Doctoral Thesis

Statistical properties of watersheds

Author(s):
Fehr, Eric Marcel

Publication Date:
2011

Permanent Link:
https://doi.org/10.3929/ethz-a-007159253

Rights / License:
In Copyright - Non-Commercial Use Permitted
Statistical Properties of Watersheds

A dissertation submitted to
ETH ZURICH

for the degree of
DOCTOR OF SCIENCES

presented by
ERIC MARCEL FEHR
Dipl. phys., ETH Zurich

born 03.08.1982
citizen of Kilchberg ZH, Switzerland

accepted on the recommendation of
Prof. Dr. H. J. Herrmann, examiner
Prof. Dr. L. Hurni, co-examiner

2011
Contents

1 Introduction 5

2 General Concepts 9
  2.1 Landscapes and Watersheds ........................................ 10
    2.1.1 What is a Watershed? ....................................... 10
    2.1.2 Relevance of Watersheds ..................................... 11
    2.1.3 Influence of the Landscape Topography ...................... 13
    2.1.4 Modeling Real Landscapes .................................... 15
    2.1.5 Creating Artificial Landscapes ................................ 16
    2.1.6 Digital Landscapes ........................................... 16
  2.2 Fractal Concepts .................................................... 17
    2.2.1 Self-Similarity ................................................ 17
    2.2.2 Deterministic versus Statistical ............................ 18
    2.2.3 Fractal Dimension ............................................. 19
    2.2.4 Self-Similarity versus Self-Affinity .......................... 20
    2.2.5 What is a Fractal? ............................................ 22
    2.2.6 Fractals in Nature ............................................ 22
    2.2.7 Practical Applications of Fractals ............................ 23
  2.3 Percolation ........................................................... 23
    2.3.1 Origins of Percolation Theory ............................... 24
    2.3.2 What is Percolation? ......................................... 24
    2.3.3 A Model for Phase Transitions ............................... 25
    2.3.4 A Model for Invasion Processes ............................... 26
    2.3.5 Recent Advances ............................................. 27
  2.4 Formal Definition of Watershed .................................... 28
3 Watershed Algorithms

3.1 Introduction ................................................. 32
3.2 Flooding ...................................................... 33
3.3 Iterative Invasion Percolation Procedure .................... 33
3.4 Equivalence of Methods ..................................... 35
3.5 Time Complexity ............................................. 35
3.6 Summary ...................................................... 38

4 Self-Similarity ................................................ 39

4.1 Introduction ................................................ 40
4.2 Method ....................................................... 40
4.3 Uncorrelated Artificial Landscapes .......................... 41
4.4 Real Landscapes ............................................. 42
4.5 Disorder Independence and Correlations ..................... 44
4.6 Extension to 3 Dimensions .................................. 49
4.7 Summary ...................................................... 49

5 Impact of Perturbations ..................................... 51

5.1 Introduction ................................................ 52
5.2 Method ....................................................... 53
5.3 Perturbing Real Landscapes ................................ 55
5.4 Perturbing Artificial Landscapes ............................ 57
  5.4.1 Effects in two Dimensional Systems without Correlations 57
  5.4.2 Finite-Size Effects .................................... 59
  5.4.3 Effect of Long-Range Correlations ..................... 61
  5.4.4 Influence of the Perturbation Strength ................. 63
  5.4.5 Relation to Invasion Percolation ..................... 64
  5.4.6 Scaling Relations .................................... 68
5.5 Impact of Perturbations in three Dimensions .............. 70
  5.5.1 Effects for Watersheds of Systems without Correlations 70
  5.5.2 Relation to Invasion Percolation ..................... 72
  5.5.3 Effect of Long-Range Correlations .................... 73
5.6 Summary ...................................................... 74

6 Corrections to Scaling ...................................... 77

6.1 Introduction ................................................ 78
6.2 Models ....................................................... 82
  6.2.1 Watershed ............................................. 82
Zusammenfassung


Zunächst stelle ich einen fortschrittlichen Algorithmus für die Bestimmung von Wasserscheiden auf digitalen Landschaften vor, welcher auf der iterativen Anwendung von Invasions-Perkolation basiert. Der Hauptvorteil meiner Methode gegenüber früheren stellt die sublineare Zeitkomplexität dar.


fechten zu Eigenschaften der Invasions-Perkolation auf.

Schlussendlich, untersuche ich systematisch die Skalenkorrekturen für die Wasserscheide, den optimal path crack und die bridge line, was für alle drei eine hochpräzise Messung der fraktalen Dimension und des Exponenten der führenden Korrektur erlaubt. Die erhaltenen Werte sind numerisch identisch, was die vermutete enge Verbindung zwischen den drei Modellen bestätigt und die Übertragung der gefundenen Eigenschaften der Wasserscheide auf die anderen Modelle ermöglicht.
Summary

The concept of watersheds arises naturally in the field of geomorphology, where the watershed is defined as the line separating adjacent drainage basins. It plays, hence, a fundamental role in water management, landslide, and flood prevention, but also in image processing. Nowadays, in particular in image processing, the interest in efficient automatic delineation of the watersheds in digital images or Digital Elevation Models (DEM) is high. But, despite this interest and the important role they play, only little is known about the properties of watersheds.

Here I first present an advanced algorithm for the determination of watershed lines on Digital Elevation Models, which is based on the iterative application of invasion percolation. The main advantage of my method over previously proposed ones is that it has sub-linear time-complexity.

Using my algorithm I am able to demonstrate, convincingly and with high accuracy, the fractal character of watersheds in real and artificial landscapes. This means that the length of the boundary of a drainage basin, the perimeter, which is often used to characterize basin properties, is not a well defined quantity, if information about the used resolution is missing.

Systematically studying small, local perturbations of the landscapes, I find that they can have a large impact on the shape of watersheds even at very long distances, hence having a nonlocal effect. These effects are shown to follow power-laws independent on the perturbation strength. In the presence of spatial long-range correlations, I could achieve a matching for the resulting scaling laws of real and artificial landscapes. Additionally, I show the relation of the studied effects to properties of invasion percolation.

Finally, I investigate systematically the corrections to scaling for the watershed, the optimal path crack, and the bridge line, yielding highly precise estimates of the fractal dimension and the leading correction exponent for each of them. The obtained values are shown to be numerically equivalent,
confirming the proposed close relation between the different models and allowing the transmission of the watershed properties to the other models.
Chapter 1

Introduction
CHAPTER 1. INTRODUCTION

The concept of watershed is used in a large variety of fields with direct relevance to human society. As the watershed separates adjacent drainage basins, it plays a fundamental role in the structure, function, and evolution of aquatic ecosystems [1, 2, 3], as well as landslide [4, 5, 6, 7] and flood prevention [6, 8, 9]. Watersheds exist on many different scales, from divides between the drainage areas of tributaries of a small stream, up to continental watersheds separating the catchments of different seas, e.g. in Europe between the Atlantic and the middle sea. Sometimes, they might also describe the boundaries between regions of different meteorology. In the mountains and at larger scales they could even separate areas of different climate, e.g. in the Alpes the northern Switzerland from the often warmer Italy. Moreover, watersheds are used extensively in Image Processing [10, 11, 12], to segment images into regions of different characteristics (color, intensity, pattern, ...). This has important applications, e.g., in Medicine [13, 14, 15, 16, 17], where for example cancer tumors need to be detected in Computer Tomography (CT) Scans. For all these applications it is, hence, crucial to understand the subtle geometrical and dynamical properties of watersheds, but, as their determination needs information about almost the entire landscape, it is far from being an easy task.

Traditional cartographical methods for basin delineation relied on manual estimation from iso-elevation lines, which required a good deal of guess work. But with the increasing amount of data and the attempt of real-time processing, the interest in automatic watershed extraction from digital elevation models (DEM) has grown [18, 19, 20, 21]. Digabel and Lantuéjoul in 1978 [22] were the first to use the concept of watershed in Image Processing by introducing the watershed transformation. In image processing one tries to simplify and/or change an image representation such that it is more meaningful or easier to analyze. This is done by image segmentation, e.g. partitioning a digital image into multiple segments (sets of pixels, also known as superpixels). In this way, one typically locates objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics as color, intensity or texture. Adjacent segments are significantly different with respect to the same characteristic(s) [12]. Although many other methods, such as clustering, histograms [23], edge-detection [24], region growing [10], level-set or graph partitioning [25], have been developed for that purpose, also watersheds are currently used. This is because, if one transforms an image
to its color gradient representation and calculates the watersheds on the so achieved DEM like data, one can identify the obtained catchment basins as the different (color) regions and their boundaries to be represented by watersheds. Due to the high interest in development of efficient algorithms, in particular in image processing or computer vision, the method of Digabel and Lantuéjoul [22] has been improved many times since its introduction and it is today at the basis of extremely powerful segmentation procedures [10]. Nevertheless, the development of even more efficient methods is needed to face future challenges in the field.

The length of the watershed of a drainage basin (the perimeter) is a value widely used to construct quantitative measures of basin shape, including circularity [26, 27, 28, 29], compactness [28, 30], and lemniscate ratio [31, 32]. In addition, the basin perimeter is used to construct other types of morphometric indices, like the texture ratio [33] and relative relief [27]. These indices are some of the core parameters used in models for the processes in fluvial geomorphology and hydrology, such that the accurate measurement of the perimeter is of critical importance for the conclusions and predictions drawn from these models. Breyer et al. [34] found evidence of fractal behavior for watersheds and pointed out that, hence, the choice of the right scales is crucial in the comparison between different drainage basins. Despite this evidence, the question, if the watershed is truly a self-similar fractal, remained open.

The location of a watershed depends only on the topography and, hence, is immobile if the landscape remains unchanged. But, if the landscape is modified, does that change the watershed? And how much? Indeed, geographers and geomorphologists have found the evolution of watersheds to be driven by local events classified as stream captures or drainage rearrangements. These events can affect the biogeography [35, 36], and may occur due to various mechanisms like erosion [37, 38, 39], natural damming [6], tectonic motion [40, 41, 42], as well as volcanic activity [43], i.e. after changing the landscape. Recently, the associated relevant mechanisms were investigated numerically and in small scale experiments [44, 45]. Such events have also been investigated in image processing, in order to circumvent segmentation failures [46, 47]. Although single events were investigated extensively, no systematic studies of the overall occurrence are available.

Recently, three new physical models, optimal path cracking, bridge percolation, and discontinuous percolation, have been introduced, which show striking similarities to watersheds. The first two of them also lead to a split-
ting of the system into distinct parts and the last one suppresses the growth of the largest cluster, delaying percolation. Optimal path cracking has been introduced by Andrade et al. [48, 49, 50] as a model for the evolution of successive optimal paths under constant failure. It describes, e.g., the breakdown of electrical or fluid flow through random media and has important consequences in other fields of science and technology, such as human transportation, fracture mechanics, or polymers in random environments, where finding the optimal path is a challenge [51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61]. More recently, Araújo et al. [62] developed a novel percolation scheme called bridge percolation, which systematically suppresses the creation of a spanning cluster. They found that the set of bridge bonds yields a fractal dimension in the same range as the optimal path crack in strong disorder and the surface of the infinite cluster in discontinuous percolation [63, 64]. How closely watersheds are related with these models, is still an open question.

The focus of this PhD thesis is to study geometrical and dynamical properties of watersheds, namely the fractal structure and the impact of small local modifications in the landscape on the location of the watershed, using a novel and efficient determination scheme, based on an iterative application of invasion percolation. Additionally, the relation to other physical models shall be investigated using corrections-to-scaling analyses. This thesis is organized as follows: In the first chapter, I give an introduction to several topics related to watersheds. After this, I briefly discuss watersheds and landscapes, in general, and I introduce the concepts of fractals and percolation, with a review of recent advances in these fields. At the end of chapter 2 a definition of a watershed on digital landscapes is formulated. In the third chapter, I develop the novel determination scheme, which I then compare with a commonly used watershed algorithm. In the following, the novel scheme is used to estimate the fractal dimension of watersheds. In chapter 5, I present a systematic study of the impact of perturbations on the watershed. Finally, the thesis closes with a chapter exploring the relation of watersheds to optimal path cracking and bridge percolation. The last four chapters 3-6 are based on the following papers: [65, 66, 67, 68].
Chapter 2

General Concepts
The aim of this chapter is not to give a thorough and detailed review of the theories behind my research, but rather to give the readers a brief overview over some concepts used later in this thesis. In the first section, I discuss the origin of the concept of watershed, its relevance to society, and emphasize the relation to the landscapes with focus on relevant mechanisms for the time evolution. After this, the concept of fractals is introduced with the basic definitions used later on. The third section considers percolation theory. Especially, the standard definition for invasion percolation is given, which is needed in chapter 3 for the there introduced algorithm. As well I resume two very recently developed percolation concepts, namely discontinuous percolation and bridge percolation, which are in close relation to watersheds, as will be discussed in detail in chapter 6. Finally, the last section gives a rigorous definition of a watershed, making use of the definitions and concepts introduced in the other sections.

2.1 Landscapes and Watersheds

2.1.1 What is a Watershed?

A watershed is the border between two neighboring river systems and, hence, corresponds to the boundary of the drainage area (drainage basin, catchment) from which the precipitated water runs into a certain river, lake, or ocean\(^1\). The word *shedding* is an old term for splitting or dividing, so the watershed is the line which divides the water. Often, this line is also called drainage divide, water divide, divide, water parting or height of land (in Canada). The watersheds can be grouped in three different hierarchical levels:

- Continental, water on each side flows into different oceans (e.g. divide between Nile and Congo river, or Danube and Rhine),

- Major, water on each side never meets again, but flows into the same ocean (e.g. Rhine and Elbe, Yellow river and Yangtze, or Danube and Dnieper),

- Minor, water parted between tributaries of the same stream (e.g. Missouri and Mississippi).

\(^1\)Note that, in North America, *watershed* has come to mean the drainage basin itself, but in this thesis it clearly denotes the boundary of these basins.
Typically one defines the watershed according to the surface flow, which in hilly terrain is dominated mainly by topographical features like hill and mountain ranges, i.e. ridges, but is less obvious in flat areas. The so called ridge watershed can be determined from pure elevation data, which is why this kind of divide is used in the processing of digital elevation models and in image segmentation. In some cases the watershed can lie within the river bed, in a marshland, or in the underground. The first case occurs when a river bifurcates, i.e. splits into two rivers like, e.g., the Orinoco does in South America. The position of the rock layers in the ground can be such that the water drains against the inclination of the surface. Hence, the actual divide, the groundwater divide, can deviate from the ridge. This is the case, e.g., between Weser and Leine at the Hoher Hagen, where the ridge watershed is located several hundred meters east of the groundwater divide. Another special case are iced watersheds in regions like the Antarctic, where glaciers dominate the water transport. Such ice-streams can wander far across ridges lying underneath. Finally, there are also regions, where a reasonable definition of watersheds fails, for example in marshland deltas like the Okavango delta or in big lake areas like the finnish lakeland. In this thesis, I do not treat these different kinds of divides and I also neglect the special cases. The focus is solely on surface flow and the watershed is defined by the topography. This choice is reasonable, as I address much larger scales (thousands of kilometers) than the typical scale of such deviations (several hundreds of meters).

### 2.1.2 Relevance of Watersheds

Apart from the water bodies themselves, watersheds are the most important objects under study in Limnology [69, 70, 71] and Hydrogeology [72, 73, 74]. These two disciplines study the water cycle above (Limnology) and below the surface (Hydrogeology) with respect to chemical and biological composition, flow, transport of contaminants, reconstruction, management, and conservation of water and water bodies, with increasing interest also in the forcing effects on the climate. Typically the length of the watershed around a drainage basin, called the perimeter\(^2\), is measured and used to characterize the shape of the basin through ratios with other quantities like

---

\(^2\)Compare also the discussion in chapter 1. For a discussion of the accurate measurement of the perimeter, I refer the reader to chapter 4 and Ref. [34].
the width of the basin or the length of the main stream \[34, 26, 27, 28, 29, 28, 30, 31, 32, 33, 27\]. These kind of ratios (morphometric indices) are important parameters used to model and predict, e.g., water runoff or sediment discharge, which are fundamental quantities in landslide \[4, 5, 6, 7\] and flood prevention \[6, 8, 9\], as well as in the management of aquatic ecosystems \[1, 2, 3\].

Especially in hilly landscapes, the ridges defining the watersheds are important landmarks, often building the boundary of geographical regions and sometimes also meteorological ones (regions of similar weather conditions). This is why they have been discussed, for example by Hartshorne in Ref. \[75\] (for the case of Silesia), as a possible basis for the delineation of political boundaries, at least where they are visible. An example for such a use is the border between Argentina and Chile, which even had a conflict, as the watershed changed its location after a huge landslide that was triggered by an earthquake \[76\]. An exception are also borders in glacier areas in the mountains, like in the Alps between Switzerland and Italy, which from time to time needs an adjustment \[77\]. Nevertheless, only few borders are chosen to follow a watershed, such that, e.g., water management has become an important matter in international politics.

Watersheds that follow ridges have, apart from being landmarks, other properties, which made them feasible as trackways for traveling and long distance trade in ancient and medieval time \[78, 79\], often called ridgeways. They are typically dry as the water drains away from them and preserve themselves without constant maintenance \[80\]. They have thinner vegetation, especially less trees than in the valleys, such that they might be safer from attack by marauders, robbers or animals \[81\]. Due to following watersheds, a further advantage of ridgeways over roads in valleys, is that they do not need to cross rivers or tributaries by fords or bridges \[80\]. In Europe, ridgeways have been in use since ancient times and in some cases are associated with Roman \[82\] and even pre-Roman usage \[83\]. A revival of long-distance footpaths in the 19th century, brought the rediscovery of some of these trails, such as the Ridgeway over the Berkshire Downs \[81\], the Old Shaftesbury Drove and the Ox Drove \[83, 84\] in Great Britain and the Rennweg in the Arnsberg forest, the Brüderstrasse between Cologne and Siegen, and the Rennsteig through the Thuringian mountains in Germany \[79\].

Last but not least, the concept of watersheds is also extensively used in the seemingly unrelated field of image processing and has been found to be in close relation with other physical models, as is discussed thoroughly in the
2.1. LANDSCAPES AND WATERSHEDS

introduction and in the chapters 3 and 6.

2.1.3 Influence of the Landscape Topography

As discussed above, the location of a watershed is mainly given by the topography of the landscape. Hence, it can also be influenced by changes in the landscape. Landscapes change shape due to several different geomorphological mechanisms like erosion [37, 38, 39], natural damming [6], tectonic motion [40, 41, 42], and volcanic activity [43]. The first two mechanisms are closely related to the drainage of precipitated water and therefore dependent on the watersheds, creating a feedback loop between landscape morphology and watershed location [85]. The latter two on the other hand, are mechanisms independent of watersheds. As I will discuss later the effect of local events classified as stream captures or drainage rearrangements, which can occur due to these mechanisms, I briefly review them in the following.

Erosion is the process in which water, ice, wind or gravity remove rock, sediment, soil or other materials from a certain region to transport it and deposit it at another place (sedimentation). Usually these materials had been weakened or loosened in advance by a process called weathering. One distinguishes many different kinds of erosion, from which fluvial erosion (rill, gully, and stream erosion) is the most important in shaping the earth’s surface. Other types are:

- Mass wasting (creep, landslides, flows, topples, slump, and falls), the movement of rock and soil down a slope due to gravity.

- Coastal erosion by hydraulic action, wave pounding, and abrasion due to the impact of waves [86].

- Glacial erosion by abrasion, plucking, and thrusting.

- Wind erosion by deflation and abrasion.

- Thermal erosion, the melting and weakening of permafrost.
Erosion rate typically depends on several geologic, climatic, and biological factors, such as type, porosity, and permeability of the rock or sediment, whether they are weathered or not, the gradient (slope) of the land, temperature on average and its variation, precipitation intensity and amount, storm frequency, ground cover by vegetation, as well as inhabiting organisms. In most cases human land use like, e.g., agriculture and road construction enhances erosion, as it reduces the vegetational ground cover or increases surface flow by compacting the loose soil (decreases the infiltration rate).

Natural damming most often occurs due to landslides blocking a river valley, but also deposits from glaciers and volcanic eruptions can disrupt drainage patterns [87, 88, 6, 76]. One of the most recent examples of landslide dams is the Usoi Dam in Tajikistan blocking the Murghab River and creating the Sarez Lake [88]. Another example is the earthquake-triggered landslide in the Andes that lead to a conflict between Argentina and Chile [76]. Sediments deposited by glaciers created, e.g., the lake of Geneva and Zurich in Switzerland. A very special case of natural damming that occurs, e.g., in the Plitvice Lakes in Croatia and constantly changes the shape of these lakes is the creation and dissolution of travertine barriers. These barriers are created by an interplay of the water, the chalk dissolved in it, the air, and plants. The chalk dissolved from the surrounding karst area is deposited on moss and algae, encrusting them. The deposited sediment is called travertine builds up barriers growing at a rate of 1 cm per year by the accumulation of encrusted plants on top of each other. These barriers are very fragile and subjected to constant changes, but do not change the overall drainage of the lakes, which is why the lakes always have been considered as one composed system of lakes. Apart from the natural occurrence, also manmade dams can affect the drainage, as for example the power plant Kaprun in the Austrian Alps close to Salzburg, which shifts the drainage of the Pasterze from south to north.

Tectonic motion is the movement of the tectonic plates, the pieces in which the earths crust, the lithosphere, is broken. It is thought to be driven by three mechanisms of which the relative importance is still unclear. These are the motion of seafloor away from the spreading ridge, drag or downward suction at the subduction zones, and forces generated due to tidal forces of the Sun and the Moon and Earths rotation. The different relative motions of the plates are responsible for the folding of mountains, earthquakes and creation of zones of active volcanism. Here in this work I focus mainly on events triggered by tectonic uplift and subsidence [40, 41, 42, 89].
Volcanic activity or volcanism, is the process in which magma, melted rock from below the solid Earths crust, is rising to the surface of the planet, where it cools down and freezes into rock again. But not always the magma reaches the surface, before it solidifies. The different rates of cooling and the mineral composition of the magma create various types of rock from obsidian (very fast), basalt (fast), tuff (fast, from explosive eruptions) to the very common granite (slow). Also the melting or heating of preexisting rocks due to the passing magma leads to new types of rock. Hence, volcanism contributes to the large variety of rock types observed on Earth. The process itself is driven by plate tectonics and thermal convection in the molten rock of the Earths mantel. Therefore, zones of volcanic activity are typically located at the boundary of the plates. Apart from the volcanos, the spots where magma exits the crust in the form of lava, many other interesting phenomena are due to volcanism, such as geysers, hot springs, and mud pots. Volcanism can also lead to earthquakes. Some of the most severe eruptions of volcanos, such as the one of Krakatoa in 1883 or the one of Mount Tambora in 1815 leading to the *The Year without Summer* [90], even had an effect on the worlds climate. In the present work, my main interest lies within those volcanic events, that lead to drainage rearrangements [43].

2.1.4 Modeling Real Landscapes

The interplay of the above described mechanisms created todays complex shape of the Earths surface, but it is difficult to distinguish just from this current state, which process at what time had dominated the evolution of the landscape. Therefore, geomorphologists have devised, apart from numerous field studies [37, 38, 39, 6, 40, 41, 42, 43], laboratory scale experiments [91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 45, 103, 104, 105] and numerical simulations [106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 44, 126]. Paola et al. present in Ref. [127] a thorough review and comparison for most of these approaches. A review of the algorithms used in the numerical models can be found, e.g., in Ref. [126]. Most of the experimental and numerical models focus on erosion, but recently also tectonic uplift and subsidence, as well as variance in the rock types and rainfall patterns have been investigated by very sophisticated experimental setups [94, 95, 96, 97, 98, 99, 100] and numerical models [44, 126]. Relevant mechanisms in stream capture, which is the type of event studied later in chapter 5, have been studied experimentally by Douglass and Schmeecke
[45], while first simulations had been devised by Howard in Refs. [106, 107, 108]. Especially the formation and evolution of river networks gained a lot of interest due to their fractal behavior (see also next section and the chapters 4 and 5) and have been the focus of a tremendous amount of numerical studies [106, 107, 108, 111, 112, 113, 114, 115, 116, 117, 119].

2.1.5 Creating Artificial Landscapes

The creation of artificial, but naturally looking landscapes is an important task in today's film industry and game design. There, the correct evolution of the landscape and the nature of the contributing mechanisms is of almost no interest. The focus lies in the right appearance on the screen. For this purpose, the algorithms used to create the landscapes, or more precisely the images, take advantage of some of the properties observed in the above discussed studies, especially of the fractal structures [128] (see also next section). In contrast to these areas and the geomorphological studies, I do not intend to reproduce the appearance of natural landscapes by my model. As I will show in chapters 4 and 5, the only property that is needed to reproduce the results observed on natural landscapes, is the presence of spatial long-range correlations. To generate artificial spatially long-range correlated landscapes, I use the fractional Brownian motion (fBm) scheme [129, 128] together with the Fourier filtering method [128, 130, 131, 132, 50] (see Ref. [50] and chapter 4 for more details).

2.1.6 Digital Landscapes

Landscapes in nature are of course continuous, but to process them numerically (as I will do) they need to be discretized. Usually, this is done through a digital elevation model (DEM), which most often is just a square lattice with the lattice spacing equal to the resolution scale $r$, which is given by the width of the studied piece of real landscape divided by the number of sites in a row (or column) of my lattice. The height stored in each site of the lattice is then given by an average of the real heights in the square area of width $r$ centered around the lattice site. For an artificial landscape, these height values at each site are generated directly (without the need of an averaging).
Such discretized landscapes can also be easily displayed as greyscale images, where a pixel corresponds to the resolution square centered at the lattice site and the grey intensity of the pixel reflects the height value stored at the lattice site.

2.2 Fractal Concepts

The concept of fractals and fractal geometry has been introduced by Benoit B. Mandelbrot [129, 133] as a theory able to describe the fact that “clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line” [133]. He discovered that natural objects, such as the coast of Britain [129], look similar on images of different magnifications of themselves and that there is no way to distinguish them if one does not know already the studied shape, i.e. they are self-similar. Even more, one cannot determine, which image has the higher magnification. His concepts proved to be valuable in diverse fields across several disciplines like geophysics, physics, biology, chemistry, and economics. Before I give you the formal definition of a fractal, I first need to go through some definitions, such as self-similarity, deterministic, statistical, fractal dimension, and self-affinity. The section closes with an overview of fractals in nature and some of its applications.

2.2.1 Self-Similarity

Self-similarity in this context means invariance under an isotropic transformation (dilation). If considering an object \( K \) formed by a set of points \( X = (x_1, x_2, x_3, \ldots, x_E) \) in \( E \)-dimensional space, a dilation (or similarity transformation) with a scaling factor \( r \), changes the coordinates to \( rX = (rx_1, rx_2, rx_3, \ldots, rx_E) \). The object \( K \) is self-similar with respect to the scaling factor \( r \) if \( K \) is the union of \( N \) non-overlapping subsets \( K_1, \ldots, K_N \) each of which is congruent to the set \( r(K) \) obtained from \( K \) by the similarity transform with \( 0 < r < 1 \). Here congruent means that \( K_i \) is identical to \( r(K) \) after possible translations and/or rotations of the set. An example of a self-similar shape is the Sierpinski gasket shown in Fig. 2.1. It is constructed from a filled triangle (Fig. 2.1a) by dividing it into four triangles of equal size and removing the central one (Fig. 2.1b). In the second step, the

\[ D_S = \frac{\ln(N)}{\ln(1/r)}. \]
same procedure is repeated for the remaining three filled triangles (Fig. 2.1c). This process is iterated indefinitely. The self-similarity of the gasket can be checked by rescaling the portion marked with a dashed square in Fig. 2.1d by a factor of 2 isotropically, which exactly matches the gasket in Fig. 2.1c.

![Fig. 2.1: Construction of the Sierpinski gasket.](image)

(a) Starting from a filled triangle, (b) a triangle, whose area is one fourth of the total, is removed from the middle. (c) In the next step, the same procedure is repeated for the remaining three filled triangles. (d) This process is repeated indefinitely.

### 2.2.2 Deterministic versus Statistical

Objects displaying such exact self-similarity, where the rescaled object $r(K)$ perfectly overlaps a piece of the original object $K$, are called *deterministically self-similar*. They can be created by a recursive application of the same rule to the next lower scale. Examples are the von Koch curve, the Sierpinski
2.2. FRACTAL CONCEPTS

gasket (see Fig. 2.1), and the Cantor set. The object $K$ is called statistically self-similar, when $r(K)$ is not congruent with the subsets $K_1, \ldots, K_N$, but shares the same statistical properties, i.e. they are identical in all statistical respects [133, 134, 135]. Most fractals in nature, such as the coast lines and the watershed (compare chapter 4), are statistically self-similar for all scaling factors $r$ above some lower cutoff given, e.g., by the smoothening of erosion or the grain size of rock and sand (micro-scale) and some upper cutoff given, e.g. by the size of the island or the planet (macro-scale).

2.2.3 Fractal Dimension

To characterize quantitatively a self-similar object and decide on its fractality, its Hausdorff-Besicovitch dimension must be measured. The volume $V(s)$ of an arbitrary object can be measured by covering it with cubes of linear size $s$ and volume $s^{d_E}$, where the embedding dimension $d_E$ is the smallest Euclidean dimension of the space in which the object can be embedded. One needs $N(s)$ cubes to cover it, so

$$V(s) = N(s)s^{d_E} \quad \text{with} \quad N(s) \sim s^{-d_f} \quad (2.1)$$

For an Euclidean object $d_f = d_E$ as expected, since the volume does not change if one changes the unit of measurement $s$. Scale-free objects are called fractals and $d_f$ is their fractal dimension\(^5\). The fractal dimension of the Sierpinski gasket shown in Fig. 2.1 is $d_f = \ln(3)/\ln(2) \approx 1.585$. Fractals, in contrast to Euclidean shapes like spheres or cubes, have, hence, no characteristic scale, meaning that they are independent of scale and scaling. In the case of the coast line this means that its length $L$ (approximately) follows a power law with the resolution $s$, at which the length is measured, as

$$L(s) \sim \frac{1}{s^{d_f-1}} \quad (2.2)$$

where $d_f$ is larger than the coast's topological dimension ($d_T = 1$). Therefore, the coast is longer the more precise its length is measured, but only as long as the resolution is in an intermediate range between the lowest scale given by the smoothening of erosion, or the grain size of rock and sand, and the largest

\(^5\)Strictly speaking, a fractal is defined by $d_E \geq d_f > d_T \forall s$, where $d_T$ is the topological dimension of the object. $d_T = 0$ for a set of disconnected points, $d_T = 1$ for a curve, $d_T = 2$ for a surface, etc.
scale given by the size of the island or the planet. The fractal dimension gives a notion of the roughness or wiggliness of the coast. An analogous scaling law can be written for any shape, Fractal or Euclidean, where $d_f$ of the former can have any real value, while for the latter $d_f$ is always integer valued. There are many methods to (numerically) measure the fractal dimension. One is the well-known box counting, where the number of boxes $N$ of equal size $s$ are counted, which are needed to completely cover the object, defining the box dimension. Another is the measurement of the correlation function and a further one, for line-like objects, the yardstick method (compass dimension). In numerical simulations, typically the resolution is kept fixed and the system size is changed instead. A corresponding measurement is then the mass of the object in terms of the basic unit cell $s N$, as function of the system size $1/s$, yielding the mass dimension. From all these measures, the corresponding fractal dimension is obtained by a fit to the logarithmic slope

$$d_f = \frac{\log(N)}{\log(1/s)}.$$  

(2.3)

In fact, for a self-similar fractals, the different measures yield the same value for the fractal dimension [129]. Note, that the fractal dimension is a statistical property, which is preserved across all scales. The appropriate method for the estimation of $d_f$ is usually determined by the particular situation, or the form of the available data, but it is reasonable to use more than one method to show consistency. In chapter 4 I apply mass scaling as well as the yardstick procedure to obtain the fractal dimension of watersheds.

### 2.2.4 Self-Similarity versus Self-Affinity

The above used power-law description for self-similar objects is isotropic. This means that changing the scale or resolution in one direction changes the system size uniformly in every spatial direction. There exist many objects, which are not similar under such isotropic re-scalings, but are under anisotropic ones, i.e. different scaling factors $r_1, \ldots, r_E$ for the different directions. The corresponding transformation of an object $K$ formed by a set of points $X = (x_1, \ldots, x_E)$ in E-dimensional space, changes the coordinates to $\hat{X} = (r_1 x_1, r_2 x_2, \ldots, r_E x_E)$ with the scaling factors $r_1, \ldots, r_E$ not necessarily equal to each other. Objects that are invariant under such a transformation are called self-affine. The simplest examples are single valued self-affine
2.2. FRACTAL CONCEPTS

Fig. 2.2: Construction of a deterministic self-affine curve. The diagonal in (a) is divided in four equal parts horizontally and replaced by the structure shown in (b). In the next iteration, the same procedure is repeated on all four segments, what leads to the structure in (c). If the section in the dotted box is rescaled isotropically, one obtains what is shown in the right, different from the shape in the previous iteration.

functions

\[ f(x) = k^{-\chi} f(kx) , \]  

(2.4)

where \( \chi \) is the Hölder exponent measuring the roughness of the function \( f(x) \). This generally formulates the fact that under a rescaling by a factor \( k \) in \( x \).
one needs to blow up vertically by a factor $k^x$. An example of a deterministic self-affine curve is shown in Fig. 2.2. As is shown there, an isotropic rescaling of a piece of the shape does not match the previous iteration. It would need an additional stretching on the horizontal axis. The simplest random, i.e. statistical, self-affine fractal is generated by a one-dimensional random walk on a lattice, in which the walker in each step moves either up or down with equal probability. In general, fractional Brownian motion generates self-affine walks and landscapes with a tunable Hölder exponent. The concept of fractals now can be extended to the more general notion of self-affine fractals, but a full description of such a self-affine objects in terms of fractal dimension is complex, as its value might depend upon the measurement technique [136]. I will show in chapter 4 that watersheds are indeed self-similar and not self-affine.

### 2.2.5 What is a Fractal?

With the above introduced terms I can define a fractal as a scale-free object with a rough or fragmented geometric shape, which is self-similar (or self-affine) and has a Hausdorff-Besicovitch dimension (i.e. fractal dimension) $d_f$. For a discussion of the definitions and the above presented concepts I refer the reader to the books of Mandelbrot [133], Feder [135], and Barabasi and Stanley [137].

### 2.2.6 Fractals in Nature

The broad variety of fields, where fractal geometries were found to be applicable is astonishing. In fact, nature presents itself to us with numerous structures with fractal characteristics, such as for plants like trees, ferns, and broccoli, whose shape can be reproduced by simple algorithms. Weather phenomena like lightening, clouds, rain, and snow flakes have a fractal pattern [133]. In biology the fractal dimension is descriptive for the branched networks of blood and pulmonary vessels, proteins (also transport and diffusion in them [138]), and the time series of heartbeats [139]. Also in physics fractal concepts apply, like for the microscopic condensation of gold particles in a thin film ([140] related to percolation, see also next section), $1/f$-noise (flow of cars [141], yearly flood levels of the Nile river [142], and many physical systems [143, 144]), crystals [145], and diffusion limited aggregation [146, 137] (describing, e.g., electrochemical deposition, fluid-fluid displace-
ment, or plant-like growth). Of major interest to us, apart from the percolation concepts introduced in the next section, is the fractal behavior observed for many geological features like coast lines [129], mountain ranges, seismic faults, craters [147], ocean waves [148], and river networks [149, 150, 151, 116]. Especially, river networks are closely related with my watersheds, as these build the boundary of the corresponding drainage areas. The scaling properties of river networks were investigated by extensive numerical simulations throughout the years [106, 107, 108, 111, 112, 113, 114, 115, 116, 117, 119]. For an introduction to the area of fractal river networks I refer to the famous book of Rodriguez-Iturbe and Rinaldo [116].

2.2.7 Practical Applications of Fractals

Additionally to provide a theoretical and quantitative description of the observations made in nature, fractal geometries have important applications in many different fields. They are used in film and game industry to produce naturally looking landscapes, plants, texture, clouds, etc. for the use in special effects (for a how to see, e.g., Ref. [128]). They are useful in signal and image compression, and for digital photographic enlargements. Soil mechanics, fracture mechanics, seismology, enzymology, turbulence and chaos, and analysis of price series in economy are some of the further areas, where fractal concepts are applied.

2.3 Percolation

In engineering and material science percolation represents the flow of fluids through porous media, but in mathematics and physics it generally refers to simplified lattice models of random systems or networks (graphs), and the nature of the connectivity in them. It is one of the simplest mechanisms giving rise to random (statistical) fractals in nature and has applications in many areas, such as epidemics, economics, complex networks, transport in porous media, fragmentation, super cooled water, gelation, galactic structures, and earthquakes. Its importance lies in the fact that it is a simple geometric model undergoing critical phase transitions.
2.3.1 Origins of Percolation Theory

The origins of percolation theory lie in the studies of Flory [152, 153, 154] and Stockmayer [155], who used it to describe how large macromolecules form from small branching polymers by successive adding of chemical bonds to the original molecules. This process leads to the formation of a network of chemical bonds spanning the entire system and is generally termed as gelation. An example of such a process is the boiling of an egg, which first is liquid, but upon heating becomes more solid-like (a gel). The name percolation, however, and several of its theoretical concepts have been introduced by Broadbent and Hammersley in 1957 in their work on the flow of liquids in porous media [156], where they asked the following question. Imagine a fluid poured on top of a porous medium. Is the fluid able to make its way from pore to pore and reach the bottom? Is it able to percolate? The fluid spreading randomly through the medium may involve two quite different types of randomness. One case is the random walk of the fluid particles in the diffusion process, and the other is, where the randomness is frozen into the medium itself. The latter is what they refer to as percolation process. The mathematical model they developed to investigate their question in a random medium is the percolation theory.

A nice introduction to percolation theory with an emphasis on critical phenomena, which occur at the percolation threshold, is given by Stauffer and Aharony [157]. The mathematical aspects are discussed by H. Kesten in Ref. [158]. The relations to disordered systems, fractals, and diffusion are explored, e.g., in the books of Bunde and Havlin [159], Pynn and Skjeltrop [134], Feder [135], and ben-Avraham and Havlin [160]. A review of applications of percolation theory can be found in Sahimi’s book [161]. It is important to note here also the relation Kasteleyn and Fortuin [162] have drawn for their percolation model to Ising and other Potts models of ferromagnets.

2.3.2 What is Percolation?

Imagine a two dimensional square lattice of size $n \times n$ (assume $n$ to be large) with the sites (vertices) connected by bonds (edges). Sites connected by one bond are called nearest-neighbor sites (in the square lattice each site has four nearest neighbors). Now, a certain fraction of sites is open, the others

---

from Lat. percolare, to filter or trickle through.
are closed. I define a cluster as a group of open sites, which are connected to each other by an unbroken chain of nearest-neighbor pairs. Percolation theory now deals with the number and properties of these clusters.

The site percolation problem as introduced by Broadbent and Hammersley [156] is defined as follows. Consider the sites of a square lattice of size $N$ to be open at random with a probability $p$ and closed with probability $1 - p$, independent on their neighbors. For large systems this means that $pN$ sites are open. While at low values of $p$ one finds only small clusters, they grow with increasing $p$ until eventually one cluster spans the system, connecting the top and bottom of the lattice. This cluster is called spanning cluster, infinite cluster or percolation cluster. The question is now, what is, for a given $p$, the probability $P_N(p)$ that such a spanning cluster exists? By Kolmogorov’s zero-one law [163], for any given $p$ and $N \to \infty$, the probability that an infinite (spanning) cluster exists is either zero or one. As for $p = 0$ all sites are closed, no spanning cluster exists and $P_{\infty}(0) = 0$. But for $p = 1$ every site is open and belongs to one single cluster and clearly $P_{\infty}(1) = 1$. Therefore, there must be a critical $p$, denoted by $p_c$, below which no spanning cluster exists ($P_{\infty} = 0$) and above which the fluid can run through the medium ($P_{\infty} = 1$). This critical value $p_c$ is called the percolation threshold and the transition from an unconnected (disordered, $P_{\infty} = 0$) to a connected (ordered, $P_{\infty} = 1$) state (phase) is the percolation transition, in analogy to a phase transition in thermodynamic systems.

### 2.3.3 A Model for Phase Transitions

The simplicity and fundamental character of percolation and the fact that it has a critical point, make it an attractive tool to study statistical physics concepts such as fractals, scaling theory, renormalization, critical phenomena, and phase transitions. One of the most remarkable features of percolation

---

7 equivalently I could call them filled and empty, occupied and vacant, red and blue, or with any other pair of words. In analogy with porous media, I stay here with open and closed.

8 Alternatively but not equivalent to this, I can consider bond percolation in which the bonds are randomly open with probability $p$ or closed with $1 - p$.

9 In simulations, one estimates the wrapping or percolation probability $P_N(p) = M(N)/N$ from an average over many realizations, where in each of which the mass $M(N)$ of the largest cluster, i.e. the number of sites in it, was determined.
processes is the *universal* scaling laws that they obey\(^{10}\). The behavior of many percolation quantities near \( p_c \) is independent on microscopic details of the system like the lattice structure or whether site or bond percolation is considered. Quantitatively, this universality is apparent in the fact that many percolation properties obey scaling laws near the percolation threshold, whose characterizing *critical exponents* depend only on the dimensionality of the system. One of those is the fractal dimension \( d_f \) of the *incipient infinite cluster* (the spanning cluster at \( p_c \)), another is the Fisher exponent \( \tau \) describing the cluster-size distribution \( n_s \sim s^{-\tau} \) and the correlation length diverges as \( \zeta \sim |p - p_c|^{-\nu} \), while the order parameter, the fraction of open sites in the incipient infinite cluster \( X_A(p) \), increases as \( |p - p_c|^{\beta} \) and is zero below \( p_c \). The remaining are the anomalous dimension \( \nu \), the exponent \( \sigma \) for the divergence of the largest cluster size, and \( \gamma \) for the divergence of the mean cluster size. In fact, most of these exponents are linked to each other by exponent (or scaling) relations, such that, e.g., from knowing \( d_f \) and \( \nu \) all others can be derived. Until today, a great deal of work has been done in finding these critical exponents, both theoretically (limited to two dimensions) and numerically.

### 2.3.4 A Model for Invasion Processes

The question raised by Broadbent and Hammersley ([156], see also above) can also be addressed by a dynamic percolation process called *invasion percolation* [164], which imitates the displacement of one fluid by another in a porous medium. Consider the case in which oil (the defender) is displaced by water (the invader). When water is injected very slowly into a porous medium filled with oil, the capillary number is small, i.e. the capillary forces dominate the viscous ones. The capillary forces are strongest at the narrowest places in the medium. Hence, if all throats are smaller than all the pores, the water-oil interface moves fast through the throats, but gets stuck entering the larger pores. Therefore, the dynamics is determined by the local pore size and can be represented by a series of discrete jumps in which at each time step the water displaces oil from the smallest available pore.

In the idealized model, the network of pores is viewed as a regular lattice (e.g. square lattice) in which sites and bonds represent the pores and throats,

\(^{10}\)This is a typical characteristics for thermodynamical and configurational systems near a critical point or a continuous (2nd order) phase transition
respectively. For convenience, the bonds (throats) are assumed to be invaded instantaneously and randomness is induced by assigning random numbers to the sites representing the pore sizes. The injection sites, typically one of the boundaries of the lattice, and extraction sites, the boundary opposite to the injection boundary, are chosen. Now, *growth sites* are identified as the sites of the defending fluid, that are neighbors to the invading fluid, i.e. the perimeter of the invasion percolation cluster. At each time step the invader advances to the growth site with the lowest random number. In this advancement, the invader can completely surround regions of the defender, thereby disconnecting that finite cluster of the defender from the extraction sites. In the case of oil, which is incompressible, such a cluster is trapped and the invader cannot penetrate this area any further. This is one origin of the phenomenon of residual oil, a great economic problem in the oil industry, and is modeled by the removal of trapped defender sites from the list of growth sites. At breakthrough, the spanning cluster of invaded sites is a fractal and, as now the invader has an open path from one end to the other, no further sites need to be invaded. This means that spanning cluster is at its critical point and stays there, which is a behavior known as self-organized criticality [165].

Important to note are the two main differences to standard percolation. First, invasion percolation will always create a spanning cluster, no finite ones and there is no analogue to the opening probability $p$ for all the pores as in standard percolation. Second, invasion percolation is a dynamic process with a well defined sequence of invaded sites. The growth of the invaded cluster is subject to local properties, but trapping introduces a nonlocal aspect to the model, as it needs global information. Indeed, with trapping the spanning cluster has a fractal dimension lower than the one for the incipient infinite cluster of standard percolation [166]. But for the invasion process defined without trapping there exist considerable evidence of equivalence to the standard percolation. As I will explain later (see chapter 3), invasion percolation without trapping is a suitable model for drainage on a (hyper-)surface, as there the defender always can escape.

### 2.3.5 Recent Advances

Although percolation is studied since decades, there are still some open questions. Most of them are related to two new concepts that have been developed recently, explosive (or discontinuous) percolation [167, 63, 64] and bridge
percolation [62], which add completely new features to the theory. They are both dealing with the systematic suppression of connectivity and are closely related to each other and to the watersheds discussed in this thesis (see chapter 6 for details on these relations). I, therefore, briefly review these two models here.

**Explosive or Discontinuous Percolation**

Recently, Achlioptas, D’Souza, and Spencer [167] have proposed a new percolation model, which showed a more pronounced transition than in the random case, being apparently discontinuous. Although this best-of-two product rule for bond selection was shown to still obey a continuous transition [168], the possible existence of a percolation model with a discontinuous phase transition was exciting. Several different models have been studied to find the main mechanisms leading to a discontinuous percolation transition, such as the Gaussian model [63, 64], the BFW model [169, 170], and the global competition model [171]. For the first of these three models, the fractal dimension of the external perimeter of the largest cluster was found to be compatible with my watersheds and with bridge percolation (compare also chapters 4 and 6).

**Bridge Percolation**

Bridge percolation is a new percolation model introduced by Araújo et al. [62] in which the creation of a spanning cluster, i.e. the connectivity, is suppressed systematically by giving the bridges, i.e. bonds that if occupied would create the first spanning cluster, a different weight. Its general idea is to appropriately incorporate the effect of local reinforcements, which lead to the above discussed discontinuous transition, into the standard percolation model. For more details on the model definition and the study of its relation to watersheds, see chapter 6. The relation of bridge percolation to discontinuous percolation is discussed in Ref. [64].

### 2.4 Formal Definition of Watershed

On my digital landscapes, i.e. square lattices, the sites $i$ represent areas with an average height $h_i$, whose edges (the bonds) connect them to their neighbors. I define the set of sites $S_k$, $k = 1, 2, \ldots, N_{\text{sinks}}$, to be the sinks of
my system\textsuperscript{11}. As in percolation, I distinguish two different states for a site, which I call again open and closed (wet and dry if you want to stay close to the rainfall and landscape picture). A cluster is again a connected set of open sites. Starting from single open site (all others are closed so far), I consider that at each time step the lowest-lying site on the perimeter of the cluster of open sites changes its state from closed to open and is added to the cluster. The perimeter is that set of closed sites in which each has an open nearest neighbor. If when starting from \(i\), sink \(S_k\) is opened before any other, then site \(i\) and the whole cluster are considered to belong to the catchment basin of sink \(S_k\), equivalently one says that \(i\) drains to \(S_k\). This procedure introduces a classification of lattice sites, i.e. a subdivision into non-overlapping subsets whose union is the entire lattice. The sets of bonds separating neighboring basins I call watersheds\textsuperscript{12}.

Throughout this thesis, I use the following simplifications. First, I restrict the number of sinks, and hence the number of drainage basins, to two \((k \equiv 2)\). Second, the sink \(S_1\) shall be the entire top row of a lattice of size \(L \times (L + 2)\) and the sink \(S_2\) the entire bottom row, while the central part of the lattice \((L \times [1 : L])\) shall be the actual landscape under study. This means that the sinks are excluded from the landscape and that the boundaries at the top and at the bottom are solid boundaries. The boundaries in the lateral direction (left and right) may be considered as periodic, although for natural landscapes obtained from satellite imagery it is more meaningful to consider them as solid as well. The introduced simplifications imply that the watershed, as defined above, is a single line of bonds separating the entire lattice into two distinct and compact parts. This allows also for the study of long-range properties of watersheds, as no branching occurs.

There are two ways to look at the watershed problem, a local and a global one. The local view resides at the watershed, focussing on its partition between two sites of different basins. The related question is, to which basin does the site on the left of the bond and to which the one on the right belong. The global view resides at the sink and asks, which sites belong to this sink. These two aspects lead to two conceptually quite different algorithms to determine the watershed. They are discussed and compared to each other in the next chapter.

\textsuperscript{11}On natural landscapes these sinks are the natural water-outlets of the terrain, as for example a lake, a river or an ocean.

\textsuperscript{12}Note, that there is also the possibility to represent the watershed as a set of sites as is discussed in chapter 6. For simplicity, I will stay here with the bond representation.
Chapter 3

Watershed Algorithms
I present a new algorithm for the determination of watershed lines on Digital Elevation Models (DEMs), which is based on the iterative application of Invasion Percolation (IIP). The main advantage of my method over previously proposed ones is that it has a sub-linear time-complexity. This enables us to process systems comprised of up to $10^8$ sites in a few cpu seconds. The method allows us to address highly complex questions concerning the geometrical and dynamical properties of watersheds, which will be discussed in chapters 4, 5, and 6.

3.1 Introduction

Traditional cartographical methods for basin delineation relied on manual estimation from iso-elevation lines and required a good deal of guess work. Modern procedures are based upon the automatic processing of Digital Elevation Models (DEM) or Grayscale Digital Images where gray intensity is transformed into height. Although many other methods have been developed for image segmentation, such as clustering, histograms [23], edge-detection [24], region growing [10], level-set or graph partitioning [25], also watersheds are currently used. One of the most popular algorithms for watershed determination [10] uses rather complicated data structures and at least one pass over all pixels in order to calculate watersheds, and is adequate for grayscale images, i.e. integer-height spaces. In this chapter, I will discuss a slightly adjusted version of this method, together with a novel technique, called iterative invasion percolation procedure (IIP), recently presented by Fehr et al. [65]. I prove the equivalence of the two methods and show the advances of the novel method over the existing ones in terms of computational effort, i.e. time complexity.

In the following, I use the definition of watershed given in section 2.4. It is important to note that, although I use all the restrictions introduced in section 2.4, the algorithms, which I introduce here, are able to deal also with multiple-sink systems (more than two sinks). This because they are not restricted to certain number of sinks/labels but are free to deal with as many as one would like to.
3.2 Flooding

I describe hereafter a slightly adjusted version of the commonly used watershed algorithm for image segmentation [10] (actually it is making use of disjoint-set data structures). In this procedure, which I call flooding, the whole lattice is flooded (or occupied) in order of increasing height, i.e. at time $t$ all sites with $h_i < h(t)$ are occupied and $h(t)$ grows site per site in time. This procedure is also related to Kruskal’s algorithm for growing MST’s [172, 173]. It is assumed that each site has a height, which is a real number, such that after sorting them, a unique sequence of sites is defined and only one site gets occupied per time step. At the beginning, each sink is assigned a label, and I will furthermore specify that a cluster of occupied sites that gets connected (through a path of occupied sites) with a sink is labelled accordingly, i.e. labels propagate from the sinks. Clusters of occupied sites not yet in contact with any sink remain unlabeled. Whenever two different labels would get in contact with each other because of the addition of a new site, the (unique) bridging site is first labelled with the label of its lowest-lying neighbor. Next all bonds connecting sites with different labels are “cut” so that labels from different sinks never mix. Bonds connecting different labels are called bridges. When this flooding procedure is completed, the different labels identify the corresponding catchment basins and the set of bridges, i.e. the set of bonds separating the different basins, identifies the watershed(s). In the described form the sinks are predefined. Of course one can reformulate this procedure to identify all intrinsic catchment basins automatically by just introducing new labels when a site gets occupied that has only non-occupied neighbors. This will in fact return all possible watersheds on the current landscape, even of the smallest basins, which is probably not what one would like to have. Unfortunately using the above described flooding algorithm one has in general to consider each site or pixel to completely determine the watershed, such that, already without considering the needed sorting, this algorithm performs at least with order $O(N)$.

3.3 Iterative Invasion Percolation Procedure

In my algorithm, that I call Iterative Invasion Percolation (IIP), a cluster is started from site $i$ and grown by adding, similar to Invasion Percolation (IP), at each step, the smallest-height site on its perimeter until the first sink is
reached (see figure 3.1). This can be done rather quickly when using binary heaps or other tree methods to sort the list of perimeter sites according to their height values. This procedure is related to Prim’s algorithm for growing Minimum Spanning Trees (MST) from $i$ [172, 173]. When implemented blindly over the whole lattice, however, this process would be inefficient since in principle a new cluster has to be grown from each site $i$. By noting that all sites occupied by a cluster at the time the sink is reached also drain to that sink, it becomes clear that one can set for each site in the cluster a link to that sink. This leads to two improvements, first, I do not need to treat again the sites contained in the cluster as they can be directly linked to the sink reached by the cluster, and second, I can stop growing clusters already when I reach a site which has an existing link to a sink. Like this, the algorithm has to visit each site only once. But still each site on the map has to be visited. Let us assume a line dividing the system is already known and I want to check if this line is really the watershed of the current system. To prove this, I actually only need to test the sites adjacent to the watershed line if they belong to different sinks on the two sides of the line. Hence the subset of the system, formed by these adjacent sites, is the minimal subset I have to test, meaning to grow IP-clusters from them, in order to clearly determine the watershed. Unfortunately I do not know in advance the exact form or position of the watershed, neither do I know any of the sites in that minimal subset. The only thing I know is, that the watershed is located somewhere in between the two sinks. This means that, whatever path I follow, to go from one sink to the other, I have to cross the watershed. The sites I visited on such a path, before the watershed traversal, belong to the sink I started from and the sites afterwards belong to the second sink. This is what I now can use to find the first two sites of the minimal subset. I follow a straight line of sites connecting the two sinks. Starting from one sink I follow the connection site by site and grow IP-clusters from these sites. I proceed until I reach the first site that drains to the second sink, which means that I have crossed the watershed (see figure 3.1). The bond between the last two sites, that drain to different sinks, is part of the watershed line and the two sites belong to the minimal subset as mentioned above. Now I can reconstruct the minimal subset and the set of bonds forming the watershed line by just following step by step the direction of the already found watershed at one of its two ends and testing the next two sites to find if the direction changes by $\pm 90$ degrees or remains the same. I just walk along the watershed and hence have only to test slightly more sites than in the minimal subset (see figure
3.4. EQUIVALENCE OF METHODS

3.1. Of course the IP-clusters for these sites have to be grown until they reach a sink or sink-connected site, but there are typically large parts of the terrain that need not be visited at all by the algorithm in order to determine the entire divide. For use in multiple sink systems one would only have to consider additionally branching and joining of the watershed lines (also only one starting bond has to be found). That only a part of the system has to be considered (compare figure 3.2) and there each site only once, is probably the biggest advantage of this procedure. The described algorithm is fast enough to allow for the determination of watersheds on lattices comprising $10^8$ sites, in a few CPU seconds on a normal workstation.

3.4 Equivalence of Methods

I now sketch a simple proof for the equivalence of these two procedures. During the flooding growth a set of unlabeled sites that become labelled by occupying site $b$ is called a lake and $b$ is called its outlet. The height of all sites in a lake is bounded by that of its outlet, i.e. $h(i) \leq h(b(i))$. Let us stop the flood procedure when site $i$ gets labelled for the first time, say with label $S_k$. Now start an IP cluster from $i$ and let it grow for as long as it takes until some sink is first occupied. During the growth of this cluster, which proceeds by always adding the lowest-lying site on the perimeter, several lakes and their outlets may be occupied. I need to demonstrate that no bridge will be ever traversed during this IP growth, and therefore that sink $S_k$ will be occupied first. During the growth of the IP cluster, a lake will be entirely flooded i.e. the water will reach the level of its outlet. Since all bridges by definition are higher than the outlet of a lake (because by definition bridges were occupied later than the outlet), clearly the next occupation will proceed through the outlet and not through the bridge. Therefore no bridge will be ever traversed thus proving the equivalence of both procedures.

3.5 Time Complexity

It is important to note that, although conceptually equivalent, the IIP and the flooding algorithms are markedly different in terms of computational performance. First, the flooding needs a sorting of the sites according to their heights, which needs at least $O(N)$ time with $N$ the number of sites in
Fig. 3.1: Different stages during the search for the main watershed (green lines), red/blue for the labels to the two sinks, green are visited but not yet labeled sites, the numbers mark the integer heights of the corresponding site and the yellow box surrounds the starting site of the current IP-cluster, right and left periodic boundary conditions are applied. Upper most and lower most bonds connect to the sinks (which are entire rows). Further the lowest two sites are already labeled. (a)-(c) shows the growth of an IP-cluster until reaching a sink. (a)-(e) search for the first watershed bond, growing IP-clusters of each site along the first column consecutively. (e)-(g) Following the watershed.

the system, i.e. the system size. Additionally it has been shown by Fredman and Saks in [174] that building the disjoint-set data structure, which the
algorithm is based on, needs in all cases at least $O(\alpha(N)N)$, where $\alpha(N)$ is the inverse Ackermann function, which is constant for almost all values of $N$. Hence the flooding scales at least with $O(N)$. In figure 3.3, for the random artificial landscape case, the number of sites visited by the IP-based algorithm is shown together with its scaling in linear dimension $L$, which clearly points out the sub-linear time-complexity of this method. The total number of visited sites $n$ scales as $n \sim L^{D_f}$ with $D_f = 1.8 \pm 0.01$. This is comparable to what is found in nature. There, river networks show fractal dimensions between 1.7 and 1.9 [116, 175, 176, 177, 150, 178]. On large scales I expect this fractal dimension to be dominated by the largest IP-cluster I grow, for which I found a fractal dimension that is in good agreement with literature and the result for the total number of visited sites. Furthermore I expect that

Fig. 3.2: Labels of the sites for a random landscape obtained during the run of the IP-based algorithm. Red labels (lower part) belong to the sink at the bottom row and the blue ones (upper part) to the sink at the upper most row. The watershed is marked in yellow. Only the colored sites are visited by the algorithm, meaning that white sites have not to be considered by the algorithm in order to obtain the right set of watershed bonds.
the algorithm is even more efficient on natural landscapes, as the largest IP-cluster will most probably follow the main stream, which is reported to have fractal dimensions between 1.0 and 1.2 [116, 175, 176, 177, 150, 178].

![Log-log plot](image.png)

Fig. 3.3: Log-log plot of the number $N(L)$ of sites visited by the IP-based algorithm, in order to determine the watersheds on artificial landscapes, as a function of linear dimension $L$ (circles). Solid line shows the least squares fit to the five last data points, which gives a dimension $D_f = 1.8 \pm 0.01$ for large scales ($N \sim L^{D_f}$).

### 3.6 Summary

I developed a new numerical algorithm of sub-linear time-complexity and showed its equivalence to a currently used watershed algorithm. Giving some qualitative arguments and analysis I pointed out the improved efficiency of the presented method. This method is of great use in answering the questions concerning the geometrical and dynamical properties of watersheds, which are addressed in this thesis.
Chapter 4

Self-Similarity
I study the morphology of watersheds in two and three dimensional systems subjected to different degrees of spatial correlations. I find the fractal dimension of the watersheds to generally decrease with the Hurst exponent, which quantifies the degree of spatial correlations. Moreover, in two dimensions, my results match the range of fractal dimensions $1.10 \leq d_f \leq 1.15$ observed for natural landscapes and the roughness is equal to unity, proving watersheds to be truly self-similar objects.

4.1 Introduction

Self-similarity is a key-feature of many fractal objects and comes along with the important physical property of scale-freeness. One of the first natural objects shown to be a fractal was the coast of Britain [129]. Since then, many other objects have been studied, most importantly river networks [116]. There have also been preliminary claims about fractality of watersheds [34] but they were restricted to small-scale observations and therefore inconclusive. This would mean that several morphometric indices [34, 26, 27, 28, 29, 28, 30, 31, 32, 33], which are used to characterize drainage basins, have to be treated carefully, as the perimeter, i.e. the watershed length, is not a well defined quantity without knowing also the corresponding measurement resolution. In this chapter I analyze the fractal structure of watersheds, based on highly accurate data obtained from extensive numerical simulations.

4.2 Method

Self-similar structures are typically characterized by their fractal dimension $d_f$, which can be defined in two ways (compare chapter 2). For a single watershed I can apply the yard stick method, but this is only feasible in two dimensions, where the watershed is a line. Measuring the length $N$ of the watershed in terms of a given yard stick size $\delta$, meaning measuring the length on a certain resolution, I can define the corresponding fractal dimension as

$$N \sim \delta^{-d_f}.$$  \hspace{1cm} (4.1)

If enough statistics is available, i.e. data for different system sizes and configurations, another option is to define $d_f$ through the scaling of the mass...
4.3. UNCORRELATED ARTIFICIAL LANDSCAPES

$M$, corresponding to the number of sites or bonds in the watershed, with the linear system size $L$,

$$M \sim L^{d_f}.$$  \hfill (4.2)

In this chapter, I discuss the fractal dimensions obtained for watershed lines in two dimensions as well as for their extension to three dimensions [65, 67]. I use real and artificial systems in the form of Digital Elevation Maps (DEM), giving the elevations (resistances) on a square (cubic) lattice. The watershed is defined as the line (surface) dividing the entire system into two sets of sites, catchments, as shown in Fig. 4.1. Each site in a catchment drains to the same boundary of a chosen pair of opposite boundaries (top-bottom in Fig. 4.1) of the DEM. Hereby, the drainage of a site is determined by following the local slope and filling the valleys (local minima), until eventually reaching one of the two boundaries. For the determination of this line I use an iterative invasion percolation procedure (IIP), as introduced in Ref. [65] (see also chapter 3). Furthermore, I study the effect of long-range correlations on $d_f$.

4.3 Uncorrelated Artificial Landscapes

I applied the IIP algorithm to two dimensional artificial landscapes, where the height of each site on the lattice is an independent random variable uniformly distributed between 0 and 1, and two sinks are defined respectively as the upper- and lowermost lines of the lattice (see Fig. 4.1). Since the watershed location only depends on the order in which sites are occupied, it is clear that any distribution of heights will produce the same statistical results, as discussed for example in [172]. Due to the high efficiency of the method I could process a huge amount of data and hence gather a lot of statistics. This enables me to very precisely estimate the fractal dimension of the watershed using the mass-scaling. I measured the total length of the resulting watersheds and averaged this value over at least $10^4$ samples for a given lattice size. The results for artificial landscapes are shown in Fig. 4.2. Although I will show later (see chapter 6), that the corrections to scaling need to be considered for an accurate estimate of the fractal dimension, my data is strikingly straight on a log-log plot. Hence, I neglect here finite size corrections and measure the fractal dimension by a least squares fit to the slope of the shown data. I find that the watershed is a fractal, i.e. $M \sim L^{d_f}$, with a fractal dimension $d_f = 1.211 \pm 0.001$. I repeated the mass-scaling analysis using the yardstick method on individual samples of $L = 1.5 \times 10^4$. 41
Fig. 4.1: Example of a watershed for uncorrelated systems with linear size $L = 129$ lattice sites in two dimensions. The upper and lower catchments drain to the top and bottom boundary, respectively.

and obtained a consistent result $d_f = 1.21 \pm 0.01$. This value of $d_f$ is close to that found for Disordered Polymers ($\approx 1.2$ [179]), “strands” in Invasion Percolation ($1.22 \pm 0.01$ [180]), paths on MST’s ($1.22 \pm 0.01$ [172]), optimal path cracks ($1.215 \pm 0.005$ [48, 50]), bridge percolation ($1.215 \pm 0.002$ [62]), and the perimeter of the infinite cluster in explosive percolation ($1.23 \pm 0.3$ [63, 64]). The roughness exponent found for the watersheds is equal to unity within the statistical error bars, supporting the fact that they are indeed self-similar fractal objects and not self-affine.

### 4.4 Real Landscapes

I also applied the IIP algorithm to study the fractal properties of watersheds on naturally occurring landscapes, which clearly have long range correlations
Fig. 4.2: The mass $M$ of the watershed of (uncorrelated) random-height maps scales with linear dimension as $M \sim L^{D_f}$, with $D_f = 1.211 \pm 0.001$. A box-counting procedure (inset) gives a consistent result. This fractal dimension is independent of the type of disorder, as long as each height is an uncorrelated random variable.

[181, 118]. Of course, if I would like to estimate the fractal dimension of a single watershed line of one given landscape, as is the case for these natural topographies, I cannot work with the statistic approach used above for the artificial ones. Therefore I have to choose another method. As I follow the watershed bond per bond and hence have the correct ordered set of bonds, I can easily apply a yard stick method and using Eq. (4.1) to estimate the
fractal dimension. I checked this procedure also on several realizations of artificial landscapes and found the newly estimated fractal dimension to agree with the above presented $d_f = 1.211 \pm 0.001$ within the error bars (see the inset of Fig. 4.2). Therefore I can assume that this equality is also true for natural landscapes. Now, using topographic data derived from the Shuttle Radar Topography Mission (SRTM) [182], I analyzed several mountainous regions and determined their watersheds within the available resolution of 3 arc-sec (i.e. roughly 90 m). As shown in Fig. 4.3, a yard-stick analysis of the watershed performed in the range $100 \text{ m} < L < 200 \text{ km}$ for the Alps and Himalayas gives fractal dimensions that are indistinguishable within their error bars, namely $D_{Al}^f = 1.10 \pm 0.01$ and $D_{Hi}^f = 1.11 \pm 0.01$, respectively. The error bars were obtained by determining the lines of maximal and of minimal slope that would still be consistent with the data in Fig. 4.3. Again the roughness exponent found for the watersheds is equal to unity within the statistical error bars, hence they are still self-similar fractal objects and not self-affine for intermediate to large scales. The origin of the upper and lower-size scales is clear. On large scales ($> 200 \text{ km}$) the watershed follows the direction of the main crust foldings, which depends on processes occurring on the scale of tectonic plates, which are non-fractal. Hence the scaling should be essentially linear ($d_f = 1$). Although beyond the resolution, at small scales ($< 100 \text{ m}$), below the size of an individual mountain, the watershed would connect peaks and troughs, which are typically self-affine. In a later study [67], I applied the same analysis also to further DEM data of real landscapes, of which the resulting fractal dimensions are summarized in Tab. 4.1.

4.5 Disorder Independence and Correlations

It is important to notice that the fractal dimension of the watershed line $d_f$ is by definition independent of the type of disorder on the landscape, as long as each height is an uncorrelated random variable because only the spatial order of the random variables (heights) in the system matters and not their relative numerical differences. The small discrepancies (around 10%) between fractal dimensions of watersheds in the natural topographical data taken from SRTM [182] (see Tab. 4.1) and the uncorrelated artificial landscapes ($1.211 \pm 0.001$ [65, 67]) is expected as in natural landscapes typically long-range correlations are present [181, 118]. In particular, long-range correlations in space have been reported in a large variety of physical, biological and geological systems.
Fig. 4.3: Log-log plot showing the yardstick results for the Alps watershed. The solid line is the best fit to the data which gives a fractal dimension of $D_{Al} = 1.10 \pm 0.01$. The inset shows my results for the watershed calculated on the Himalayas. The fractal dimension in this case is $D_{Hi} = 1.11 \pm 0.01$.

Spatial long-range correlated distributions can be obtained with fractional Brownian motion (fBm) [129, 128]. Similarly to previous studies [66, 86, 130, 131, 132, 50, 187, 188, 189, 190, 191, 192, 193], I use the Fourier
Table 4.1: Fractal dimension of watersheds for natural landscapes obtained from satellite imagery [182]. I added here the values presented in Ref. [65] for the Alps and the Himalayas for completeness. The fractal dimensions are obtained using the yardstick method. The error bars are of the order of 2%.

<table>
<thead>
<tr>
<th>Landscape</th>
<th>$d_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpes</td>
<td>1.10 [65]</td>
</tr>
<tr>
<td>Europe</td>
<td>1.10 [67]</td>
</tr>
<tr>
<td>Rocky Mountains</td>
<td>1.11 [67]</td>
</tr>
<tr>
<td>Himalayas</td>
<td>1.11 [65]</td>
</tr>
<tr>
<td>Kongo</td>
<td>1.11 [67]</td>
</tr>
<tr>
<td>Andes</td>
<td>1.12 [67]</td>
</tr>
<tr>
<td>Appalachians</td>
<td>1.12 [67]</td>
</tr>
<tr>
<td>Brazil</td>
<td>1.12 [67]</td>
</tr>
<tr>
<td>Germany</td>
<td>1.14 [67]</td>
</tr>
<tr>
<td>Big Lakes</td>
<td>1.15 [67]</td>
</tr>
</tbody>
</table>

filtering method [128, 130, 131, 132, 50], which allows to control the nature and the strength of correlations. A detailed description of this method can be found, e.g., in Ref. [50]. In brief, the desired correlated distribution can be introduced by generating Fourier coefficients in the reciprocal space according to a power-law spectral density. For each frequency in the reciprocal space, I calculate these Fourier coefficients through a random phase in the interval $[0 : 2\pi)$ and an amplitude $\left(\sqrt{k_1^2 + \cdots + k_d^2}\right)^{-2H-d}$, where $k_i$ are the frequency indices of the discrete Fourier transform, $d$ the spatial dimension, and $H$ the Hurst exponent. The inverse Fourier transform is applied to obtain the distribution in real space. Finally, I normalize the spatial domain distribution in the range $[0 : 1]$ to represent the correlated topology, characterized by the Hurst exponent $H$. Three different categories of fBm surfaces can be distinguished: $0 < H < 1/2$, $H = 1/2$, and $1/2 < H < 1$ [194]. The correlations between the increments are persistent (positive) for $1/2 < H < 1$, meaning that sites with similar height tend to cluster together, leading to rather smooth surfaces. The opposite is true for $0 < H < 1/2$, where the correlations of the increments are anti-persistent (negative), resulting in surfaces that seem to oscillate more erratically. For $H = 1/2$, the classical Brownian motion is recovered, where the increments are uncorrelated but the obtained heights are still correlated. The uncorrelated distribution of
heights is solely obtained for a constant spectral density, i.e. $H = -d/2$ ($H = -1$ and $H = -3/2$ in two and three dimensions, respectively).

Fig. 4.4: (color online) The fractal dimension $d_f$ of the watershed as a function of the Hurst exponent $H$ of the system in two (red stars) and three dimensions (green open circles) according to left and right hand axes, respectively. In two dimensions, each point corresponds to an estimate of the fractal dimension by fitting the power law in Eq. (4.2) to the watershed masses for system sizes $L = \{5, 17, 65, 257, 1025, 4097\}$. For each system size, the mass was averaged over $10^4$ landscape realizations. In three dimensions, the system sizes $L = \{5, 9, 17, 33, 65, 129, 257\}$ were used with the same number of realizations. The lines show the fractal dimension of the watershed obtained for the Alps (solid) and close to the Big Lakes (dotted), characterizing, according to the left hand axes, the range of estimates for natural landscapes as summarized in Tab. 4.1. In the typical range of natural landscapes, $0.3 < H < 0.5$, the simulation results are in agreement with the natural ones.

In Fig. 4.4 I plot the dependency of the fractal dimension of the watershed
on the Hurst exponent, measuring the degree of correlations. A monotonic decrease of $d_f$ with $H$ is observed, in line with what was observed for the optimal path crack in $d = 2$ [50]. Considering the typical range of $0.3 < H < 0.5$ for natural landscapes [118], my simulation results in two dimensions are in agreement with the values for the fractal dimension of watersheds in natural landscapes measured from satellite imagery, as listed in Tab. 4.1. The lines in Fig. 4.4 show the fractal dimension obtained for the watershed of the Alps (solid line) and the one close to the Big Lakes (dotted line), characterizing the range of values obtained for natural landscapes.

![Example of a watershed for uncorrelated systems with linear size $L = 129$ lattice sites in three dimensions. The upper and lower catchments drain to the top and bottom boundary, respectively.](image)

Fig. 4.5: Example of a watershed for uncorrelated systems with linear size $L = 129$ lattice sites in three dimensions. The upper and lower catchments drain to the top and bottom boundary, respectively.
4.6 Extension to 3 Dimensions

I now extend the concept of a watershed to a three dimensional system, in which the values at the sites no longer represent heights but for instance resistances. The watershed is now a surface, as shown in Fig. 4.5, that divides the system into two parts. Similarly, each site in one part drains to the same boundary of a chosen pair of opposite boundaries (top-bottom in Fig. 4.5) of the DEM. The drainage of a site is determined by following the lowest gradient in the resistances and filling the regions of local minima, until eventually reaching one of the two boundaries. Again, the IIP procedure [65] can be used to determine the watershed surface numerically. Alike the two dimensional case, I use Eq. (4.2) to estimate the fractal dimension. For uncorrelated three dimensional systems, I find $d_f = 2.48 \pm 0.02$, again similar to what is found for optimal path cracks [48, 50], bridge percolation [62], and the surface of the clusters in explosive percolation [63, 64]. In Fig. 4.4, I show that the fractal dimension of the watershed for three dimensional systems also decreases continuously with the Hurst exponent $H$.

4.7 Summary

Using the highly efficient iterative invasion percolation procedure, I was able to measure the fractal dimension of watersheds in two and three dimensions to high accuracy. Applying the same algorithm to real landscapes I could prove that watersheds in nature indeed are self-similar objects, whose variety of fractal dimensions can be explained by the long-range spatial correlations present in the landscapes. Furthermore, introducing long-range spatial correlations into my model landscapes by using fractional Brownian motion, I was able to match the fractal dimensions of natural landscapes. The similarity of the obtained values for watersheds on uncorrelated landscapes to other physical models will be investigated later (see chapter 6).
Chapter 5

Impact of Perturbations
I study watersheds in two and three dimensional systems subjected to different degrees of spatial correlations. The response of these objects to small, local perturbations is investigated with extensive numerical simulations and, in two dimensions, compared to the one obtained for natural landscapes. I report that the watershed is strongly affected by local perturbations. For perturbed two and three dimensional systems, I observe a power-law scaling behavior for the distribution of areas (volumes) enclosed by the original and the displaced watershed, and for the distribution of distances between outlets. Finite-size effects are analyzed and the resulting scaling exponents are shown to depend significantly on the Hurst exponent, but to be independent on the strength of perturbation. The intrinsic relation between watersheds and invasion percolation, as well as relations between exponents are confirmed.

5.1 Introduction

So far I have considered that the properties of the landscape, and consequently of the watershed, are static, i.e., do not change in time. However, landscapes might change due to several phenomena such as erosion, natural damming, tectonic motion, and volcanic activity. Such changes in the landscape are known to trigger local events called stream capture [40, 37, 38, 41, 42, 43, 39], which can affect the watershed and the biogeography [35]. Recently, the associated relevant mechanisms were investigated numerically and in small scale experiments [44, 45]. Finally, the problem studied here is also of interest to image processing, in order to circumvent segmentation failure [46, 47].

But, how sensitive are watersheds to such slight localized modifications of the landscapes? Can these perturbations produce large, non-local changes in the watershed? In this chapter, I investigate the effects of such local topographical changes like landslides or tectonic motion for watersheds on real and artificial landscapes, as well as for its extension to three dimensions. In fact, I show that the same type of topological perturbation can indeed trigger non-local effects of any length scale, i.e., following power-laws distributions. For illustration, as shown in Fig. 5.1, I obtain after a local height change of less than 2 m at a location (cross) close to the Kashabowie Provincial Park, some kilometers North of the US-Canadian border, a substantial displacement in the watershed (dark (blue) line), which encloses together with
5.2 Method

This approach is based on a perturbation scheme, where a local event is induced by changing the height \( h_k \) at a single site \( k \) of the system to \( h_k + \Delta \), with \( \Delta \) being the perturbation strength. It will be shown later (see section [53].

Fig. 5.1: Example of the watershed between US and Canada, close to the big lakes (light (red) line). Also shown is the resulting change in the watershed (dark (blue) line) due to a perturbation of 2 m at a spot (cross) close to the border, near Thunder Bay. The watershed displacement encloses an area of about 3730 km\(^2\). The dot marks the new outlet of the area after perturbation. The inset shows the same area on a larger scale.

the original watershed (light (red) line) an area \( N_s \sim 3730 \text{ km}^2 \). Here a model is developed to provide a qualitative and quantitative description of this phenomenon.
5.4.4 and Ref. [66]) that, without loss of generality, I can fix the perturbation strength to the height difference between the highest and lowest height of the landscape, which corresponds to a perturbation of infinite strength for the landscape under investigation. With this choice all possible changes are obtained within the DEM. As shown in Fig. 5.2, I quantify this response by the area, i.e. the number of sites, $N_s$, enclosed by the resulting watershed (light (green) line) after the perturbation at site $k$ (cross) and the watershed of the original landscape (dark (red) line). The water can only escape from this area through one single site, which I call outlet. In this scheme, two outlets can be considered, one before (cross) and another after the perturbation (dot). The former always coincides with the perturbed site $k$ and can be connected to the latter, inside the enclosed area, by an invasion percola-
tion cluster, whose mass I denote by $M$. The number of enclosed sites $N_s$, the mass $M$ of the connecting cluster and the distance $R$ between the two outlets (dotted line in Fig. 5.2) are measured. After that, the original landscape is restored by resetting the height at $k$ to its initial value. Except of those sites located on the original watershed, this procedure is repeated for every site $k$ in the landscape. In the following, I consider only perturbations actually leading to a displacement of the watershed, i.e. $N_s > 0$. To reduce finite-size effects, I explicitly exclude those perturbations, where the changed areas touch the borders of the system and, therefore, the original and the perturbed watershed are always overlapping at the boundaries. Additionally, in all definitions hereafter, I consider only those perturbations leading to a displacement of the watershed. From the obtained set of measures $N_s$, $M$, and $R$, I calculate the distribution $P(N_s)$ of the numbers of enclosed sites (areas) $N_s$, the distribution $P(M)$ of the clusters mass, and the probability distribution $P(R)$ of the Euclidean distance $R$ between the two outlets. To investigate the dependence of $N_s$ and $M$ on the distance $R$, I define the average $\langle N_s \rangle$ and distribution $P(N_s|R)$ of areas associated with a distance $R$, as well as the average $\langle M \rangle$ and distribution $P(M|R)$. For the dependence of $M$ and $N_s$ on each other I study the average mass associated with an area $N_s$.

All these distributions and measures were sampled for each configuration, then averaged over 2000 realizations of systems with size $L = 513$ and over 4000 for $L = 129$ and 257.

## 5.3 Perturbing Real Landscapes

First, I study several natural landscapes, from mountainous (e.g. Rocky Mountains) to rather flat landscapes (e.g. US-CAN, Kongo and Germany). The DEM data was obtained from the SRTM-project [182], where for each set I used a size of 2700 km$\times$2700 km (except 1080 km$\times$1080 km for Germany), and a resolution of 540 m, defining the size of a site. Hence, the physical size of the 8 data sets are large enough, so that finite size effects emerging from the DEM boundaries could not be detected. As shown in Fig. 5.3 for Rocky Mountains, Andes and Appalachian (unshifted); Brazil and Europe (shifted by a factor of 100 for better visibility); US-CAN, Kongo and Germany (shifted by a factor of 10000), I find the distribution of areas to follow a power-law, $P(N_s) \sim N_s^{-\beta}$, with $\beta = 1.65 \pm 0.15$ for all landscapes. The probability distribution $P(R)$ of outlet distances $R$ also obeys a power-law.
The distribution $P(N_s)$ is shown for various regions: Rocky Mountains, Andes and Appalachian (unshifted); Brazil and Europe (shifted by a factor of $10^2$ for better visibility); US-CAN, Kongo and Germany (shifted by $10^4$). All data sets have a resolution of 540 m. The solid line shows the fit to the Andes data with a power-law of exponent $-1.65 \pm 0.15$. The inset shows $\langle N_s \rangle$ as function of $R$ for the Rocky Mountains at resolutions of 270 m (squares), 540 m (circles) and 1350 m (triangles). The solid line has slope 2.

$P(R) \sim R^{-\rho}$, with $\rho = 3.1 \pm 0.3 \approx 2\beta$ (see Fig. 5.4), and displays an upper cutoff in the range $50 \text{ km} < R < 500 \text{ km}$ for the studied landscapes. This cutoff is independent on the resolution and could be due to a length scale arising from tectonics. The value of $\rho$ implies $\langle N_s \rangle \sim R^2$, which agrees well with my data (inset of Fig. 5.3). The distribution for a given distance $R$ scales as $P(N_s|R) \sim N_s^{-\alpha}$ with $\alpha = 2.3 \pm 0.2$ (see inset of Fig. 5.4).
5.4. Perturbing Artificial Landscapes

5.4.1 Effects in two Dimensional Systems without Correlations

In order to understand these power-laws and the dependence between $N_s$ and $R$, I study artificial landscapes, where the local heights are generated using fractional Brownian motion (fBm) on a square lattice [195, 86] (see also section 4.5). This model incorporates spatial long-range correlations to the system that are controlled by the Hurst exponent, $H$. I first consider the case of uncorrelated landscapes, which is a special case of the fBm model with $H = -1$. In Fig. 5.5, I present the results obtained for several system

---

Fig. 5.4: Distribution of outlet distances $R$ for various natural landscapes. The solid line is the best fit to the data of the power-law, $P(R) \sim R^{-\rho}$, with $\rho = 3.1 \pm 0.3$. The inset shows the distributions $P(A|R)$ associated with the outlet distance $R = 0.54$ km. The solid line is the best fit to the Andes data of the power-law, $P(A|R) \sim R^{-\alpha}$, with $\alpha = 2.3 \pm 0.2$. 

---

5.4. Perturbing Artificial Landscapes
sizes, using the same procedure as for the natural landscapes. The probability density $P(R)$ again follows a power-law $P(R) \propto R^{-\rho}$, without upper cutoff as in real landscapes. I estimate $\rho = 2.21 \pm 0.01$ using the scaling $P(R) = L^\rho f[RL]$, where $L$ is the linear dimension of the landscape. For the distribution $P(N_s) = L^\beta f[N_sL^2]$, I obtain an excellent data collapse for $\beta = 1.16 \pm 0.03$ (see the inset of Fig. 5.5). In the case of the distribution $P(N_s|R)$ at a fixed outlet distance, I again find a power-law $P(N_s|R) \sim N_s^{-\alpha}$ (see Fig. 5.6). Finite size scaling analysis yields an exponent $\alpha = 2.23 \pm 0.03$ independent on the value of $R$. Assuming that $R$ describes the extension
of \(N_s\) in every direction and considering the similarity of the exponents, the relation \(\rho = \alpha\) seems reasonable, but as I will show later in section ??, this relation does not hold for higher dimensions. From these results, and from the ones obtained for real landscapes, I conclude that \(P(N_s)\), \(P(R)\), and \(P(N_s|R)\), as well as \(\langle N_s \rangle\), follow power laws of the form

\[
P(N_s) \sim N_s^{-\beta},
\]

\[
P(R) \sim R^{-\rho},
\]

\[
\langle N_s \rangle \sim R^\sigma,
\]

\[
P(N_s|R) \sim N_s^{-\alpha},
\]

The area \(N_s\) was rescaled by \(L^2\), indicating that \(\sigma = 2\). This is also supported by the measured value \(\beta = 1.16 \pm 0.03\), which agrees well with the relation \(\beta = \rho/2 \approx 1.11\). Furthermore, the areas are by definition compact. The strong finite-size effects observed for the average area, make the numerical estimation of \(\sigma\) a difficult task, which is addressed in the following.

### 5.4.2 Finite-Size Effects

Hereafter, I present a functional description for the average number of enclosed sites \(\langle N_s \rangle(R, L)\) that captures the dependence on \(R\), including also the finite-size effects. I can write the average for a given distance \(R\) as the first moment of the distribution \(P(N_s|R)\), i.e.,

\[
\langle N_s \rangle(R) = \int dN_s N_s P(N_s|R).
\]  

The distribution obtained from my simulation has a lower and an upper cutoff, as can be seen from Fig. 5.6. Considering that those perturbations with areas touching the boundary are excluded, I need to determine the scaling of the cutoffs numerically. From the data collapse achieved using the scaling \(P(N_s|R) = L^{2n} f[N_sL^2]\), one can see that the upper cutoff indeed scales as \(L^d\). Similarly, the lower cutoff follows \(R^d\) as obtained from the data collapse shown in the inset of Fig. 5.6. Applying both cutoffs as bounds for the integral on the right hand side of Eq. (5.2), as well as for the normalization.
Fig. 5.6: Data collapse for the distribution \( P(N_s|R) \) with \( R = 1 \) (filled) and \( R = 10 \) (open) and system sizes \( L = \{129, 257, 513\} \) (triangles, circles, and squares, respectively), using the scaling \( P(N_s|R) = L^{2\alpha} f[N_sL^2] \). The solid lines represent fits to the data of a power law with exponent \( \alpha = 2.23 \pm 0.03 \).

The inset shows the collapse of the same data for \( L = 513 \), when the x-axis is rescaled by \( R_d \), visualizing the lower cutoff from the power-law behavior. Each data point corresponds to an average over 2000 realizations of linear system size \( L = 513 \) and over 4000 for \( L = 129 \) and 257. The error bars are smaller than the size of the symbols.

integral for \( P(N_s|R) \), I obtain

\[
\langle N_s \rangle (R, L) = \frac{\int_{R^d}^{L^d} dN_s N_s^{1-\alpha}}{\int_{R^d}^{L^d} dN_s N_s^{-\alpha}}
\]

\[
= C \left( \frac{1 - \alpha}{2 - \alpha} \right) \left( \frac{L^{d(2-\alpha)} - R^{d(2-\alpha)}}{L^{d(1-\alpha)} - R^{d(1-\alpha)}} \right).
\]
5.4. PERTURBING ARTIFICIAL LANDSCAPES

Fig. 5.7: The average number of sites (area) $\langle N_s \rangle$ enclosed by the original and the perturbed watershed at an outlet distance $R$ for system sizes $L = \{129, 257, 513\}$ (triangles, circles, and squares, respectively). The lines show the expression given by Eq. (5.3). Each data point corresponds to an average over 2000 realizations of linear system size $L = 513$ and over 4000 for $L = 129$ and 257. The error bars are smaller than the size of the symbols.

This result matches, within the error bars, my simulation data as shown in Fig. 5.7. In the limit $L \to \infty$, only possible because $\alpha > 2$ [66], it reduces to $\langle N_s \rangle(R) = C(1 - \alpha)/(2 - \alpha)R^d$ and, therefore, $\sigma = 2$, as observed numerically.

5.4.3 Effect of Long-Range Correlations

In the following, I show that the exponents can be matched quantitatively by tuning the Hurst exponent $H$ to introduce spatial long-range correlations, as present in real geological systems. The exponents $\alpha$, $\beta$ and $\rho$ were calculated for several values of $H$ (see Fig. 5.8). As shown in Fig. 5.8, I observe that
Fig. 5.8: The exponents $\alpha$ (squares), $\beta$ (circles) and $\rho$ (triangles) are shown for several values of the Hurst exponent $H$. Each point results from a similar study as done for the uncorrelated landscapes. The exponents for the natural landscapes (open symbols), all corresponding to Hurst exponent values in the range $0.3 < H < 0.5$, are consistent with my model. The average angle $\theta$ (in radians) between the outlets from the center of mass is shown too (crosses). The insets depict schematic shapes of the areas and positions of the two outlets for small (left) and large (right) values of $H$.

Both $\beta$ and $\rho$ increase with $H$. Furthermore, the relationship $\beta = \rho/2$ is maintained, since the areas remain compact in the entire range of $H$ values. Around $H = -0.5$, $\alpha$ starts to deviate from $\rho$ and for $H > 0$ I observe $\alpha$ to decrease. Previously, I had assumed $R$ to reflect the extension of the area, i.e., the outlets to reside typically on opposite sides of the area. To check whether this is still valid, I measured the angle $\theta$ between the lines connecting the center of mass of the area with the two outlets. I observe the average angle to decrease as a function of $H$ (see Fig. 5.8). This implies...
that, on average, the two outlets approach each other with increasing $H$ (see also the insets of Fig. 5.8), so that $R$ is no longer representative of the area extension. Therefore, with $H \to 1$, fixing the distance $R$ no longer restricts the areas entering the distribution $P(N_s|R)$, such that the exponent $\alpha$ decreases with $H$ and approaches $\beta$. For all considered landscapes the value $\sigma = 2$ has been obtained. Finally I find good quantitative agreement with the exponents obtained from the natural landscapes, which are known to have a Hurst exponent inside the range $0.3 < H < 0.5$ (see Ref. [118] and references therein). Hence, except for the upper cutoff in $R$, my model provides an excellent quantitative description of the effects observed on natural landscapes.

5.4.4 Influence of the Perturbation Strength

Next I analyze quantitatively the impact of the perturbation strength $\Delta$ on the watershed. In Fig. 5.9 the number of perturbed sites $N$ that change the watershed is shown for uncorrelated, for artificial correlated (with $H = 0.3$) and for natural landscapes (Andes). In all three cases, $N$ is found to increase linearly with the applied perturbation strength, $N \sim \Delta$. This indicates that changes on the watershed can be observed even for infinitesimally small perturbations. Additionally, in both cases where correlations are present, $N$ is observed to reach a plateau. As already stated, when $\Delta$ is equal to $h_w$, this corresponds to the largest relevant perturbation, so that $N(h_w) = N_{\text{max}} \ll L^2$ indicates that many perturbations never change the watershed at all. It is clear that $\Delta > |h_j - h_i|$ is needed, where $h_i$ and $h_j$ are the heights of the outlets of the area. Therefore, if the distribution $p_o(h)$ of outlet heights is known, one obtains,

$$N(\Delta) = \frac{2N_{\text{max}}}{L^2} \int_0^\Delta \int_{\Delta'}^{h_w} p_o(h) p_o(h - \Delta') dh d\Delta'.$$

(5.4)

For landscapes with uniformly distributed heights, I find $p_o(h)$ to be still a uniform distribution. Then I obtain from Eq. (5.4),

$$N(\Delta) = (h_w\Delta - \Delta^2/2)2N_{\text{max}}/(L^2h_w^2),$$

(5.5)

which is in excellent agreement with my data (see Fig. 5.9), where an approximately linear behavior can be observed for $\Delta < h_w$. The observed power-laws are maintained for all values of $\Delta$, as can be clearly seen for $\langle N_s \rangle$ in the inset.
Fig. 5.9: Dependence of the number of perturbed sites \( N \) that promote changes on the watershed on the perturbation strength \( \Delta \) applied for uncorrelated (squares), Andes (triangles) and fBm landscape with \( H = 0.3 \) (circles). The solid line corresponds to the analytic relation obtained from Eq. (5.4) for uncorrelated landscapes. The inset shows the average area \( \langle N_s \rangle \) as a function of the distance \( R \) between the outlets for \( \Delta/h_w = 1, 0.016 \) and 0.00025 (pluses, crosses, and squares, respectively), and \( L = 513 \).

I conclude that infinitesimally small perturbations have qualitatively the same effect on the watershed as any larger perturbation strength \( \Delta \).

### 5.4.5 Relation to Invasion Percolation

In the case of uncorrelated landscapes \( (H = -1) \), for a given area \( N_s \), the corresponding invasion percolation (IP) cluster is obtained by starting the penetration process from one outlet to another, always growing along the lowest site in the perimeter. The area \( N_s \) can therefore be understood as
the envelop of this IP-cluster. The size-distribution $P(M|R)$ of the number of sites (mass) of the IP-clusters between two sites at a fixed distance $R$ is known to follow a power-law $M^{-\alpha^*}$ with $\alpha^* = 1.39$ [196]. Note that, for comparison of my results to Araújo et al. [196], $P(M|R)$ needs to be divided by $M$, as I grow the IP-cluster starting from the outlet at the watershed, which is always the highest of the $M$ sites of the cluster. Hence I expect

$$P(M|R) \sim M^{-(\alpha^*+1)},$$

what is indeed in good agreement with my data (see Fig. 5.10). In a similar way as for the area $N_a$, I find also the distribution $P(M)$ and the average

$$P(M|R) \sim M^{-(\alpha^*+1)},$$

what is indeed in good agreement with my data (see Fig. 5.10).
mass $\langle M \rangle$ to obey power laws of the form

\begin{align}
P(M) &\sim M^{-(1+\beta^*)} & (5.7a) \\
\text{and} \quad \langle M \rangle &\sim R^{\sigma^*}, & (5.7b)
\end{align}

where I choose Eq. (5.7a) to have a similar form as Eq. (5.6). Using the scaling $P(M) = M^{-(1+\beta^*) f[M L^{df^*}]}$, I obtain a clear data collapse for $\beta^* = 0.21 \pm 0.02$ (see Fig. 5.11).

Fig. 5.11: Data collapse of the size distribution $P(M)$ of the mass $M$ of the invasion percolation cluster connecting the two outlets for uncorrelated landscapes of three different system sizes $L = \{129, 257, 513\}$ (triangles, circles, and squares, respectively). The line represents a power-law fit to the data for the largest landscape (squares) yielding an exponent $1 + \beta^* = 1.21 \pm 0.02$. Each data point is an average over 2000 realizations of systems with size $L = 513$ and over 4000 for $L = 129$ and 257. The error bars are smaller than the size of the symbols.

From percolation theory, I expect $\sigma^* = d_f^*$, with $d_f^* = 91/48$ the fractal
dimension of IP clusters in two dimensions [196, 157], but like the average area $\langle N_s \rangle$ the average mass shows strong finite-size effects. Analogously to Eq. (5.2), I can define the average mass $\langle M \rangle$ as the first moment of the distribution $P(M|R)$. With a (finite-size) scaling analysis similar to the one used to get the cutoffs in $P(N_s|R)$, I find the lower and upper cutoff of $P(M|R)$ to scale as $R^{d_f}$ and $L^{d_f}$, respectively, where $d_f$ is the fractal dimension of the invasion percolation clusters. Using Eq. (5.6) together with Eq. (5.2) with $M$ instead of $N_s$ I obtain

$$\langle M \rangle(R, L) = \frac{\int_{R^{d_f}}^{L^{d_f}} dM M^{-\alpha^*}}{\int_{R^{d_f}}^{L^{d_f}} dM M^{-\alpha^*-1}}$$

$$= C \left( \frac{\alpha^*}{\alpha^* - 1} \right) \left( \frac{L^{d_f(1-\alpha^*)} - R^{d_f(1-\alpha^*)}}{L^{-d_f\alpha^*} - R^{-d_f\alpha^*}} \right).$$

In Fig. 5.12 the matching, within the error bars, of this result with the simulation data is shown. For $L \to \infty$ the result reduces to $\langle M \rangle(R) = \alpha^*/(\alpha^* - 1)R^{d_f}$, as $\alpha^* > 1$, yielding $\sigma^* = d_f = 91/48$, as expected. The relation of the area and the mass is analyzed with the average mass $\langle M \rangle$ associated to a given number of enclosed sites $N_s$. From the dependence of $\langle N_s \rangle$ and $\langle M \rangle$ on $R$ I expect

$$\langle M \rangle \sim N_s^{\frac{d_f}{\alpha^*}},$$

which matches the value $0.95 \pm 0.01$ obtained from the power-law fit to the data plotted in the inset of Fig. 5.12. Based on this result, using Eq. (5.9) together with Eq. (5.1a) and (5.7a) I expect $\beta^* = \sigma \beta/d_f^* - 1 \approx 0.22$, which matches my result from the data collapse of $P(M)$ shown in Fig. 5.11.

Using now the Eqs. (5.1d), (5.6), and (5.9), I can relate my exponent $\alpha$ of the distribution of areas at fixed distance to $\alpha^*$ by

$$P(N_s|R) = P(\langle N_s \rangle(M)|R) \propto \langle N_s \rangle^{-\alpha} (M) \propto M^{2\alpha/d_f^*} \propto P(M|R).$$

I obtain $\alpha = \frac{d_f^*}{2}(\alpha^* + 1) \approx 2.266$, which is very close to what is measured ($\alpha = 2.23 \pm 0.03$). Therefore, I can relate my results on uncorrelated landscapes to the sub-critical point-to-point invasion percolation process [196] and to the mass distribution of avalanches that occur during the IP-cluster growth [197, 198, 199].
Fig. 5.12: The average mass $\langle M \rangle$ of the invasion percolation cluster connecting the two outlets with a distance $R$ for uncorrelated two dimensional landscapes of sizes $L = \{129, 257, 513\}$ (triangles, circles, and squares, respectively). The lines show the predictions according to Eq. (5.8) for different system sizes. The inset shows the average mass varying with the number of enclosed sites. The line is a power-law fit to the data yielding an exponent $0.95 \pm 0.01 \approx d_f/2$. Each data point corresponds to an average over 2000 realizations of systems with size $L = 513$ and over 4000 for $L = 129$ and 257. The error bars are smaller than the size of the symbols.

### 5.4.6 Scaling Relations

Similar to $\alpha$ and $\alpha^*$ I can relate $\beta$ and $\beta^*$ by replacing in Eq. (5.10) $P(N_s|R)$ and $P(M|R)$ with $P(N_s)$ and $P(M)$, respectively. These two exponent relations together with the relation between $\beta$, $\rho$, and $\sigma$ discussed in sections
5.4. PERTURBING ARTIFICIAL LANDSCAPES

5.3, 5.4.1 and 5.4.3, I conjecture the following three scaling relations, namely

\[ \alpha = \frac{d_f^*}{\sigma} (1 + \alpha^*) , \quad (5.11a) \]
\[ \beta = \frac{d_f^*}{\sigma} (1 + \beta^*) , \quad (5.11b) \]
\[ \text{and} \quad \rho = \sigma \beta . \quad (5.11c) \]

These relations appeared to be valid for artificial landscapes of any degree of correlation, as well as for natural landscapes. In the next section, I check them for an extension of watersheds to three dimensions.

Fig. 5.13: The largest of all observed changed volumes (dark, blue) attached to the perturbed watershed (light, red) for the same uncorrelated three-dimensional system of linear size \( L = 129 \), as used in Fig. 4.5. The upper and lower catchments drain to the top and bottom borders, respectively.
5.5 Impact of Perturbations in three Dimensions

5.5.1 Effects for Watersheds of Systems without Correlations

In this section I extend the previous concepts to three dimensional systems. As introduced in chapter 4, the watershed in three dimensions is a surface that divides the system into two parts. Similar to two dimensions, a perturbation is induced by changing the local resistance $r_k$ at site $k$ to $r_k + \Delta$, where $\Delta$ is the perturbation strength. I quantify the impact on the watershed surface, as before, by the number of sites $N_s$ enclosed by the original and
5.5. IMPACT OF PERTURBATIONS IN THREE DIMENSIONS

the perturbed watershed, which now corresponds to a change in volume. As an example, the largest of all changed volumes observed in my simulations is shown in Fig. 5.13 (attached to the perturbed watershed) and in Fig. 5.14 (without a watershed, for better visibility). The outlets of this volume, before and after perturbation, are determined as well as their distance $R$ and mass $M$ of the invasion percolation cluster connecting them. The original system is then restored by resetting $r_k$ to its initial value. This procedure is repeated for all sites $k$ of the system, except those located on the watershed. Again, the averages $\langle N_s \rangle$, $\langle M \rangle$, and distributions $P(R)$, $P(N_s)$, $P(N_s|R)$, $P(M)$, and $P(M|R)$ are sampled and averaged over 2000 configurations. Similar to the two dimensional case, as shown in Fig. 5.15 for the distribution $P(N_s)$ of volumes, I find these quantities to follow power laws of the form introduced in Eqs. (5.1a)-(5.1d), Eq. (5.6) and Eqs. (5.7a)-(5.7b). The values of all the exponents estimated for uncorrelated systems are summarized in Tab. 5.1.

Table 5.1: Summary of the exponents numerically obtained in this study. Estimates are given for watersheds on natural and uncorrelated artificial landscapes in two dimensions, as well as for uncorrelated artificial systems in three dimensions. For exponents obtained in previous works the corresponding citations are given. The range $1.10 - 1.15$ of fractal dimension for natural landscapes is obtained from Tab. 4.1.

<table>
<thead>
<tr>
<th>Type</th>
<th>Eq.</th>
<th>natural</th>
<th>artificial (uncorr.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>(4.2)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$d_f$</td>
<td>(4.2)</td>
<td>1.10 - 1.15</td>
<td>1.211±0.001 [65]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>(5.1d)</td>
<td>2.3±0.2 [66]</td>
<td>2.23±0.03 [66]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>(5.1a)</td>
<td>1.65±0.15 [66]</td>
<td>1.16±0.03 [66]</td>
</tr>
<tr>
<td>$\rho$</td>
<td>(5.1b)</td>
<td>3.1±0.3 [66]</td>
<td>2.21±0.01 [66]</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>(5.1c)</td>
<td>2 [66]</td>
<td>2 [66]</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>(5.6)</td>
<td>–</td>
<td>1.39±0.03 [196, 66]</td>
</tr>
<tr>
<td>$\beta^*$</td>
<td>(5.7a)</td>
<td>–</td>
<td>0.21±0.02 [67]</td>
</tr>
<tr>
<td>$\sigma^*$</td>
<td>(5.7b)</td>
<td>–</td>
<td>91/48 [157]</td>
</tr>
</tbody>
</table>

Although the volumes $N_s$ are still compact, I find $\sigma = 2.48 \pm 0.02$ by an analysis of $\langle N_s \rangle$ and $P(N_s|R)$ similar to the one performed in two dimensions using Eq. (5.3) in the limit $L \to \infty$. The obtained value of $\sigma$ is close to the fractal dimension of the watershed, what suggests that the mass of the changed volume is dominated by its surface, in contrast to what I observe
in two dimensions. This is also confirmed by a box counting analysis of the largest of all changed volumes, as depicted in Fig. 5.16.

### 5.5.2 Relation to Invasion Percolation

In agreement with the findings by Lee in Ref. [200], that the size distribution of sub-critical point-to-point invasion percolation, as introduced by Araújo et al. [196], is independent on the dimensionality of the system, I obtain $\alpha^* = 1.4 \pm 0.1$ for uncorrelated systems, similar to the value found in two dimensions. This confirms the relation to invasion percolation drawn in Ref. [66]. Inserting the estimates of $a^*_f$, $\sigma$, and $\alpha^*$, given in Tab. 5.1, into
5.5. IMPACT OF PERTURBATIONS IN THREE DIMENSIONS

Fig. 5.16: Number of cubic boxes $C$ of size $\varepsilon$ covering the changed volume, shown in Fig. 5.13 and 5.14, as obtained from a box counting method. Each data point consists of a single measurement. The line corresponds to $\varepsilon^{2.48}$.

Eq. (5.11a), I find $\alpha \approx 2.48$ matching with the value $2.4 \pm 0.1$ obtained numerically, hence validating the conjectured relation. Similarly, I also observe that my results are consistent with Eq. (5.11b).

5.5.3 Effect of Long-Range Correlations

Again, as in 2D, I investigate the effect of long-range spatial correlations on the exponents by applying my perturbation scheme to systems generated with fractional Brownian motion. Similar measurements as performed in Fig. 5.15 to obtain $\beta$ for uncorrelated systems, are repeated for several different values of the Hurst exponent $H$, to estimate the dependence of the exponents $\alpha$, $\beta$, $\sigma$, and $\rho$ on $H$ as shown in Fig. 5.17. I find $\beta$ and $\rho$ to increase with increasing correlation degree, while $\sigma$ stays constant. This confirms the relation $\rho = \sigma \beta$ (Eq. (5.11c)), independent on the degree of correlations. I observe $\alpha$ to
decrease for $H > 0$ and to approach $\beta$, similar as found in two dimensions (compare Section 5.4.3). I assume that, again like in two dimensions, on average, the two outlets approach each other with increasing $H$, so that $R$ is no longer representative of the volume extension. Therefore, with $H \to 1$, fixing the distance $R$ no longer restricts the volumes entering the distribution $P(N_s|R)$, such that the exponent $\alpha$ decreases with $H$ and approaches $\beta$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{exponents.png}
\caption{The exponents $\alpha, \beta, \rho, \sigma$ (squares, circles, triangles up, and triangles down, respectively) as a function of the Hurst exponent $H$ for perturbations in 3D. Each data point consists of a similar measurement as performed in Fig. 5.15 to obtain $\beta$ for uncorrelated systems.}
\end{figure}

### 5.6 Summary

In summary, I was able to show that small and localized perturbations can have a large impact on the shape of watersheds even at very long distances, hence having a non-local effect. The distribution of changes $P(N_s)$ is found
to decrease as a power-law with exponent $\beta = 1.65 \pm 0.15$ on all studied real landscapes from mountainous (e.g. Rocky Mountains) to rather flat (e.g. US-Canadian border). By applying perturbations to model landscapes with long-range correlations, I determined the dependence of the scaling exponents on the Hurst exponent, finding good quantitative agreement with real landscapes, for which $0.3 < H < 0.5$. The obtained exponents $\alpha$, $\beta$ and $\rho$ are independent of the perturbation strength $\Delta$. I found the lower and upper bounds of the distribution of areas to scale with distance and system size with an exponent $\sigma = 2$ in two and $\sigma = 2.48 \pm 0.02$ in three dimensions. Considering these cutoffs, I was able to derive a function describing the average area, which includes the finite-size effects. In a similar way, by integrating $P(M|R)$, I obtained a function for the average mass $\langle M \rangle$ of the invasion percolation cluster. In the extension of the perturbation study to three dimensional systems, I observed the changed volumes to be dominated by their surface, i.e., the watershed, resulting in a value of $\sigma$ close to the fractal dimension of the watershed itself. The distributions and averages for the three dimensional case are shown to follow power laws as in two dimensions. Indeed, I found in the uncorrelated case an intrinsic relation to invasion percolation. Finally, my results in three dimensions are consistent with the conjectured relations between the exponents.
Chapter 6

Corrections to Scaling
I study the corrections to scaling for the mass of the watershed, the bridge line, and the optimal path crack in two and three dimensions. I disclose that these models have numerically equivalent fractal dimensions and leading corrections-to-scaling exponents. I conjecture all three models to possess the same fractal dimension, namely, \( d_f = 1.2168 \pm 0.0005 \) in 2D and \( d_f = 2.487 \pm 0.003 \) in 3D, and the same exponent of the leading correction, \( \Omega = 0.9 \pm 0.1 \) and \( \Omega = 1.0 \pm 0.1 \), respectively. The close relations between watersheds, optimal path cracks in the strong disorder limit, and bridge lines are further supported by heuristic arguments.

### 6.1 Introduction

A fractal dimension, consistent with the one of watersheds (discussed in chapter 4), has also been found for optimal path cracks in the limit of strong disorder. Optimal path cracking has been introduced by Andrade et al. [48, 49, 50] as a model for the evolution of successive optimal paths under constant failure. It describes, e.g., the breakdown of electrical or fluid flow through random media and has important consequences in other fields of science and technology, such as human transportation, fracture mechanics, or polymers in random environments, where finding the optimal path is a challenge [51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61]. More recently, Araújo et al. [62] developed a novel percolation scheme called bridge percolation, which systematically suppresses the creation of a spanning cluster. They found that the set of bridge bonds yields a fractal dimension in the same range as the watershed and the optimal path crack in strong disorder (see also Fig. 6.1). The watershed (WS), the optimal path crack in strong disorder (OPC), and the bridge line (BL) in bridge percolation (BP) correspond to a set of sites or bonds splitting the system into two distinct parts and seem closely related to classical percolation. Despite these similarities and the broad relevance of the models, no detailed studies of the relation between them are available. Due to the ubiquity of the fractal dimension, also relations to other physical models have been proposed, such as optimal paths [49, 179, 180, 166, 201], the shortest path in loopless invasion percolation [180], the infinite cluster in multiple invasion percolation [202], and the surface of the infinite cluster in discontinuous percolation [63, 64]. Despite the similarity in two dimensions, optimal paths [179, 180, 166, 201] differ from watersheds, optimal path cracks, and bridge percolation in higher dimensions, as has been shown by Araújo.
et al. [62].

Fig. 6.1: Mass $M$ of the watershed (WS site/bond), the main crack (MC), and the bridge line (BL) as function of the system size $N$, defined as the number of sites (bonds) in the system, in both two and three dimensions. The errorbars are smaller than the symbols. The lines show the fractal dimensions obtained in this work.

In this chapter, I explore the relation between the watershed [65, 66, 67], the main crack (MC) of the optimal path crack in strong disorder [48, 49, 50], and the bridge line of bridge percolation [62] by investigating numerically the fractal dimension and the corresponding corrections to scaling for two and three dimensional uncorrelated systems. Due to the numerical difficulty in obtaining sufficient statistics, I omit a discussion of the surface of the infinite cluster in discontinuous percolation [63, 64]. The fractal dimension $d_f$ is defined through the scaling of the mass $M$, corresponding to the number of sites or bonds in the object, with the linear system size $L$,

$$M \sim L^{d_f}.$$  \hspace{1cm} (6.1)
Table 6.1: Number of samples used to obtain the average mass of the bridge line (BL), the watershed (WS sites/bonds), and the main crack (MC) for different system sizes $L$ in two and three dimensional systems.

<table>
<thead>
<tr>
<th>$L$</th>
<th>BL</th>
<th>WS site</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2D</td>
<td>3D</td>
<td>2D</td>
</tr>
<tr>
<td>4</td>
<td>$1.01 \times 10^{11}$</td>
<td>$6.87 \times 10^{10}$</td>
<td>$2.51 \times 10^8$</td>
</tr>
<tr>
<td>8</td>
<td>$1.20 \times 10^{10}$</td>
<td>$8.59 \times 10^9$</td>
<td>$6.36 \times 10^7$</td>
</tr>
<tr>
<td>16</td>
<td>$1.13 \times 10^9$</td>
<td>$1.07 \times 10^9$</td>
<td>$4.29 \times 10^7$</td>
</tr>
<tr>
<td>32</td>
<td>$3.09 \times 10^9$</td>
<td>$1.34 \times 10^8$</td>
<td>$3.69 \times 10^7$</td>
</tr>
<tr>
<td>64</td>
<td>$9.74 \times 10^8$</td>
<td>$1.67 \times 10^7$</td>
<td>$3.26 \times 10^7$</td>
</tr>
<tr>
<td>128</td>
<td>$1.01 \times 10^9$</td>
<td>$2.09 \times 10^6$</td>
<td>$1.81 \times 10^7$</td>
</tr>
<tr>
<td>256</td>
<td>$8.57 \times 10^8$</td>
<td>$2.62 \times 10^5$</td>
<td>$6.54 \times 10^6$</td>
</tr>
<tr>
<td>512</td>
<td>$2.12 \times 10^8$</td>
<td>$1.31 \times 10^5$</td>
<td>$1.09 \times 10^6$</td>
</tr>
<tr>
<td>1024</td>
<td>$2.68 \times 10^7$</td>
<td>4608</td>
<td>$2.04 \times 10^5$</td>
</tr>
<tr>
<td>2048</td>
<td>$1.95 \times 10^6$</td>
<td>–</td>
<td>48564</td>
</tr>
<tr>
<td>4096</td>
<td>$5.15 \times 10^5$</td>
<td>–</td>
<td>13607</td>
</tr>
<tr>
<td>8192</td>
<td>$1.28 \times 10^5$</td>
<td>–</td>
<td>3450</td>
</tr>
<tr>
<td>16384</td>
<td>56847</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>32768</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$L$</th>
<th>WS bond</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2D</td>
<td>3D</td>
</tr>
<tr>
<td>4</td>
<td>$1.61 \times 10^{10}$</td>
<td>$1.61 \times 10^{10}$</td>
</tr>
<tr>
<td>8</td>
<td>$8.05 \times 10^9$</td>
<td>$2.01 \times 10^9$</td>
</tr>
<tr>
<td>16</td>
<td>$2.01 \times 10^9$</td>
<td>$2.51 \times 10^8$</td>
</tr>
<tr>
<td>32</td>
<td>$5.03 \times 10^8$</td>
<td>$3.14 \times 10^7$</td>
</tr>
<tr>
<td>64</td>
<td>$1.25 \times 10^8$</td>
<td>$3.93 \times 10^6$</td>
</tr>
<tr>
<td>128</td>
<td>$1.03 \times 10^9$</td>
<td>$4.91 \times 10^5$</td>
</tr>
<tr>
<td>256</td>
<td>$2.59 \times 10^8$</td>
<td>$1.22 \times 10^5$</td>
</tr>
<tr>
<td>512</td>
<td>$6.48 \times 10^7$</td>
<td>$1.22 \times 10^5$</td>
</tr>
<tr>
<td>1024</td>
<td>$4.24 \times 10^7$</td>
<td>65536</td>
</tr>
<tr>
<td>2048</td>
<td>$1.03 \times 10^7$</td>
<td>–</td>
</tr>
<tr>
<td>4096</td>
<td>$2.70 \times 10^6$</td>
<td>–</td>
</tr>
<tr>
<td>8192</td>
<td>$7.14 \times 10^5$</td>
<td>–</td>
</tr>
<tr>
<td>16384</td>
<td>$1.76 \times 10^5$</td>
<td>–</td>
</tr>
<tr>
<td>32768</td>
<td>33248</td>
<td>–</td>
</tr>
</tbody>
</table>
Due to the finite system size, corrections to scaling arise \cite{203,204} masking the true asymptotic behavior. Hence, the estimated $d_f$ can be improved by describing the size dependence of the mass as

\begin{equation}
M_L = L^{d_f} C_L \, ,
\end{equation}

where the general form for the corrections to scaling $C_L$ is

\begin{equation}
C_L = a_{00} + a_{01} L^{-1} + a_{02} L^{-2} + a_{03} L^{-3} + \ldots + a_{11} L^{-\Omega_1} + a_{12} L^{-\Omega_1-1} + a_{13} L^{-\Omega_1-2} + \ldots + a_{21} L^{-\Omega_2} + a_{22} L^{-\Omega_2-1} + a_{23} L^{-\Omega_2-2} + \ldots + a_{n1} L^{-\Omega_n} + \ldots ,
\end{equation}

with non-universal coefficients ($a_{ij}$). The exponents fulfill $\Omega_1 < \Omega_2 < \ldots < \Omega_n$ and are non-analytic (non-integer). They are usually independent on the geometry of the lattice and only depend on the dimensionality \cite{203,204}. Finding the same non-analytic corrections-to-scaling exponents for all three models will give another hint for the close relation between them. But, in general, the precise estimation of corrections to scaling is a difficult task.

Numerical studies typically measure the leading correction exponent, a sub-leading correction exponent, or an effective exponent arising from the sum of two or more correction-to-scaling terms \cite{205}. Hence, a reliable estimate of the leading correction exponent depends on both the method and the precision of the data. I first have a look at the functional form of the corrections to scaling that can be considered for the individual models given the available statistics, using a simple fitting and checking which amplitudes in Eq. (6.3) are small. Using this and truncating terms with an order (exponent) equal or larger than 3, I define my effective corrections-to-scaling ansatz. By defining a fit quality, I identify the leading correction exponent (highest maximum of the quality) and obtain a highly accurate estimate for the fractal dimension $d_f$. I cross check the obtained results with a careful analysis of the local logarithmic slopes as suggested by Ziff \cite{206,207}. This method uses the fact that for large enough system sizes the higher order terms are negligible, such that the local logarithmic slope of the corrections to scaling should converge to the leading correction exponent.

The chapter is organized as follows. In Sec. 6.2 I describe the models. Section 6.3 introduces the corrections to scaling and summarizes the obtained results. The relations between the models are discussed in Sec. 6.4 and conclusions are drawn in Sec. 6.5.
6.2 Models

In the following, I give a brief overview of the watershed (WS), optimal path cracking (OPC), and bridge percolation (BP), focusing on the role of percolation in the numeric procedures used to determine the watershed, the main crack (MC), and the bridge line (BL). For simplicity, the description is given for two-dimensional systems (square lattices), where they lead to lines. The extension of the discussed models to higher dimensions, where they lead to (hyper)surfaces, is straightforward and has been done in Refs. [62, 50, 67].

6.2.1 Watershed

Watersheds are the lines separating adjacent drainage basins and play a fundamental role in many fields [1, 2, 3, 4, 5, 7, 6, 8, 9]. Recently, a novel and efficient algorithm has been proposed to determine watersheds on digital elevation maps (DEM) [65], which has been used afterwards to investigate numerically several properties of these objects [65, 66, 67]. Up to now, I defined the watershed as a set of bonds. It is also possible to define the watershed solely on sites, thereby using a very similar numerical technique to determine it as presented in chapter 3. To focus on the differences between these two options, I give in the following an alternative definition to the one given at the end of chapter 2.

I consider uncorrelated artificial landscapes mapped on a square lattice as a digital elevation map, where each site $i$ represents an area and the bonds are the connections to neighboring areas from and to where water can flood. The lattices are of linear size $L$ with free boundary conditions in the vertical direction and periodic boundary conditions in the horizontal one. The height $h_i$ at each site $i$ is drawn randomly from a uniform distribution in such a way that $h_i > 0$ and $h_j \neq h_i, \forall j \neq i$. A site with more than one nearest neighbor lower in height than the site itself is called diversion site. I define sets of sites $S_k, k = 1, 2, \ldots, N_{\text{sinks}}$, to be the sinks of the system. On natural landscapes these sinks are the natural water-outlets of the terrain, such as lakes, rivers, or oceans. For simplicity, I consider again only two sinks, namely, the row of sites above the top row of the lattice and the one below the bottom row, i.e., two new rows are defined, with $h = 0$, to describe the sinks.

I use an invasion-percolation-like scheme to associate each site to a sink, where the two sets of sites, associated to one of the two sinks, are separated by the watershed. Algorithmically, the association of a given site to a sink
of the system proceeds via the growth of a runoff cluster seeded in the site under consideration. Eventually, the runoff cluster connects each site to at least one sink. At each time step, this cluster grows only to the lowest-lying site on its perimeter defined as the set of unoccupied sites with, at least, one nearest neighbor occupied by the cluster. Typically, this leads to local growth from any occupied site to its nearest neighbor with lower height, in analogy to water flowing downhill and the cluster corresponds, e.g., to a flooded region. In the case of a diversion site, i.e., a site with more than one nearest neighbor lower in height, this can lead to cluster growth in more than one direction. I can define the watershed according to two different cluster-growth or run-off schemes, a strict one excluding diversion, what leads to a definition on bonds (WS bond), and a diverting one, which leads to a definition on sites (WS site). In the following, I give a detailed description of both.

The diverting run-off scheme includes all lower-lying nearest neighbors of a diversion site, but follows them separately. This scheme, therefore, consists of a hierarchical tree of runoff clusters, where each lower-lying neighbor \( n_j \) of a diversion site \( j \) is the starting point of a new cluster, i.e. a new branch, and I call \( j \) the parent of such a branch. A branch ends either if it reaches a sink, if it grows up to a height equal to its parent diversion site \( j \), or if it grows to a new diversion site. In the case of all branches with parent \( j \) stopping due to growing again to the parents height, \( j \) is no longer considered a diversion site and all these branches are joined with the parent branch of \( j \) to form a single branch which is then grown further until one of the stopping conditions is fulfilled again. This procedure is repeated until all branches stopped growing. If from a branch the sink \( S_k \) is reached before any other, then the entire branch is considered to belong to the catchment basin of sink \( S_k \), equivalently one says that this branch drains to \( S_k \). For any diversion site \( j \) there are now two possibilities, either all branches developing from it drain to the same sink \( S_k \), i.e. \( j \) and its parent branch drain to \( S_k \), or its branches drain to different sinks. In the latter case, the diversion site \( j \) and all its parent branches as well as their parent diversion sites cannot be attributed to one of the catchments solely. In fact, they need to be considered part of both catchments at the same time or neither of them. This procedure introduces a classification of lattice sites, i.e. a subdivision into three non-overlapping subsets whose union is the entire lattice. Two are the sets of sites which uniquely belong to one catchment only and the third is the one in which all sites are considered to be part of both basins. The latter subset of sites separates the two other subsets and I name it the watershed on sites.
(WS site).

With the strict run-off scheme one considers only the lowest-lying (unoccupied) nearest neighbor \( n_j \) of a diversion site \( j \), starting only one new runoff cluster, i.e. one new (active) branch with parent \( j \). This active branch stops growing if a new diversion site is reached, if a sink is reached, or if the branch grows up to the height of its parent \( j \). In the latter case, the abandoned branch is joined with the parent branch of its parent site \( j \) to form together a single branch. If there are any other unoccupied lower nearest neighbors of the parent \( j \), then a new branch is started from the lowest of them. If there are no more lower unoccupied nearest neighbors of \( j \), then the parent branch of \( j \) becomes active again and \( j \) is no longer considered a diversion site. This reduces the tree to a simple chain of branches. The procedure stops if a sink is reached. If when starting from \( i \), sink \( S_k \) is reached by the active branch before any other, then site \( i \) and the entire chain of branches are considered to belong to the catchment basin of sink \( S_k \), equivalently one says that \( i \) drains to \( S_k \). This procedure introduces a unique classification of lattice sites, i.e. a subdivision into two non-overlapping subsets whose union is the entire lattice. To the set of bonds separating the two basins I call watershed on bonds (WS bond).

The mass \( M \) of the watershed (WS) is defined as the number of bonds (sites) forming the watershed. Although, the height values could be interpreted as a third dimension and, hence, the watershed as a line embedded in 3D space, studying the scaling of the mass with system size, corresponds to the analysis of the fractal properties of the projection of the watershed to the two-dimensional space.

### 6.2.2 Optimal Path Crack

The optimal path crack (OPC) was introduced by Andrade et al. [48, 49, 50] and it is obtained in the following way. I start with a square lattice of size \( L \) using free boundary conditions in the vertical direction and periodic boundary conditions in the horizontal one. A random energy is assigned to each site and the energy of any path in the system is defined as the sum of the energy of its sites. In particular, the optimal path is the one among all paths connecting the top and bottom boundary of the system with the lowest total energy. Once the first optimal path is determined, the site in the optimal path having the highest energy is identified and removed. This is equivalent to impose an infinite energy to this site. Next, the optimal path
6.2. MODELS

is calculated among the remaining accessible sites of the lattice, from which the highest energy site is again removed. The process continues iteratively until the system is disrupted and no further path can be found. The set of removed sites then defines the optimal path crack (OPC). The OPC is dependent on the type of disorder, but in the limit of strong disorder, it is localized in a single line, denoted as the main crack (MC), with mass $M$ given by the number of cracked sites. From this point on, I consider the OPC only in the limit of strong disorder and, for simplicity, just refer to it as main crack (MC).

6.2.3 Bridge Percolation

Bridge percolation is a new percolation model introduced by Araújo et al. [62] in which the creation of a spanning cluster, i.e. the connectivity, is suppressed by giving the bridge bonds a different weight. In the random occupation of bonds, the bridge bonds are those, which, when occupied, create a spanning cluster, i.e. a cluster connecting top and bottom edge of the system. In the following, I focus solely on the case where bridges are never occupied, i.e. having vanishing probability of being occupied. While the original studies were done for bond percolation, I consider here site percolation. Similarly, as in the bond case, I start with an empty square lattice of size $L$, choose sites uniformly at random and occupy them. If two neighboring sites are occupied, they are considered as connected and belong to the same cluster. In contrast to standard site percolation, every time the occupation of a site would lead to a spanning cluster, this bridge site is blocked. The process proceeds until all sites are occupied or blocked and the system is disrupted into two parts. The separating line is formed by the set of bridge sites and its mass $M$ is given by the number of sites in the line. Cieplak, Maritan, and Banavar [179] have studied this line in a different context and argued that the occupation procedure is equivalent to randomly assign an energy to each site, rank order them by increasing energy, and occupy them according to their rank (the bridges are defined similarly). In fact this is an alternative method to determine watersheds as argued by Fehr et al. [65] presenting it as the flooding method (see chapter 3). In the following I will call this object the bridge line (BL).
6.3 Corrections to Scaling

I perform extensive numerical simulations of the described models measuring
the mass $M$ of the watershed (WS), the bridge line (BL), and the main crack
(MC) for different (linear) system sizes $L$. For details about the considered
system sizes and the corresponding number of samples, see Tab. 6.1. The
obtained masses are shown in Fig. 6.1 as a function of the system size $N$,
namely $N = L^d$ for sites and $N = d L^d - (2d - 1)L^{d-1}$ for bonds (the second
term arises due to the solid walls in the vertical direction), where $d$ is the
dimensionality of the system. Although this is not visible in Fig. 6.1, the
masses of the BL and MC are equal within the error bars. Those of WS
site and WS bond are different from the masses of BL and MC. Neverthe-
less, I observe all of them to follow very similar scaling behaviors. The true
asymptotic behavior for the mass scaling, Eq. (6.1), is masked by corrections-
to-scaling arising due to finite system size [203, 204]. Hence, the estimate
of the fractal dimension $d_f$ can be improved by considering these corrections
explicitly, see Eq. (6.3). In the following, I first analyze the general ansatz
to find the number of distinguishable correction exponents and if there are
vanishingly small amplitudes. This results in simplified functional descrip-
tions of the corrections to scaling in 2D and 3D, which are then studied by
two different techniques in order to obtain highly accurate estimates of the
exponents.

6.3.1 Ansatz for Corrections to Scaling

To understand the structure of the data, I study least-square fits of different
truncated versions of Eq. (6.3) to the corrections to scaling $C_L = M/L^{d_f}$ in
2D, where $d_f \approx 1.217$ has been chosen such that $C_L$ converges to a constant
value for large $L$ (see Fig. 6.2). This choice of $d_f$ is consistent with the more
precise estimates obtained later. Using different numbers of exponents $\Omega_n$
and varying numbers of expansions, I find for all models that, with the current
precision, I cannot resolve correction terms of an order higher than $1/L^2$. In
the following, I therefore truncate the expansions by setting $a_{ij} = 0 \forall j > 2$. For
the case of WS site, I obtain reasonable fits down to fairly small $L$ using a set
of two exponents ($n = 2$), yielding $\Omega_1 \approx 0.6$ and $\Omega_2 \approx 0.9$, while $a_{12}$ seems
to be small and also the amplitudes of the analytic terms seem to be small
and unresolvable ($a_{01} \approx 0, a_{02} \approx 0$). It is important to note that, despite
these findings, $\Omega_2$ is still compatible with unity. For WS bond, MC, and BL
Fig. 6.2: Corrections to scaling $C_L = M/N^{d_f/d}$ of the watershed (WS site/bond), the main crack (MC), and the bridge line (BL) as a function of the system size $N$, defined as the number of sites (bonds) in the system, in two dimensions. The fractal dimension $d_f \approx 1.217$, consistent with the more precise estimate obtained later, has been chosen such that $C_L$ converges to a constant value for large $N$. The error bars are typically much smaller than the symbols. The lines show fits of truncated versions of Eq. (6.3) to the data, which is divided here by $a_{00}$ to show the matching of the scaling behavior of the different models for large $N$. 
I obtain similar results, although $a_{11}$ seems to be very small in all the three models. As shown in Fig. 6.2, my fits match $C_L$ fairly well for the models. Hence, defining $\omega \equiv \Omega_1$ and $\Omega \equiv \Omega_2$ the (visible) corrections reduce to

$$C^{2D}_L = a_{00} + a_{11}L^{-\omega} + a_{21}L^{-\Omega} + a_{22}L^{-\Omega - 1}, \quad (6.4)$$

with $\omega \approx 0.6$ and $\Omega \approx 0.9$, while $a_{11}$ is large only for the WS site model. The latter fact will be discussed in section 7.

In 3D, I find by a similar study, that the corrections to scaling of all four models can reasonably well be described by a single correction term such that I can write

$$C^{3D}_L = a_{00} + a_{11}L^{-\Omega}, \quad (6.5)$$

with $\Omega \approx 0.9$, but compatible with unity. A simple least-squares fit of the ansatz given by Eq. (6.4) (Eq. (6.5) in 3D), to the data to obtain the coefficients, $d_f$, $\Omega$, and/or $\omega$ directly can be ambiguous. Dependent on the choice of the initial values for the fit parameters (the coefficients and exponents), a fit could even lead to an estimate of $\Omega$ or $\omega$ reflecting higher order corrections instead of the leading ones. To overcome this and improve the accuracy, I discuss, in the following, a method that explores the parameter space by varying the exponents in a given range and analyzing the quality of the corresponding fits. The results from this method are then cross checked with a second method, which allows to estimate the leading correction from the convergence of the local logarithmic slopes in the reduced mass $M L^{-d_f}$.

### 6.3.2 Fit Quality Method

The output of a fit of the ansatz in Eq. (6.4) or in Eq. (6.5) to the data of the reduced mass $M L^{-d_f}$ can be sensitive to the initial conditions. I, therefore, perform a more systematic study as follows. To have a good control over the actual fitting, I use Eq. (6.2), (6.4), and (6.5) in the following form,

$$C_L(\alpha) = M L^{-\alpha}, \quad (6.6a)$$

$$C^{2D}_L(\alpha) = a_{00} + a_{11}L^{-\omega'} + a_{21}L^{-\Omega'} + a_{22}L^{-\Omega' - 1}, \quad (6.6b)$$

$$C^{3D}_L(\alpha) = a_{00} + a_{11}L^{-\Omega'}, \quad (6.6c)$$

in 2D and 3D respectively, with fixed values of $\alpha$, $\omega'$, and $\Omega'$ and estimate the Quality $Q = n/\chi^2$, where $n$ is the number of degrees of freedom of the
Fig. 6.3: (a) Inverse of the quality $Q$ as a function of $\Omega'$ for different values of $\alpha$, as obtained from fits of the ansatz, Eq. (6.6), to the reduced mass $M L^{-\alpha}$ for the watershed on bonds in 2D with sizes as indicated in Tab. 6.1. The vertical lines give the position of the global minimum in $1/Q$ (solid) and the estimated error (dashed). (b) For the same system as in (a), the minimum value $1/Q_{\text{loc}}$ as a function of $\alpha$ is shown, where $Q_{\text{loc}}$ is obtained from curves $1/Q(\Omega')$ for a given $\alpha$, like those shown in (a). The vertical lines highlight the value of $\alpha$ at the global minimum $1/Q_{\text{max}}$ (solid) and the estimate for the error (dashed). The error bars are determined from the width of the minima. The vertical lines show the estimate $d_f = 1.2168 \pm 0.0005$ for the fractal dimension of the watershed on bonds and the horizontal ones the corresponding leading correction $\Omega = 0.95 \pm 0.05$. These exponents were obtained from the analysis of a single model (WS bond). By combining the results for different models, I obtain more accurate estimates for the exponents.
Fig. 6.4: Inverse of the quality at the minimum $1/Q_{\text{loc}}$ as a function of $\alpha$ for the different models in 2D. The inset shows for each model the inverse of the quality $Q$ as a function of $\Omega'$ with $\alpha$ fixed to its value at the global minimum $1/Q_{\text{max}}$. The vertical lines show the averages $d_f = 1.2168 \pm 0.0005$ and $\Omega = 0.9 \pm 0.1$ of the estimates for the fractal dimension and for the leading correction, respectively.
6.3. CORRECTIONS TO SCALING

fit, i.e. the number of system sizes used in the data (see Tab. 6.1) minus the number of fit parameters (here the number of resolvable amplitudes), and $\chi^2$ is the (weighted) mean square deviation of the fit. The quality $Q$ is a function of $\alpha$, $\omega'$, and $\Omega'$, but, as the terms of $\omega$ only have visible amplitudes for the WS site model in 2D, I drop hereafter the dependence of $Q$ on $\omega'$ and fix $\omega' = 0.6$ (changing this value had no effect on the results presented here).

Now, $Q$ should be maximal for $\alpha = d_f$ and $\Omega' = \Omega$, as the leading correction gives the dominant contribution compared to higher order ones. As a matter of convenience, I use the inverse of the quality $1/Q$, which is minimal for $\alpha = d_f$ and $\Omega' = \Omega$. The procedure to obtain $d_f$ and $\Omega$ for a given model is to measure the inverse quality $1/Q(\alpha, \Omega')$ of a fit of the proper ansatz to the data. I first choose a value of $\alpha$ and then derive $1/Q$ as a function of the exponent of the leading correction, scanning in the range $0 < \Omega' < 2$ with a step size $\delta \Omega' = 0.015$. The obtained curve, see for example Fig. 6.3(a) for WS bond, typically has a (local) minimum $1/Q_{\text{loc}}(\alpha)$, which marks the best fit of the leading correction $\Omega_{\text{loc}}(\alpha)$ for the chosen $\alpha$. The error $\Delta \Omega_{\text{loc}}(\alpha)$ is estimated from the width of the minimum. In two dimensions, analyzing these minima $1/Q_{\text{loc}}(\alpha)$ by varying $\alpha$ in the range $1.212 < \alpha < 1.220$ with steps of size $\delta \alpha = 0.00025$, yields an estimate of the global minimum $1/Q_{\text{max}}$ and the fractal dimension $d_f$. The error bar in $d_f$ is also determined from the width of the minimum [see, e.g., Fig. 6.3(b)]. I repeated this analysis for the watershed on sites, the main crack, and the bridge line (see Fig. 6.4) and the corresponding estimates are summarized in Tab. 6.2. The obtained values all agree with each other within the error bars. The ones for the MC, due to the low statistics, seem to differ more. Nevertheless, based on the similarity of the numerical values, I estimate by combining the results for all models that in 2D $d_f = 1.2168 \pm 0.0005$ and $\Omega = 0.9 \pm 0.1$ for all models.

The values and error bars have been obtained from the intersection of the estimated intervals for the individual models as given in Tab. 6.2. The value obtained for $\Omega$ is close to unity, which suggests that the leading correction is likely to be the analytic correction $\Omega = 1$.

I applied a similar analysis to the data obtained in three-dimensional systems, scanning $\Omega'$ in the range $0 < \Omega' < 2$ with a step size $\delta \Omega' = 0.015$ and $\alpha$ in the range $2.450 < \alpha < 2.535$ with steps of size $\delta \alpha = 0.0025$. As before, the detailed analysis is done like is shown in Fig. 6.3). For the case of the main crack in 3D, no conclusive results could be obtained with my method, but the obtained masses are within their error bars equivalent to those measured for the bridge line. I show in Fig. 6.5 only the results obtained
Table 6.2: The fractal dimension $d_f$ and the exponent of the leading correction $\Omega$ of the bridge line (BL) and the watershed (WS sites/bonds) for 2D and 3D, as obtained from a similar analysis as done in Fig. 6.3 for the WS bond case. The main crack (MC) result is only shown for 2D.

<table>
<thead>
<tr>
<th>model</th>
<th>$d$</th>
<th>$d_f$</th>
<th>$\Omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WS bond</td>
<td>2</td>
<td>1.2168±0.0005</td>
<td>0.95±0.05</td>
</tr>
<tr>
<td>WS site</td>
<td>2</td>
<td>1.21705±0.00075</td>
<td>0.91±0.19</td>
</tr>
<tr>
<td>BL</td>
<td>2</td>
<td>1.21655±0.0015</td>
<td>0.87±0.08</td>
</tr>
<tr>
<td>MC</td>
<td>2</td>
<td>1.21655±0.0045</td>
<td>0.86±0.11</td>
</tr>
<tr>
<td>WS bond</td>
<td>3</td>
<td>2.4865±0.0025</td>
<td>0.96±0.10</td>
</tr>
<tr>
<td>WS site</td>
<td>3</td>
<td>2.4865±0.0025</td>
<td>0.98±0.09</td>
</tr>
<tr>
<td>BL</td>
<td>3</td>
<td>2.4878±0.0025</td>
<td>1.06±0.16</td>
</tr>
</tbody>
</table>

for the watershed on bonds, on sites, and the bridge line. Like in 2D, the obtained estimates for $d_f$ and $\Omega$ agree within the error bars. Therefore, I estimate $d_f = 2.487 ± 0.003$ and $\Omega = 1.0 ± 0.1$ for three dimensions. As in 2D, the value of the leading correction is likely to be the analytic one $\Omega = 1$. Given this, for 2D and 3D, I also analyzed the data fixing $\Omega = 1$. The obtained values for the fractal dimensions and their error bars are consistent with the ones reported in Tab. 6.2, therefore, the possibility of $\Omega$ being analytical cannot be discarded.

The estimates of the fractal dimension for the different models are in agreement with the ones found in previous works for the watershed ($1.211 ± 0.001$ [65] and $2.48 ± 0.02$ [67]), the main crack ($1.215 ± 0.005$ and $2.46 ± 0.05$ [48, 49, 50]), the bridge line ($1.215 ± 0.002$ and $2.498 ± 0.005$ [179, 180, 62]), and the perimeter of the infinite cluster in discontinuous percolation ($1.23 ± 0.03$ [63] and $2.5 ± 0.2$ [64]). The value $1.211 ± 0.001$ given in Ref. [65] for the fractal dimension of the watershed in two dimension seems to underestimate the error bar.

### 6.3.3 Local Logarithmic Slope

Another approach to estimate the leading correction-to-scaling exponent $\Omega$ is to calculate the local logarithmic slope of the reduced mass $C_L(\alpha) = M_L L^{-\alpha}$,
Fig. 6.5: Inverse of the quality at the minimum $1/Q_{\text{loc}}$ as a function of $\alpha$ for the watershed on bonds, the watershed on sites, and the bridge line in 3D. The inset shows for each model the inverse of the quality $Q$ as a function of $\Omega'$ with $\alpha$ fixed to its value at the global minimum $1/Q_{\text{max}}$. The vertical lines show the averages $d_f = 2.487 \pm 0.003$ and $\Omega = 1.0 \pm 0.1$ for the estimates for the fractal dimension and for the leading correction.
i.e.,

\[ \Omega_{\text{est}}(L, \alpha) = \log \left( \frac{C_L(\alpha) - C_{L/2}(\alpha)}{C_{L/2}(\alpha) - C_{L/4}(\alpha)} \right) / \log(2). \]  

(6.7)

Taking \( L \) relatively large such that higher order corrections are negligible, \( \Omega_{\text{est}} \) converges to \( \Omega \) for \( \alpha = d_f \) (see, e.g., Refs. [206, 207]). Due to the uncertainty \( \Delta M_L \) in the average of the mass \( M_L \), there are in the estimate of the local slope systematic errors of the form

\[ \Delta \Omega_{\text{est}}^2(L) = \sum_{k=1,2,4} \left( \frac{d\Omega_{\text{est}}}{dC_{L/k}} \frac{\Delta C_{L/k}}{C_{L/k}} \right)^2 \]

\[ = \left( \frac{(\Delta C_L)^2 + (\Delta C_{L/2})^2}{(C_L - C_{L/2})^2} \right) \]

\[ + \left( \frac{(\Delta C_{L/2})^2 + (\Delta C_{L/4})^2}{(C_{L/2} - C_{L/4})^2} \right) \]

\[ + \left( \frac{(\Delta C_{L/2})^2}{(C_L - C_{L/2})(C_{L/2} - C_{L/4})} \right), \]

(6.8)

where \( \Delta C_L = L^{-\alpha} \Delta M_L \). I omitted here the \( \alpha \) dependence. This error heavily depends on the precision of the single mass measurements and, therefore, statistics considerably higher than for the fit quality method are needed, especially for the larger system sizes. I focused mainly on improving the statistics for the watershed on bonds and for the bridge line, where larger systems can be addressed.

In Figs. 6.6 and 6.7 I show \( \Omega_{\text{est}} \) with \( \alpha = 1.2168 \) and 2.487 for two- and three-dimensional systems, respectively. In both figures, only values of \( \Omega_{\text{est}} \) with \( \Delta \Omega_{\text{est}} < 1 \) are shown, except those for the MC, which are shown for completeness, but without their error bars. In the limit of large \( L \), I find for WS bond, MC, and BL data an agreement with the range of values for \( \Omega \) obtained from the fit quality method, which corroborates my numerical results. For the WS site model in 2D \( \Omega_{\text{est}} \) is consistent with \( \omega = 0.6 \), while in 3D it agrees with the other models. I cross checked also by applying other methods like, e.g., the one used in Refs. [208, 209] and found results consistent with the ones presented here.
6.4. RELATION BETWEEN THE MODELS

Fig. 6.6: Estimated leading correction $\Omega_{est}$ as defined in Eqs. (6.7) and (6.8) from the mass data of the bridge line, the watershed on bonds (sites), and the main crack in 2D. The value of $\alpha$ is fixed to 1.2168, the fractal dimension estimated by the fit quality method. For better visibility, the data of each model is shown with connecting lines and data points with $\Delta \Omega_{est} > 1$ have been removed. The values for the main crack (MC) are shown for comparison, but without their error bars. The horizontal lines give the value (solid) and error bar (dashed) for $\Omega$ as estimated by the fit quality method, as well as the value for $\omega$ (dotted).

6.4 Relation between the Models

The agreement of the fractal dimension and the leading correction exponent for all models within their error bars raises the question how closely related they are. All three models are based on percolation concepts. Below, I compare the models pairwise, based on the algorithmic details, focusing mainly on the sequential addressing of the sites. Both, the bridge line and the watershed are independent on the type of disorder [65], unlike the optimal
path crack, which I consider only in the strong disorder limit.

6.4.1 Bridges and Cracks

The random occupation procedure in bridge percolation, as already argued by Cieplak, Maritan, and Banavar [179], can be interpreted as rank sites by increasing order in the energy and iteratively occupy them according to their position in the rank. At every step, each occupied site has a lower energy than any unoccupied one. In strong disorder, the energy of any path

Fig. 6.7: Estimated leading correction $\Omega_{est}$ as defined in Eqs. (6.7) and (6.8) from the mass data of the bridge line, the watershed on bonds (sites), and the main crack in 3D. The value of $\alpha$ is fixed to 2.487, the fractal dimension estimated by the fit quality method. For better visibility, the data of each model is shown with connecting lines and data points with $\Delta \Omega_{est} > 1$ have been removed. The values for the main crack (MC) are shown for comparison, but without their error bars. The horizontal lines give the value (solid) and error bar (dashed) for $\Omega$ as estimated by the fit quality method.
6.4. RELATION BETWEEN THE MODELS

is dominated by the one of the site with the largest energy and, therefore, a path of occupied sites, has always lower energy than any path containing unoccupied ones. Occupying the first bridge site would lead to a spanning cluster (SC) and for the first time enable paths that connect the two opposite borders. The bridge site, as being the last occupied one in such a path, has the largest energy of all sites in it and characterizes the energy of the path. The optimal path is one of those paths, as their energy is lower than any other connecting path passing through unoccupied sites. This means that the first optimal path is cracked at the bridge site. Proceeding with the occupation of sites following the rank, the next time connecting paths are obtained is when the next bridge site is occupied. Again, the energy of the new optimal path is dominated by the energy of the current bridge site. As before, the crack appears at the bridge site. In this picture, the optimal paths always crack at bridge sites, until the system is completely disconnected. I, therefore, conjecture that the bridge line and the optimal path crack are identical.

6.4.2 Watershed and Cracks

The difference between WS bond and the other models is obvious, as it is defined on bonds where the others are on sites only. I might compare it with the bridge line in the bond percolation scheme, but, as I do not study that one, I drop this discussion here and just state that in general the watershed bonds are attached to bridge sites. The relation of the watershed on sites to the bridges and cracks is more subtle. In Fig. 6.8, I compare the cumulative distribution $P(m_s > m)$ of masses obtained for BL, WS bond, and WS site of 2D systems of size $L = 128$ and $m_s$ denotes the mass measured for a single system. Although for all the peak is at $m \approx L^{d_f}$, supporting the observed similarity of the fractal dimension, they have significantly different tails, where the one of WS site even follows a power law with an exponent $-1.8 \pm 0.1$. Similar observations have been made, e.g., in Ref. [210]. The origin of this power-law tail lies deep in the definition of the watershed on sites, namely in the fact that entire branches in the diverting runoff scheme can be part of the watershed. I will explain this here for the representative system depicted in Fig. 6.9. First, I start with the BL, occupying the sites in increasing order of the heights, so 1, 2, 3, . . . , 9. The first percolating cluster I would obtain when 6 is occupied, which is therefore a bridge site and the same applies to 7 and 8, while 9 just belongs to the bottom part. For WS site, I find that starting from 6 three branches develop, one to 1 and the bottom
sink, another to 2, passing to 4 and reaching the bottom sink (passing 1 and directly from 4), and a third to 3 and the top sink. Hence, from 6 both sinks can be reached and it is therefore part of the watershed. The same is true for 8, two going to the top (3 and 5) and one to the bottom (2). If I now start my runoff scheme from 9, I see that initially it diverts into four branches, where three are part of the basin of the bottom sink. But the branch going upwards, is split at 7 into three sub-branches (to 2, 5, and 6), the branch from 2 again reaches the bottom sink, but the one growing from 5 is part of the top basin. Hence, 7 is part of both (or neither) basin, so it is part of the watershed and, by definition, also its parent 9 has to be considered part of the watershed. Similarly, this can be deduced from the sub-branch to 6. The watershed of this system, therefore, consists of the BL and an overhang of one additional site. In general, such overhangs can be larger than one site.
Fig. 6.9: Representative system with $L = 3$, where each square cell represents a site of the lattice. The numbers in the lower left corner of each cell give the height, i.e. the sequence of occupation. The BL and the WS site model are applied on this system, where the two sinks are chosen as the rows below the bottom row and above the top row of the lattice. The letters in the upper right corner of each cell mark the bridge sites (B) and the watershed sites (W). According to the WS site model, the sites with arrows pointing up or down belong to the basin of the top or bottom sink, respectively.

but, as I will explain later, all bridge sites are always part of the watershed.

As I have conjectured, the main crack and the bridge line are identical, such that discussing the relation of WS site and BL is equivalent to discuss the relation of WS site and MC. Considering the elevations of a landscape to correspond to energies, e.g. potential energy, its watershed and its optimal path crack can be compared. I have defined that a site belongs to the watershed, when the invasion percolation clusters grown from two lower nearest neighbor sites do reach the opposite borders (catchments). As both clusters, by definition, do not cross the watershed, the watershed site separating the two, has a larger height (energy) than any site in both clusters. Therefore, for each watershed site (also the overhangs) there are always paths which consist solely of sites with lower energy than the watershed site, connecting it to either border. In the strong disorder limit, the energy of these paths is dominated by the largest local energy, i.e. the energy of the watershed site. Hence, every path connecting the two borders has at least the energy of the site where it crosses the watershed and the optimal path, the one of lowest energy, crosses the watershed at its lowest site. From the same arguments it follows that it is then the watershed site which is removed by the OPC model. After this, the next optimal path will cross the watershed at the next site in increasing order in the energy and cracks at the watershed as well. Until the
final disruption of the system, in strong disorder, every crack appears at the watershed site. Hence, the MC is also a subset of the watershed.

From these findings it follows that the power-law tail in the mass distribution for WS site arises due to the existence of overhangs. I know from previous studies, that the finite-size cut-off of distributions which follow a power law can heavily affect the scaling behavior of the moments of this distribution [67]. The average mass \( M \) is the first moment of the mass distribution \( P(m_s) \) (the derivative of the distribution shown in Fig. 6.8) and, therefore, its scaling behavior is affected by the cut-off \( L^2 \) of its power-law tail. As I based my analysis of the corrections to scaling on \( M \), also \( C_L \) might be affected. I observe the upper cut-off of the tail to scale with \( L^2 \) and the lower cut-off to scale with \( L^{d_f} \). Therefore, the functional form of the tail of the cumulative distribution is given by

\[
P(m_s > m) \propto m^{-1.8} L^{1.8 d_f}.
\]

To quantify the contributions of the overhangs to \( C_L \), I derive here, similar as it was done in Ref. [67], the contribution of the power-law tail between its cut-offs \( C_{\text{tail}} \) which scales as

\[
C_{\text{tail}} \sim \int_{L^{d_f}}^{L^2} m_s P(m_s) dm_s,
\]

\[
C_{\text{tail}} \sim \int_{L^{d_f}}^{L^2} m_s \left( \left. \frac{d}{dm} \right|_{m=m_s} P(m_s > m) \right) dm_s,
\]

\[
C_{\text{tail}} \sim \int_{L^{d_f}}^{L^2} m_s \left( \left. \frac{d}{dm} \right|_{m=m_s} m^{-1.8} L^{1.8 d_f} \right) dm_s,
\]

\[
C_{\text{tail}} \sim L^{d_f} (L^{-0.6} - \text{const}), \quad (6.9)
\]

what leads to a contribution of order \( L^{-0.6} \) to the corrections to scaling of WS site. Although it is only a rough estimate, the similarity of this contribution to the value I found for the leading correction (\( \omega \approx 0.6 \)) is striking. The other models have no such overhangs and therefore the corresponding amplitude is very small. Together with the fact that for these other models the amplitudes of the \( \omega \) correction are small, this suggests that this term in \( C_L \) of WS site only arises due to the overhangs. Apart from this I find the corrections to scaling of all models to be in agreement with each other. Furthermore, in 3D no such power-law tail is observed for the watershed on sites and all models hence have similar distribution of masses.
6.5 Summary

I obtained from a correction-to-scaling analysis, with high precision, an estimate for the fractal dimension of the watershed (WS), the bridge line (BL), and the main crack (MC). I found these fractal dimensions to be, within the error bars, in agreement with each other. Furthermore, the three models have within error bars also the same leading correction to scaling exponent. These results are also corroborated by the analysis of the local logarithmic slopes in the limit of large system sizes. I estimate for all models $d_f = 1.2168 \pm 0.0005$ and $\Omega = 0.9 \pm 0.1$ in two dimensions and $d_f = 2.487 \pm 0.003$ and $\Omega = 1.0 \pm 0.1$ in three dimensions. The equivalence between the models is also supported by heuristic arguments. The estimated values agree with the fractal dimensions obtained in previous studies for the watershed [65, 66, 67], the optimal path crack [48, 49, 50], and the bridge line [179, 180, 62], as well as with the ones found for the perimeter of the infinite cluster in discontinuous percolation ($1.23 \pm 0.03$ [63] and $2.5 \pm 0.2$ [64]).
Chapter 7

Conclusion
In this thesis, I introduced a novel and advanced numerical algorithm for the determination of watersheds in digital representations of landscapes, i.e. digital elevation model (DEM). In fact, the algorithm can be applied to any DEM like data to localize watershed like lines dividing the system into distinct parts. The extension to higher dimensions is straightforward and has been done for three dimensions. Although, I proved the equivalence to one of the most popular watershed algorithms, the *flooding*, my method outranges the current watershed algorithms in terms of efficiency, due to its sub-linear time-complexity. So far, I restricted myself to the study of systems, described by square (cubic) lattices, with only two catchments. A remaining task, therefore, is the extension of the algorithm to treat multiple-sinks in complex geometries.

Making use of the high performance of my method, I was able to measure the fractal dimension of watersheds in two and three dimensions to high accuracy. Applying the same algorithm to real landscapes I could prove that watersheds in nature indeed are self-similar objects, whose variety of fractal dimensions can be explained by the long-range spatial correlations present in the landscapes. Furthermore, introducing long-range spatial correlations into my model landscapes by using fractional Brownian motion, I was able to match the fractal dimensions of natural landscapes. Unanswered is the question, how the fractal structure of the watershed affects the transport properties in the basin and across the basins.

In chapter 5, I addressed the time evolution of watersheds due to slight landscape modifications. I was able to show that, indeed, small and localized perturbations can have a large impact on the shape of watersheds even at very long distances, hence having a non-local effect. The distribution $P(N_s)$ of areas enclosed by the perturbed and the original watershed is found to decrease as a power-law with exponent $\beta = 1.65 \pm 0.15$ on all studied real landscapes from mountainous (e.g. Rocky Mountains) to rather flat (e.g. US-Canadian border). By applying perturbations to model landscapes with long-range correlations, I determined the dependence of the scaling exponents on the Hurst exponent, finding good quantitative agreement with real landscapes, for which $0.3 < H < 0.5$. The obtained exponents $\alpha$, $\beta$ and $\rho$ are independent of the perturbation strength $\Delta$. I found the lower and upper bounds of the distribution of areas to scale with distance and system size with an exponent $\sigma = 2$ in two and $\sigma = 2.48 \pm 0.02$ in three dimensions. Considering these cutoffs, I was able to derive a function describing the average area, which includes the finite-size effects. In a similar way, by
integrating $P(M|R)$, I obtained a function for the average mass $\langle M \rangle$ of the invasion percolation cluster. Extending the perturbation study to three dimensional systems, I find the distributions and averages to follow power laws similar as in two dimensions, with similar dependence of $\alpha$, $\beta$, and $\rho$ on the Hurst exponent. This confirmed the conjectured relations between the exponents and reveals for uncorrelated systems an intrinsic relation to invasion percolation. In contrast to two dimensions, I observed the changed volumes to be dominated by their surface, i.e., the watershed, resulting in a value of $\sigma$ close to the fractal dimension of the watershed itself. As a followup of the work presented here, it would be interesting to see how the impact of perturbations could be related to other physical models, such as optimal path cracks [48, 50], bridge percolation [62], and the surface of explosive percolation clusters [63, 64]. How my findings can potentially be applied to face the future challenges in, e.g., water management or flood prevention, is an open question as well. Also missing is an experimental verification of my results. Further, it is still unknown whether my obtained power-laws match the distribution of watershed rearrangements, which occurred naturally.

I obtained from a correction-to-scaling analysis, with high precision, an estimate for the fractal dimension of the watershed (WS), the bridge line (BL), and the main crack (MC). I found these fractal dimensions to be, within the error bars, in agreement with each other. Furthermore, the three models obey very similar corrections to scaling. These results are also corroborated by the analysis of the local logarithmic slopes, which I found to agree in the limit of large system sizes. Based on the similarity of the obtained fractal dimensions and leading corrections, I estimate for all models $d_f = 1.2168 \pm 0.0005$ and $\Omega = 0.9 \pm 0.1$ in two dimensions and $d_f = 2.487 \pm 0.003$ and $\Omega = 1.0 \pm 0.1$ in three dimensions. The close relation between the models is also supported by heuristic arguments. The estimated values are in line with the fractal dimensions obtained in previous studies for the watershed [65, 66, 67], the optimal path crack [48, 49, 50], and the bridge line [179, 180, 62], as well as with the ones found for the perimeter of the infinite cluster in discontinuous percolation (1.23 ± 0.03 [63] and 2.5 ± 0.2 [64]). It would be interesting to know if this perimeter also obeys similar corrections to scaling as I have found. Another future work would focus on the duality of the discussed models and optimal paths in two dimensions by comparing my results with their corrections to scaling.
Acknowledgment

Though only my name appears on the cover of this dissertation, I would not have come so far without the thorough support of many great individuals.

I wish to express my gratitude to my supervisor, H.J. Herrmann. I was fortunate to have a supervisor who gave me not only the freedom to explore ideas on my own, but also strict enough boundaries to keep me upright and working. With his guidance, he helped me to realize my true potentials. Much of my understanding and skills in scientific research I own only due to his constant and constructive criticism.

Fortunately, I had the chance to collaborate with many other professors and colleagues, from whom I learned a lot. My special thanks go to J.S. Andrade, C. F. Moukarzel, P. Grassberger, D. Kadau, N.A.M. Araújo, M. Fuhr, K. J. Schrenk, S. D. da Cunha, L. R. da Silva, and E. A. Oliveira.

I have benefited a lot from very fruitful discussions with J. S. Andrade, P. Grassberger, and K. J. Schrenk, who also shared their datasets and/or codes with me. I acknowledge exciting discussions with other great researchers: D. García-Castellanos and R. Hidalgo.

I am grateful to N.A.M. Araújo and D. Kadau for their constant support through elaborated and constructive comments and suggestions during the writing process for this thesis and the preceding papers. I also want thank them for the long and numerous discussions, that helped me to improve my knowledge in the field. Whenever, I had doubts or questions on anything, they took their time to help me out.

I would like to thank K. Kovacs and T. Burgener, who helped me with their technical support and L. de Arcangelis, E. Del Gado, and T. Shinbrot for their continuous interest in my work. I also thank T. Kesselring and M. Mendoza for fruitful and funny discussions in our office.

I very much enjoyed my stay within the group of computational physics here at the institute for building materials, due the warmly welcome in every
office whenever I dropped by. Especially, I would like to thank A. Weinberger for her optimism, her open heart, and administrative support.

I acknowledge financial support from CNPq, CAPES, FUNCAP, and the CNPq/FUNCAP-Pronex grant.

Some of the research in this thesis uses data from the satellite radar topography mission (SRTM) [182], which can be obtained from http://srtm.csi.cgiar.org/.

Last but not least I thank my family and my friends for their endless support, especially my parents and their neighbors D. & G. Bernhard-de Caro, who welcomed me at any time during day and night. Not to forget the people from the Tanzquotient of VSETH, who helped me to find again joy and relaxation, after a hard day of work.
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


CV

Eric Marcel Fehr

Date of birth 03.08.1982
Place of birth Dielsdorf, Switzerland
Citizenship Kilchberg ZH, Switzerland

Education

Aug 1995 - Jan 2002 Kantonsschule Zürcher Unterland in Büäch

Military Service

Feb 2002 - Okt 2002 Rekrutenschule and Unteroffiziersschule in Fribourg as Nachrichtensoldat

Study of Physics

Aug 2003 - Aug 2008 Diploma study of physics with additional specialization in computer science at ETH Zurich
May 2008 - Aug 2008 Diploma thesis in the computational physics group at the institute of building materials (IfB)
Sep 2008 - Aug 2011 PhD study in physics at ETH Zurich
List of publications

New efficient methods for calculating watersheds

E. Fehr, D. Kadau, J. S. Andrade Jr., and H. J. Herrmann
Impact of Perturbations on Watersheds

J. S. Andrade Jr., S. D. S. Reis, E. A. Oliveira, E. Fehr, and H. J. Herrmann
Ubiquitous Fractal Dimension of Optimal Paths

E. Fehr, D. Kadau, N. A. M. Araújo, J. S. Andrade Jr., and H. J. Herrmann
Scaling Relations for Watersheds

E. Fehr, K. J. Schrenk, N. A. M. Araújo, D. Kadau, P. Grassberger, J. S. Andrade Jr., and H. J. Herrmann
Corrections to Scaling for Watersheds, Optimal Path Cracks, and Bridge Lines
arXiv:1203.3038

List of conference talks

E. Fehr, D. Kadau, N. A. M. Araújo, J. S. Andrade Jr., and H. J. Herrmann
Impact of Perturbations on Watersheds
Journée de Physique Statistique, Paris (January 2011)

E. Fehr, K. J. Schrenk, N. A. M. Araújo, D. Kadau, C. Moukarzel, J. S. Andrade Jr., and H. J. Herrmann
Floods, Dams, and Watersheds
International Workshop on Coping with Crises in Complex Socio-Economic Systems, Zurich (June 2011)