GIS-based land-use-suitability mapping: cognitive processes and designing instructions that lead to expertise

R. Balzarini,[*,†] M. Ney,[*] and P-A.Davoine[*]

[*]Laboratoire d'Informatique de Grenoble, 681 rue de la Passerelle 38400 Saint Martin d'Hères, France
[†]C. S. A.Moiroux, 64 chemin des Mouilles 69134 Ecully, France

Abstract. To enhance the learning and teaching of geographic information systems in higher-education Earth sciences, we present ongoing research that aims to identify the strategies and concepts that underlie the suitability-mapping process. Linking quantitative and qualitative data analysis, our research clarifies the differences between experts and students and suggests instructions to help students acquire problem-solving expertise.

Keywords: cognitive processes, verbal data analysis, suitability mapping

1. Introduction

In the geosciences, geographic information systems (GISs) are especially useful in two areas: digitization and suitability-vulnerability mapping. Land-use-suitability problems are often solved with multicriteria spatial analysis (MCSA) methods. These methods combine various criteria to obtain a map that indicates areas that are more or less apt to solve a location problem. The decision analyst must make various choices that lead to a host of cartographic issues: the standardization method for each criterion, criterion layers, decision rules, and weights for relative importance (Malczewski 2004).

Making suitability maps implies applying a cognitive process that involves the representation, interpretation, and (mathematical) treatment of geographic data and requires a deep knowledge of GIS.

From an educational point of view, our purpose is to examine which cognitive processes underlie the visual thinking, strategies, and cartographic skills required by MCSA tasks and how these processes evolve in going from a novice to an expert. Understanding the transition from novice to expert is a prerequisite for developing effective learning environments for students at all levels (Petcovic & Libarkin 2007).

This research is based on a quantitative-based qualitative approach to analyzing verbal data. The main goal is to understand how to represent the knowledge used in the pertinent cognitive processes. In quantifying the qualitative data, the researcher searches the data for patterns and trends and then categorizes these according to codes or concept indicators (Chi 1997).

In this paper we first present the methodology adopted; then we show the preliminary results and the didactical scaffolding that help students acquire expertise in suitability mapping.

2. Background

2.1. Suitability land-use mapping: semantics and cognition

The process of making suitability maps involves overcoming conceptual obstacles, which we summarize here.

We first consider the mathematical procedures (i.e., the GIS-MCSA approaches), including weighted summation and Boolean operations. These are easy to implement within the GIS environment by using map-algebra operations and cartographic modeling. The principle of the method is also easy to understand and intuitively appealing to decision makers. However, GIS implementations of the weighted-summation procedures are often used without a full understanding of the assumptions underlying this approach. In addition, the method is often applied without complete insight into the meanings of two critical elements of the weighted-summation model: the weights assigned to attribute maps and the procedures for deriving commensurate attribute maps (Malcewski 2006).

For perceptual organization and categorization related to functional-representation concepts of map syntactic, we rely on Tversky and Hemenay's basic-level's categories theory (1984). These basic-level categories, including events as well as objects, are categories for which we can form a single image. Categorization is strongly linked to data classification for choropleth maps (Mac Eachren 1995). The map resulting from MCSA is a choropleth: it displays quantities (plethos) relative to areas (Khore) via a graduated color scale. There are both mathematical and graphical issues to solve when making choropleth maps. Implementing such a map depends primarily on the choice of discretization method; that is, how to divide the statistical series to map into classes or intervals (Béguin & Pumain 2007). In this map, all map units falling into a particular category are depicted with identical symbols; this perspective is based on the acceptance that, in classical categories, any element must be as representative of the category as any other element. The cartographic concept of a choropleth map is based on the classical theory of categorization being "correct." The understanding derived from choropleth maps depends on how categories are interpreted by users (MacEachren 1995).

To interpret and use suitability maps, we rely on Marr's (1985) visual map processing and MacEachren's (1995) perceptual organization, categorization, and judgment. Human vision is good at extracting shapes from a visual scene, assessing depths and relative size, and noticing movement. A key feature is that the visual system should emphasize contrast more than absolute illumination and higher acuity for color hues than for color value. Thus color value and saturation can be ordered whereas hue cannot. A second key is the system's ability to group the elements that neurological image processing renders into "objects."

Continuing in terms of interpretation, another feature of land-use-suitability mapping is that it must be "checked" repeatedly throughout the process: its validity is determined by its correlation with the representation of the actual terrain. Indeed, suitability maps are often superimposed

(draped) over base maps, 3D model of terrain, virtual globes, or maps with administrative boundaries. Our study is particularly concerned with superpositions with relief, which are created by shading on DEM. Their interpretation and reading is then linked to the perception of depth of a two-dimensional (2D) or three-dimensional (3D) scene: for that we rely on the taxonomy of depth cues provided by Kraak (1988) and particularly on the "pictorial" cues that are related to the object's structure and the way the structure organizes visual input. By using shading and/or color and shadow, many cartographers create an effective plan-view relief representation that suggests depth in a non-perspective approach.

One last specificity of suitability maps concerns their use for decision making. As Jankowski and Nyerges (2001) showed, maps play only a limited support role in various stages of the decision process and a reduction of cognitive complexity is needed.

These theoretical aspects are treated in our approach, which strives to highlight the key differences between experts and novices in interpreting a suitability map.

2.2 Expert-novice continuum: some cognitive models

The cognitive sciences have a rich research tradition that has examined expertise across a variety of fields, compared the characteristics of experts and novices, and considered how expertise is acquired. Abstract thinking skills, problem-solving strategies, storage and recall of a wide array of information, and ability to work flexibly within a domain of knowledge all exemplify what it means to be an expert (Petcovic & Libarkin 2007).

We rely on MacEachren's map schemata as structures for representing and organizing concepts that link together cognitive processing of map-derived information, the roles of knowledge, experience, practice, and training on the part of map readers (MacEachren 1995). There are significant differences in schemata available to domain specialist versus novices, so applying appropriate schemata requires learning and practice (MacEachren 1995).

In a map-reading task where the map corresponds to a known landscape, experts might be expected to encode it not just as separate "chunks," but within an overall template that incorporates the relationships between the groups of objects viewed (Kent & Chang 2008). For the experienced map reader, an extensive vocabulary exists that defines and labels complex entities (Edwards, in MacEachren 1995).

Specifically, experts may be (1) focusing on the distinctive features of a display to establish how it may differ from the norm, (2) identifying what is familiar and typical and that therefore requires minimal processing, and (3) spatial-feature matching either the geometric or symbolic information on the map with geographic-feature matching in the landscape being represented (Chang et al. 1985). Another main strategy for approaching spatial-problem solving was highlighted by Crampton (in MacEachren 1995): experts used active self-analysis and error-prevention with progressive repetition to focus in on a solution.

The study of expertise can be used to improve instruction to develop the visualization skills and schemas necessary for parsing complex spatial information.

3. Methodology

3.1. Goals of current study

Conceptualization and visualization of suitability maps involves a complex interaction of skills and mental schemas. The purpose of this study is to assess the concepts, actions, and strategies that underlie the mental processes in a suitability-mapping task. Specifically, our study addresses one main question: What operations and concepts are used by experts and students when solving suitability-map problems?
We used the design-based-research approach (Edelson 2002) to design experiments to shed light on this research question. This approach employs multiple methods allowing cycles between theories or models and field observations and experiments, and it includes descriptive data representations as well as quantitative and qualitative analysis.

3.2. Participants

We conducted our survey with two third-year-undergraduate student populations, each well distinct and composed of GIS and MCSA novices. The students were majoring in geoscience (i.e., geologists); 15 were engineers specializing in environmental science. The students worked in groups, we analyzed three groups of geologists and three groups of engineers.
Three specialists participated in the experiment as experts: a geophysicist, a cartographer, and a GIS engineer. They were questioned individually at their work place and were asked to complete a predesigned activity ("contrived tasks"). The approach used by an expert to perform a task gives invaluable insight into how experts reason. Both the experts and the students were given the same suitability map task.

3.3. Learning sequence: site location problem

We designed a GIS learning sequence in which student subjects were given a typical site-location problem to solve. To favor authentic learning (Simon 1962), students were asked to work in genuine professional situations. They had to perform a feasibility study and locate suitable sites for a ski resort in the southern French Alps. The approach used to solve the problem is based on multicriterion methods and requires tools for spatial analysis. More precisely, it involves analyzing environmental criteria, making surface analysis calculations, reclassifying data, and creating suitability models. The GIS used was ArcGIS 10.0. The sequence took 17 hours and was divided into five sessions of three hours and one session of two hours (S6) over a period of two weeks.

3.4. Data Collection

To understand how an apprentice learns a trade, one possibility is to observe the learner in context. We collected three types of data from the learning sequence:

- *productions*: maps made by students or experts;
- *verbal data*: students' verbal interactions during key moments of decision making (audio and video records) and experts' "thinking-aloud" protocols (video records);
- three open-ended *questionnaires*.

We present herein an analysis of the verbal data, which were transcribed. For each expert, we analyzed three hours of recordings; for each of the six student groups we analyzed 2.5 hours of recordings.

3.5. Verbal-Data Analysis

Our approach relies on Chi's (1997) method of quantifying a qualitative analysis of verbal data.

Verbal analysis is a methodology for quantifying the subjective or qualitative coding of the *contents* of verbal utterances. Instead of representing the ideal knowledge, the goal of the method here is to understand what a learner knows and how that knowledge influences the way the learner reasons and solves problems, be it correct or incorrect. The method of coding and analyzing verbal data consists of the following eight steps:

1. Reduce or sample the protocols; that is, reduce data by selecting a particular activity;
2. Segment the reduced or sampled protocols;
3. Develop or choose a coding scheme (taxonomic categories scheme);
4. Operationalize evidence in the coded protocols;
5. Depict the mapped formalism;
6. Seek pattern(s) in the mapped formalism;
7. Interpret the pattern(s) and its (their) validity;
8. Create interrater reliability.

For all the processes that were applied to resolving the problem of site location, we retained four fundamental tasks:

T0 = Data standardization and classification
T1 = Weighted Sum
T2= Map Analysis
T3= Site Choice

Once the corpus to be coded was decided, we then had to segment the verbal utterances to identify the unit of analysis. The defining cut can occur at many points, revealing units of varying granularity, such as a proposition, a sentence, an idea, an interchange as in conversational dialogue, or an episode.

We developed our formalism in an interactive bottom-up and top-down process: categories were derived from the subjects' explanations and interactions and were enhanced by some theoretical background. We defined five categories of "objects": algebraic, color semiotics, orographic, geographic, and analytic concepts.

In the segment protocols we searched for tags or keywords, such as "mountain" or "valley" for the orographic object's category and "ski resorts" or "names of localities" for the geographic object's category.

Two coders worked on the data (the first author of this paper and an independent coder), each making a complete analysis of the data. Each disagreement between the coders was considered and reanalyzed in the segment protocol. The rate of agreement between the coders was 87.8%.

The analysis is based on a taxonomy founded on the characterization of tasks, actions, operations, and conceptual objects: tasks (*Figure 1*) are divided into actions, each action is characterized by one or more operations, and each operation is characterized by one or more conceptual objects.

Figure 1 presents an example taken from the segmentation of task T0: action 1 has one operation (1.1) and one object (A1), action 2 has two operations (2.1 and 2.2), operation 2.1 has an object (A1), and operation 2.2 has three objects (A1, R1, G1).

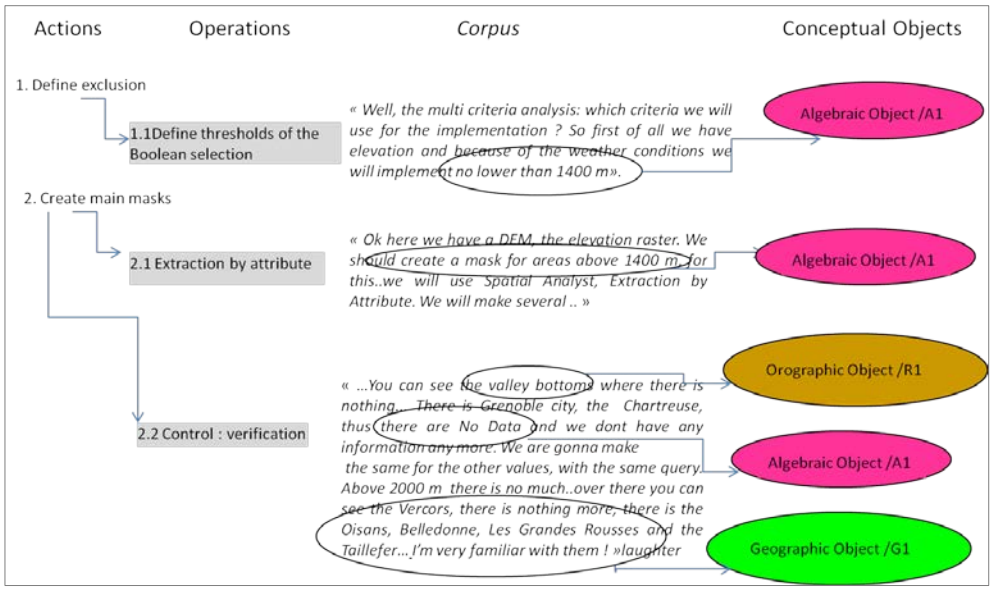

Figure 1. Example of segment protocols from sampled corpus.

4. Preliminary results from qualitative analysis

A Taxonomic categories scheme

- Objects

We established a set of codes that fit a taxonomic categories scheme specific to land-use suitable mapping.
Our taxonomy is composed of five categories, each representing a conceptual object. Each object is defined by descriptors that are assigned a code. Specifically,

1. The algebraic object, which encompasses all the verbal elements pertinent for rendering the Geographic Information data in mathematical form. This object has six descriptors, which are identified in *Table 1*.

Codes	Descriptors	Example quotes
A1	exclusion, selection, Boolean	*"At first look, we can exclude certain things, make a mask: let's say the slope is between 15° and 60° and the altitude is between 1000 and 3000 m" [e1]*
A2	Thresholds, number and order of classes, digitization, synthesis	*"What scale for the classes, I don't remember any more … 1, 2, 3? I'll make a hierarchy" [e1]*
A4	Hierarchy of criteria, representation of weights, correlation between criteria	*"So altitude is important, 3, slope as well, 3, distance, 2, orientation, 1. I'll do it in percent, it'll be easier to interpret" [e1]*
A5	Weighted sum, combinations	*"Try one with altitude 40 (A) plus 30 (P) plus 20 (E) and plus 10 (CLC)" [g1]*
A6	Measurements (areas, perimeters, pixels)	*"We could chose with respect to the area, define a minimum number of pixels, then vectorize to calculate the area" [e1]*
A7	Vector intersections	*"To cross check the data we'll use intersection tools" [e2]*

Table 1. Algebraic objet and its descriptors.

2. The color object, which encompasses all the verbal elements related to the variable of visual color. This object has five variations (see *Table 2*)

Codes	Descriptors	Example quotes
V1	Hue and values	*"The colors are good because red means high, strong ; green means weak" [e2]*
V2	Contrast, disassociation, and degradation	*"Which map is the clearest? There it's the clearest, a lot of blue jumps out, the other is more fuzzy but there is less blue" [e3]*
V3	Color harmony (the palette)	*"To rasterize it SP1 I chose the blue-yellow-red palette" [e3]*
V4	Edges, boundaries	*"You have to clearly differentiate between light and dark red" [g4]*
V5	Presentation and transparence	*"You can act at the level of the transparencies, but transparencies on black and white … whistle" [g3]*

Table 2. Color semiotics objet and its descriptors.

3. The orographic object, which encompasses all the verbal elements related to the geomorphological description of the terrain. This object has five variations (see *Table 3*).

Codes	Descriptors	Example quotes
R1	All references to knowledge of the terrain (alpine)	*" The Alpes du Grand Serre has 10 times less snow than the Col de Porte, which is buried" [g1]*
R2	Common nomenclature: mountain, lowlands, valley, slope	*"We have nothing in the mountains and everything in the lowlands" [g2]*
R3	Specialized nomenclature : ridge, thalweg, cone	*"I don't think it would be good to cross a red zone, that means you're crossing a thalweg; a valley" [g3]*
R4	Representations of relief: contours or DEM	*"Don't we have a contour map?" [g1]* *"I'm going to do a shading to see the relief" [e1]*
R5	Lithology, cliffs	*"It (the calculation) removes cliffs, but look at Saint-Eynard, it leaves it (in green) and it's too steep" [e1]*

Table 3. Orographic objet and its descriptors

4. The geographic-feature object, which encompasses all the expressions related to the geographic elements of the region being analyzed. This object has six variations (see *Table 4*).

Codes	Descriptors	Example quotes
G1	Known places, ski resorts, towns, communities	*"Below 2000 there's not much left, you can see there's nothing left in the Vercors; that leaves the Oisan, Belledonne, the Grandes Rousses, and the Taillefer" … laughter [e2]*
G2	Road network	*"Just look at the roads" [g2]*
G3	Drainage network	*"Because you're next to water there, you can put in pumps" [g5]*
G4	Representations: topographical map, background map	*"I need a map background, I'm using the Scan 100" [e2]*
G5	Green spaces: forests, protected areas	*"I've got to have the protected areas" [e2]*
G6	Infrastructure: ski lifts	*"You can't put your ski lifts like that all along your thing, and how do you go down, you need things that converge toward the lifts" [g3]*

Table 4. Geographic feature objet and its values.

5. The analyzed-suitable areas object, which encompasses all the expressions related to observing one or more zones with the goal of evaluating capability. This evaluation consists of six types of observations (see *Table 5*).

Codes	Descriptors	Example quotes
An1	Suitable areas and correspondence between criteria, weights, and the goals of the problem	*"When the slopes are very steep then their weights have to be smaller" [e1]*
An2	Suitable areas and correspondence with the legend, colors, and values	*"We have the value 8 (legend) in white and 1 in black, low is not good and high is good" [g3]*
An3	Suitable areas and correspondence with the nature of the terrain	*"Bourg d'Oisans is there and the Deux Alpes … besides, look, it's there (zoom) and it's in green!" [e1]*
An4	Suitable areas and restriction character of the map	*"This map has a few less red zones, more precise, it's a little better, then there's this one that adds some zones, you go there the thing there it'll increase, so this one won't" [g5]*
An5	Suitable areas compared on different suitability maps	*"The other is fuzzier but it has less blue, it restricts the zones more and luckily the zones are the same (proper zones on both maps)" [e3]*
An6	Suitable areas and in need of more information	*"There we're on quite a few green zones, it's hard, we'd have to make a mask again with higher-value zones and then compare with the scree slope and the protected areas" [e1]*

Table 5. Analyzed-(suitable) area objet and its descriptors.

- Actions

To identify the solution path, we defined a formalism to capture the sequence of actions and operations that are used in the mapping process.
All the actions stemming from solution processes for experts and students were indexed. In all, 30 actions were identified; each characterized by one or more operations. Control operations were of particular interest to us, so we present them briefly below.

- Control operations

We categorized control errors based on Ohlsson's Theory of Learning from Error (1997). "Suitability mapping is a sequential-choice task characterized by sequentiality, multiplicity, and effect orientation. [...] What does it mean to commit an error in a sequential-choice task? It is an action that is not on the path to the intended goal. This action is inappropriate or useless in a specific context and it involves environmental effects, called error signal. Since knowledge guides the action, to correct an error is to improve future performance by revising the faulty knowledge structure."

We summarize hereunder the three control-error categories that we defined:

1. Verification: actions that validate or anticipate a choice; for example, by comparing the same zone on several maps or different zones that have the same type of terrain, or by verifying the algebraic results in the legend. Some examples of quote are "*I sum the two, that way that in 2 will be the good one" [e1]; "It's very sensitive to altitude there. It gives rather the change in altitude, which is artificial. The altitude is exaggerated; we should have put more classes in altitude to attenuate.*" [e1]

2. Diagnostic: actions that lead to recognition of sources of error; for example, the inappropriate use of a function or a request, the inappropriate definition of the threshold of classes, or the incorrect attribution of weighting to the criteria. Some examples of quotes are "*I made a mistake when I wrote ... wait, I'll do it one more time. I put parentheses around the expression ... I don't remember if it's AND or OR" [e1]; "there is a problem there, we still have the same thing so it doesn't mean anything! Stop! Everything is wrong! It's this one that's incorrectly classified!*" [g2]

3. Correction: actions allowing intervention to correct the error; for example, by modifying the function, the expression, the order of the classes, or the hierarchy of the criteria. Some examples of quotes are "*I remain persuaded that we have to redo the classification from 1 to 9 for each, because we're not working on the same values" [e3]; "We have to redo everything. Should we remove the orientation or this from the sum? Ok, that's much better!*" [g2]

5. Preliminary results from quantitative analysis

To highlight the patterns, the quantitative analysis included counting the instances of objects, control operations, and actions based on presence, absence, recurrence, or comparison between experts and students (See *Fig.2*). We discuss these patterns below based on the details of the operations undertaken (not shown here).

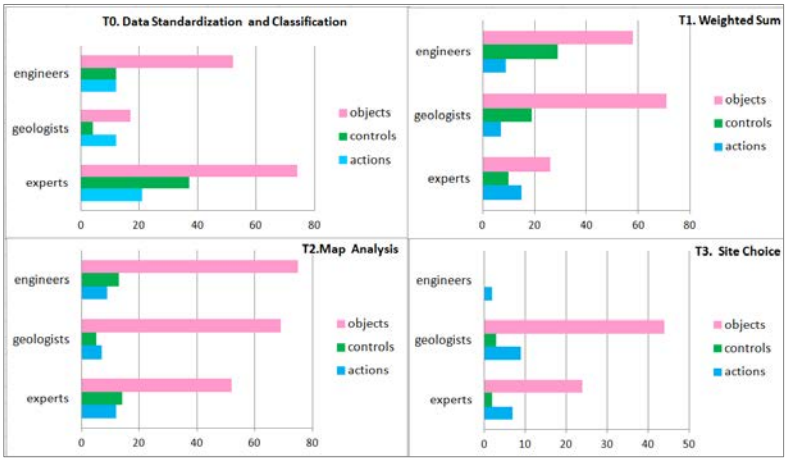

Figure 2. Summary of variable instances for the four main tasks.

For the phase where data is standardized and digitized (T0), we find in general that the experts use more actions and objects than the students. For example, the action of exclusion by thresholds (T01), which corresponds to creating masks and selecting which data to treat (chunking theory), is used by all three experts but by none of the students (T01Experts = 5 occurrences and T01Students = 0 occurrences). See *Fig. 3*.

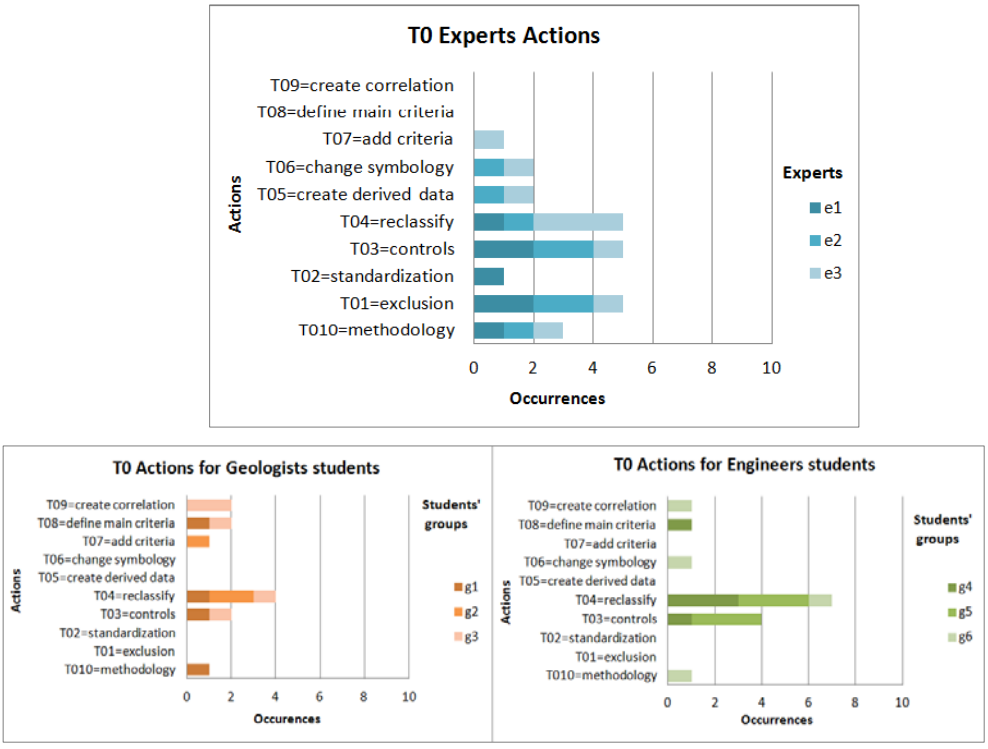

Figure 3. Occurrences of the 10 actions of task T0 for experts and for students

In addition, control operations (and in particular those of verification and diagnostic, CV = 11 and CD = 14, respectively) undertaken at this point by the experts are more important than those undertaken by the students (CExperts = 37, CGeologists = 4, and CEngineers = 12 for the engineers). See *Fig. 4*.

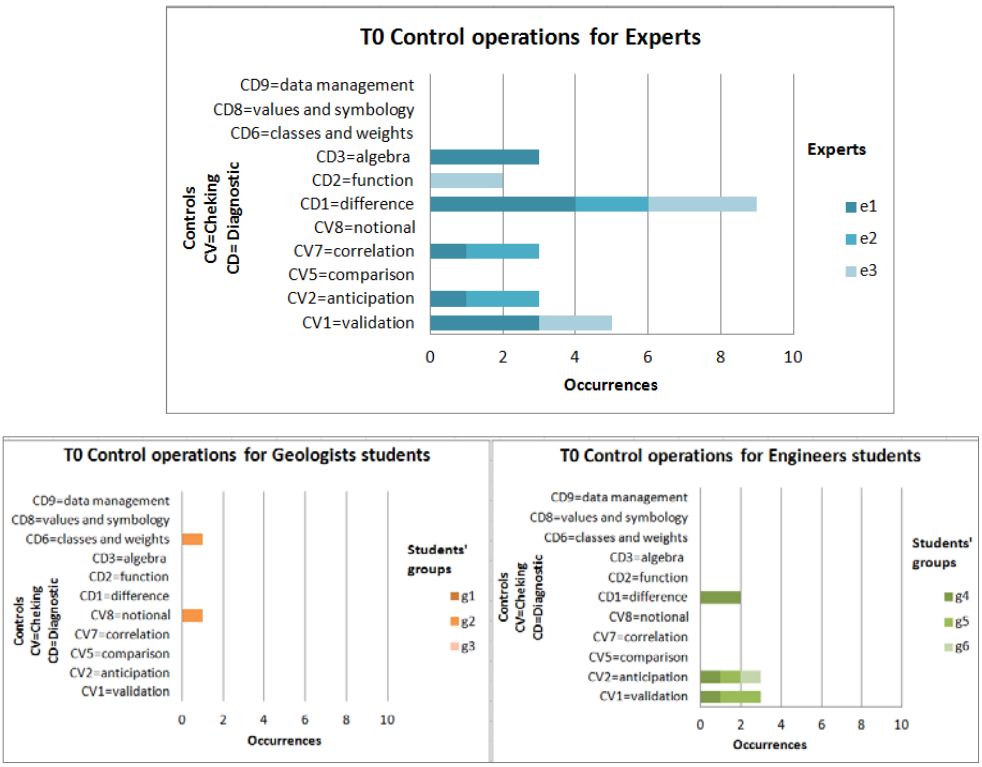

Figure 4. Occurrences of T0 control operations for experts and for students.

Thus, the experts put controls in place very early in the problem-solving process. In this phase of data preparation, the experts use more objects than do the students; in particular the algebraic object (AEx = 45, AGeol = 11, AEng = 39) and the geographic object (GEx = 9, GGeol = 0, GEng = 3). *See Fig. 5*.

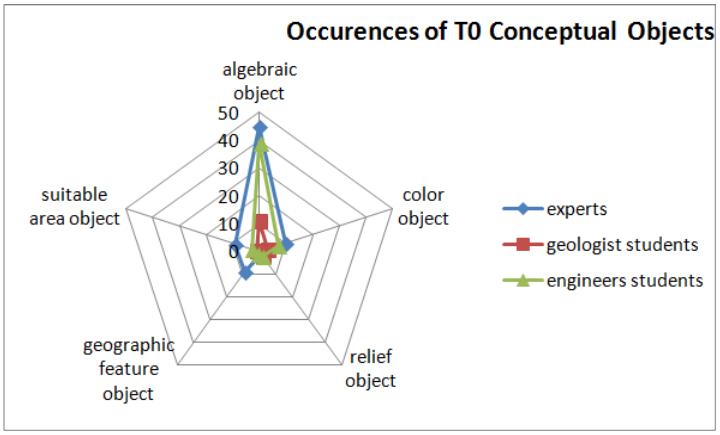

Figure 5. Occurrences of T0 conceptual objects for experts and for students.

In the phase where weights and weighted sums are allocated for the criteria (T1), the students undertake more control operations than the experts because they spend more time correcting mistakes.
In T2, we find that the experts use visual-effect tools to improve the analysis (zoom, transparence) (T22 Experts = 2, T22students = 0). They also use control actions such as verification with the relief and detection of threshold effects. These actions are less used by the students.
In T3, one or more sites are chosen by the experts with respect to controls. The choice is based especially on the addition of information and the detection of know zones (T32Experts = 3 and T32 students = 1).

6. Conclusions

The preliminary results of our study reveal the differences in strategies (i.e., the series of actions and operations) employed by experts and by students and track the paths taken to solve the problem. We extract from our data a model consisting of five categories of conceptual objects required to solve the problem. Finally, we proposed a model consisting of three categories of control operations. The differences between experts and novices are thus described at the level of concepts, operations, and controls.

The field of GIS education should consider how expertise is related to teaching and learning. Some operations or methods of execution as well as concepts that experts use and that may be absent in the strategies that students use suggest that a different type of pedagogical scaffolding should be provided. We suggest the following:

- methodological scaffolding: create raster mask as “partitioning”;
- thematic scaffolding: provide rules and notions in semiotics (use of colors);
- strategic scaffolding: suggest controlling for conflicts, errors, or anomalies (threshold effect);
- technical scaffolding: learn to manipulate visual-effect tools (relief, transparency, virtual globe) in GISs to enhance visual analysis.

We are currently testing some of these scaffoldings.
Strategies that help students acquire expertise in problem solving require explicitly teaching expert strategies, using of real-world problems, and organizing collaborative groups to encourage metacognition (Petcovic & Libarkin 2007).

References

Béguin M, Pumain D (2007) La représentation des données géographiques. Armand Colin, Paris

Chang KT, Lenzen T, Antes J (1985) The effect of experience on reading topographic relief information: analyses of performance and eye movements. The Cartographic Journal, 22:88–94

Chi THM (1997) Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. The Journal of the Learning Sciences 6(3):271–315

Edelson D (2002) Design research: What we learn when we engage in design. *Journal of the learning Sciences*, 11(1):105–121

Hmelo-Silver CE (2004) Problem-based learning: What and how do students learn? Educational Psychology Review, 16:235–266

Jankowski P, Nyerges T (2001) GIS-supported collaborative decision making: results of an experiment. Annals of the Association of American Geographer 91(1):48–70

Kent R, Chang P CH (2008).... Expertise in a Map Reading Task: The Role of Schemas in the Processing of Topographical Relief Information. Proceedings of the 30th Annual Meeting of the Cognitive Science Society, Washington

Kraak MJ (1988) Computer-assisted cartographical Three-Dimensional Imaging Techniques. Delft: Delft University Press

MacEachren AM (1995) How maps work: representation visualization and design. New York: Guilford Press

Malzcewski J (2004) GIS-based land-use suitability: a critical overview. Progress in Planning, 62:3–65

Malzcewski J (2006) GIS-based multi criteria decision analysis: a survey of the literature. International Journal of Geographic Information Science, 20(7): 703–726

Marr D (1985), Vision: The philosophy and the approach. Issues in Cognitive Modeling. London: Erlbaum

Newell A, Simon HA (1972) Human problem solving. Englewood Cliffs, NJ: Prentice Hall

Ohlsson S (1996) Learning from Performance Errors. Psychological Review, 241–262

Petcovic HL, Libarkin JC (2007) Research in Science Education: The Expert-Novice Continuum. Journal of Geoscience Education 55(4):333–339