SELECTIVE SAMPLING FOR DIGITAL TERRAIN MODELLING

B. Makarovič
Photogrammetric Dept., ITC, The Netherlands
Commission IV, Working Group IV-8

ABSTRACT

Selective extraction and sampling of distinct morphometric features are primary ingredients of Composite Sampling for DTMs. The aim is to attain a balance between Selective and Progressive Sampling. Feature extraction is subjective and therefore needs to be systemised. Attention is given to main structural schemes concerning the basic information and control data, and the semantic and metric domains. Structures should be introduced in the data preparation stage, during feature extraction and before sampling. They provide a frame of reference for sampling and further processing.

INTRODUCTION

The paper aims at systemising the procedure for manual Selective Sampling of distinct morphometric features and "anomalous" terrain regions (Σ-data). Such data represent a geometric framework for subsequent manual, semi-automatic or automatic non-selective sampling to form a complete DTM. Data on distinct morphometric features have high information content - they tend to model the skeleton of terrain relief. The modelling can be full or partial; in the latter case, additional, more, homogeneous data have to be sampled, providing the so called "filling data". These can represent contour lines, parallel profiles or point grids.

Composite Sampling (CS) comprises both selected and filling data [1] [2]. Selective sampling is subjective and therefore inconsistent. Progressive Sampling (PS) excessively densifies the point grid in the vicinity of terrain breaks, despite the implied objective decision criteria. By unifying both sampling approaches into Composite Sampling, their advantages are preserved and deficiencies suppressed.

The actual problem, however, is deciding to what extent Σ-data should be sampled, i.e., to attain an effective trade-off among data integrity, accuracy and sampling effort. To this end, however, the procedure for Selective Sampling first needs to be systemised and the Σ-data purposefully structured.

For geo-information systems, a functional distinction can be made between the basic information and the control data. Basic information contains semantic and metric components. Control data are needed to define the basic information and for internal process control; they refer to semantic and metric domains. Information can be structured according to these four entries for all process stages (figure 1). Thus, structures can provide an orderly framework for information extraction, sampling, management and processing, conversion to other structures, and for representation and communication.

Domain / Function	Semantic	Metric
Basic (contents)	Key-items Attributes	Key-items Attributes
Control (specifications)	Inputs/Outputs Process control	Inputs/Outputs Process control

Fig. 1: Main entries for structuring

In the following sections, the four main schemes for structuring (figure 1) are outlined in the context of Selective Sampling for DTMs.

II. BASIC SEMANTIC INFORMATION

The rules for structuring are inherent in the hierarchical tree for structuring items of semantic information. These rules represent a part of the control data (Section IV). The semantic information refers to various features (or phenomena) in the terrain relief. The structure of key-items refers to regions and networks, which can be further differentiated (figure 2). Subdivision should be sufficiently detailed to permit comprehensive modelling of terrain relief.

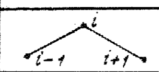
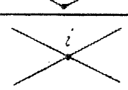
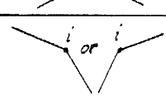
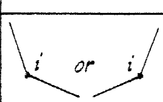
Description / Break point	Code	Definition	Property
Peak		$h_i-h_{i+1}>0$ and $h_i-h_{i-1}>0$	Local maximum
Pit		$h_i-h_{i+1}<0$ and $h_i-h_{i-1}<0$	Local minimum
Pass		As for peak in one direction and for pit in the perpendicular direction	Min. & max. in two perpendicular directions
Convex points		$h_{i-1}-h_i>h_i-h_{i+1}\begin{cases}h_{i-1}-h_i\geq 0;\ h_i-h_{i+1}>0 \text{ or}\\ h_{i-1}-h_i<0;\ h_i-h_{i+1}\leq 0\end{cases}$	Up-down transition from gentle to steep slope
Concave points		$h_{i-1}-h_i>h_i-h_{i+1}\begin{cases}h_{i-1}-h_i>0;\ h_i-h_{i+1}\geq 0 \text{ or}\\ h_{i-1}-h_i\leq 0;\ h_i-h_{i+1}<0\end{cases}$	Up-down transition from steep to gentle slope

Table 1: Description of distinct points

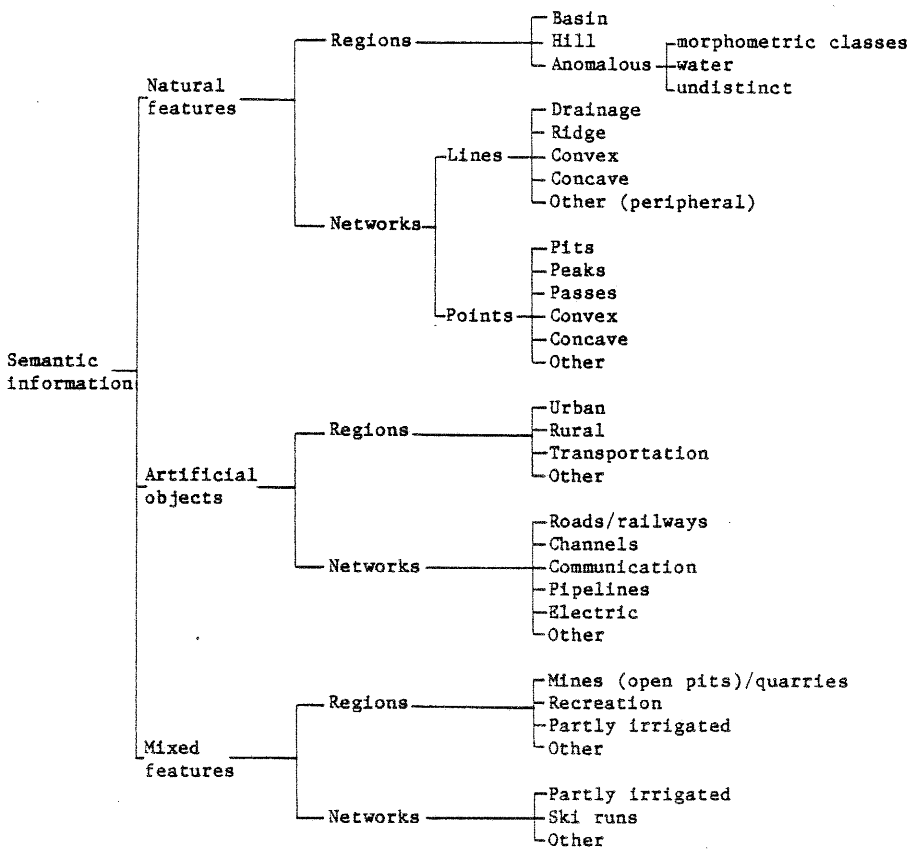


Figure 2: Structural scheme for semantic information (key-items)

The scheme in figure 2 is general and thus also applicable to other features in planimetric mapping. Thus feature extraction and sampling (measurement) for planimetric mapping and DTMs can be carried out together; thereby avoiding duplication.

Boundaries between less-distinct regions can be fluid and therefore need not be located accurately.

Attributes of the key-items differ widely and will therefore not be considered further here.
The basic entities of Σ-data are regions, lines and points.
Descriptions of these entities are summarised in tables 1, 2 and 3.

Line type \ Description	Code	Definition	End-nodes
Ridge line	r	Line segment/string of $\wedge$-points	Passes or convex points where $+\delta h \gg Th$
Drainage line	d	Line segment/string of $\vee$-points	Passes or concave points where $\lvert -\delta h \rvert \gg Th$
Convex line	n	Line segment/string of ⌝ or ⌜ -points	$\wedge$-points or convex points where $+\delta h \gg Th$
Concave line	u	Line segment/string of ∟ or ⌟ -points	$\vee$-points or concave points where $\lvert -\delta h \rvert \gg Th$
Other lines	o	Real or ficticious boundaries of regions	

Table 2: Description of lines

Region type \ Description	Code	Definition	End-nodes
Hill region	H	Contiguous surface - prevailing convex shape	Bounded by strings of local minima
Basin region	B	Contiguous surface - prevailing concave shape	Bounded by strings of local maxima
Other regions	A	Real or ficticious surfaces	To be excluded or handled differently.

Table 3: Description of regions

Break-point and break-line are defined in the annex.

III. BASIC METRIC INFORMATION

The structure concerns metric contituents of information which should be defined before sampling. The metric key-items refer to location in space and other information derived from (or related to) locations. Time is not involved explicitely.
The key-items (locational data) can be supplemented by description of the neighbourhood, units, etc. Such descriptions can be regarded as attributes. The rules for structuring basic metric information (structural scheme) are a part of the corresponding control data (Section V). Figure 3 shows a simplified scheme for structuring basic metric information - in the context of Selective Sampling for DTM.

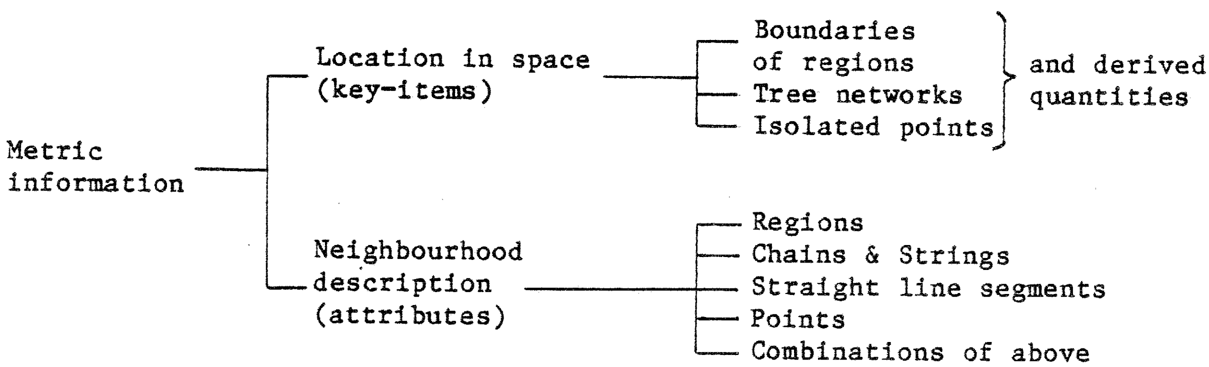


Figure 3: Structural scheme for basic metric information

Some neighbourhood-relations are summarised in tables 4, 5, 6 and 7.
The relations among data entities can be encoded and/or the entities can be mutually linked by pointers.

Region \ Region	Basin	Hill	Anomalous
Basin B	neighbours	duals	-
Hill H	duals	neighbours	-
Anomalous A	unrelated, neighbours, or nested		

Table 4: Relations between regions

Network \ Network	Drainage	Ridge	Other
Drainage d,u	neighbours	complementary	-
Ridge r,n	complementary	neighbours	-
Other o	miscellaneous		

Table 5: Relations between networks

Region \ Network	Drainage d,u	Ridge r,n	Other o
Basin B	inside	bounding	miscellaneous
Hill H	bounding	inside	-"-
Anomalous A	-	-	bounding

Table 6: Relations between regions and networks

Line \ Point	Pit and concave point	Peak and convex point	Pass
Drainage d,u	bounding	non	bounding
Ridge r,n	non	bounding	bounding
Other o	weak or no relation		

Table 7: Relations between lines and points

IV. CONTROL DATA FOR SEMANTIC DOMAIN

These data contain specifications, structural data (data entities, structural schemes), etc., for the basic semantic information (input and output) and for processes in the semantic domain. The control data can be structured according to the scheme in figure 4; they can be further differentiated. Strategy, procedures and parameters need be defined for all process stages, including evaluation of output.

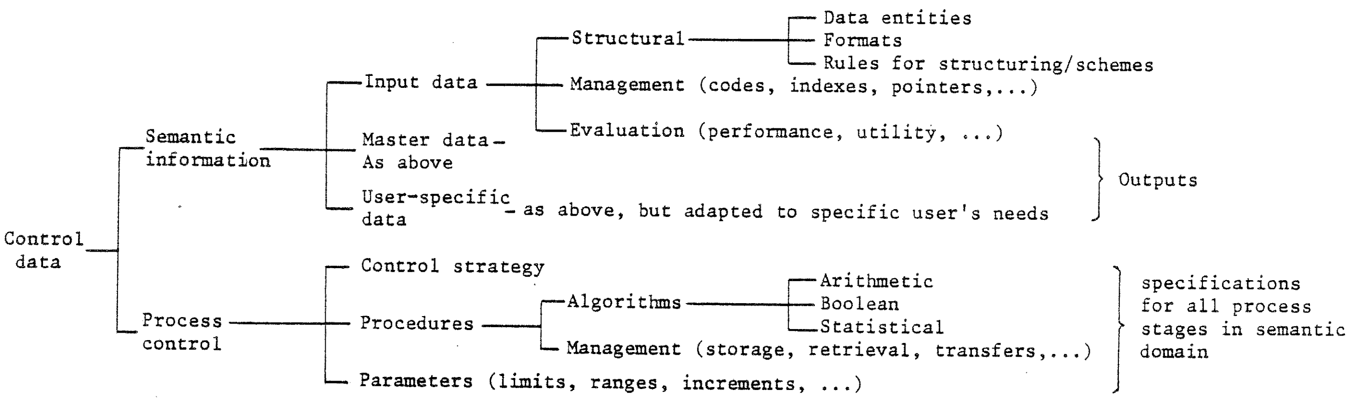


Figure 4: Structural scheme for control data in semantic domain

Distinct morphometric features should be structured at the extraction stage. Since manual extraction does not rely on algorithms, the procedures need be systemised to permit creation of structured data sets at sampling.

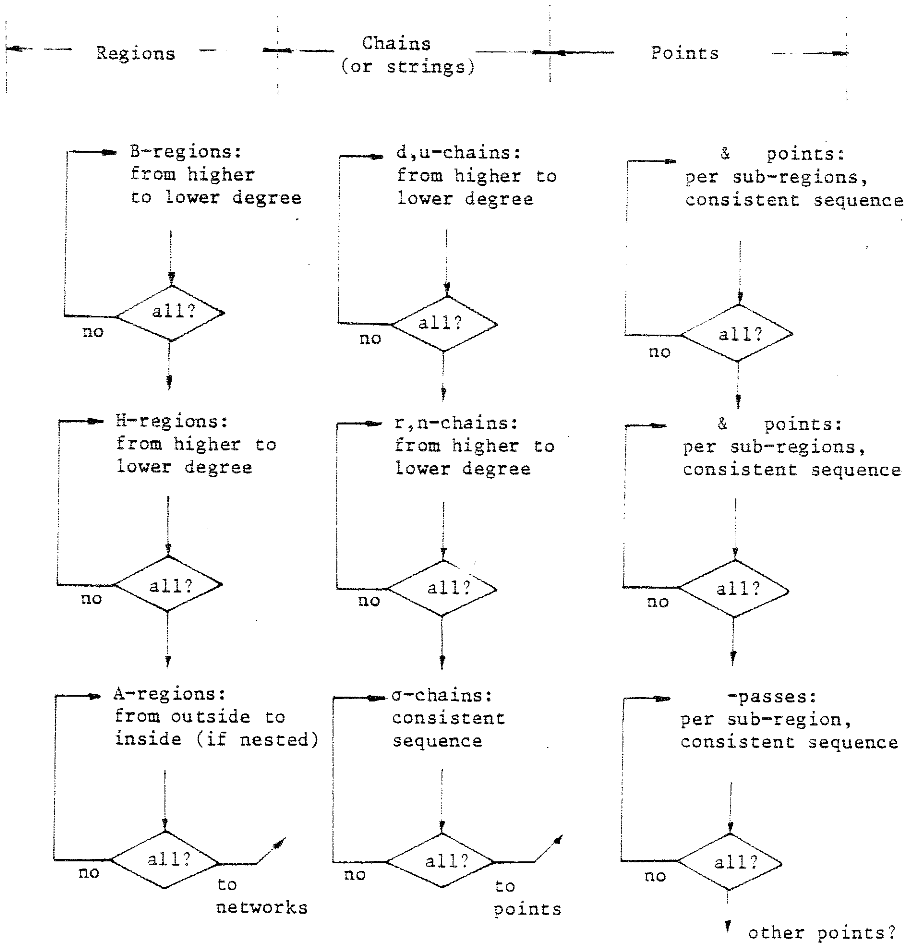


Fig. 5: Indexing and sequence of operation

In the following, the extraction procedures are outlined for regions, networks, and discrete points. Some general guidelines are "from regions to networks", "from large to small" and "consistent sequence". Coding and indexing of data entities should reflect these guidelines (figure 5).

1. Regions - segmentation and structuring

Regions can be extracted segmented and structured hierarchically. The corresponding coding and indexing should be orderly, i.e, from large to small entities, and in each hierarchical level the sequence should be consistent (e.g. clockwise, as shown in figure 6).

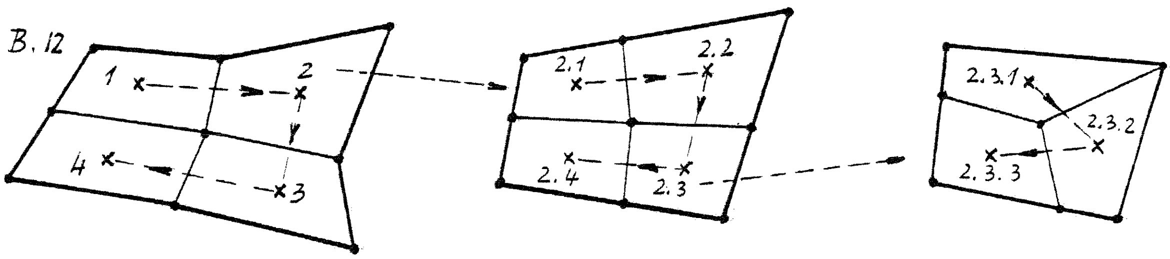


Figure 6: Sequence of extracting regions (3 levels)

The corresponding code(s) and indexes can be tabulated (Table 8); such a table indicates topological relationships.

Region type & no.	B12								
2nd degree regions	1	2						3	4
3rd degree regions	...	2.1	2.2	2.3			2.4	...	...
4th degree regions	...	...	...	2.3.1	2.3.2	2.3.3	...	...	...

Table 8: Structure of regions

Control data for regions also contain the rules for segmenting and structuring. A first degree basin region (B) is bounded by primary (first degree) ridge lines and their extensions of maxima ("pseudo" ridge-lines). For segmenting a B-region, secondary ridge lines need to be defined and extended to the nearest junctions of drainage lines or the nearest passes (figure 7). Further segmentation (into sub-sub-regions) is analogous.

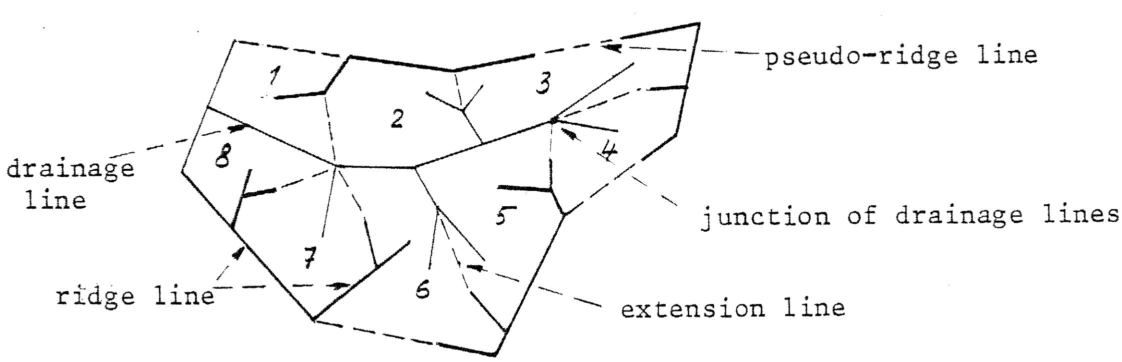


Figure 7: Segmentation of a region

Lay-out of the extension lines is subjective - there is some freedom in segmenting regions into comparable entities. the Sub-regions obtained are bounded by ridge lines, their extensions, and by drainage (or shore) lines of the same and higher degree.

A first degree hill region (H) is delimited by primary (first degree drainage lines) and their extensions of minima ("pseudo" drainage lines). Segmenting a hill region is the same as a basin region, but the roles of drainage and ridge lines are reversed. Thus extensions of drainage lines are connected to the nearest junctions of ridge lines or to the nearest passes. For segmenting regions in the context of CS, the extension lines need not be located accurately.

Anomalous regions (A) are bounded by their "peripheral" lines. These lines need not be located accurately except where they coincide with distinct discontinuities in terrain slope or with planimetric features (e.g., shore line). Peripheral lines serve two purposes. The first is to exclude anomalous regions from (further) sampling, and the second is to provide a datum for subsequent Progressive Sampling. In the latter, it is possible to prevent omissions of locally significant variations in terrain relief. Hence inclusion of peripheral lines increases the integrity of DTM data.

2. Networks - segmentation and structuring

Networks and regions are interrelated (table 6). Networks are composed of chains, and chains of line segments or continuous strings of points. They can be extracted and sampled systematically, e.g., according to lay-out of the regions containing networks. The following procedure can be used:

- Delimit, encode and index the regions containing networks (figure 8: hill region H.3).
- Identify, encode and index the main network, its main (1st degree) chain (stem r.3) and its main branches (2nd degree: 3.1, ..., 3.4). Proceed gradually to lower degree branches (3.1.1, ..., 3.4.2) of the same tree.
- Initiate extraction (and index), e.g., at the most north west end of the main ridge chain (or at the highest point of the main drainage chain), and continue towards its opposite end.
- Extract branches, e.g., from "inside (stem) to outside", first RH (odd indexes) then LH (even indexes) branches. Continue with the lower degree branches.
- Proceed similarly with lower degree entities (networks, chains, line segments, points) inside the region.

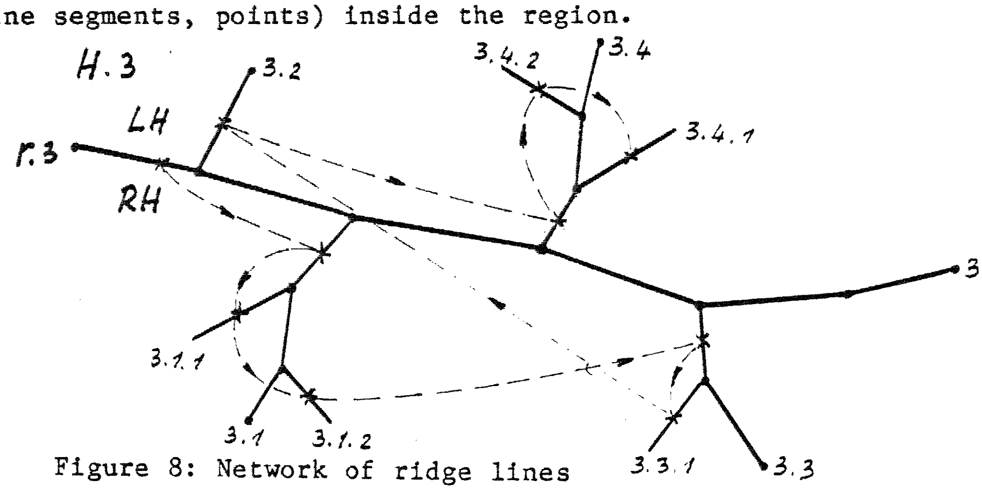


Figure 8: Network of ridge lines

In figure 8, the sequence is indicated by the stripped line. The corresponding code and indexes are presented in table 9.

Region	H.3					
1st degree: stem	r.3					
2nd degree branches	3.1		3.3	3.2	3.4	
3rd degree branches	3.1.1	3.1.2	3.3.1		3.4.1	3.4.2
4th degree branches	...			...		

Table 9: Structure of a ridge network

This table indicates topological relationships.
The sequence of indexing data entities in a network and their extraction should be consistent; this is represented symbolically in figure 9.

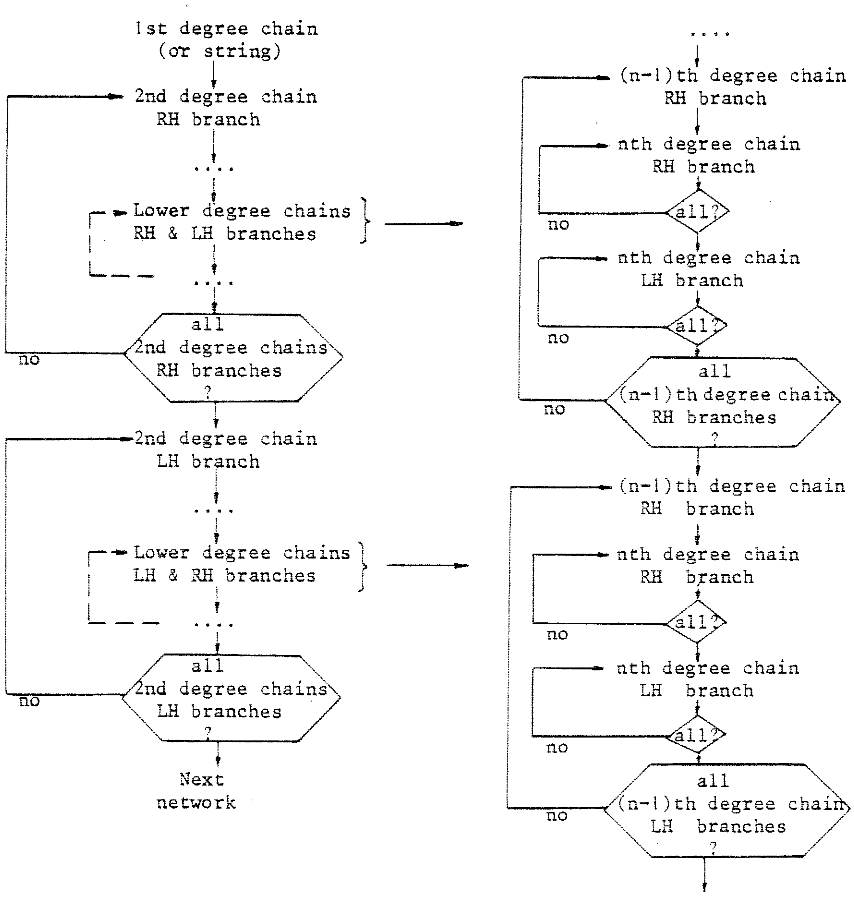


Fig. 9: Indexing data entities in a network and sequence of extraction

Consistency improves the integrity of the data, permits partitioning the work into smaller units, and allows easier inspection.
Distinct morphometric features usually do not form complete networks. Disconnected parts, isolated chains, line segments and/or points often occur. They can be entered in the hierarchical structure (table 9) at suitable levels, and do not necessarily have predecessors at higher levels and/or successors at lower levels.

V. CONTROL DATA FOR METRIC DOMAIN

These data contain specifications and some other information about the basic metric data (input and output; section III) and about the corresponding processes. A simplified structural scheme for such control data is shown in figure 10.

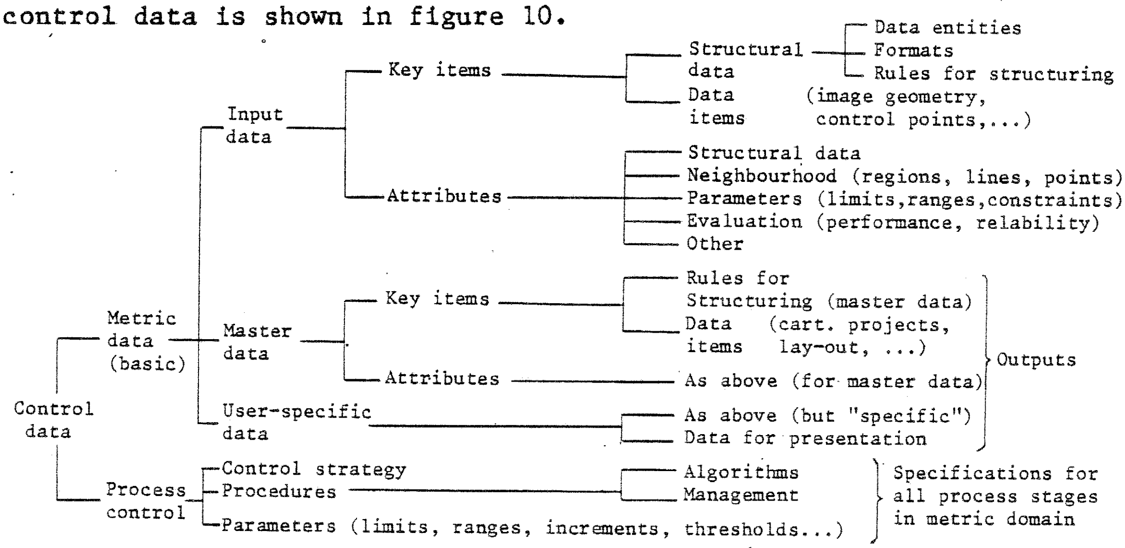


Fig. 10: Structural scheme for control data in metric domain

The categories of specifications for input and output are similar, but, obviously, the actual data differ. Data for process control refer to the strategy, algorithms and corresponding data management in all process stages, i.e., extraction, sampling, processing, presentation and evaluation of results.

Some metric data have a double role; although they are essentially "basic", they can also provide some process control. Examples are pre-positioning for measurement and tracking control at profiling in photogrammetric plotters, selective retrieval or exclusion of data at processing, windowing and positioning at display, etc.

VI. OPERATIONAL ASPECTS

In principle, extracting distinct features and delimiting anomalous regions, and the corresponding segmentation, coding and indexing, should be carried out in a separate preparation stage, before sampling. An "index chart" should be prepared as a guide for the operator for sampling. When the amount of Σ-data is small, however, extraction and sampling can be combined in one stage.

In the context of Composite Sampling, three types of regions to be delimited from the more homogeneous area can be differentiated. These are:

1. Regions to be excluded from sampling; nested regions of different degrees, e.g., lakes with islands with lakes, are special cases.
2. Regions to be fully sampled selectively, i.e., covered merely by Σ-data.
3. Regions in which Selective Sampling (of Σ-data) supplements PS (or vice-versa).

In cases 1 and 2, such regions (to be excluded from PS) need to be delimited and coded accordingly.
For CS, case 3 is most pertinent. Regions and their structures serve two purposes:

- to guide the operator for systematic extraction and sampling of Σ-data;
- to provide the data required for a proper response at PS to local changes in terrain relief; e.g., a peak of a hill in a plane needs to be supplemented by its ("peripheral") foot-line.

Σ-data can form an autonomous (sub-)set of the DTM, to be preserved for further uses. Such information should be provided with codes, indexes and topological data (Section III). One potential application is generalisation of Σ-data for representation at smaller scales, where regions and branches of lower degree(s) of network can be omitted.
When Σ-data serve merely to supplement CS, complete indexing and topological description are not necessary. Indexing can be partly replaced by a consistent sequence of sampling Σ-data, and the reserve: if Σ-data are indexed the sequence need not be (strictly) maintained. In the latter case, however, there is a greater risk of disregarding significant features.

CONCLUSION

A balance between Selective and Progressive Sampling enhances the integrity of the resulting DTM and permits efficient processing. More experience is needed, however, to gain insight in how a balance can be established. Good understanding of PS is a prerequisite, however. Manual extraction and sampling of distinct morphometric features for a DTM needs be systemized to be effective. The process should be time-efficient, the data structure functional, and the information sufficiently accurate. Integrity of the Σ-data should be tuned to properties of Progressive Sampling. The latter tends to compensate considerably for omissions at Selective Sampling.

The information can be structured according to four main entries, i.e., the basic information and control data, and further according to the semantic and metric domains. Such a conceptual differentiation is helpful for structuring and analysis, despite the fact that in practice some processes involve both domains. The basic information is problem (or user) oriented, whereas the control data are solution (or information producer) oriented. Nevertheless, some basic metric information can also provide process control, i.e., it has a double role.

Information should be structured during the feature extraction stage, when coding and indexing items. Some basic guidelines for extracting data entities are: "first regions then networks", "from coarse to fine" and "consistent sequence". Codes and indexes of data entities can be tabulated; such tables imply topological description. The sampling sequence should preferably comply with the system of coding and indexing. This may contribute to the integrity of the Σ-data, speed of operation and to ease of inspection. Sampling should preferably be interactive; stereo-superposition of sampled data on an image-pair would facilitate effective inspection and editing.

REFERENCES

[1] Makarovič, B.: "Composite Sampling for DTMs", ITC Journal 1977-3, Enschede, Netherlands.

[2] Makarovič, B.: "From Progressive to Composite Sampling for DTMs", Geo-Processing, 1, 1979, Amsterdam, Netherlands.

ANNEX

Definition of break-point and break-line

A quantitative definition of a break-point (discontinuity in slope) in terrain relief can be based on a symmetric triplet of points (fig. 11a). The midpoint k is located at the anticipated break and the two wing-points k-1 and k+1 perpendicular to the tangent at k of the break-line. The interval $\Delta\ell$ can be a multiple (e.g. n=3) of the standard planimetric error s of a map series ($\Delta\ell$=n.s) or $\Delta\ell$ equals the spacing of the finest grid used at PS. In the latter case, the triplet (k-1, k, k+1) is replaced by a group of five points, centered on the anticipated break-point (fig. 11b).

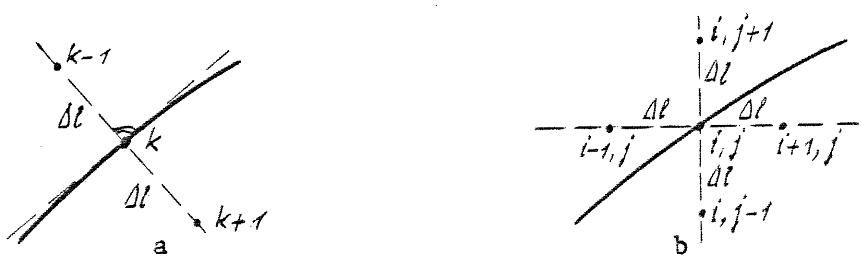


Figure 11: Definition of break-point

After measuring heights of these points, the second difference can be calculated: $\delta h = h_{k+1} + h_{k-1} - 2h_k$;

or $\delta h_x = h_{i+1,j} + h_{i-1,j} - 2h_{i,j}$; $\delta h_Y = h_{i,j+1} + h_{i,j-1} - 2h_{i,j}$

and $\delta h = \sqrt{\delta h_x^2 + \delta h_Y^2}$

By determining a suitable threshold Th, a break-point is defined by $|\delta h| \geq Th$. A break-line is a continuous string of break-points.
Such a quantitative measure, however, should be applied in practice only in doubtful situations and/or for training purposes.