C. Beukenkamp, M.R. Hendriks, S.M. de Jong, D. Karssenberg, E.J. Pebesma, H. van Steijn, E. Weiss (2010) Préalpes du Sud: Manual for the field course for first-year students of the BSc programme "Earth Sciences". Faculty of Geosciences, Utrecht University, 135p.

[2] Burrough, P.A. and R.A. McDonnell (1998) Principles of geographical information systems. Oxford University Press, Oxford, 333p.

[3] Caers, J. and T. Zhang (2004) Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. In: Integration of outcrop and modern analog data in reservoir models, AAPG Memoir 80, pp. 383–394

[4] Guardiano, F. and R.M Srivastava (1993) Multivariate geostatistics: Beyond bivariate moments, in A.Soares [ed], Geostatistics-Troia, Vol.1, pp.133-144

[5] Hagen, A. (2002) Multi-method assessment of map similarity. 5th AGILE Conference on Geographic Information Science, Palma, Spain.

[6] Hagen-Zanker, A. (2006) Map comparison methods that simultaneously address overlap and structure. Journal of Geographical Systems, 8, 165-185

[7] Liu, Y. (2006) Using the Snesim program for multiple-point statistical simulation. Computers & Geosciences, 32, pp. 1544-1563

[8] PCRaster Team (2011) PCRaster Documentation, release 3.0.1. Available online at http://pcraster.geo.uu.nl/documentation/PCRaster/PCRaster.pdf

[9] Rawls, W.J. and D.L. Brakensiek (1983) Proceedings of the American Society of Agricultural Engineers Conference on Advances in Infiltration. American Society of Agricultural Engineers, St. Joseph, MI, pp. 102-112.

[10] Schuur, J. (2009) Automated hydro-morphological classification of a part of the Buëch catchment, Southern France. MSc Thesis, Faculty of Geosciences, Utrecht University, 119p.

[11] Strebelle, S. (2002) Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology, 34(1), pp. 1-21

[12] Viera, A.J., J.M.Garrett (2005) Understanding interobserver agreement: the Kappa statistic. Family Medicine, 37(5), pp. 360-363