Identification of landforms in digital elevation data

Nora Pacheco Blázquez
MSc Mathematics & Computer Science
Session 2010/2011

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

(Signature of student)_____

ABSTRACT

Vagueness is a widespread problem in the field of geography. Dealing with vague concepts can be very troublesome and challenging so techniques are required to handle it efficiently.

Landform features are inevitably vague in many ways, especially when establishing their boundaries. This paper tackles the identification of these features and the study of their properties by which they can be categorized. It especially focuses on the feature *mountain* although other landform features have also been studied.

A software tool has been created in order to allow the visualisation of topographic maps as well as the pertinent features. The system provides the user with a set of different functionalities which the user can make use of only by providing digital elevation data. These functionalities allow the user to explore a particular terrain and the features that are part of it.

The new tool has been evaluated by comparing the results with real maps and accurate results have been obtained when identifying peaks which has been the core of the project. It has been shown that the two created models can be potentially extrapolated to other datasets whilst maintaining their quality .

Keywords: topography, vagueness, geographic information systems (GIS), landforms, software application.

ACKNOWLEDGEMENT

Firstly, I would like to thank my supervisor Dr. Brandon Bennett for his guidance, help and patience in our meetings throughout the project, as well as my assessor Dr. Hamish Carr for his feedback.

I am also very thankful to my parents for their emotional and economic support during this year. There are no words to describe what they have done and they would do for me.

Finally, I would like to thank everyone that has helped me and motivated me. Especially to my colleagues on the Masters course , for providing me moments that I will never forget. Thanks.

Table of contents

ABSTRACT I
ACKNOWLEDGEMENT II
1.INTRODUCTION 1
1.1. Presentation of the project 1
1. 2. Aim of the project 2
1.3. Objectives 2
1.4. Minimum requirements 2
1.5. Deliverables 3
1.6. Relevance 3
1.7. Methodology 4
1.7.1 Project Management 4
1.7.2 Project plan 6
1.7.2.1 Key activities performed 7
1.7.2.2 Milestones 7
1.8 Scope 8
1.9 Report structure 9
2. BACKGROUND 11
2. 1. Introduction 11
2.2 Vagueness 11
2.3 Mountains as a main target features 12
2.4 What makes a mountain a mountain? 13
2.5 Valleys and other land-form features 14
2.6 Previous work 16
2.7 Contour trees 17
3. METHODOLOGY 23
3.1 Overview 23
3.2 Data collection 23
3.2 Data analysis 24
3.3 Project methodology 24
3.3.1. Phase 0: Learning and background research 24
3.3.2: Phase 1: Design of a solution 25
3.3.3. Phase 2: Development of prototypes 25
3.3.3.1. Prototype 1 (Period: 13/04/2011 – 27/04/2011) 25
3.3.3.2. Prototype 2 (Period: 30/05/2011 – 28/06/2011) 25
3.3.3.3. Prototype 3 (Period: 29/06/2011 – 28/07/2011) 26
3.3.3.4. Final prototype (Period: 29/07/2011 – 18/08/2011) 27
3.3.4 Phase 3: Testing and Evaluation 28
4. DESIGN AND IMPLEMENTATION 29
4.1 Technology 29
4.1.1 Programming Language 29
4.1.2 Java Libraries 29
4.2 Design 30
4.2.1 Packages structure/ Classes 30
4.2.2 Input/ output data 31
4.3 Functionalities 31
4.3.1 Setting the colour threshold 31
4.3.2 Main characteristics of the data 31

4.3.3 Candidate peaks 32
4.3.4 Highest contour including two peaks 32
4.3.5 Parent peaks 33
4.3.6 Closest higher peak 34
4.3.7 Classifier factors 35
4.3.7.1. Factor 1: Steepness 35
4.3.7.2. Factor 2: Difference in prominence from its closest peak 35
4.3.7.3. Factor 3: Horizontal distance to a higher peak 36
4.3.7.4. Factor 4: Absolute height in relation to the average height of the surroundings 36
4.3.7.5 Disjunctive and Conjunctive Combination 37
4.3.8. Global models 37
4.3.9 Valleys, rivers and lakes 38
4.4 Visualization window 41
5. TESTING 42
5.1 Program details 42
5.2 Testing plan 42
5.3 Performance measures 42
5.3 Performance results and determination of parameter values 44
5.4 Visualization results 47
5.4.1 Candidate peaks 47
5.4.2 Correctly and incorrectly classified peaks 47
5.4.3 Contours 48
5.4.4. Lakes 48
6. PROJECT EVALUATION 50
6.1 Comparison with minimum requirements 50
6.2 Evaluation of the final prototype 50
6.3 Discussion of results 53
6.4 Limitations 55
7. CONCLUSIONS AND FUTURE WORK 57
7.1 Conclusions 57
7.2 Future work 57
8. REFERENCES 59
APPENDIX 1: PERSONAL REFLECTION 63
APPENDIX 2: INTERIM REPORT 65
APPENDIX 3: INITIAL PROJECT SCHEDULE 66
APPENDIX 4: REVISED PROJECT SCHEDULE 67
APPENDIX 5: MANUAL 68
APPENDIX 6: LOCATION OF LAKES 72

1.INTRODUCTION

1.1. Presentation of the project

People tend to think of geographical features as fixed and well-defined rather than vague and subjective. However, a feature might be characterised differently depending on many factors, such as scale, location, surroundings and even cultural differences. A couple of geographical features that exhibit a high degree of vagueness are mountains and valleys. What is a mountain? The answer might be straightforward: "It is an elevation rising from its surroundings". Nevertheless, there are many aspects that have to be taken into consideration. Supposing a certain elevation to be a mountain, it begs the following questions: What are the boundaries of the mountain? Which areas lie within these boundaries and which areas form part of the surrounding landscape? At their base, mountains have no distinguishable boundaries, so demarcating the extent of a mountain is problematic.

One significant factor is the topographic prominence (Helman [20]). This is an objective measurement that gives us information about the significance of a summit in relation to the topographic profile of its surroundings. It is also known as the relative height and it provides a way of hierarchizing a set of peaks among a mountainous region. Prominence measures the distance between one peak and the lowest contour line that encircles it and does not encircle a higher peak. It might be equivalently defined as the least vertical distance one would have to descend before beginning the ascent of something higher. It is important to take into account that relative height is different from absolute height. We are familiar with the concept of elevation, the measure of vertical space between sea level and a particular point, nevertheless this is very limited since the surrounding of the mountains is totally ignored. The dividing line between a set of peaks belonging to a mountain and a set of different mountains is not always clear.

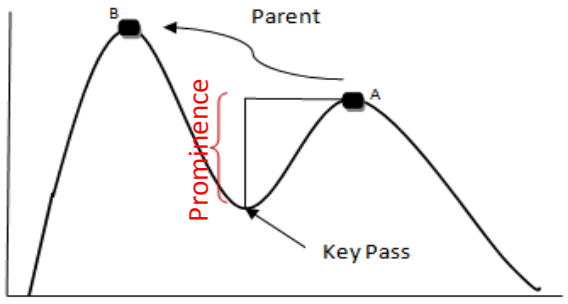


Figure 1. Peak B is the parent peak of peak A. The prominence of peak A is showed and it is thought of as the difference in height between the peak and the key pass.

1. 2. Aim of the project

This project sets out to explore different techniques of identification of landforms using digital elevation data and to develop an application capable of helping us to identify unambiguously distinct geographical features. The outcome is an application whereby one can visualise different landform features as the set of samples that have been used to test the application.

1.3. Objectives

The main objectives of this project are:

1) To acquire knowledge about the problem from previous academic studies.

2) To propose feasible approaches through the creation of several algorithms that will be useful for the identification of landform elements.

3) To implement the corresponding algorithms and develop an application which will be tested and refined. The usefulness of the proposed algorithms will be evaluated.

1.4. Minimum requirements

As it was stated in the minimum requirements form and afterwards in the interim report, the minimum requirements are:

1. Analysis of relevant geometric and other characteristics that may be used to identify landform features.
2. Design of algorithms based on relevant characteristics.
3. A tool for the visualisation of results.
4. Analysis of behaviour using real elevation data.

The design of the algorithms has been carried out by exploring the relevant characteristics. A particular characteristic was added to the set of relevant features according to the results obtained by using it. The visualisation tool has also been created at the same time, making it easier to see the results and decide whether a characteristic was important or not. These decisions were made using the same set of data. However, other datasets have been used during the testing and evaluation stage. This is useful to check how well the chosen characteristics might be extrapolated.

1.5. Deliverables

Accompanying the pertinent report, a software application will be submitted. The application will be able to be executed on any kind of machine independently of operative system that it has. The corresponding libraries will be included so that the user can run the application having to carry out the minimum number of modification to the actual delivery. Data files will be also required to run the application so some sample files will be included. Most of these sample files will be the ones used to test the application and therefore the images appearing in the final report will be obtained by using some of these data sets. A manual will be also submitted, describing the steps to be followed to execute the corresponding functionalities provided by the application.

A breakdown of the deliverables can be seen below. They have been distributed according to the different stages of the project. Note that most of them are submitted as part of the project report.

Stage	Deliverables
Research	Literature research Technologies research Statistical methods research
Design	Methodology choice discussion Implementation choice discussion Structure of Java classes discussion
Implementation	Software application Implementation plan discussion Faced problem discussion
Testing	Test plan Test result and comments on it. Data files
Evaluation	Evaluation plan Evaluation results with corresponding comments. Overall system evaluation
Writing-up	Project report Manual of the application.

1.6. Relevance

This project builds on skills and knowledge acquired from some of the MSc modules taken at Leeds University. Techniques for knowledge management (COMP5390M) and Computational Modelling (COMP5320M) make the author of this dissertation familiar with some statistical concepts that were used in the evaluation stage. The latter also provided some knowledge regarding data storing structures. Machine Learning (COMP5425) and Scheduling (COMP5920M) were useful as I could acquire expertise in designing algorithms and in methods to evaluate their performance. In general, all the modules that were taken

during the first two semesters were valuable to hone the research and especially the documentation skills which have been put into practice carrying out this project.

1.7. Methodology

1.7.1 Project Management

A system development is characterised by the combination of different tasks that must be accomplished to ensure the success of the development process. However, the system developments can be differently classified by the control of the timing and the ordering of activities.

The process management is a crucial aspect that must be borne in mind. The Software Engineering Institute of Carnegie Mellon University points out that the quality of a system is highly influenced by the quality of the process used to acquire, develop, and maintain it. To control the process of developing there is a wide variety of standard frameworks, some of which are commonly used by organizations and project teams. However, the variety of these frameworks begs the questions of which one should be selected given a particular project. The scale and the type of project are the main aspects that should be considered while deciding the methodology that suits it best.

Waterfall methodology [34], Spiral model [5] and Rapid Application development [28] are some of the most famous software development methodologies. However, after examining the characteristics of this project they have been ruled out.

The waterfall model was discounted due to the lack of flexibility that is clearly required in the project. It needs as much flexibility as possible, allowing moving backwards at any stage of the process which is unfeasible considering the waterfall methodology. The other drawback of this model is related to the non-simultaneity of the development and testing processes. The performance of the system cannot be tested until it has been almost fully coded which is totally unsuitable when the development and testing stages necessarily overlap.

The Spiral model has also been discounted since it does not count on firm deadlines. That is not advisable when the project has some fixed deadlines as in the case at hand. Moreover, this model requires a high degree of expertise due to the difficulty that entails the application of it to any given project.

Finally, the Rapid Application Development (RAD) might seem the most appropriate due to the rapidity in which the software can be created and the ease with which the application might be modified

since it is split into smaller segments that can be treated independently. Against the waterfall model, the system can be tested before it is completely finished. Nonetheless, these factors may lead to very low quality systems and subsequently the system is bound to fail. However, this is not the only problem that is entailed by RAD. Good productivity is obtained at the expense of scalability due to the fact that application is designed as a full system from the start.

After considering and discounting the previous methodologies, I found another alternative that enhanced the achievement of the project requirements. The most convenient method to this project is called *prototype development*, more concretely the '*evolutionary prototyping*' [10] that is the most convenient to this project. The decision of selecting the prototype was based upon the fulfilment of the following aspects:

- Flexibility throughout the life-cycle of the project.
- Ease of production of different versions and solutions for a given problem.
- Possibility of moving backwards when the achieved performance is not suitable.
- Simultaneity of the development and testing stages.
- Possibility of demonstrating the running of some functions on intermediate versions that will be included in the final prototype.

With evolutionary prototyping, one starts by designing and implementing the most prominent parts of the program in a prototype and then gradually refines the prototype until achieving the final product. The prototype becomes the application that one eventually releases.

The model does not only increase communication between users and the system developers but also allows the developers to find out more about the needs of the users in advance and acquire users' feedback quickly [43]. This is especially suitable for us, considering the supervisor as the user and the student as the system developer.

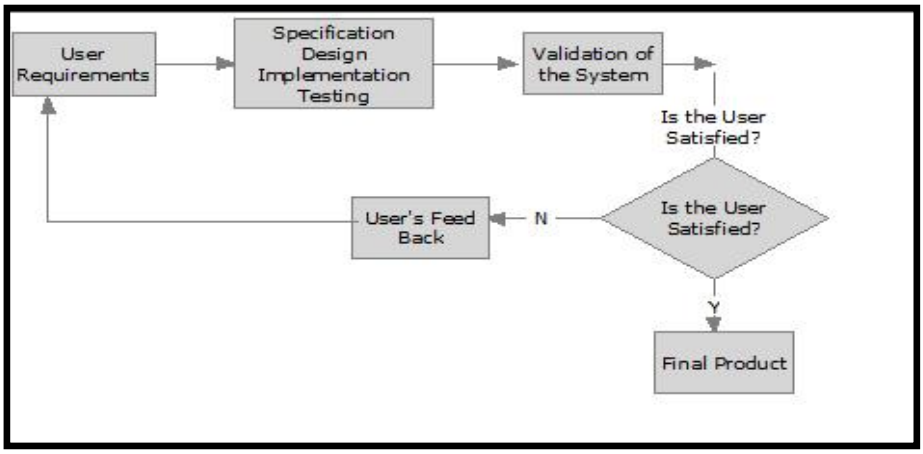


Figure 2. This diagram shows the steps carried out by the evolutionary prototype development.

1.7.2 Project plan

Planning is the key to a successful project, especially when the projects is not straightforward. Often planning is ignored in favour of getting on with the work, however there are many factors that have to be taken into account such as:

- Identification of tasks that will be required.
- How long each task will take.
- Identification of dependencies between tasks.
- Specific requirements for each stage.

Regarding the importance of the planning, A.A.Milne said the following:

"Organizing is what you do before you do something,

so that when you do it, it is not all mixed up"

Tasks are the activities that have to be completed in order to achieve the project goal and they are identified by their start and ends points. The end of each stage should be determined to avoid delays. Suppose that only the submission date is fixed beforehand. The project may be released on time but with problems due to the lack of intermediate control points which help to predict how well the system would meet the minimum requirements.

The milestones are important checkpoints for a project and they also make easy the identification of risk areas. There might be a stage in which one knows what to do but one has to learn how to do it. It is not easy to estimate how long the learning process will take.

There are different tools to visualise a project plan. One of the most simple to understand and easy to construct is the Gantt Chart [13]. The Gantt chart is made available in the appendix 3. The Gantt chart shows the duration it takes to accomplish each task. There are 7 tasks, excluding the *"Revision and Exams Periods"* that is also shown in the corresponding chart and each task is subdivided into different subtasks when necessary. In a Gantt chart each task occupies one row and the expected time for each task is represented by the horizontal bar. Recall that tasks might run in parallel, sequentially or overlapping.

1.7.2.1 Key activities performed

Background Research:

This phase involved researching of the problem areas and developing ideas for solving the problems identified as well as the performance methods that can be used.

Development:

This phase of the project covered all the activities of the evolutionary prototype methodology outlined in above in 1.7.1. Each prototype involved three different activities: analysis, coding and testing.

Testing:

The performance of the system was measured. Tables and other visualisation methods were used to help the reader to get a good understanding.

Evaluation:

The assessment of the final prototype was involved in this phase.

***Assessment and Reflection*:**

This phase involved analysing and writing-up of the evaluation results and subsequent write-up of project experiences and lessons learnt.

Write-up:

As it can be seen in the Appendix 3, report writing started the last week of July with the draft chapter that had to be submitted. The write-up was performed concurrently with other activities such as the development and the evaluation process. However, two weeks before the final submission were designated only for the writing task.

1.7.2.2 Milestones

Milestones are thought of as project checkpoints to validate how the project is progressing. They are essential to manage and control the progress, nevertheless there is no task associated with them although their preparation might entail significant work. They are useful when one wants important events which are not tasks to appear on the project timeline.

The following milestones have been identified, along with their completion dates. They have been

selected even thought they did not have official deadlines associated to them.

1. Returning completed preference forms: 11.02.2011
2. Devising aim and minimum requirements : 11.03.2011
3. Background research: 01.08.2011
4. Interim report: 17.06.2011
5. Table of contents: 28.07.2011
6. Draft chapter: 3.08.2011
7. Complete implementation: 17.08.2011
8. Complete testing: 24.08.2011
9. Finish write-up: 26.08.2011
10. Final deadline of hard copy report submission: 01.09.2001
11. Final deadline of online report submission: 05.09.2011

These milestones will be showed in the Gantt graph (Appendix 3). Each milestone is represented as a *'Milestone X'*.

1.8 Scope

This report presents a problematic issue regarding the field of geography. Vagueness is widely involved when identifying geographic features and dealing with it is challenging in any field. The report and the overall project are focused on mountains which are one of these features whose boundaries cannot be clearly identified. A set of different parameters has to be studied to demarcate the boundaries of mountains. Due to the fact that the concept of vagueness cannot be removed from the field of geography, the main objective is to develop a framework within which the variability in meaning of the vague term 'mountain' can be accounted for in a way that corresponds well with natural usage of the term. Many academics have researched this issue (such as [3], [12], [17] and [41]).

Ideally the classification algorithm should work well on a wide range of data incorporating a variety of different landscapes and types of mountains. However, it is likely that different choices of parameter values

would work better on different types of data. A further complication is that the classification of what counts as a mountain depends to a significant extent on the interests and attitude of the person who is making the classification.

The scope of the system is multi-disciplinary. It can be used by people with different backgrounds although the most common user will be related to geographic studies. It may be useful for earth surface-based investigations, archaeology studies, urban planning or statistical analysis.

It also has a non-academic target. Hikers and climbers might be interested in knowing the topographic profile of a particular zone to carry out some kind of outdoors activity. They might need a simple way of knowing what areas are suitable for the particular activity they are going to do.

1.9 Report structure

In light of the objectives of the project, the report will be structured in the following chapters:

Chapter 2: Background research

This chapter describes the area of the problem. It familiarizes the reader with the concept of vagueness and it explains previous work that has been done in the field. The research is mainly focused on the work related to the identification of mountains but some research is also done regarding other land-form features such as valleys. In this chapter an approach based on contour tress will be presented, explaining how it might be applied to deal with the vagueness.

Chapter 3: Design and methodology

The third chapter firstly documents the type of data required to run the application, how this was obtained and how the data is analysed by the system. Secondly the project methodology is also discussed regarding the prototyping framework that was selected. The evolution of the project will be explained in terms of the phases which the project went through.

Chapter 4: Implementation

This chapter details the different frameworks that have been used to develop the software application as well as the programming language that was chosen. It also explains the storing structures that have been created to achieve the objectives. The functionalities of the project are described by using examples and screenshot that will help the reader to understand how the application works.

Chapter 5: Testing

This chapter focuses on the testing of the application. It documents how the test process has been carried out, not only the planning but also the measures that have been taken into consideration

Chapter 6: Project evaluation

The sixth chapter details a comparison between the results and the minimum requirements. It also provides an overall evaluation of the performance of the application as well as the scheduling that has been followed. Finally, the testing results are discussed.

Chapter 7: Conclusion and future work

Suggestions for improvements and ways to further the project are provided in the last chapter. A general balance of overall work which has been done will be a given as conclusion.

Appendixes

The appendix mainly contains my *personal reflection* commenting on everything that I have experimented during the implementation of this project. The comments on the *interim report* given by the assessor and supervisor can be also found in the appendixes. The Gantt graphs are also included as well as tables of results, screen shots or any other image considered to be useful for the understanding of any part of the project.

2. BACKGROUND

2. 1. Introduction

The origin of the research into techniques for describing landforms lies in geomorphology and in soil science (Maxwell [27] and Cayley [9]). Geomorphology is the measurement of the shape of the earth's surface where it is possible to assign a location of the landscape to a named feature depending on the local form of its surface.

The characterisation of landforms is useful for analysis, for inferring terrain characteristics and the distribution of water and soils. Partitioning soil-landscapes into landform elements is very important for an effective management of units for precision farming, for the application of biological methods and to provide a base for benchmark soil testing (e.g. MacMillan et at. 1998). This characterisation is also useful as it gives a precise and repeatable human-understandable definition of the different elements of a landscape. These definitions are not easy to find, especially regarding the identification of mountains and their corresponding extent (cf. Fisher et al. 2003).

2.2 Vagueness

The first written references related to the concept of vagueness goes back to the beginning of the 20th century, concretely to 1923 by Bertrand Russell [39]. In 1937, Virgil Aldrich [1] pointed out the different kinds of vagueness and propounded a list of vague concepts. A concept is considered to be vague if it is not clear, precisely determined or indefinite in shape, form or character.

People prefer precise information over vague information but the vagueness problems arises in many fields. It is so widespread that it is sometimes said that the world might itself be vague rather than considering vagueness as a deficiency in the mode of describing the world [15].

Stanford Enclyclopedia of Philosophy[1]states that vagueness is defined as the possession of borderline cases which may belong or may not to its extension. For instance, the concept *'box'* divides the world into the two different sets. The set of elements which are boxes and the set whose elements are not boxes. However, the concept *'big box'* is controversial. What makes a box big? How big does it have to be? What is

[1] http://plato.stanford.edu/

the threshold we should use to decide whether an box is big or not. The sets of 'big boxes' and 'small boxes' cannot so clearly determined.

Many topographic feature definitions are ambiguous. As Kweon and Kanade [22] stated, natural language definitions of topographic features have the substantial drawback that such definitions either use terms which are not exactly defined but are assumed to be generally understood, or they end up in circular definitions.

Mountains and valleys exhibit a high degree of vagueness. Mountains have a vague spatial extent and it is often hard to accurately establish where the mountains itself is geographically based. It is well-known that a determined feature might be characterised differently depending on many factors; such as scale, location, surroundings and even cultural differences play a role here.

The aim of this project is to create a solution that is able to classify landform elements automatically and can be applied to a wide variety of landscapes without any kind of modification. But what kind of knowledge should be applied? Because of lack of agreed definitions one cannot use explicit knowledge Typically, tacit knowledge has been applied to face this paradigm but it is actually considered to be very poor and inexact. Tacit knowledge covers all the knowledge that is hard to articulate with formal language. The ability to ride a bicycle is the most common example of this kind of knowledge.

2.3 Mountains as a main target features

Summits and peaks are relatively straightforward to define. A mountain might be defined as its peak, yet it is not considered to be an acceptable definition, summit is not a synonym for mountain. The mountains are not only peaks and not all peaks belong to mountains. One can be on the mountain without ever reaching the summit. Many definitions of what a mountain is have been given, often reflecting that it varies depending on the context. For instance, a definition of a mountain in England is likely to differ from the definition used in the Alps.

Mountains are also vague in a philosophical sense (Smith and Mark [41]). Mountains do exist but are different compared to everyday objects. These have determinate boundaries. They are detached objects separated from each other by *bona fide* boundaries. Although the boundaries between the mountain and the air above is clearly determined, when one proceeds towards the foot of the mountain there is no distinguishable candidate boundary. If one is on the summit, how long does one have to go down until being able to say that one is not on the mountain anymore? A fine scale analysis allows us to define mountain regions and from each of these, to create *parent-child* relationships. In such a way that the

highest point of the region would be the parent of each possible peak that might appear within the extent of the mountain.

2.4 What makes a mountain a mountain?

Mountains are produced by forces in the Earth that trigger changes in the Earth's crust. The crust is broken into large sections called plates and when they crash into to one another they cause some areas to rise and others to sink. The shape of mountains is produced by chemical and physical erosion, by the action of wind, rain, frost and other natural forces. People are familiar with the concept of 'mountain' and their way of perceiving the terrain determines what a mountain is and what is not.

There is no universal definition of a mountain. Numerous definitions of what constitutes a mountain have been proposed but they diverge a lot of from one to another making very difficult to achieve agreement in description [19]. In the *Oxford English Dictionary*[2] a mountain is defined as "a natural elevation of the earth surface rising more or less abruptly from the surrounding level and attaining an altitude which, relatively to the adjacent elevation, is impressive or notable."

Different aspects have to be borne in mind to define/concrete what a mountain is. Hereafter, the aspects are explained:

1) **Kind of soil**

   Mountainous areas are characterised by the lack of fauna and vegetation cover to protect the ground against erosion. The slope make them unstable surfaces and especially vulnerable to erosion because of the large amount of water and its high speed. Due to low temperatures, the soil of mountains forms very slowly and consequently it is shallow, rocky and usually not solid.

2) **Form**

   a) **Height**

      Mountains are a type of landform that is characterized by a higher elevation in comparison to the surrounding areas. Commonly the height of a point is considered to be the distance between that point and the sea level; however a problem appears regarding underwater mountains. If one takes this into account, the highest mountain in the world is not the Everest. It is the Mauna Kea, a dormant volcano of the island of Hawaii. It is the highest island-mountain

[2] http://www.oed.com/

in the world. Its height is 4205 metres over sea level but nearly 10.000 metres if one considers the distance from the summit to its base [47].

b) Shape

Some definitions of mountains include that a mountain should have a significant slope. The slope is the measure of steepness of a feature to the horizontal plane. The slope can be calculated by comparing the height of a point with the height of the surroundings and it can be clearly estimated by looking at contour maps.

3) Context

When examining a candidate mountain, the nature of the surroundings of the mountain under study is crucial. Some of the factors that might be considered are:

a) The horizontal distance to the nearest higher peak.
b) The neighbouring landforms in general.
c) The prominence of the mountain above the key pass regarding the nearest higher peak.

With regards to the previous aspects, the system developed in this project studies the following characteristics (ordered by locality):

1) Absolute height **(2.a)**
2) Steepness **(2.b)**
3) Prominence **(3.c)**
4) Horizontal distance**(3.a)**

2.5 Valleys and other land-form features

A similar argument can be applied regarding the identification of valleys. Intuitively, it is very easy to say where a valley is but describing the spatial extent of the feature is far more complicated. Valleys are lowlands between mountains and hills. They can be seen as natural networks that transport water to lakes and oceans.

Most of the approaches related to the computation of *valleyness* start from the notion of valley floors (e.g. Straumann and Purves 2008). The valley floors are the areas bordering *thalwegs* (valley way), considering a *thalweg* as the deepest continuous inline within a valley or watercourse system and where the current (if there is one) is fastest. Its extent might be calculated by using concavity notions. Starting from the valley floor and proceeding upwards the surface is usually concave. The floor of the valleys varies

not only in width, which is normally determined by the steepness of the nearest mountains, but also in shape. Valleys are characterised into two groups: V-shaped valleys and U-shaped valleys (Figure 3). The river-valleys are usually V-shaped although its exact shape depends on the steep gradients of the surroundings. On the other hand, the U-shaped valleys are those corresponding to glacial valleys most of which were V-shaped before the erosion.

Figure 3. An irregular V-shaped valley produced by stream erosion; B, the same valley after it has been occupied by a glacier. Note the smooth topography and U-shaped form.

This lack of similarity between different valleys make it even harder to identify them. Another controversial issue related to the demarcation of valleys concerns the use of different names when referring to the same feature. *Dales* and *hollows* are some of the terms that refer different kind of valleys. A *hollow* is a 'small valley' but how small does the valley have to be to be considered as a *hollow*?. If the identification of valleys is not straightforward, the decision of whether a valley can also be called hollow or dale is even more complicated.

From a hydro-geological point of view, the identification of valleys is becoming increasingly important as groundwater reserves are stored beneath the valley surface. As climate change progresses, the lack of freshwater is starting to be a problem and therefore the location of the valleys and consequently of the water sources are under study [30].

We might think that the vagueness is limited to land features but there are many other cases in the sea. Given an extension of sea close to the coast (see Figure 4), which of the red lines or any other should we consider as the boundary? A bay is thought of as an area of water surrounded mainly by land. Its limits used to be selected having into account the extension in square metres (between the 10km^2 and 100km^2). The size of the extension is arbitrary but even if a particular extension and the corresponding points in land (A, B, C, X, Y, Z) are selected, it is not clear how the dividing line should be traced. The vagueness still remains here.

Figure 4. Different delimitation lines between the bay and the sea.

2.6 Previous work

Today most studies focus on continuous models, quantitative definitions and fuzzy classes. There is an increasing interest in assessing class membership for each grid cell in a matrix. The class membership is considered as the degree at which a cell is likely to belong to a defined class. But how can we determine the number of classes and what they are? The classifiers can be based on expert knowledge and informed judgement to define the number of classes, their main characteristics and the values associated to the boundaries between different classes. Another way to determine the number of classes is through statistical approaches such as k-means algorithm. K-means algorithm might identify a suitable number of classes for a given landscape. Nevertheless, the drawback here would be that the number of classes and their definitions would be optimised only for a particular size. Namely, the number of classes would vary depending on the kind of landscape one is considering and what is interesting is a classification procedure capable of classifying landform entities for a wide variety of types and scales of landscape. It is not clear that the classes produced by a statistical technique such as K-means would correspond to the classification of features used by humans.

Regarding the demarcation of mountains, some researchers have used using fuzzy theories (e.g. Fisher et all. 2003). In fuzzy set theory, when an object exactly matches the concept is assigned a membership of 1. On the other hand, when it does not have any similarity, the membership will be 0. Two different techniques have been used to determine the membership values. The first technique assigns the values based on *priori* knowledge given a particular metric property (Cheng and Molenaar 1999). The second technique uses surface characteristics such as slope and curvature which determine the membership values (MacMillan et al 1998). But one factor that has to be taken into consideration is the scale, which is understood as the combination of spatial extent and detail. A different scale might radically change the membership values of a region. Wood (1999) stated that a location may belong to very different classes just by varying the corresponding scale one is considering. But how has this problem been tackled? Multi-scale analysis approaches (Fisher et al. 2003) have been carried out obtaining good results when modelling objects which are vague due to scale reasons. This kind of analysis is based on the assessment of the membership value of the morphometric classes for a given location and for a particular scale. In such a way the final membership value of a morphometric class is the weighted average of those obtained by varying the scale. The accuracy of the model was assessed by comparing the obtained results with landscape objects which have an established proper name which might be found on a map. However, it is not clear whether all the features whose names appear in a map actually deserve to have a proper, established name.

Are the features with established names a good baseline to determine whether a location belongs in a morphometric class? Is it possible to find a region that meets the requirements that characterise a peak even though it does not have an established name? Many researchers state that these features are not a good baseline to decide whether a feature has been correctly characterised. Human subject experiments have been carried out (e.g. Straumann and Purves 2008) as an evaluation method. A set of pictures was shown to different subjects who where then asked to determine whether a certain landscape belongs to a morphometric class. Straumann and Purvas state that any study of landform characterisation should consider human subject experiments as a valid benchmark in spite of the laboriousness that is entailed.

2.7 Contour trees

Another approach that has been used to the extraction not only of mountains but also extracting land-form features in general is the *topographic change tree* also known as contour tree. The contour tree is a connectivity tree of regions separated by contours and it was first used in 1963 by Boyell and Ruston [6]. They were the pioneers in using graph theory to manipulate digitalized contour lines. They used the fact that the areas defined by the lines of the contour maps are sorted by inclusion. Hence, once one knows the contour line that has been crossed, the most effective way of identifying the line that will be crossed next is using the contour tree.

Many other authors, since then, have gone into the applications of contour trees in depth [24], [25], [29] and [45]. Roubal and Poiker [38] were the first ones in pointing out the applicability of this data structure in the automatic labelling of contour lines.

Two regions separated by the same contour line have different average height. The relation between average height of adjacent contours is a partially ordered relation (it is reflexive, anti-symmetric and transitive) [14]. Therefore the set of regions limited by contour lines is a partially order set (also called *poset*) and it has two subsets of special interest. These subsets are the maximal set and the minimal set. An element is maximal when there is not an element larger than it in the set of elements under study. Similarly, an element is minimal when there is not an element smaller than it. Any *poset* can be represented by an directed graph and it can be fully defined by finding the subsets of minimal or maximal values (one of them would be enough) [42]. Below, it can be seen a contour map with its corresponding contour tree, where the areas labelled with A, C and F represent the minimal elements and those labelled with I and K represent the maximal elements.

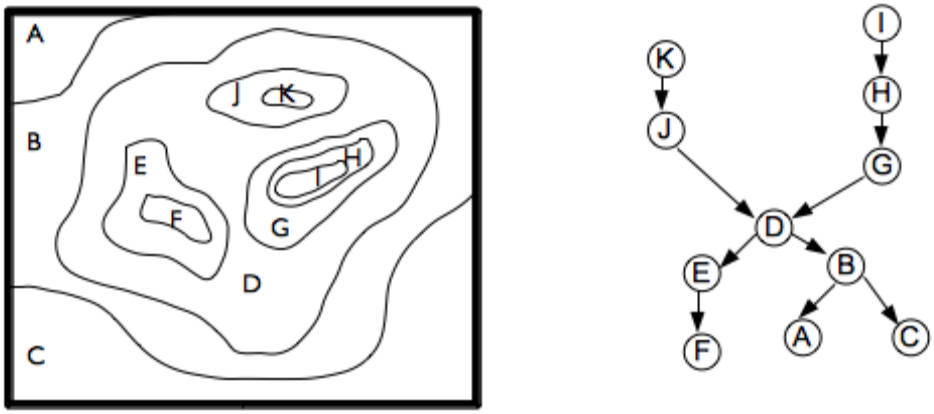


Figure5. Left: Contour map where each region has been labeled with a different letter. Right: Contour tree obtained from the contour map on the left. Each region is represented by a node. Connections between nodes represent adjacency between regions in the map.

The work of Freeman and Morse (1967) [18] was also very relevant regarding the use of contour trees as a tool for dealing with the profile search problem. In their approach, the notation of the graphs used was different from that used previously. The contour lines were mapped into nodes and the areas between contour lines into edges. Nevertheless in Figure 5, an inverse notation is used. The nodes represent the regions and the edges represent the contour lines instead. These two different representations are equivalent but one might be more efficient and convenient regarding the data structures and techniques in an implementation [40].

Kweon and Kanade [22] also used the contour maps to identify terrain features. They analysed the contour trees that were obtained from the contour maps (see Figure 6). They were able to extract peaks and pits as well as ridges and ravings. Using their method, peaks and pits are identified by finding the nodes in the tree which have no descendants. When a node without descendants is found, the height of this node and the height of its parent node is compared. Therefore if the height of the parent node is higher it will be a pit otherwise it will be a peak. Note that the parent node represents the contour line that contains the contour under study and does not contain any other contour. The ridges and ravines are identified by examining the properties of the lines that form the contour map. The contour lines that belong to ridges or ravines have a shape of U or V and therefore these features can be extracted by studying the local maxima (or minima) of the curvature of the contour. Recall that the shape of V corresponds to the local maximum in curvature of the contour.

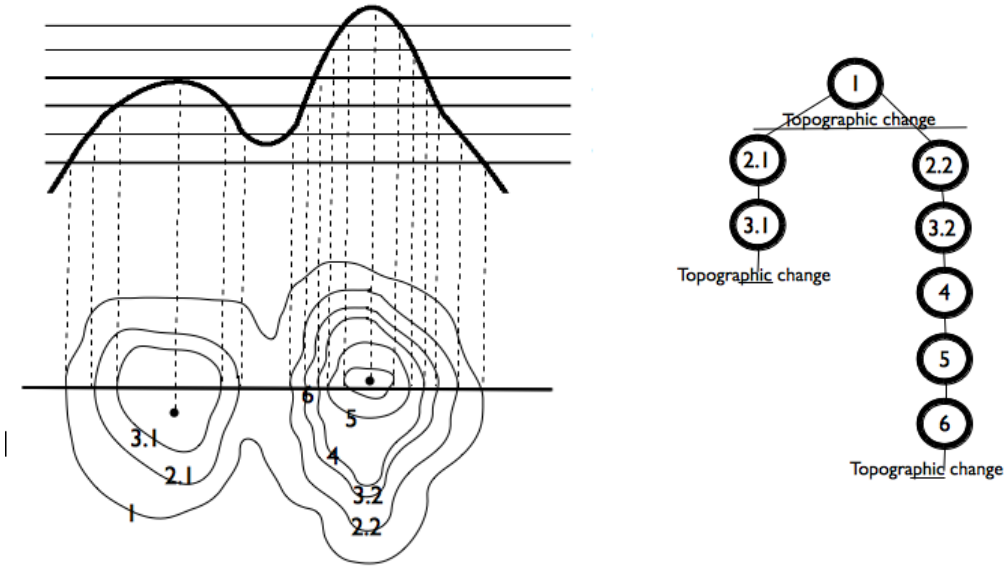


Figure 6. Left: Terrain profile along with its contour map. Right: The contour tree associated to the map. Note that the nodes represent contour lines whereas the edges represent regions between contours.

A large volume data set cannot be represented by contour trees in a very compact way. The problem lies in the size of the resulting trees. They often contain many thousands of nodes which is prohibitive due to the fact that graph drawing techniques fail to produce satisfactory results .

Efficient algorithms for computing contour trees have been published, for example Carr et al. in 2003 [8] and Pascucci and Cole-McLauglin in 2002 [35]. Carr et al. showed that contour trees can be computed by a simple algorithm that merges two trees in any number of dimensions. These two trees are respectively called the *join tree* and *split tree*. The *join tree* is a graph that encapsulates all the joins in the contour tree whereas the *split tree* is a graph that encapsulates all the splits in the contour tree. The first one is created starting at the highest peak and adding points in order of height. Therefore different areas are selected and they end up joining as the height is decreasing. When two different areas join together, the saddle point (also called col or mountain pass) is found and they are treated as only one area using the saddle point as reference point for a possible further join. When all the regions in the map under study have been selected and no more joins are possible to carry out, the *join tree* is obtained. Likewise, the *split tree* can be created but starting from the bottom up. Namely, selecting the lowest points and adding new areas while increasing the height used as a threshold. An example can be seen below (Figure 7 and 8).

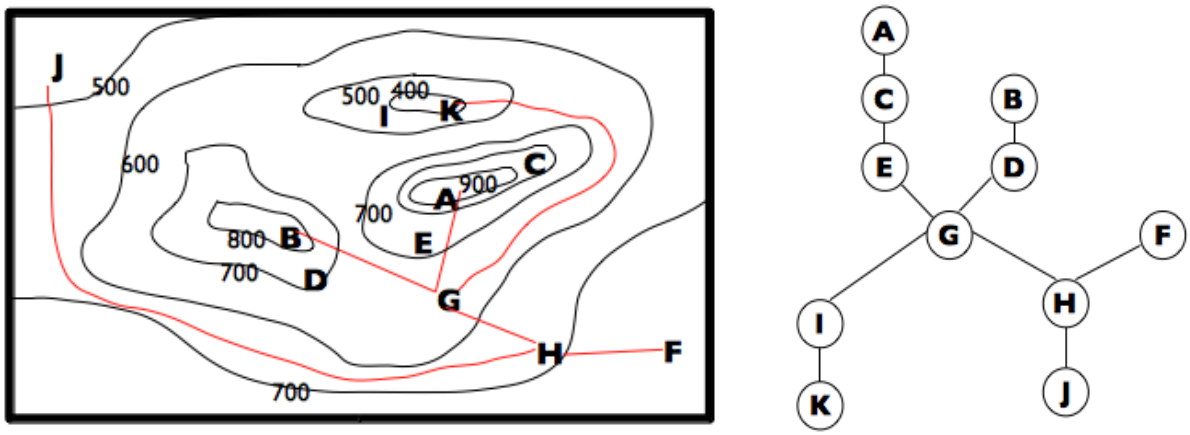


Figure 7. Left: Contour map with the simplified contour tree in red. Right: Non- simplified contour tree associated to the map on the left.

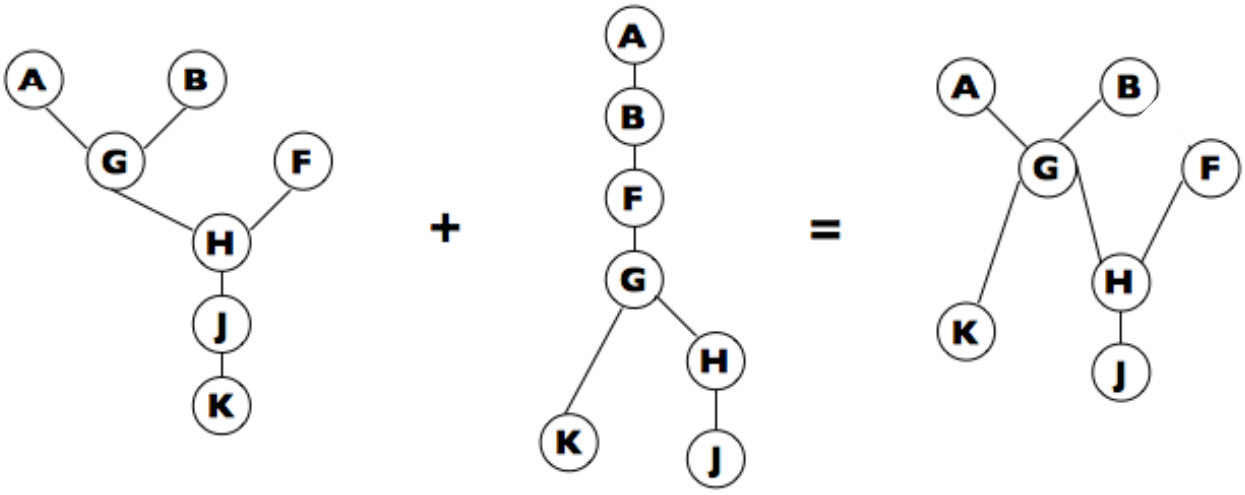


Figure 8. Join tree of the contour map of Figure 7, its split tree and the simplified tree after merging.

The contour tree is a very useful tool of visualization but it also has a drawback. When representing data sets with complex topology, the contour tress obtained are extremely large involving millions of edges and therefore unfeasible to visualize. Most of these edges represent non-significant features which have been introduced due to noise. In order for the contour tree to continue being useful for visualizations, a simplification technique should be applied. The small non-important features should be removed from the tree revealing the underlying major structure. Simplification algorithms are been given by Takahashi et al. [44], Pascucci and Cole-McLauglin [35] and by Carr et al. [7]. The latter created a general mechanism that simplifies the tree by applying two different operations: *leaf pruning* and *vertex reduction*. Leaf pruning selects a low importance node and removes it from the trees and therefore all contours corresponding to the pruned edge are also discarded. This is equivalent to flattening the region of the data represented by the leaf and edge that have been removed. Vertex reduction is the second simplification operation which

eliminates connectivity regular points in the contour tree. Given a vertex connected to a parent vertex an a child vertex through two edges, this will be removed and the two incident edges will be replaced by only one edge joining the corresponding parent and child. This kind of operations is always preferred over the leaf pruning since it does not modify the field. A simple example can be seen below (Figure 9). The process is followed from left to right and up to bottom. Leaf pruning has been applied in stage 1, 2, 4 and 6 and vertex reduction in stage 3,5 and 7.

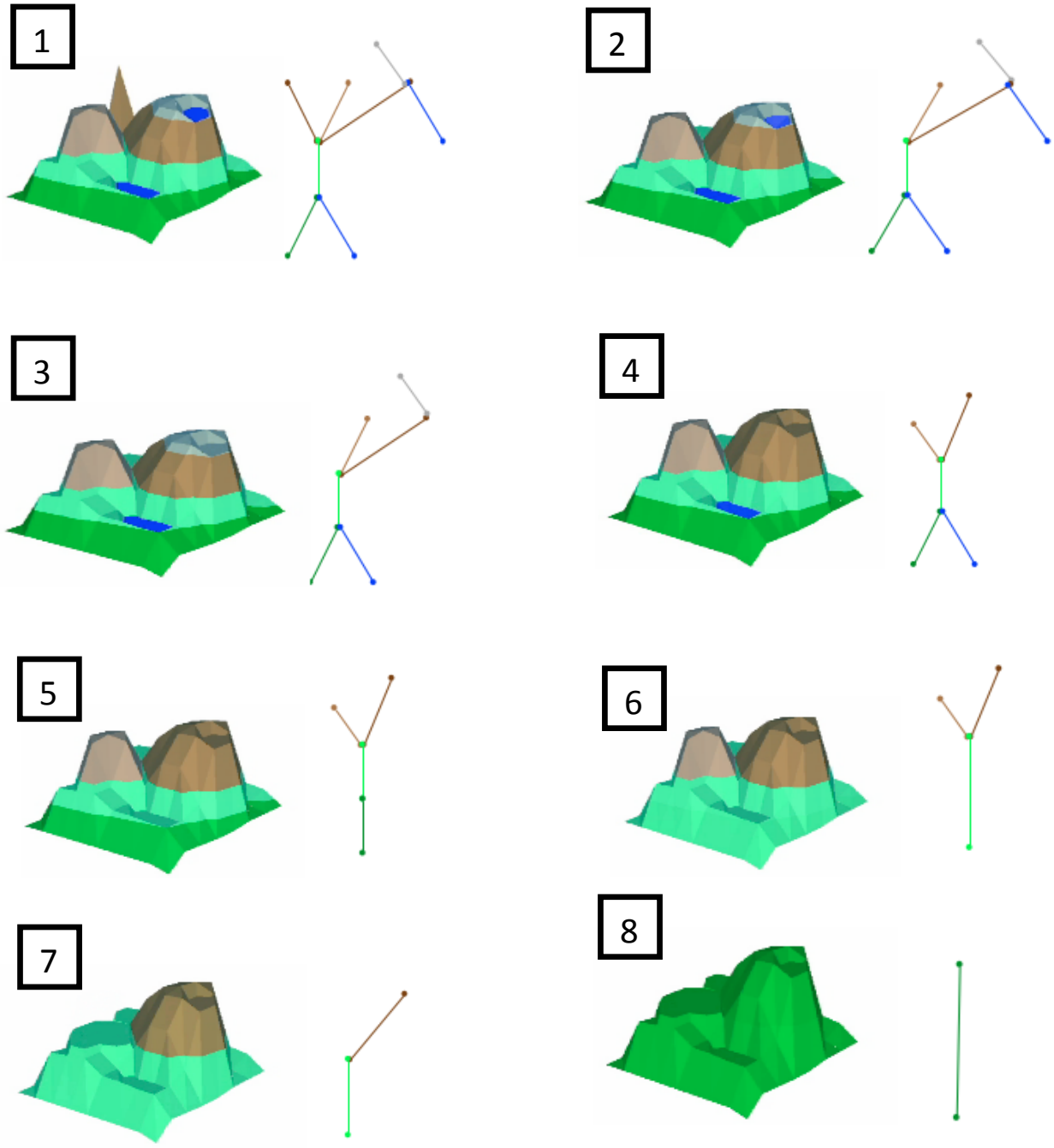


Figure 9. Process of simplification through leaf pruning and vertex reduction.
Source: http://www.csi.ucd.ie/staff/hcarr/home/research/simplification/simplification.html

However these techniques are not able to deal with uncertain contour trees although another approach has been proposed. By computing multiple versions of a data set through grayscale morphology Kraus [21] visualizes uncertain structures in contour trees. In his work, he visualizes the contour trees by combining morphologically filtered versions of a volume data set. The visualization of multiple contour trees in a single image allows us to visually distinguish the more and the less certain parts of the contour tree. In this way, the work of Kraus demonstrates that uncertainly visualization is feasible even for very large graphs.

3. METHODOLOGY

3.1 Overview

During the development of this project, different metrics related to terrain characteristics have been formulated and subsequently combined to classify mountains from a set of candidate peaks.

The visualisation tool, which is one of the final products, has been used in experimentation with the classification metrics, and enable the developer to get visual feedback on whether a measure is likely to be a good classifier of mountains. The performance of the different classifier methods have been measured using statistical concepts such as precision, recall, accuracy and F-measure. In order to describe the quality of the results, these must be compared using real data. The list of Wainwright Peaks[46] has been used as a ground-truth reference.

3.2 Data collection

Datasets are required to run the created application. The dataset has to be chosen by the user and it can be changed during the running of the application. I have obtained the corresponding datasets from Ordnance Survey [33]. Ordnance Survey is the national mapping agency of Great Britain. It produces a large variety of maps and digital mapping products which are used by public and private companies that need accurate and reliable geographic framework to deliver efficient services. The main advantage of Ordnance Survey is that it has available a huge range of mapping data that can be downloaded for free. Everyone from companies, universities or even ramblers are able to make use of the OS OpenData service with which one can access the most detailed mapping datasets of Great Britain. I have used Land-Form PROFILETM datasets which provide detailed height data defining the physical shape of the landscape of Great Britain. To be more precise, the data I have been using is Digital Terrain Model (DTM). The DTM has a horizontal grid interval of 10 meters and accuracy between 2.5m and 5m. This Ordnance Survey product is available through Digimap which is a service available to UK Education institution and provides us with a collection of EDINA services. EDINA is a UK national academic data centre that delivers access to a set of online services through a UK academic infrastructure.

3.2 Data analysis

Each dataset consists of a total of 160801 values representing heights which can be seen as a matrix of 401 columns and 401 rows. Therefore the surface that is studied with each file is approximately a 4x4 km square. The datasets have been examined to identify properties and relations between the heights that can help us to identify mountains and their extent, peaks and so on.

In order to analysis the data, different algorithms were designed with regards to the following:

- Steepness of a particular point in relation to its surroundings.
- Difference in prominence between a peak represented by a cell and the nearest peak. See the figure below to clarify.

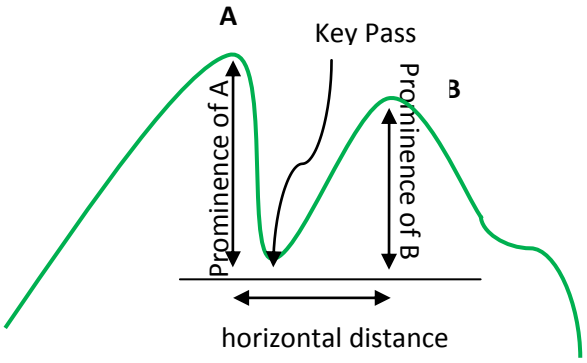


Figure 10. It shows some of the measures that have been used to carry out the identification of peaks.

- Horizontal distance to the nearest peak.
- The relation between the prominence of a cell and its absolute height.
- The relation between the average height of the dataset and the height of the cell.
- The relation between the average height of the nearest 100 cells (1 square kilometre).

3.3 Project methodology

The work that has been accomplished can be divided into different phases.

3.3.1. Phase 0: Learning and background research

The first meetings with the supervisor of this project Dr. Brandon Bennett were focused on the decision of what problem was going to be tackled. During the first weeks, many articles regarding the problem of vagueness were studied. Most of these papers were related to geographical vagueness which was very useful to acquire an understanding of the problem and the notion of vagueness. Previous projects under the supervision of Dr. Brandon Bennett were carried out in relation to the identification of forests and desserts. The code of the latter as well as the data needed to run it was given to the author of this project which was very helpful to assess the amount of work that should be done.

After several meetings, the mountains were chosen as main landform feature to be studied. The study of valleys and other possible features was postponed until the end of the project, once the minimum requirements were already met.

Once the mountains were chosen, the concept of *topographic prominence* was examined. Other aspects related to the prominence such as parent peaks, subpeaks or key pass were studied. While reading articles on previous works regarding the study of mountains, the concept of contour maps appeared as a powerful technique to study relations between mountains and their corresponding peaks. The techniques proposed on the papers recommended were examined as well as their advantages and drawbacks.

3.3.2: Phase 1: Design of a solution

In order to identify mountains, peaks and several aspects related to them, a software tool had to be developed. Java was selected as programming language after assessing the advantages (see 4.1.1). However, a data source was required and after examining the different possibilities Ordnance Survey [33] was chosen as data provider (see 3.1). The application had to be able to visualize different datasets and provide the user with a set of different functionalities. With that purpose, a couple of Java Libraries were used (see 4.1.2)

3.3.3. Phase 2: Development of prototypes

As it was discussed before (see 1.7), the *evolutionary prototyping* [10] was the methodology that best suits this project. A summary of the prototypes that have been carried out is given below.

3.3.3.1. Prototype 1 (Period: 13/04/2011 – 27/04/2011)

The first prototype was focused on adding the Java libraries to the application and the visualization of the dataset. No other functionality was added at this point. In order to make one of the libraries work (*Piccolo* [37]), some research had to be done. Different examples using the library were examined and the documentation about the framework was studied.

In order to visualize the dataset, the input file had to be parsed. The parser reads the file, interprets the bulk of numbers and creates an ordered structure containing all the values that can be used for further analysis.

3.3.3.2. Prototype 2 (Period: 30/05/2011 – 28/06/2011)

The search for candidate peaks was carried out in the second prototype. It was determined that these peaks were a good baseline to identify actual peaks. Most of the functionalities that are provided by the application are based upon the location of the candidate peaks.

A slider was also added to the main window to increase or decrease the value of the threshold used to paint the map (see Figure 11). At this point, the slider was not used properly and it took a long time to see the new results once the value of the slider had been changed by the user. A further improvement was required.

An exploratory approach to detect actual peaks was carried out in this stage and it was done in relation to how steep the peaks were. The first calculations related to the contours were also made here. As we can see in the Figure 11, a contour has been drawn in white. The height of the contour and the prominence of the selected peaks were given to the user. However, the selected peaks were not shown and so the user had to remember the peaks that he had selected. This problem was fixed on the next prototypes.

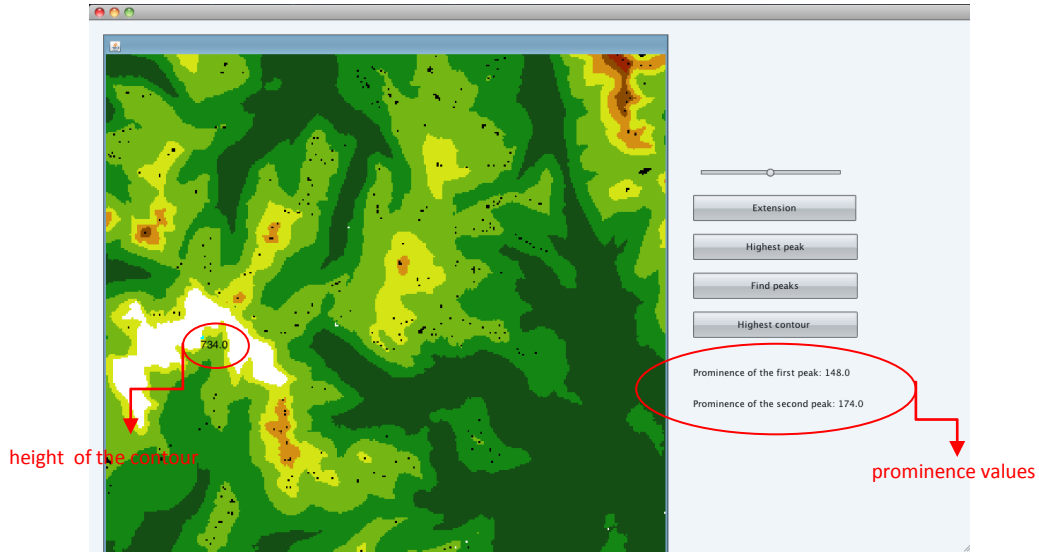


Figure 11. Snapshot of the second prototype. The white area represents the highest contour containing two points. The black points represent the candidate peaks. This was the state of the application the 28th of June.

3.3.3.3. Prototype 3 (Period: 29/06/2011 – 28/07/2011)

The main characteristic of the third prototype is that the different classifier methods were added to the applications. Not only the one based on the steepness factor but also the other three methods (see 4.3.7.1 to 4.3.7.4). The combination methods were also created (see 4.3.7.5) as well as two initial versions of the models (see 4.3.7.6). Note that these methods were activated and tested internally and therefore there is no button associated to them on the visualization window (Figure 12).

The algorithm to find the parent of a particular peak was created in this stage and the corresponding button was added to activate it.

At this stage, the application was also able to generate a list of peaks along with their closest higher peak. This list is stored internally because the output file functionality was added during the development of the last prototype.

As it can be seen below (Figure 12), a menu bar was also incorporated. This bar allows the user to change the dataset that is being analyzed.

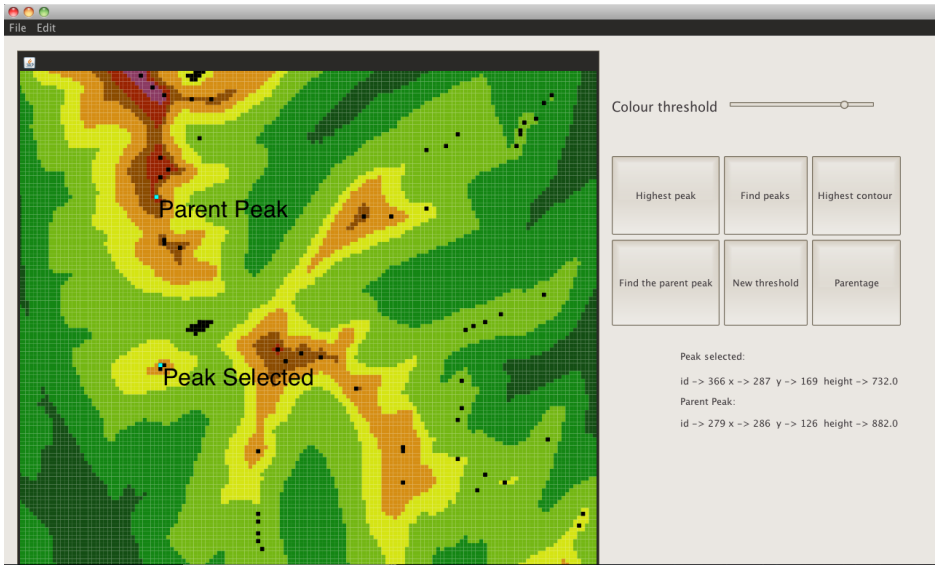


Figure 12. Snapshot of the third prototype. A peak has been selected and its corresponding peak has been found.

3.3.3.4. Final prototype (Period: 29/07/2011 – 18/08/2011)

The fourth or final prototype is the version of the application that has been submitted as part of the deliverables. In general terms, it consists of:

- An algorithm to generate the candidate peaks.
- An algorithm to obtain the main characteristics of the data: maximum height, minimum height, average height and the standard deviation.
- Four different algorithms to detect actual peaks in the set of candidate peaks. Each algorithm uses one of the properties that have been considered as relevant to the purpose:
    - Property 1: Steepness
    - Property 2: Prominence
    - Property 3: Horizontal distance to the closest peak
    - Property 4: Absolute height in relation to the average height of the surroundings.
- Disjunctive and conjunctive combination algorithms.
- Two models (see 4.3.7.6). Both models use sets of parameters whose optimal values were determined experimentally.
- An output file with the main activities performed along with the results of the executions.
- An algorithm to generate the record of peaks with the nearest higher peaks

- An algorithm to find the parent of each peak.
- An algorithm to generate the highest contour that contains two given peaks.
- An algorithm to obtain the candidate valley floors and potential rivers.
- Options to change the value of the default parameters.
- Option to change the data file under study.

All these functionalities can be accessed through the main visualization window (see Figure 24).

3.3.4 Phase 3: Testing and Evaluation

The testing stage has been carried out simultaneously with the development of the software application (see Section 5.2). Not only the classifier methods were tested but also the visualization of the results by comparing the obtained maps with actual maps. However, the evaluation stage was accomplished when the development stage had almost been finished. The accuracy of the models was measured by using different data sets (see Chapter 6). The ability of detecting the highest points of the Lake District was also studied.

4. DESIGN AND IMPLEMENTATION

4.1 Technology

In this section some technological aspects are commented upon regarding the programming language chosen and the libraries that have been used.

4.1.1 Programming Language

The application created in this project has been developed in Java. Java was chosen due to its platform independency and the ease with which one can deal with graphical user interfaces. It is object oriented which favours the creation of interrelated data structures and provides a way of establishing dependencies and hierarchy between objects. The rich set of Java libraries which make the work easier was another reason.

4.1.2 Java Libraries

Two Java libraries have been used in this project. The first one is called Substance. Substance is a toolkit of type "Look and Feel" whose aim is to change the visual appearance of graphic user interfaces. These kind of libraries allow us to obtain more professional, modern and attractive styles on our applications changing the default aspect provided by Java Swing. The main advantage is that these changes can be carried out by adding a couple of code lines. Substance stands out from other similar libraries in that it counts on a set of different skins that can be applied in a very easy way.

The second one is called Piccolo [37]. Piccolo is a visualization toolkit created by the University of Maryland that allows us a way to create complex graphic applications in Java. Piccolo makes it easy to draw on screen, to control the elements that have already been drawn and to control the interaction with the user's mouse. It also provides a set of appealing effects such as animation, events management and the use of zoomable user interfaces (ZUI). ZUI is a type of interface that allows the user to zoom in, to get more detail, or zoom out to get an overview of the elements that make up the canvas provided by ZUI (see Figure 13). The main advantage is that we can do this without worrying about the low level details.

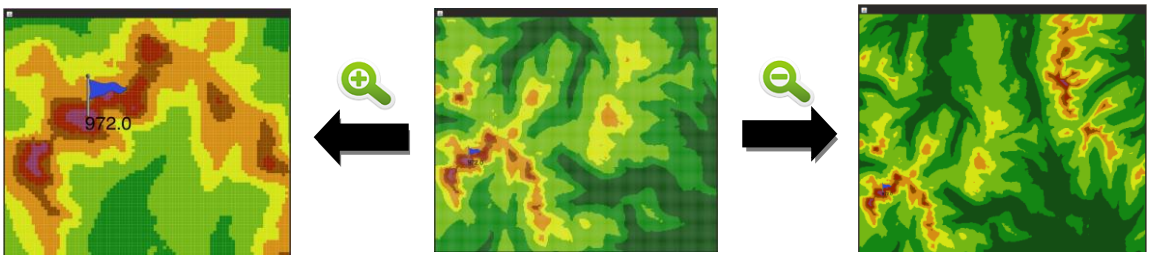


Figure 13. Zoom-in, Zoom-out

4.2 Design

4.2.1 Packages structure/ Classes

The internal structure of the application is composed of four different packages which contain classes according to their functionality. They are explained individually as follows:

- **Package GUI**: All the classes related to the visualization are included in this package. Not only the JFrame that appears in the execution of the software but also the class that connects the application with Piccolo.
- **Storage structures**: It contains the classes required to store the information related to the peaks. There is a class *'Peak'* which contains mainly the locations which a peak consists of. Another important class in this package is the class *'Peakness'* which manages large collections of peaks.
- **Engine:** It contains two relevant classes. The first one deals with the parsing of the file. This class is the responsible for transforming the vast information contained in the input file into an ordered set of numbers. The second one contains all the algorithms and the auxiliary methods needed. It extends PCanvas that combined Piccolo with Java Swing.
- **Lists of Wainwrights:** In order to carry out the testing and evaluation stages, information regarding the real peaks had to be stored. The *Wainwrights* are the fells of the Lake District in northwest England detailed by A. Wainwright [46]. One of the lists was used in the testing and development stage whereas the other remaining two were used only in the evaluation stage.

The relation of the packages can be seen below (Figure 14). As one can see, the package *GUI* only has access to the package *Engine* which has access to any other package. The package that contains the lists of actual peaks only makes use of the storage structures which does not need access to any package.

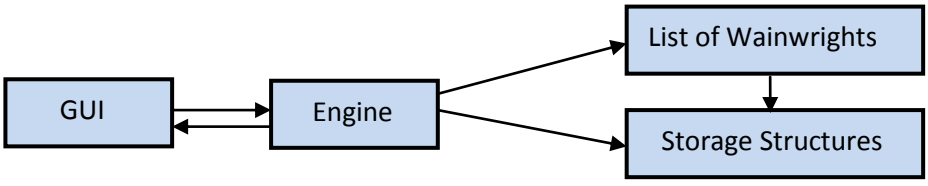


Figure 14. Packages structure

4.2.2 Input/ output data

As it was stated before, the application requires height data to be processed. The files used contain 160801 values which are divided into 401 sets by a new line character. Each set also contains 401 values separated by a space character and which are associated to points with the same latitude.

The file has to be chosen by the user and he or she is able to change it easily to a new data file. An output file is generated at the end of every execution. The main actions carried out by the user, together with the main results are saved in this file.

4.3 Functionalities

The developed application presents a set of functionalities which are explained with detail below.

4.3.1 Setting the colour threshold

By using a slider the user is able to change the threshold that adjusts the colours associated with every point of the map (see Figure 15). It can be useful to study more deeply the topographic profile of the terrain. The range of colours used (see Figure 16) starts with blue (points of zero height) and ends with purple (highest points).

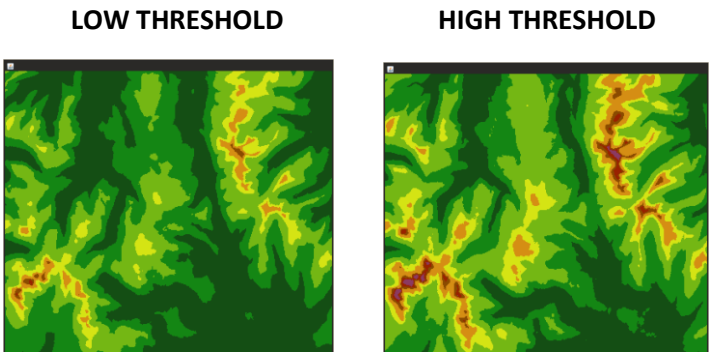


Figure 15.Maps of the same terrain but drawn using different colour threshold.

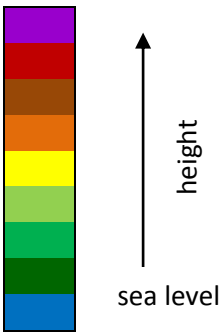


Figure 16. Range of colours used

4.3.2 Main characteristics of the data

The user might be interested in knowing the average profile of the dataset. The application developed in this project provides the user with general information about it. The largest and lowest heights are given as well as the average and the standard deviation. The standard deviation shows the dispersion of the data from the average. A low deviation means that the values of the points are close to the average and therefore the terrain is even. On the other hand, if the terrain is hilly or mountainous, the standard deviations will be large. For instance, an average of 378 and deviation of 204 is obtained by analyzing the sample dataset. However when analysing other datasets such as the one that contains the far Eastern fells of the

Lake District, the standard deviation presented is 134 metres and an average of 359. Note that the sample dataset that has been mentioned corresponds to the most mountainous area of the Lake District.

4.3.3 Candidate peaks

Every further analysis that can be carried out by using the application is based on the candidate peaks of the area under study. The search of candidate peaks has been limited to areas whose height is higher than the following threshold:

$$threshold = minh + \frac{2}{5} \bullet (maxh + minh)$$

Where *maxh* and *minh* are respectively the maximum and minimum height of any cell in the dataset.

In the data set used as sample the value of this threshold is 402 metres height. The selection of this threshold was determined by experimentation. All the analysis is carried out based upon a set of candidate peaks. This set is obtained by selecting those cells whose height is greater than the previously mentioned threshold and also greater or equal than the height of its neighbour cells. Once this is done, a merge process is required. This is basically the merge of neighbour cells which have the same height. In the left of both Figure 17 and 18, the grey cell and the black one represent different peaks that become into one after merging. Recall that the merging is carried out not only when the cells are adjacent but also when they are diagonal neighbours (see Figure 18).

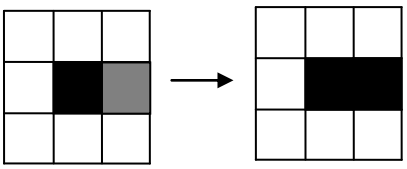


Figure 17. Merging of adjacent cells

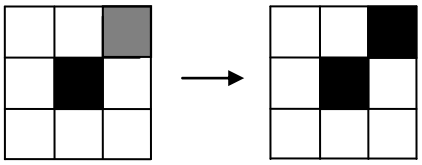


Figure 18. Merging of diagonal neighbours.

Note that the set of actual peaks is a subset of the candidate peaks set.

4.3.4 Highest contour including two peaks

A contour map is a tool that provides us information about the terrain profile. Given two different peaks, one might be interested in knowing what the highest contour containing both peaks is. The contour is obtained based on the height of the saddle point between two peaks. The saddle point is the lowest point of the highest route from one peak to the other. The height of the set of points included in the contour is higher or equal than the height of the saddle point.

The prominence of one peak can be calculated in relation to any other peak. In this case, the prominence is thought as the difference in height between the peak and the saddle point obtained by calculating the highest contour containing both peaks.

The application obtains such contour as follows:

1) Selection of the pertinent peaks.

2) Calculation of the saddle point.

3) Calculation of the extent of the contour.

The user is the one responsible for selecting the peaks. Once the peaks are chosen, the saddle point is assessed. This is done by using two different sets of locations. Initially, one set containing the location associated to one of the peaks and the other set containing the location of the second peak. The size of the sets increases by adding lower locations until both sets meet at one point which is the saddle point. Hereafter, the pseudo-code is presented:

```
Peak1 := peak selected(); Peak 2:= peak selected();
Creation and initialization of the corresponding sets.
while (!commonPoint)
        Locations are added to each set.
Initialization of the set that represents the contour
while (!contour line achieved)
        locations are added to the contour.
        Only those locations whose height is higher than the height of the common point
```

4.3.5 Parent peaks

Identifying the parent peak of a particular peak in a terrain is not a straightforward task especially when there are several higher peaks in its surroundings. The parent of a peak is the higher peak whose base contour surrounds the peak and no other peak. It is obtained by calculating the key pass between the peak and any other peak. The highest key pass is selected and the parent peak will be the peak that produced the highest key pass. Estimating the amount of close peaks to be studied is not easy either. How close do they have to be? Should the peaks that are very far away from the peak, which is under study, be considered?

When determining the parentage, the system only takes into consideration a pair of peaks whose distance from each other is not larger than two kilometres. In the case the parent of a peak is not found, the peak is considered as the parent of all the hierarchy of lower peaks. Only the peaks within the region enclosed by a circle of radius 2 kilometres centred at the peak are studied. The size of this region has been chosen experimentally based on the quality of the results and the running time required.

A peak can have at most one parent and any number of children. The children of one peak may also be the parents of their own child peaks. Therefore, a peak might be the grandparent of other peaks establishing a hierarchy between peaks (see Figure 19). It should be remembered that a peak only has one parent peak whereas it can be the parent of many peaks.

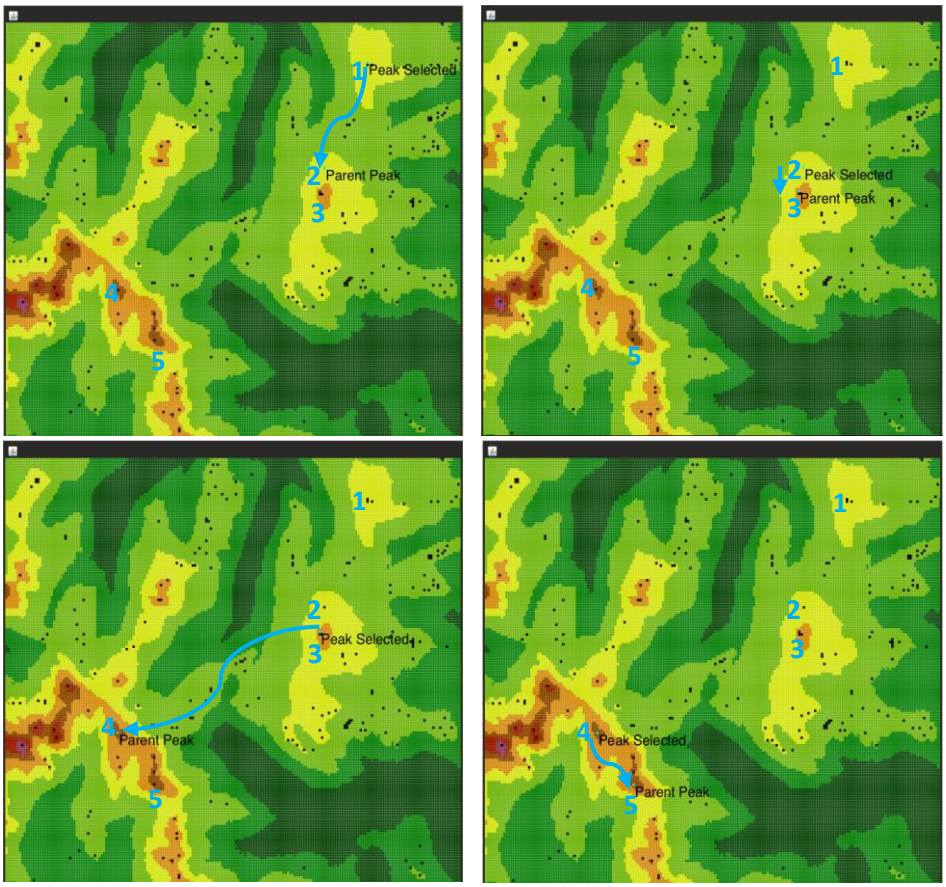


Figure 19. These images show the hierarchy of peaks by starting with the peak labeled with 1. The following chain 1 ->2->3 -> 4-> 5 is obtained where the peak 5 is the parent of the peak 4; the peak 4 is the parent of the peak 3 and so on.

4.3.6 Closest higher peak

Depending on the characteristics of the terrain examined and the radius of exploration, the calculation of the parent peaks might take a long time. Therefore, the system also provides a record of proximities between mountains. In this record each peak is listed along with its nearest higher peak. This report is given within the output file but the user can see these relationships by clicking on the corresponding buttons of the visualization window.

One might think that given a mountain, the closest higher peak is the parent peak. However, it is not correct. In many cases, the nearest peak to a particular peak whose height is higher is not its parent peak.

For instance, the figure 20 shows a typical situation with four different peaks. Suppose the peak 3 is being studied. Its nearest higher peak is the number 1. However its parent peak is the number 2. The orange contour line containing both peaks 2 and 3 is associated to a higher height than those represented by the purple contour line containing peak 3 and peak 1 (and also 2).

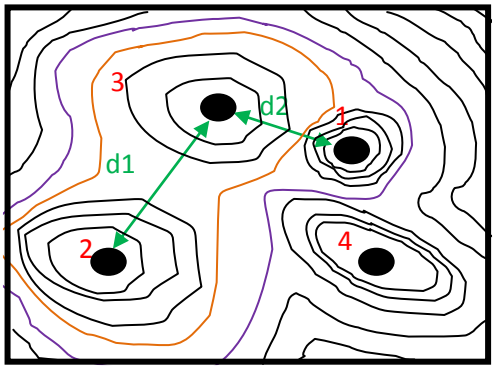


Figure 20. Contour map showing the difference between the parent peak and the closest higher mountain.

4.3.7 Classifier factors

The set of candidate peaks is generally too large and contains many locations that do not correspond with actual peaks. Several factors have been studied with regards to relevant characteristics to the identifications of mountains.

4.3.7.1. Factor 1: Steepness

The steepness is a crucial factor that must be taken into account. The application developed considers a candidate peak to be steep if there is at least one neighbour cell whose difference in height between it and the candidate peak is larger than a particular threshold. See the pseudo code below:

```
For each peak p
    For each neighbour Cx of p
        if (height(p) – height(Cx)) > threshold
            p is selected
```

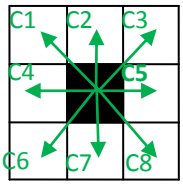


Figure 21.The candidate peak represented in black and its neighbours whose difference in height is checked.

4.3.7.2. Factor 2: Difference in prominence from its closest peak

As it was stated in the beginning of this report, the prominence is a significant factor that informs about how relevant a mountain is regarding the elevations that surround it. It is ever more important than the

absolute height. The application makes use of this factor and it calculates the difference in prominence between a particular peak under study and its closest peak. See pseudo code below:

```
for each peak p
    peak_aux := closestPeak(p)
    if (prominence(p)-prominence(peak_aux)) > threshold
        p is selected
```

4.3.7.3. Factor 3: Horizontal distance to a higher peak

A candidate peak situated very close to another peak of higher height is unlikely to be an actual peak. For instance, a set of candidate peaks might belong to the extent of a mountain but only one (the summit) is considered as peak. For instance, in the figure 22, a terrain profile with five potential peaks is showed. However, only the highest one has been registered as peak.

Figure 22. Terrain profile along with five candidate peaks

The distance between two peaks is measured as the Euclidean distance between them. In the same way as the previous factor, the distance value is also compared to a determined threshold. The value of this threshold might vary a lot if the window size is modified. See pseudo code below:

```
for each peak p
    peak_aux := closestHigherPeak(p)
    if (euclideanDistance(p,peak_aux)) > threshold
        p is selected
```

4.3.7.4. Factor 4: Absolute height in relation to the average height of the surroundings

The first tests carried out by using several combinations of factors showed that the peaks in highlands were well identified. However, those located in lowlands were very hard to identify. This suggested that a new measure should be added to the set of already considered factors. The new measure had to make it easier to detect the low peaks and so some considerations had to be made on the nature of the surroundings. Therefore the fourth factor was created regarding the absolute height of each peak and the average height of the locations included in an area of one kilometre square centred at the peak. This factor

on its own provides poor results but helps to detect low peaks when using the model (see 4.3.7.7). The model uses the difference in height between the summit and the average of the surroundings. A threshold is also used when the factor is used on its own.

4.3.7.5 Disjunctive and Conjunctive Combination

After using the factors individually and obtaining not very good results, different experimentations were made by combining two different factors. Although it has been proved that it is not the best way of detecting actual peaks, they system provides the possibility of carrying out the disjunctive and conjunctive combinations. The user can choose the two factors that he is interested in combining (see Figure 23).

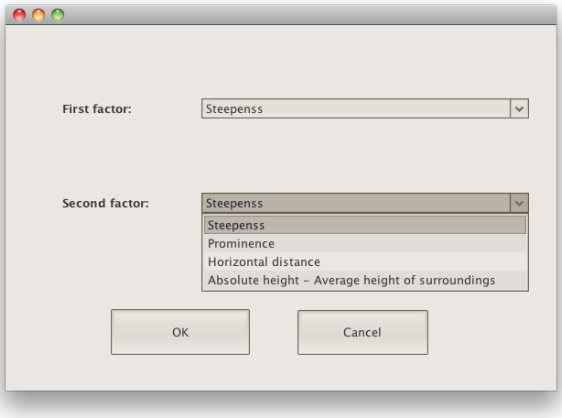


Figure 23. Window that is prompted when the user selects combining operations.

Disjunctive combination

The disjunctive combination of two factors represents the OR operation of the potential peaks selected by the methods that make use of them. The peaks which are selected as actual peaks are those that have been selected by at least one of the methods.

Conjunctive combination

The conjunctive combination of two factors represent the AND operation of the potential peaks selected by each method. Therefore, the peaks that are selected are those detected by both factors.

4.3.8. Global models

Global models are multiple combinations of factors to identify actual peaks within a bulk of candidate peaks. After experimenting (see section 5.3), two global factors are given to the user. These global models

can be selected by clicking on the buttons located on the main visualization window. The equation which characterizes the first model makes use of the prominence factor, the steepness factor as well as the horizontal distance factor. The relation between the parameters has been chosen by experimentation and it is the following:

$$promFact + (steepnessFact \bullet k_1) + (promFact \bullet horizDistanceFact \bullet k_2) > threshold$$

Where k_1 and k_2 and *threshold* are three parameters which can take any value. Note that the factor that represent the horizontal distance to a higher peak (*horizDistanceFact*) is multiplied by the prominence factor (*promFact*). Therefore, it will only have significance if the prominence is also high.

The second equation also uses the fourth factor (the factor that relates the absolute height with the average height of the surroundings) and a new parameter k_3 is added to the model. It is as follows:

$$promFact + (steepnessFact \bullet k_1) + (promFact \bullet horizDistanceFact \bullet k_2) - (distanceToAverage \bullet k_3) > threshold$$

Initially the parameters take the default values which are the values which provided best results. However they can be changed regarding the users needs.

4.3.9 Valleys, rivers and lakes

For the purposes of this dissertation an approach to detect valleys has been carried out using the created application. As seen in the image below (figure 24) the system is able to find candidate valley floors (dark blue areas). The valley floors are the lowest part of the valleys. The points that belong to the valley floors have been selected by comparing their height with the height of their neighbours. A cell is considered as part of a valley floor if its height is lower than or equal to the height of its neighbours. The light blue lines represent the path that the water would follow from the summit of every candidate peak to the floor of a particular valley. Therefore these blue lines might be interpreted as potential rivers. They have been calculated by taking the path of maximum slope from each peak until reaching a valley floor. To generate a candidate river, a peak is chosen as its start and the height of its neighbours will then be examined. The neighbour of lowest height will be selected, added to the path and its neighbours will examined. Iteratively, this process will be carried out until finding a cell that belongs to the sets of locations of valley floors. It should be noted that the set of valley floors is a subset of the candidate valley floors. Additional criteria are needed to identify which ones should be considered as valley floors. However, the size of an actual valley

floor might contain locations close to the dark blue areas. It depends on what criterion is chosen to determine the boundary between the valley floor and the rest of the valley. Again, vagueness is present.

This approach is not only useful to identify valleys and rivers but also lakes. As it can be seen in Figure 24, there are some dark blue areas whose height is not especially low. They can be considered as water containers. In fact, if one checks the actual locations of lakes in that area of the Lake District, there is a strong correlation between the dark blue areas and the lakes (see section 5.4.4).

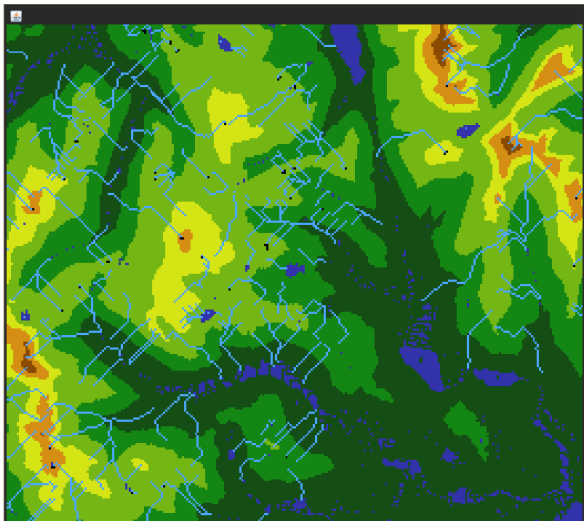


Figure 24. Image obtained by running the application. The light blue lines represent potential rives whereas the dark blue areas correspond to potential valley floors.

Delimitating the extent of a valley is a far more complicated task. An approach was taken regarding the delimitation of valleys but instead of using gradient notions as it has been used previously [42], the '*peakness*' surrounding the valleys has been utilised. Valleys are generally characterised by the mountains or elevations that encircle them and therefore it would be interesting to know what locations belong to the valley and which to the surrounding elevations. To carry out the search of the extent of a particular valley, the closest peak from the valley floor has to be found. Once this peak is located and its characteristic are known, the assessment of the extent is carried out. Starting from the valley floor, locations are added to the valley extent having into account their height. These locations will be added until achieving:

1) Half the height of the closest peak over the valley floor.
2) The height of the key pass between the peak previously found and its closest higher peak.

The size of the valley extent varies considerably depending on the terrain characteristic. For instance, when selecting the valley whose floor corresponds with the Thirlmere Lake, the extents can vary from the one that has been drawn in Figure 25 to the one in Figure 26.

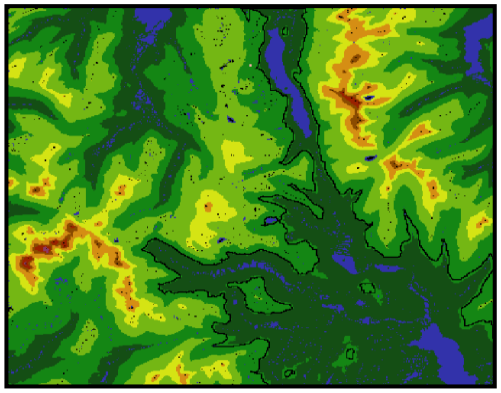


Figure 25. Map that shows the boundaries of the valley by using the first method.

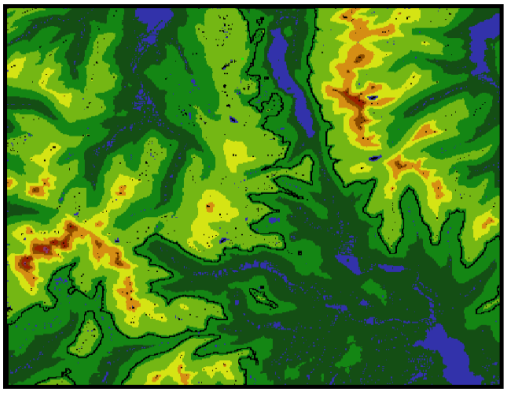


Figure 26. Map that shows the boundaries of the valley by using the second method.

As it can be seen in the maps above, there is a large low region which has been fully included in the area drawn in the Figure 26 and partially in the Figure 25. If the threshold used to delimitate the extent of the valleys was increased, three different valleys would be obtained in the way represented by the following image (Figure 27). This raises the question ´which threshold is the most suitable?´. By using Google Street View[3], the separations made by the red lines were studied. It was observed that both zones 1 and 2 seemed to be part of a big valley and in the same way, both zones 2 and 3 also seemed to be part of the same valley. It is obvious that this depends on the criteria one uses to decide where the boundaries of the valleys should be. One might say that this large valley is actually a main valley (zone number 2) with two side valleys (zones number 1 and 3).

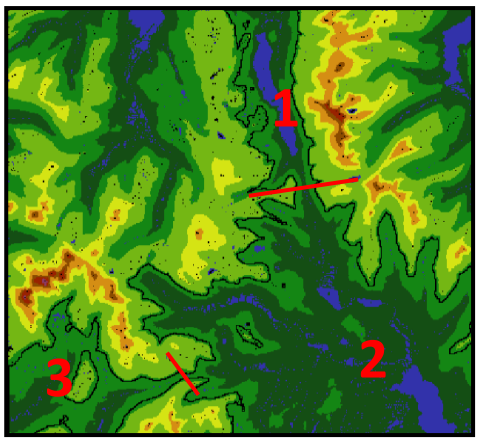


Figure 27. Three different valleys that might be obtained by changing the value of the threshold that demarcates the extent of the valleys.

[3] Website: www.**google**.com/**streetview**

4.4 Visualization window

The main visualization window can be seen below (Figure 28). It consist of a frame wherein the corresponding maps are shown as well as a set of buttons that are located on the right. These buttons can activate any of the classification methods, the drawing of the candidate peaks, the search of contours, search of parent peaks as well as the search of valleys, rivers and lakes. On the top part of the window, there is a menu with three different options. The first one is called *'File'* which allows the user to change the dataset that is going to be analysed. The second one is labelled as 'Edit' and is used to change the value of the classifier parameters. The last one corresponds with the help information that guides the user to make the most of the application.

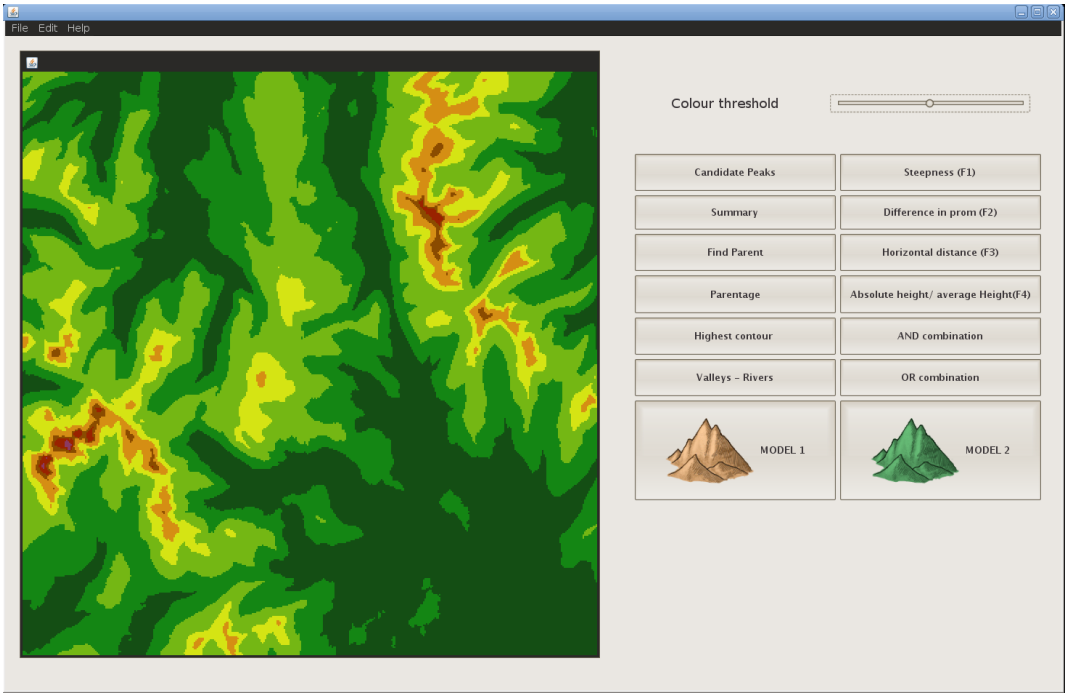


Figure 28. Final version of the main visualization window.

5. TESTING

5.1 Program details

The aim of the testing stage was to check how well the different classification techniques performed. It is a classification task and therefore the predicted class (selected peaks) should be compared with the actual class (actual peaks). It can be seen as a search of actual peaks in the search space (set of candidate peaks).

The testing process was carried out using the same dataset. The parameter values of the classifier models were established observing the terrain characteristics. The sample dataset was considered representative enough to get models which could be effectively extrapolated.

5.2 Testing plan

Due to the fact that the evolutionary prototyping [10] was chosen as main development methodology, the testing process was carried out at each stage of the development period. Each prototype was tested before continuing onto and creating the next prototype. When a new functionality was added to the current prototype, it was fully tested and its behaviour was analysed to avoid future problems.

5.3 Performance measures

The performance technique used is based on the distribution of elements of the search space in four different sets: true positives (TP), false positives (FP), true negatives (TN) and false negative (FN). Regarding this project, these concepts are defined as follows (see Figures 29 and 30):

- The set of true positives is the set of peaks that have been correctly selected as actual peaks.
- The set of false positives is the set of peaks that have been selected as actual peaks but they are not.
- The set of true negatives is the set of elements that have been selected correctly discarded as actual peaks.
- The set of false negatives is the set of elements that are actual peaks but have not been considered as such.

The elements that have been correctly classified belong to the set of true positives or to the set of true negatives whereas those that have been incorrectly classified belong to the set of false positives or false negative. The four sets can be explained in term of a confusion matrix (see Figure 30).

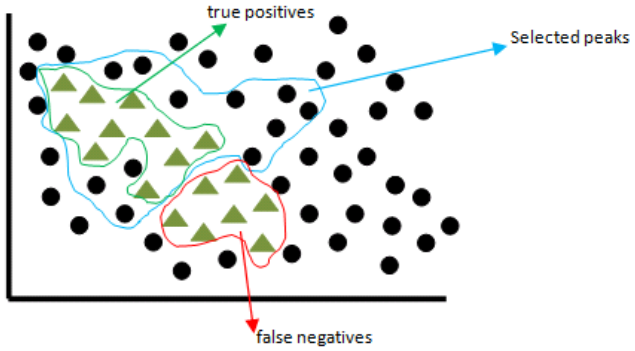


Figure 29. Search space. The green triangles represent the set of actual peaks whereas he remaining elements (black points) are the candidate peaks that are not actual peaks.

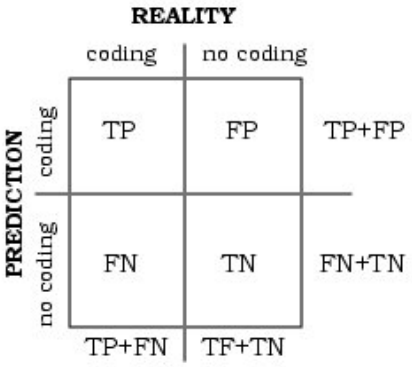


Figure 30. Confusion matrix or contingency table showing the relations between the four outcomes.

Precision, recall accuracy and F-measure are terms which evaluate the quality of the classification method. They relate to the value of the four previously mentioned sets (FP, TP, FN, TN).

In the context of the project, precision is the measure that determines the ability of the classifier to obtain peaks that are actual peaks. It is the percentage of actual peaks obtained in the set of selected peaks. The equation that describes the precision is as follows:

$$\text{Precision} = \frac{TP}{TP + FP}$$

The recall is the ability of a classifier method to detect all the actual peaks. Its equation is the following:

$$\text{Recall} = \frac{TP}{TP + FN}$$

The accuracy refers the number of elements (peaks) correctly classified in relation to the total number of elements (candidate peaks).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

The latter is the harmonic mean, called F-measure, that combines the previous concepts: the values of the precision and recall. The equation is as follows:

$$F = 2 \times \frac{precision \times recall}{precision + recall}$$

5.3 Performance results and determination of parameter values

In this section of the project, several result tables are presented showing the evolution of the performance. These tables have been used during the development of the prototypes and they have been very helpful to determine the models and to establish the parameter values that take part in them. Initially, poor results were obtained by using the corresponding factors individually. Then, the tables were updated regarding the results obtained by using the disjunctive and conjunctive combination of factors. The performance was still not good enough and therefore new combinations were considered. These combinations made use of parameters whose values were studied to create the models.

First of all, the characteristics of the dataset under study have to be given:

Number of candidate peaks = 503 (after merging, see section 4.3.3)

Number of actual peaks = 89

The three tables below represent the results obtained by using the parameters individually. They were obtained during the development of the third prototype. The first one was obtained by using the method that made use of the *steepness factor* (see section 4.3.7.1), the second one is related to the *difference in prominence factor* (see section 4.3.7.2) and the last one was created in relation to the *horizontal distance factor* (see section 4.3.7.3).

Threshold	True Positives (tp)	True Negatives (tn)	False Positives (fp)	False Negatives (fn)	Precision	Recall	F	Accuracy
15	75	145	274	14	0.215	0.843	0.342	0.433
20	54	216	203	35	0.210	0.607	0.312	0.531
25	39	284	135	50	0.224	0.438	0.297	0.636
28	31	320	99	58	0.238	0.348	0.283	0.691
30	27	337	82	62	0.248	0.303	0.273	0.717
32	21	350	69	68	0.233	0.236	0.235	0.730
33	21	359	60	68	0.259	0.236	0.247	0.748
34	19	361	58	70	0.247	0.213	0.229	0.748
35	19	367	52	70	0.268	0.213	0.238	0.760
Max	0	419	0	89	-	-	-	0.825

Table 1. Results obtained by using the *steepness factor* (First factor).

Threshold	True Positives (tp)	True Negatives (tn)	False Positives (fp)	False Negatives (fn)	Precision	Recall	F	Accuracy
3	22	254	165	73	0.118	0.232	0.156	0.537
5	19	261	158	76	0.107	0.200	0.140	0.545
10	17	320	99	78	0.147	0.179	0.161	0.656
15	4	388	31	91	0.114	0.042	0.062	0.763
20	0	410	9	95	-	-	-	0.798
Max	0	419	89	0	-	-	-	0.825

Table 2. Results obtained by using the difference in prominence factor (Second factor)

Threshold	True Positives (tp)	True Negatives (tn)	False Positives (fp)	False Negatives (fn)	Precision	Recall	F	Accuracy
30	42	251	168	47	0.200	0.472	0.281	0.577
40	33	277	142	56	0.189	0.371	0.250	0.610
41	33	281	138	56	0.193	0.371	0.254	0.618
42	33	284	135	56	0.196	0.371	0.257	0.624
50	32	293	126	57	0.203	0.360	0.259	0.640
51	32	294	125	57	0.204	0.360	0.260	0.642
Max	0	419	0	89	-	-	-	0.825

Table 3. Results obtained by using the *horizontal distance factor* (Third factor).

If one looks at the last row of each table one realises that the best accuracy is obtained when any peak of the search space is considered as an actual peak. This means that using only one parameter will produce poor results as the amount of incorrectly classified peaks will be larger than the amount of correctly classified peaks.

The results obtained by the disjunctive and conjunctive combination of factors were similarly insufficient. The disjunctive combination provided a high number of identified peaks. However, the set of selected peaks was also very big and therefore the global accuracy was poor. Regarding the conjunctive combination, the results were a little better. Although it was simple to get a good precision, it was difficult to obtain a good recall.. A good recall was only obtained when the number of selected peaks was very large and therefore the precision was very low. Putting it simply, the higher the recall, the lower the precision. In the same way, the higher the precision, the lower the recall.

Since poor results were obtained by applying the previous methods, different multiple combinations of factors were used. The best combinations were summarized in the two given models (see section 4.3.7.6).

As it was stated, the first model makes use of three parameters: k_1, k_2 and the threshold.

The best values for those parameters were obtained after a long period of experimentation. Lots of different sets of values were used until selecting those that provided the highest value of accuracy. The table below shows the results obtained for some parameter values. The row painted coloured in green presents the parameter values that provide the best F-measure and accuracy. Therefore the default values are 0.4 and 0.8 for k_1 and k^2 respectively and 20000 for the threshold.

Note that the result tables given in this section are all of them obtained by analyzing the sample dataset.

k_1	k_2	Threshold	TP	TN	FP	FN	Precision	Recall	F	Accuracy
0.4	4	2400	74	281	133	15	0.357	0.831	0.5	0.706
0.4	4	8800	65	369	45	24	0.591	0.730	0.653	0.863
0.4	4	24000	45	403	11	44	0.804	0.506	0.621	0.891
0.4	4	25200	43	403	11	46	0.796	0.483	0.601	0.887

0.4	4	28400	40	407	7	49	0.851	0.449	0.588	0.889
0.4	5	24000	46	402	12	43	0.793	0.517	0.626	0.891
0.4	5	20000	53	395	19	36	0.736	0.596	0.658	0.891
0.7	4	20000	58	382	32	31	0.644	0.652	0.648	0.875
0.7	4	24000	54	391	23	35	0.701	0.607	0.651	0.885
0.4	6	24000	51	401	13	38	0.797	0.573	0.667	0.899
0.4	7	24000	54	396	20	34	0.750	0.607	0.671	0.895
0.4	8	24000	55	394	20	34	0.733	0.618	0.671	0.893
0.4	**8**	**20000**	60	392	22	29	0.732	0.674	**0.702**	**0.899**
0.4	8	15000	63	377	37	26	0.630	0.708	0.667	0.875
0.4	8	18000	62	388	26	27	0.705	0.697	0.701	0.895

Table 4. Some results obtained by using the first model. The green row is the one that presents the best results.

After experimenting with the first model, the author of this dissertation realised that the peaks located in low areas were very difficult to identify. Although almost every peak in the highlands were identifiable, only a few in the lowlands were identified as actual peaks. Therefore another factor was required to get better results. To improve the results for the second model, the author of this dissertation decided to add a factor that relates the height of a peak with the average height of its surroundings (see section 4.3.7.4). The results can be observed in the table below (Table 5) and again the row that presents the best results is coloured in green. The parameter values in this row are the default values of the model.

k_1	k_2	K_3	Threshold	TP	TN	FP	FN	Precision	Recall	F	Accuracy
5	7	40	10000	51	403	11	38	0.823	0.573	0.675	0.903
5	7	30	10000	52	401	13	37	0.800	0.584	0.675	0.901
5	7	30	8000	56	399	15	33	0.789	0.629	0.700	0.905
5	7	30	12000	49	402	12	40	0.803	0.551	0.653	0.897
5	7	20	12000	51	402	12	38	0.810	0.573	0.671	0.901
5	7	15	12000	53	401	13	36	0.803	0.596	0.684	0.903
5	**8**	**5**	**12000**	58	396	18	31	0.763	0.652	**0.703**	**0.903**
12	7	30	10000	57	395	19	32	0.750	0.640	0.691	0.899
5	7	30	7000	57	392	22	32	0.722	0.640	0.679	0.893
5	9	30	7000	60	387	27	29	0.690	0.674	0.682	0.889
5	7	40	8000	53	398	16	36	0.768	0.596	0.671	0.897
8	7	30	8000	57	393	21	32	0.731	0.640	0.683	0.895
8	9	30	10000	60	392	22	29	0.732	0.674	0.702	0.899
8	9	30	7000	61	379	35	28	0.635	0.685	0.659	0.875

Table 5. Some results obtained by using the second model. The green row is the one that presents the best results.

5.4 Visualization results

The software application created in this project is a visualization tool that provides information about a particular terrain. Getting a suitable visualization of the datasets and the corresponding results is essential to the success of the application.

5.4.1 Candidate peaks

The best way to check how good the visualisation results are is to compare the maps generated by the application with actual maps. The main aspect to check is whether the distribution of candidate peaks is similar to the distribution of actual peaks. To that end, the produced maps were exhaustively analyzed and the actual peaks were uniquely identified and associated to points on the map.

An example is given below (Figure 31) through two pictures. The one on the left corresponds to a map produced by the application and the second one is an actual map of the central area of the Southern Fells in the Lake District. In the picture of the left, the peaks that are pointed with arrow represent the actual peaks which can be seen in the picture of the right.

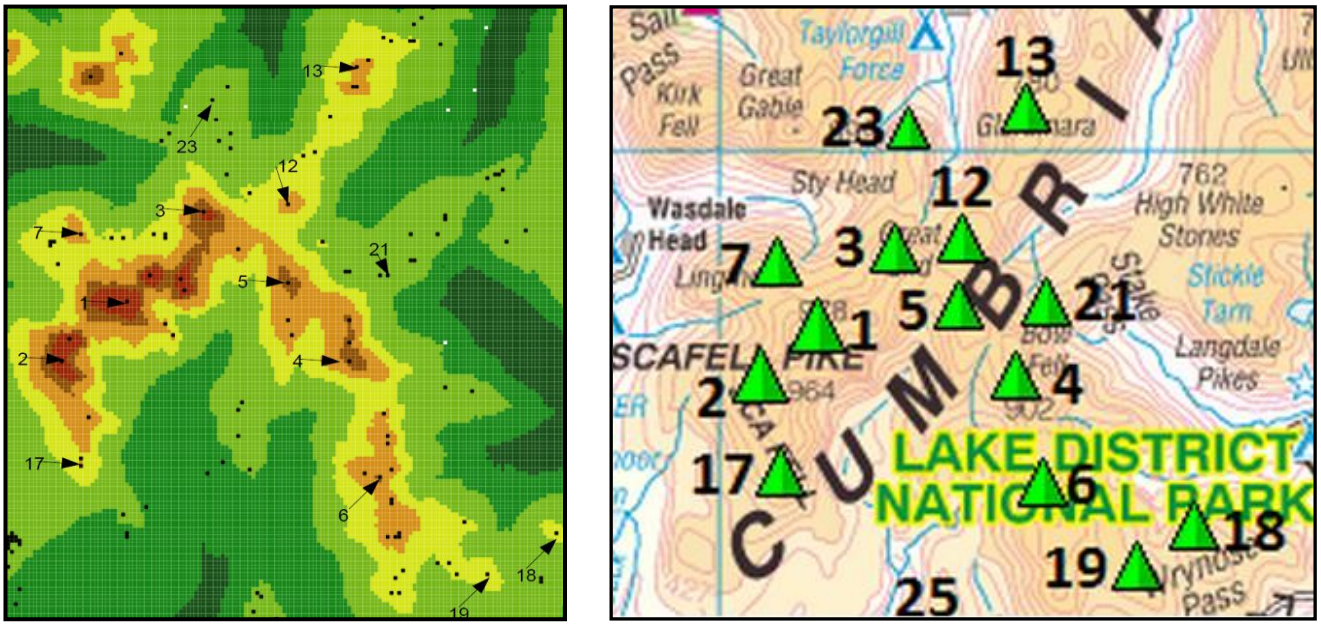


Figure 31. Comparison between a generated map and an actual map. They represent the central area of the Southern Fells in the Lake District. The peaks labeled with a '1' is the Scafel Pike.

5.4.2 Correctly and incorrectly classified peaks

After a correct visualization of peaks, a way of representing the set of selected peaks was required. Knowing what peaks were incorrectly selected as actual peaks was especially necessary throughout the testing stage. I was useful to know which parameter value had to be increased or decreased. when display-

ing the result of a classifier method, the peaks that have been correctly classified are coloured in cyan and the incorrectly classified peaks are colour in red (see Figure 32).

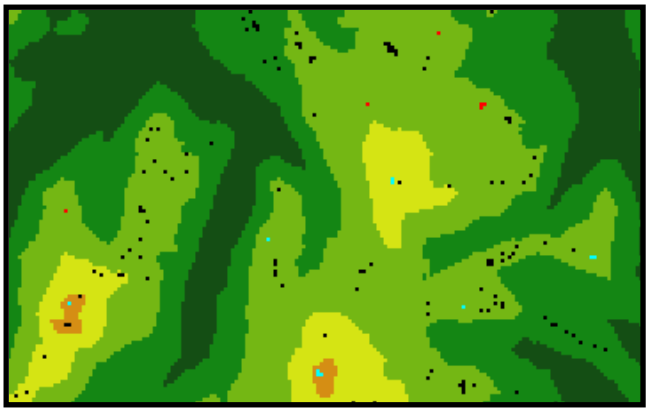


Figure 32. Region of the map showing the six correctly classified peaks (cyan colour) and four that have been incorrectly classified (red colour).

5.4.3 Contours

Regarding the drawing of contours, there are two different options of visualization. The user is able to choose between them based on his preference. The first option (see Figure 11) colours any cell included in the contour whereas the second option (see Figure 33) only colours its boundaries. These drawing were considered to be clear enough to inform the user about the location and shape of the corresponding contour.

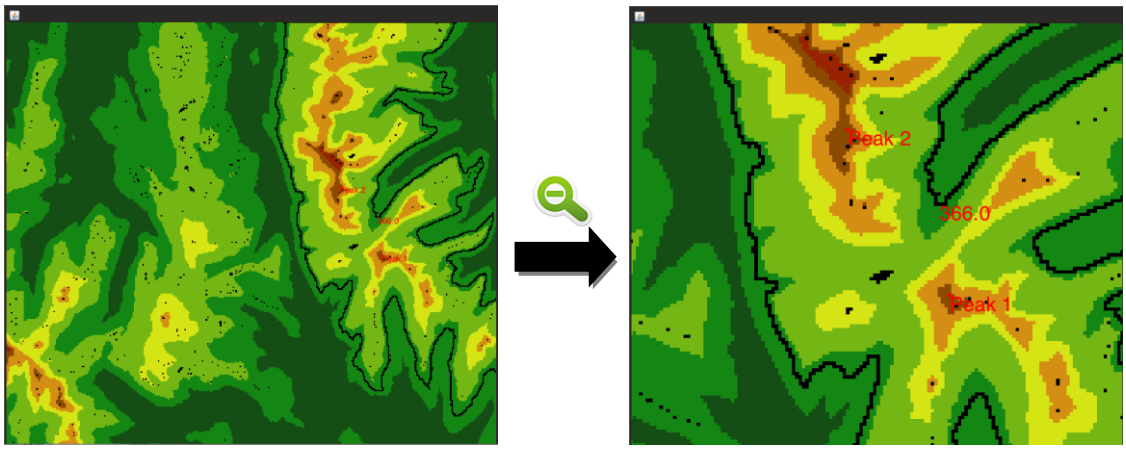


Figure 33. Contour drawn using the technique that only colours the contour boundary. On the image shows the complete map whereas on the left the image has been obtained by zooming-in to get a more detailed image of the region of interest.

5.4.4. Lakes

Lakes are bodies of water that presents a surface of uniform height. Most lakes are fed and drained by rivers and streams and the natural ones are generally located in mountainous areas but they can also be found in basins where the water does not have access to the sea. As it was commented before (see section

4.3.7.7), some large areas of same height were found in the sample dataset. Most of them were the end of potential rivers whose origins were situated at candidate peaks. These areas were identified on maps as actual lakes. The image below (Figure 34) shows a portion of the sample dataset with the named lakes that have been identified. Note that there is one lake whose name was unknown but which could be seen on maps. Appendix 6 contains two maps whose lakes have also been identified. These maps represent the two datasets used for evaluation.

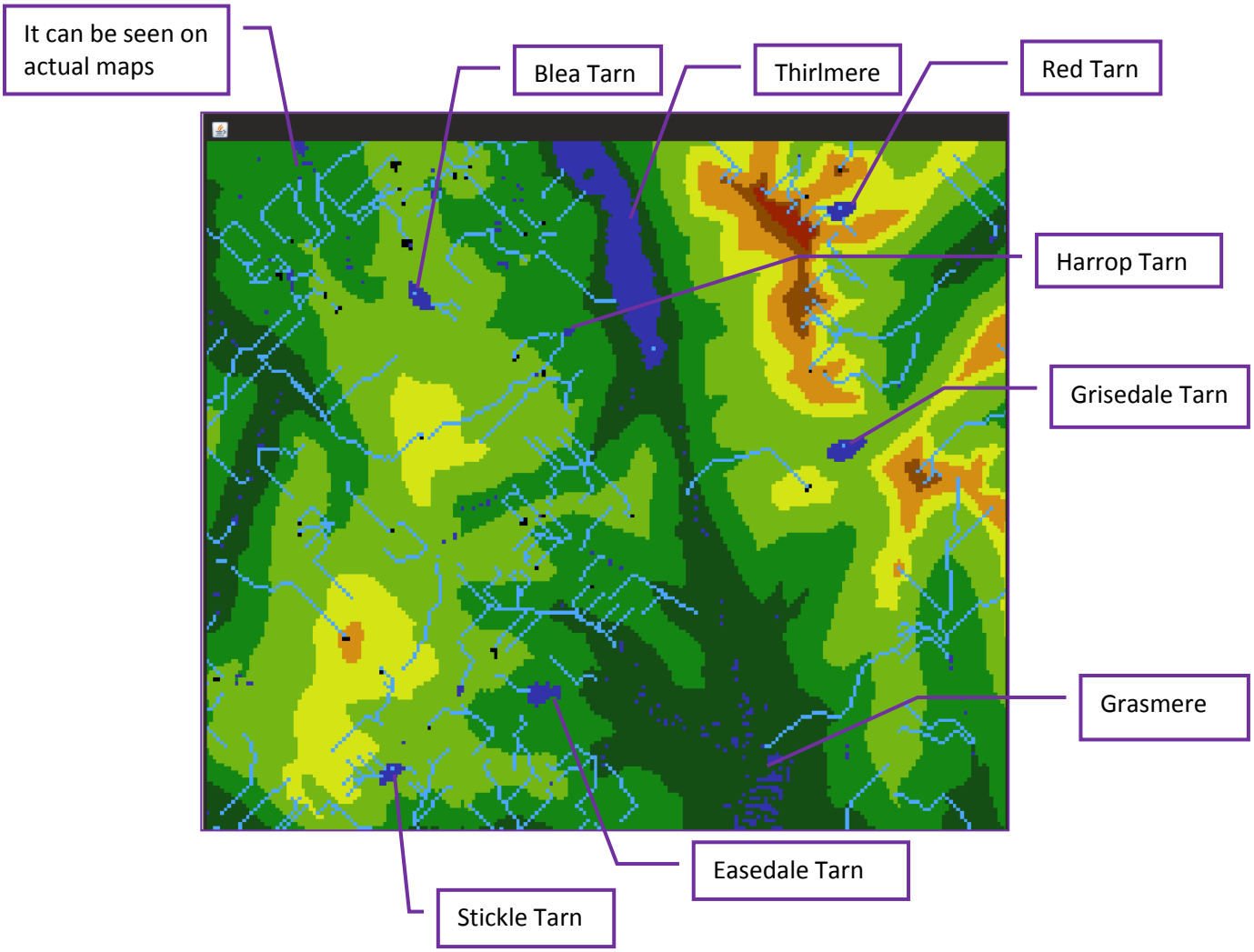


Figure 34. Picture of the map generated by the application that shows the relationship between the found areas and the actual lakes.

6. PROJECT EVALUATION

6.1 Comparison with minimum requirements

This section provides a comparison between the minimum requirements that were established in the beginning of the project and the corresponding justification of their achievements . See the table below:

REQUIREMENT	JUSTIFICATION
Analysis of relevant geometric and other characteristics that may be used to identify landform features.	Chapter 2 describes the concept of vagueness and analyses the different approaches that have been taken to deal with it. Relevant characteristics have been studied mainly regarding the identification of mountains (see section 2.4).
Design of algorithms based on relevant characteristics.	Chapter 4 presents the different algorithms that have been created and it describes the functionalities that the software application created provides.
A tool for the visualisation of results.	The application created is constituted by a set of functionalities whose results can be visualised. To prove this, snapshots have been presented in this dissertation. This tool provides the user with several methods to study the terrain characteristics.
Analysis of behaviour using real elevation data.	The behaviour of the application has been analysed not only during the development stage using the sample dataset but also during the evaluation process (see section 6.2) using other datasets.

Table 6. A table gives a comparison between the results obtained regarding each requirement.

6.2 Evaluation of the final prototype

This section describes the evaluation of the product produced. Evaluation describes how well the software developed performs. The evaluation of the final prototype has been carried out by using different datasets and by observing the results of the two models that were created. The parameter values that are involved in the models were fixed in relation to the sample dataset therefore the results obtained using

other datasets are likely to be poorer. The results of the models on two datasets have been registered and they are as follows:

- **First dataset:**

  Candidate peaks = 190; Number of peaks = 30

	Model 1	Model 2
Selected peaks	33	38
TP	17	20
TN	147	142
FP	13	18
FN	13	10
Precision	0.567	0.526
Recall	0.567	0.667
F	0.567	0.588
Accuracy	0.863	0.853

Table 7. Comparison of results of the first dataset by using the models.

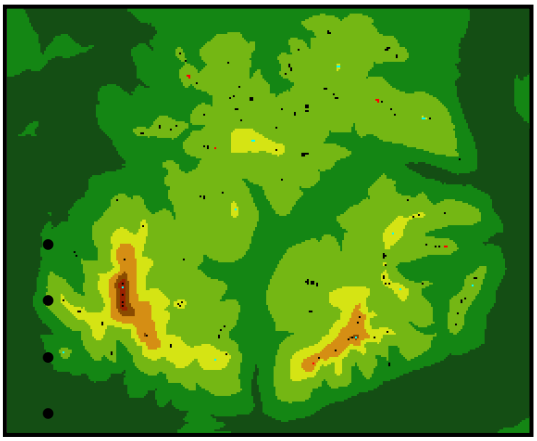


Figure 35. Area central of the first dataset. The Northern Fells of the Lake District.

- **Second dataset:**

  Candidate peaks = 256

  Number of peaks = 33

	Model 1	Model 2
Selected peaks	20	35
TP	15	19
TN	218	207
FP	5	16
FN	18	14
Precision	0.750	0.543
Recall	0.455	0.576
F	0.566	0.559
Accuracy	0.910	0.883

Table 8. Comparison of results of the second dataset by using the models.

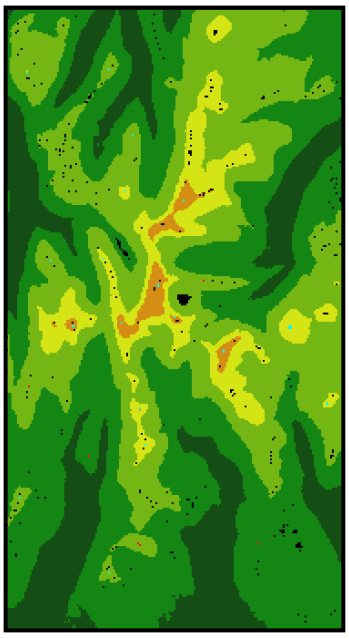


Figure 36. Image corresponding to the second dataset.. The Far Eastern Fells of the Lake District.

These results will be discussed in the following section (see 6.3).

Another way of evaluating the models is by checking if the most important peaks of a particular terrain are identified. The most important peaks in a terrain are usually the highest ones and so the easiest to

identify as actual peaks. The datasets that have been studied covered almost $10m^2$ and they contain most of the highest peaks. The 50 highest peaks are listed in the table below (Table 9) which peaks are included were included in the datasets under investigation. If they have been included, the table shows whether or not they have been correctly classified by the models. Note that both models present the same result. 35 out of 50 highest peaks were in the examined datasets, of which 28 out of 35 have been correctly classified as actual peaks. The misclassification of the remaining seven peaks is likely due to their close proximity to a higher, correctly classified peak, which they were classified by the models.

D = Development dataset
E1 = First evaluation dataset
E2 = Second evaluation dataset
NS = Not studied

		IDENTIFIED	SET
1	*Scafell Pike, 978 m (3,210 ft)*	yes	D
2	*Scafell, 965 m (3,162 ft)*	yes	D
3	*Helvellyn, 951 m (3,118 ft)*	yes	D
4	*Skiddaw, 931 m (3,054 ft)*	yes	E1
5	*Great End, 910 m (2,986 ft)*	yes	D
6	*Bowfell, 902 m (2,960 ft)*	yes	D
7	*Great Gable, 899 m (2,949 ft)*	yes	D
8	*Pillar, 892 m (2,926 ft)*	-	NS
9	*Nethermost Pike, 891 m (2,923 ft)*	yes	D
10	*Catstycam, 889 m (2,917 ft)*	no	D
11	*Esk Pike, 885 m (2,903 ft)*	yes	D
12	*Raise (Lake District), 883 m (2,896 ft)*	yes	D
13	*Fairfield, 873 m (2,863 ft)*	yes	D
14	*Blencathra, 868 m (2,847 ft)*	yes	E1
15	*Skiddaw Little Man, 865 m (2,837 ft)*	no	NS
16	*White Side, 863 m (2,831 ft)*	no	D
17	*Crinkle Crags, 859 m (2,818 ft)*	yes	D
18	*Dollywaggon Pike, 858 m (2,815 ft)*	yes	D
19	*Great Dodd, 857 m (2807 ft)*	-	NS
20	*Grasmoor, 852 m (2,795 ft)*	-	NS
21	*Stybarrow Dodd, 843 m (2,772 ft)*	yes	D
22	*St Sunday Crag, 841 m (2,759 ft)*	yes	D
23	*Scoat Fell, 841 m (2,759 ft)*	-	NS
24	*Crag Hill, 839 m (2,753 ft)*	-	NS
25	*High Street, 828 m (2,717 ft)*	yes	E2
26	*Red Pike (Wasdale), 826 m (2,710 ft)*	-	NS
27	*Hart Crag, 822 m (2,697 ft)*	no	D
28	*Steeple, 819 m (2,687 ft)*	-	NS
29	*Lingmell, 807 m (2,648 ft)*	yes	D
30	*High Stile, 807 m (2,648 ft)*	-	NS
31	*Old Man of Coniston, 803 m (2,635 ft)*	-	NS
32	*Kirk Fell, 802 m (2,631 ft)*	yes	D
33	*High Raise, 802 m (2,631 ft)*	yes	E2
34	*Swirl How, 802 m (2,631 ft)*	yes	D
35	*Green Gable, 801 m (2,628 ft)*	no	D
36	*Dove Crag, 792 m (2,598 ft)*	no	D
37	*Rampsgill Head, 792 m (2,598 ft)*	no	E2
38	*Grisedale Peak, 791 m (2,595 ft)*	-	NS
39	*Great Carrs, 785 m (2,575 ft)*	yes	D
40	*Allen Crags, 784 m (2,572 ft)*	yes	D
41	*Thornthwaite Crag, 784 m (2,572 ft)*	yes	E2
42	*Glaramara, 783 m (2,569 ft)*	yes	D
43	*Kidsty Pike, 780 m (2,559 ft)*	no	E2
44	*Dow Crag, 778 m (2,552 ft)*	-	NS
45	*Harter Fell, 778 m (2,552 ft)*	yes	E2
46	*Red Screes, 776 m (2,546 ft)*	yes	D
47	*Grey Friar, 773 m(2,536 ft)*	yes	D
48	*Sail, 773 m (2,536 ft)*	-	NS
49	*Wandope, 772 m (2,533 ft)*	-	NS
50	*Hopegill Head, 770m (2,526 ft)*	-	NS

Table 9. The highest 50 peaks of Lake District. The rows coloured in white contain the peaks that are beyond the boundaries of the studied maps and those coloured in red represents the peaks which have not been classified as actual peaks.

As many peaks in the evaluation datasets are not listed above (they are not among the 50 highest), the following tables (Table 10 and 11) have been created showing the highest peaks of those datasets. In both cases, only 10 out of 15 peaks have been correctly classified. If the recall was increased, a larger

amount of peaks would have been identified, but the precision and the accuracy would likely have decreased.

		IDENTIFIED
1	*Skiddaw, 931 m (3,054 ft)*	yes
2	*Blencathra, 868 m (2,847 ft)*	yes
3	*Skiddaw Little Man, 865 m (2,838 ft)*	no
4	*Carl Side, 746 m (2,448 ft)*	no
5	*Long Side, 734 m (2,408 ft)*	no
6	*Lonscale Fell, 715 m (2,346 ft)*	yes
7	*Knott, 710 m (2,329 ft)*	yes
8	*Bowscale Fell, 702 m (2,303 ft)*	yes
9	*Great Calva, 690 m (2,264 ft)*	yes
10	*Ullock Pike, 690 m (2,264 ft)*	yes
11	*Bannerdale Crags, 683 m (2,241 ft)*	yes
12	*Bakestall, 673 m (2,208 ft)*	no
13	*Carrock Fell, 663 m (2,175 ft)*	yes
14	*High Pike, 658 m (2,159 ft)*	yes
15	*High Pike, 658 m (2,159 ft)*	no

Table 10. Table that shows the 15 highest peaks of the first evaluation dataset. The central part of this dataset corresponds to the Northern Fells of the Lake District.

		IDENTIFIED
1	*High Street, 828 m (2,717 ft)*	yes
2	*High Raise, 802 m (2,631 ft)*	yes
3	*Rampsgill Head, 792 m (2,598 ft)*	no
4	*Thornthwaite Crag, 784 m (2,572 ft)*	yes
5	*Kidsty Pike, 780 m (2,559 ft)*	no
6	*Harter Fell, 778 m (2,552 ft)*	yes
7	*Caudale Moor, 763 m (2,503 ft)*	yes
8	*Mardale Ill Bell, 760 m (2,493 ft)*	yes
9	*Ill Bell, 757 m (2,484 ft)*	yes
10	*The Knott, 739 m (2,425 ft)*	no
11	*Kentmere Pike, 730 m (2,395 ft)*	no
12	*Froswick, 720 m (2,362 ft)*	no
13	*Branstree, 713 m (2,339 ft)*	yes
14	*Yoke, 706 m (2,316 ft)*	yes
15	*Gray Crag, 699 m (2,293 ft)*	yes

Table 11. Table that shows the 15 highest peaks of the second evaluation dataset. The left part of this dataset corresponds to the Far Eastern Fells of the Lake District.

6.3 Discussion of results

When analysing tables 6 and 7 shown in the previous section, one realises that the results obtained by using the second model are slightly worse. This might seem contradictory due to the fact that the second model was created with the aim of identifying peaks of a low height. It is clear than the models are going to perform differently depending on the profile terrain but the low performance of the fourth factor (included in the second model and not included in the first one) means that it is not a relevant factor even though the second model performs better on the development dataset.

Although good results have been obtained by using both models on different datasets, there are some problems that the models are not able to resolve. The first problem is mainly associated with the identification of peaks on, or near to, the boundaries of the map. Those points are very difficult to identify because most of the classifier factors take into account the surroundings of the peaks but only one portion of the surroundings is available. Another problem is that candidate peaks are often selected as actual peaks. This usually happens when a definite actual peak is located near to the candidate peak, but is located outside the boundaries of the maps. An example of is given below; a peak was incorrectly classified in a portion of the North of the Western Fells map (Figure 37). That peak was given a high value (achieving the threshold)

as it was very far away from another peak even though its prominence were not very large. However, this peak is very close to an actual peak which is outside the map.

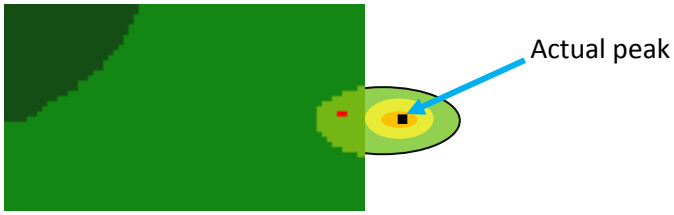


Figure 37. Area included in the first evaluated dataset. It corresponds to the North part of the Western Fells. The red point corresponds to a misclassified peak.

Regarding the first dataset used during the evaluation stage, there are 6 peaks just on the boundaries of the map and two more very close to them. These eight peaks represent a high percentage of the incorrectly classified peaks, 8 out of 14. If they were not considered, an accuracy of 91.2% is recorded against an accuracy rate of 87.6% and 86.1% obtained before. They are incorrectly classified by using both models but if a larger map was used, they would not be selected as actual peaks. There is no way of knowing how far away from the boundaries a candidate peak should be in order for the classifier factors to work properly.

The second problem is related to the distinction between the features that have an established name and those do not. The list of actual peaks used during the testing and evaluation was created by the British fell walker Alfred Wainwright [46] and it includes a set of 214 named fells. The problem lies in that there might be fells without an established name whose characteristics such as the height or the shape are very similar to those of named fells. The name of a fell might have been given with regards to any other factor not related to topographic characteristics. The current names of mountains and features in general comes from the ancient folks. The people that inhabited a particular geographical landscape historically have been responsible for naming the lands that belonged to their territories. Originally, they named the mountains using factors such as hydrology, zoology or botany. For instance, a mountain might have been named if it was home to cattle or to nomads. However, an elevation with the same topographic characteristic might lack of an historically recognised name because its terrain was not suitable for cattle breeding in the past.

The list of Wainwright peaks used as a baseline to evaluate the models has been compared to the Database of British Hills (DoBH)[4]. The Wainwright peaks are mainly just a subset of the hills that can be found on the DoBH. The relation between both sets of peaks can be observed below (Figure 38). The figure shows two images of the central area of the Lake District. The green triangles represent the Wainwright peaks

[4] Website: http://www.biber.fsnet.co.uk/database_notes.html

while the blue ones represent the peaks only registered on the DoBH. As is evident, there is a number of peaks that are not included in the Wainwright list. Therefore, some of the peaks that have been misclassified by the models might have been registered on the DoBH. In other words, the number of false positives (FP) might be lower due to the fact they belong to the set of peaks recorded on the DoBH which are not contained on the Wainwright list.

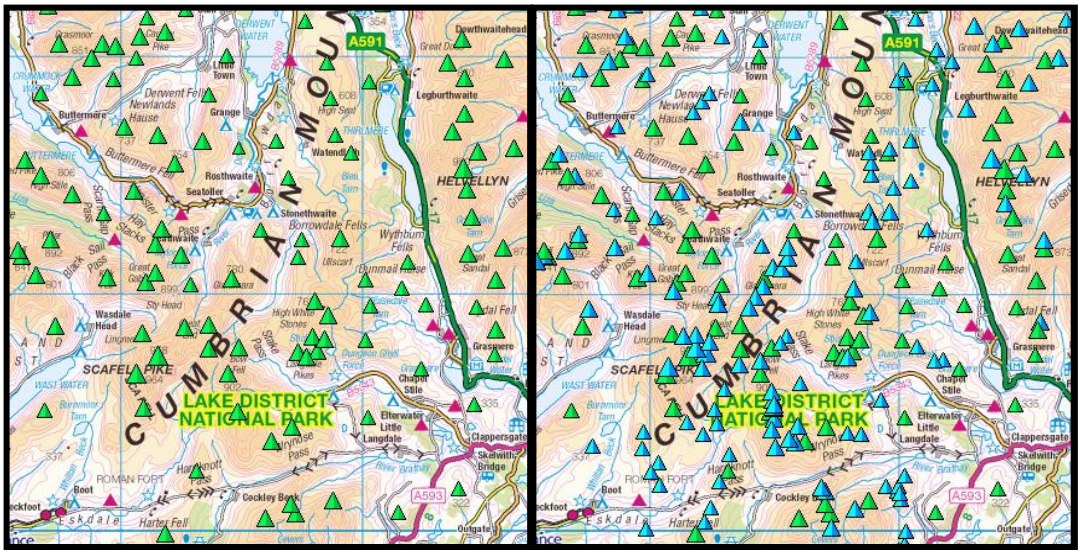


Figure 38. Maps of the central are of the Lake District in England. Left: The map shows the peaks registered by Alfred Wainwright . Right: The map shows the peaks registered on the Database of British Hills. Source: Ordnance Survey [33].

6.4 Limitations

This section details the limitations of the project and they are outlined in the following points:

- **Wait time**

  Depending on the functionality that the user is interested in executing, the wait time can vary a lot according to the complexity of the calculations. There is no need to have an application that provides results quickly. However, waiting time has been a problem throughout the testing and evaluation stage. Comparing results or just checking the running of the application was a very time consuming task. One of the most time consuming task is the assessment of the parent peak of a particular peak. For instance, there are 508 candidate peaks and so the 507 different key passes have to be calculated in order to obtained the parent peak. This was infeasible so the search was lightly restricted to peaks situated not very far away from the peak under study. Despite this restriction, the parent peaks take a considerable time to be assessed.

- **Problem of the dataset boundaries**

  As it was previously commented (see section 6.3), the correct classification of candidate peaks on the boundaries or close to them is a challenging task. This problem might be responsible for poor results. As a solution, larger maps might be used, but this would entail longer wait times.

- **Vagueness**

  The application created deals with concepts that are vague in many aspects. Vagueness is a persistent problem in the field of geography and its analysis is not a straightforward task. Although it is impossible to create an exact method to identify landform features, good approaches might be given.

- **Manual evaluation**

  In the beginning of the testing stage, the performance of the different classifier techniques was evaluated manually. When the selected peaks were shown on the screen, they were compared to the corresponding map and those that matched were considered as actual peaks correctly classified. This process was very slow due to the fact that each peak had to be checked individually. Considering this and the wait time until getting the results, the testing process became totally infeasible. For this reason, the lists of Wainwright peaks were stored on the system and the checking process was done internally. Although the automatic checking was faster than the manual one, the adding of the peaks was a burden. The exact locations of each peak in the 401x401 matrix had to be determined and this was subjective especially when there were several candidate peaks very close to each other. Even if the lists of peaks are properly stored in the application, one does not know if the lists contain all the peaks that they should (see second problem commented in section 6.3). The lists of Wainwright peaks are not entirely accurate.

7. CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This project has presented a solution to identify landform features. It is focused on the identification of aspects related to mountains such as: their peaks, their extents and their parent peaks. An approach regarding valleys have been also given as well as the identification of potential rivers and lakes.

The application is mainly a visualization tool which provides a set of functionalities to the user. Its ease of use is the main advantage in relation to other geographic information systems (GIS) although obviously less powerful. The evaluation of the project was made mainly by comparing the results with actual maps. Figures were obtained regarding the identification of actual peaks and good values of accuracy were recorded even using different datasets.

Overall, the project was completed on time and the minimum requirements have been fulfilled. The functionalities related to mountains were the first ones to be added to the application and those related to valleys were added as an extension when the minimum requirements had already been met.

7.2 Future work

There are many possibilities in extending the work described in this report and they are strongly connected to the limitations commented before (see Section 6.4). They are summarized by the following points:

- Deeper experiments might be done to give a more accurate solution. Regarding the models, the parameter values were chosen after a lot of experimentation. However, the results of every possible combinations were not compared.
- Further studies regarding the factors that are pertinent to the identification of mountains. The interpretation of the results indicates that the fourth factor (see Section 4.3.7.4) was not very relevant therefore some other factors favouring the identification of low peaks might be added.
- More efficient algorithms might be created. Not only to reduce the wait time problem but also to deal with the problem of the identification of peaks close to the boundaries. Some thresholds might

be used to indicate the minimum distance that there should be between a particular peak and the boundary of the map.

- Advanced techniques to deal with the demarcation of the extent of valleys. The identification of their floors is very simple but determining the boundaries is far more complicated. Although the vagueness problem is present, some models might be created to handle it efficiently.
- 3D visualization might be added to the application. At the moment, the terrains can only be seen in 2D, however some Java libraries or other tools might be use to get an improved visualisation.

8. REFERENCES

1. Aldrich, V.C. (1937). ”Some meanings of Vague”. Analysis, Vol. 4, No. 6 (Aug., 1937), pp. 89-95 Published by: Blackwell Publishing on behalf of The Analysis Committee.

2. Bangay, S; De Bruyn, D; Glass, K. (2010). Minimun Spanning Trees for Valley and Ridge Characterization in Digital Elevation Maps. Afrigraph, pp 73-82.

3. Bennett, B. (2001). What is a forest? On the vagueness of certain geographic concepts. Topoi. 20(2), pp 189-201.

4. Blyth, S. G., B; Lysenko, I; Miles, L; Newton, A. (2002). "Mountain Watch". Cambridge, UK, UNEP World Conservation Monitoring Centre.

5. Boehm B, "A Spiral Model of Software Development and Enhancement ", ACM SIGSOFT Software Engineering Notes", "ACM", 11(4):14-24, August 1986

6. Boyell, R.L; Ruston, H. (1963). Hybrid Techniques for Realtime Radar Simulation. Proceedings 1963 FaIL Joint Computer Conference. Las Vegas, Nevada: 445-458.

7. Carr, H; Snoeyink, J; Van de Panne, M.(2010). "Flexible isosurfaces: Simplifying and displaying scalar topology using the contour tree." Comput. Geom. Theory Appl. 0925-7721 43(1): 42-58.

8. Carr, H; Snoeyink, J; Axen, U. (2003). "Computing contour trees in all dimensions." Comput. Geom. Theory Appl. 0925-7721 24(2): 75-94.

9. Cayley, A. (1859). On contour lines and slope lines: Phil. Mag., v. 18, p. 264-268.

10. Centre for Technology in Government (1998). A Survey of System Development Process Models. University at Albany, National Historical Publications and Records Commission.

11. Chaundry, O; Mackaness, W. (2008). "Making Mountains Out of Molehills." Transactions in Gis 12: 567-589.

12. Cheng, T; Molenaar, M. (1999). "Diachronic Analysis of Fuzzy Objects." Geoinformatica 3(4): 337-355.

13. Clark, W; Gant, H. (1922). The Gantt chart, a working tool of management. New York, Ronald Press.

14. Deshpande, J. V. (1968). "On continuity of a Partial Order." Proceeding of the American Mathematical Society 19(2): 383-386.

15. Evans, G. (2011). "Can There Be Vague Objects?" Oxford Journals. Oxford University Press 38.

16. Fisher, P. Named Valleys & Hollows- Research Requirements, Ordnance Survey.

17. Fisher,P; Wood,J; Cheng,T. (2003). Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. Royal Geographical Society.

18. Freeman, H; Morse, S.P. (1967). "On Searching a Contour Map for a Given Terrain elevation profile." Journal of the Franklin Institute 248(1).

19. Gerrard, J. (1990). Mountain environments: an examination of the physical geography of mountains.

20. Helman, A. (2005). *The Finest Peaks–Prominence and Other Mountain Measures*. Trafford Publishing.

21. Kraus, M. (2010). Visualization of Uncertain Contour trees. Aalborg East, Denmark, Department of Architecture, Design, and Media Technology.

22. Kweon, S. Kanade,T. (1994). "Extracting Topographic Terrain Features from Elevation Maps." CVGIP: Image Understanding 49(2): 171-182.

23. Liu, X. Ramirez, J.R. Automated Vectorization and Labelling of Very Large Raster Hypsographic Map Images Using Contour Graph. Ohio, The Ohio State University, Columbus.

24. MacMillan R.A; Pettapiece, W.W; Nolan, S.C; Goddard, T.W.(1998). "A generic procedure for automatically segmenting landforms into landform element using DEMs, heuristic rules and fuzzy logic." ELSEVIER Fuzzy Sets and Systems(113): 81-109.

25. Mark, D. M. (1978). Topological Properties of Geographic Surfaces: Applications in Computer Cartography. Harvard Papers on Geographic Information System. Cambridge, Massachusetts. 5: 11.

26. Mark, D. M. (1986). Two Contour-Tagging Algorithms Based on the Contour Enclosure Tree. Abstracts of the Canadian Cartographic Association Meeting. Burnaby BC.

27. Maxwell, J. C. (1870). On hills and dales: Phil. Mag.,v. *40,* p. 421-427.dales: Phil. Mag., v. *40,* p. 421-4

28. McConnell, S. (1996). *Rapid Development: Taming Wild Software Schedules*, Microsoft Press Books

29. Merril, R. D. (1973). "Representation of Contour and Regions for Efficient Computer Search." Communications of the ACM 16(2): 69-82.

30. Meyland, S. J. (2008). Rethinking Groundwater Supplies in Light of Climate Change: How Can Groundwater be Sustainably Managed While Preparing for Water Shortages, Increased Demand, and Resource Depletion? Forum on Public Policy. New York, Institute of Technology.

31. Morse, S. P. (1968). "A Mathematical Model for the Analysis of Contour-Line Data." Journal of the ACM 15: 2.

32. Morse, S. P. (1969). "Concepts of Use in Contour Map Processing." Communications of the ACM 15(2): 205-220.

33. Ordnance Survey: http://www.ordnancesurvey.co.uk/oswebsite/P

34. Parnas, D. L.; Clements, P. C. (1986). *A rational design process: How and why to fake it*, IEEE Transactions on Software Engineering 12(2), 251—257.

35. Pascucci, V; Cole-McLaughlin, K. (2002). Efficient Computation of the Topology of Level Sets. In Proceedings of Visualization 2002.

36. Pascucci, V; Cole-McLaughlin, K; Scorzell, G. (2004). Multi-resolution computation and presentation of contour trees. In Proceedings of the IASTED conference on Visualization, Imaging and Image Processing(VIIP 2004).

37. Piccolo. Human computer Interaction Lab. University of Maryland. http://www.cs.umd.edu/hcil/piccolo/index.shtml

38. Roubal, J; Poiker, T.K. (1985). Automated Contour Labelling and the Contour Tree. Proceeding, AutoCarto, 7. Washington D.C., ACSM-ASPRS.: 472-481.

39. Russell, B. (1923). "Vagueness". *The Australasian Journal of Psychology and Philosophy* **1** (2): 84-92. ISSN 1832-8660. Retrieved 15th August, 2011, from: http://www.webcitation.org/5lNl9fNiP.

40. Sircar, J. K; Cebrián, J.A. (1987). Creación de Modelos Topográficos Digitales (MTDs) a partir de curvas de nivel rasterizadas. Conferencia Internacional de Cartografía de la ACI. U. C. d. Madrid. Morelia, Mexico, Anuales de Geografía de la Universidad Complutense. 10: 13-34.

41. Smith, B; Mark, D. (2003). Do Mountains Exist? Towards and Ontology of Lanforms. Buffalo, New York, Departments of Philosophy and Geography, University of Buffalo.

42. Straumann, R. Purse, R; (2010). Computation and Elicitation of Valleyness. Spatial Cognition and Computation: An Interdisciplinary Journal. Zurich.

43. Symonds, M. (2011, 29 June, 2011). "Project Planning Essentials." Retrieved 30th July, 2011, from http://www.projectsmart.co.uk/project-planning-essentials.html.

44. Takahashi, S; Fujishiro, I; Takeshima, Y. (2005). Interval volume decomposer: A topological approach to volume traversal. Visualization and Data Analysis 2005 (Proceeding of the SPIE).

45. Van Krevel, M; Van Oostum. R; Bajaj, C; Pascucci. C; Schikore,D. (1997). Contour trees and small seed sets for isosurface traversal. Proceedings of the thirteenth annual symposium on Computational geometry. Nice, France, ACM: 212-220.

46. Wainwright, A. (1955-1966). Pictorial Guide to the Lakeland Fells.

47. Wolfe, E.W; Wise, W.S; Dalrymple, G.B. (1997). The geology and petrology of Mauna Kea Volcano, Hawaii; a study of post-shield volcanism. U. S. G. S. P. Paper: 19.

APPENDIX I: PERSONAL REFLECTION

This project was the most challenging part of the Masters programme. It required a huge amount of background research which was something new for me. It was not easy to find relevant information when searching through articles and conferences paper. However my supervisor Dr. Brandon Bennett and my assessor Dr. Hamish Carr provided me with some relevant papers which made me familiar with vagueness and different techniques to deal with it. As advice for new students I would like to recommend them to attend the Literature Research Meeting. Most of the students did not attend and although I did, I underestimated its usefulness.

At the beginning, the design and implementation was a little chaotic as I did not know what exactly I had to do and how to do it. The aim of the project was too general and I was confused about where I had to start from. The meetings with my supervisor were very helpful to decide what to do next and how to deal with new problems that appeared. He always provided me very useful ideas.

A lot of time was wasted trying to use other Java Libraries. Some of them were not suitable for the project and others were too complicated. I was going to need a long period of time just learning how to use them without even knowing if the facilities that they were going to provide me were useful for the project.

I consider that the testing was one of the most consuming and boring stages mainly because of the wait time needed at each running. It was quite de-motivating when poor results were obtained after working many hours on it.

The project schedule was a very important part of the project. Probably one of the most important factors to guarantee the success of the project. I established the planning in a very pessimist way, having into account that some unforeseen problems might appear. Almost a week before the deadline was aimed to tackle unexpected issues. I reckon that having a very optimistic plan might be dangerous especially toward the end of the project.

Writing-up the project report was very time-consuming. As a non-English speaker, the time that I spent writing was far more than the time I had needed doing it in my language and the results quite worse. At the beginning of the course I was really convinced that I was not going to be able to do it but now after a huge effort I am writing the last part of the report. I can say that I feel very relived and I consider that I have gained in self confidence.

Personally I think that this project was a great opportunity to increase my organization skills as well as my ability to face new challenges. It has been really hard to work during the summer months but I have

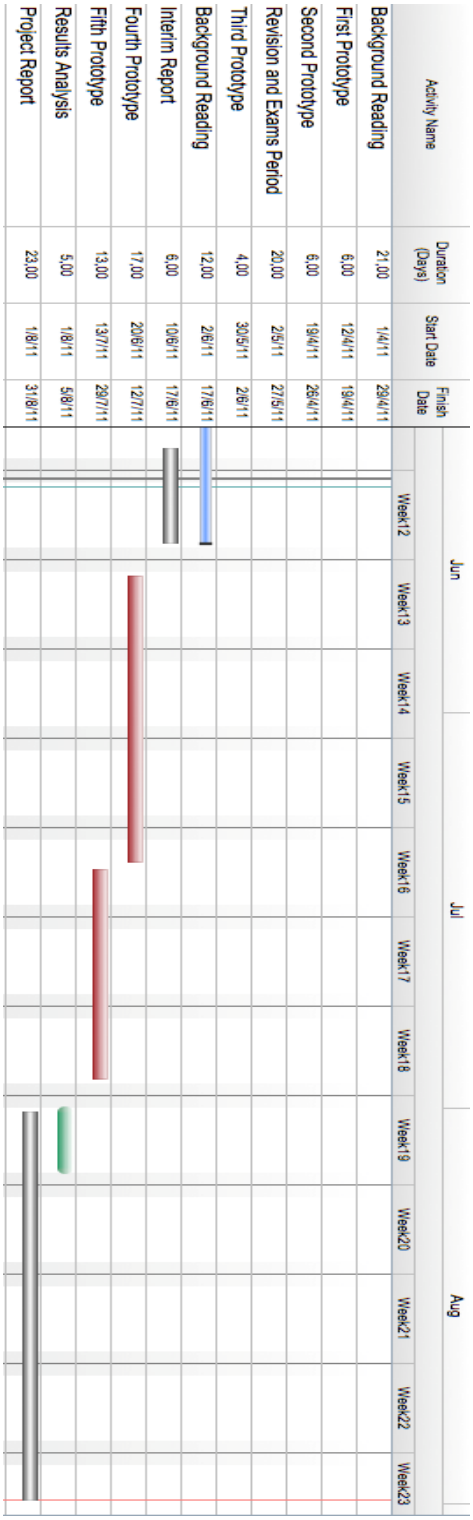
learnt many things such as how to approach a big background study, how to deal with new tools and how to create a development planning. It was the first time that I have accomplished a project of this size and I consider that I am now better prepared to carry out similar projects in the future. I learn from my mistakes and I consider that if I had the chance to start now, the global result would be better and probably less time would be required.

APPENDIX 2: INTERIM REPORT

Attached as hard copy

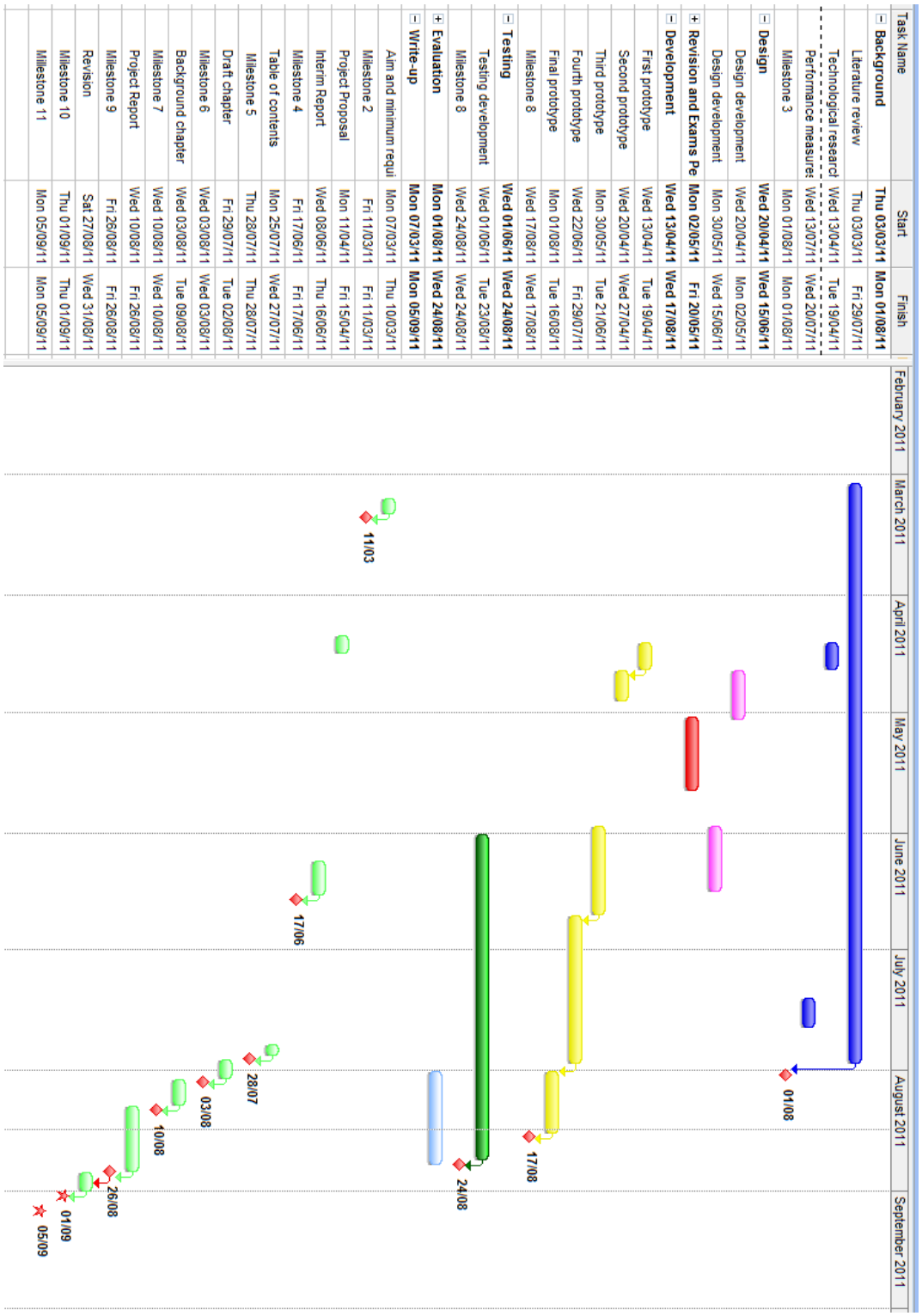
APPENDIX 3: INITIAL PROJECT SCHEDULE

The project planning was initially as follows:



APPENDIX 4: REVISED PROJECT SCHEDULE

The revised project planning is the following:



Task Name	Start	Finish
Background	**Thu 03/03/11**	**Mon 01/08/11**
Literature review	Thu 03/03/11	Fri 29/07/11
Technological researcl	Wed 13/04/11	Tue 19/04/11
Performance measurel	Wed 13/07/11	Wed 20/07/11
Milestone 3	Mon 01/08/11	Mon 01/08/11
Design	**Wed 20/04/11**	**Wed 15/06/11**
Design development	Wed 20/04/11	Mon 02/05/11
Design development	Mon 30/05/11	Wed 15/06/11
Revision and Exams Pe	**Mon 02/05/11**	**Fri 20/05/11**
Development	**Wed 13/04/11**	**Wed 17/08/11**
First prototype	Wed 13/04/11	Tue 19/04/11
Second prototype	Wed 20/04/11	Wed 27/04/11
Third prototype	Mon 30/05/11	Tue 21/06/11
Fourth prototype	Wed 22/06/11	Fri 29/07/11
Final prototype	Mon 01/08/11	Tue 16/08/11
Milestone 8	Wed 17/08/11	Wed 17/08/11
Testing	**Wed 01/06/11**	**Wed 24/08/11**
Testing development	Wed 01/06/11	Tue 23/08/11
Milestone 8	Wed 24/08/11	Wed 24/08/11
Evaluation	**Mon 01/08/11**	**Wed 24/08/11**
Write-up	**Mon 07/03/11**	**Mon 05/09/11**
Aim and minimum requi	Mon 07/03/11	Thu 10/03/11
Milestone 2	Fri 11/03/11	Fri 11/03/11
Project Proposal	Mon 11/04/11	Fri 15/04/11
Interim Report	Wed 08/06/11	Thu 16/06/11
Milestone 4	Fri 17/06/11	Fri 17/06/11
Table of contents	Mon 25/07/11	Wed 27/07/11
Milestone 5	Thu 28/07/11	Thu 28/07/11
Draft chapter	Fri 29/07/11	Tue 02/08/11
Milestone 6	Wed 03/08/11	Wed 03/08/11
Background chapter	Wed 03/08/11	Tue 09/08/11
Milestone 7	Wed 10/08/11	Wed 10/08/11
Project Report	Wed 10/08/11	Fri 26/08/11
Milestone 9	Fri 26/08/11	Fri 26/08/11
Revision	Sat 27/08/11	Wed 31/08/11
Milestone 10	Thu 01/09/11	Thu 01/09/11
Milestone 11	Mon 05/09/11	Mon 05/09/11

APPENDIX 5: MANUAL

A brief manual explaining how to use the application is presented in this appendix.

Firstly, the user must open the file *'Landforms.jar'* to run the application. When the visualization window is shown, a set of functionalities are available for the user. They can be activated as follows:

Candidate Peaks

They can be visualized only by clicking on the button labeled with *'Candidate Peaks'.* They are drawn as black points.

Summary of the terrain characteristics:

This includes the calculation of the highest and lowest location as well as the average height and the standard deviation. This functionality can be executed by clicking on the button *'Summary'*. The highest and lowest points will be drawn on the map with a flag and their corresponding height, while the values of the average height and the standard deviation are shown in text format under the *'buttons'* section.

Finding the parent peak of a particular peak

This functionality can be activated by clicking on the button '*Find Parent'*.

The set of candidate peaks appears on screen and the user will have to select a particular point by clicking on one of the candidate peaks. A window (see Figure I) will be prompted to confirm that a particular peak has been correctly selected. This is one of the most time-consuming functions and it may take a few minutes. When the parent peaks is calculated, it will be coloured in cyan as well as the peaks that was selected.

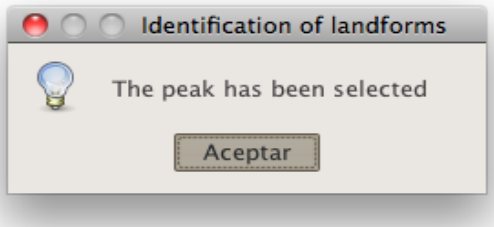


Figure I. Window that is prompt when a peak is selected.

List of peaks along with their closest higher peak

To obtain this record the button *'Parentage'* should be clicked. The record will be stored on the output file however they can be found by using the mouse. When the button has already been clicked and the

record has already been generated, the candidate peaks will be shown on the map. The user can select one of the peaks and its closest higher peaks will be coloured in pink. This new peak can then be clicked and its closest higher peaks will be shown and so on. In this way, a hierarchy of peaks is established until obtaining the highest peak on the map.

Highest contour containing two peaks

This function can be executed by clicking on the button labeled as *'Highest Contour'.* After clicking on it, a window will be shown asking the drawing mode (see Figure II). Once the drawing mode is selected, the candidate peaks will be shown (see Figure III) and two peaks should be selected. When the peaks are selected, the confirmation message will be shown (see Figure I). Depending on the size of the contour calculated, this may take a few minutes.

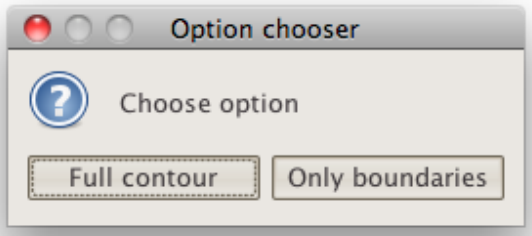


Figure II. Window that provides the user with two visualization options.

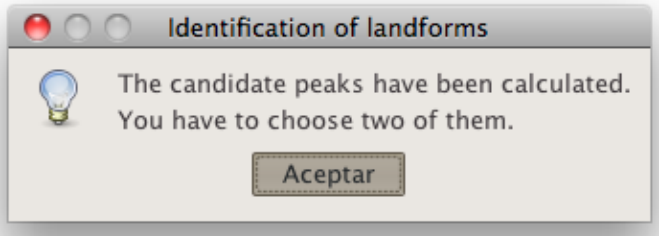


Figure III. Window that is prompt when the user wants to visualize the highest contour contained two peaks

Classifier factors

Different classifier factors can be selected individually by clicking on the first four buttons in the right-hand column. When the calculations are done, the selected peak will be shown in cyan whereas the non-selected candidate peaks will remain in black.

Combination of factors

The fifth and the sixth buttons in the right-hand column allow for combinations of two different factors. When one of these buttons is clicked, a window will be shown so that the user can choose between the four factors. The running time required will be longer than the time needed to execute one of the classifier methods that use only one factor .

Models

The models can be activated by the corresponding buttons labeled with *'Model 1'* and *'Model 2'*. The user will have to wait until the selected peaks are shown on the map. The cyan points represent the correctly classified peaks whereas the red ones represent the incorrectly classified peaks.

Valleys and Rivers

To visualise the potential floor valleys, the user should click on the button labeled with *'Valleys-Rivers*'. Once the user has clicked, a window will be prompted (Figure IV) asking the user if he or she also would like to visualize the river. If the user selects *'Yes'*, both the rivers and floor valleys will be shown. To obtain the delimitation of a valley corresponding to a valley floor, the user should click on the blue area that represents the floor and wait for the results. Depending on the extent of the valley, the wait time will vary.

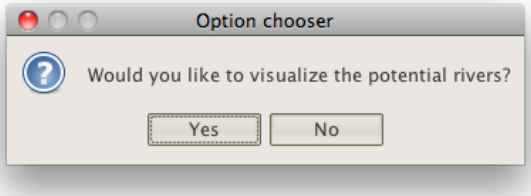


Figure IV. Window which will be prompt when clicking on the '*Valleys-Rivers*' button.

Change of the dataset

To change the dataset under study, click on the *'File'* menu of the bar located on the top of the visualization window. Then, click on *'Open file of data'* and select the file stored on the user's computer.

Change the parameter values

The *'Edit'* menu of the bar provides the user with a set of options to change the parameter values of the models as well as the value of the methods that use the classifier factor individually.

Help

To visualize the help menu, click on *'Help'* option in the toolbar on the top part of the main window. Different tags will be provided to the user (Figure V).

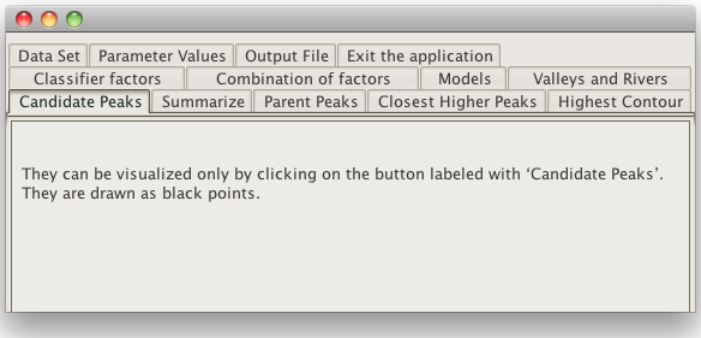


Figure V. Help window

Output file

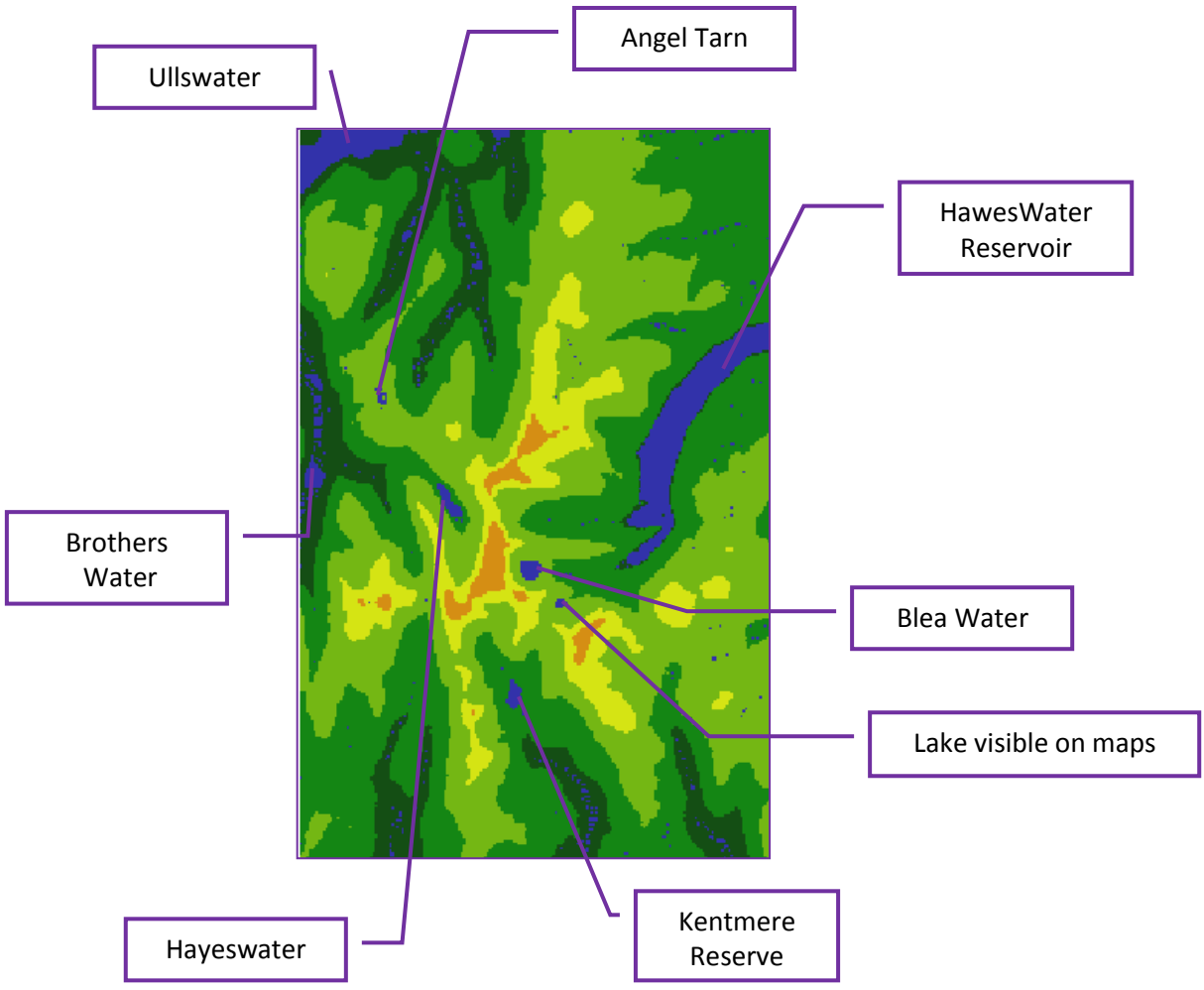
A record of the results obtained as well as the main characteristics of the functionalities executed by the user is stored on the file *'Output.txt'.*

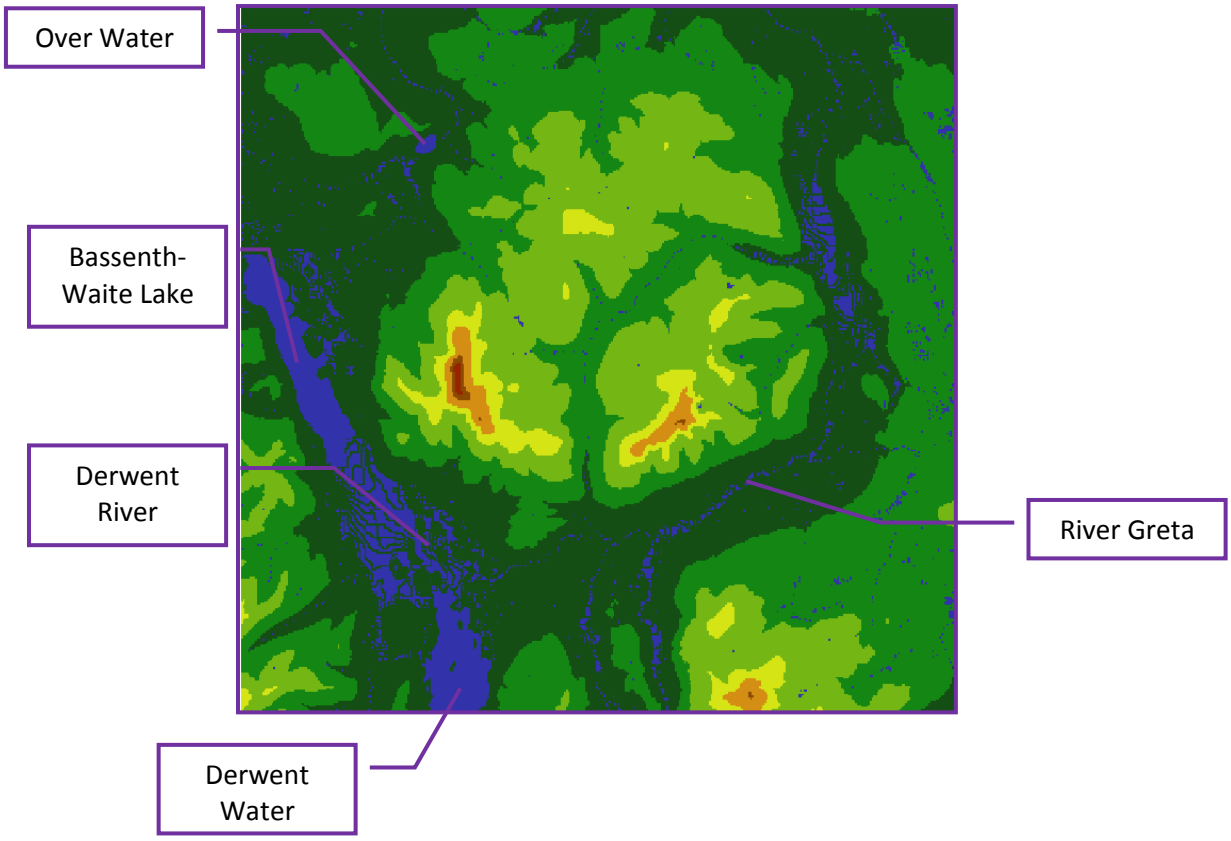
Exit

To close the application, click on the *'Exit'* option in the *'File'* menu.

APPENDIX 6: LOCATION OF LAKES

The following images represent the two dataset that have been used for the evaluation of the system. Their corresponding lakes have been identified.




Over Water
Bassenth-
Waite Lake
Derwent
River
Derwent
Water
River Greta