

Evaluation of Image Pattern Feature Extraction From Digital Elevation Model for Hydrogen Seeps Detection

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of
Master of Science (M.Sc.)

PARIS-LODRON UNIVERISTY SALZBURG (PLUS)

Faculty of Digital and Analytical Sciences

Department of Geoinformatics

and

UNIVERSITY OF SOUTH BRITANNY (UBS)

Faculty of Sciences and Engineering Sciences

Mathematics, Computer Science, Statistics Department

Data Science Specialization

PLUS Supervisor:

Dr. Zahra Dabiri

UBS Supervisor:

Prof. Dr. Sébastien Lefrère

submitted by

Rodrigo Brust Santos

Salzburg & Vannes, June 2024

With the support of Erasmus+ programme of the European Union. This Master's Thesis has been developed in the framework of the Erasmus Mundus Joint Master Degree (EMJMD) "Copernicus Master in Digital Earth", jointly coordinated by Paris-Lodron University Salzburg, Department of Geoinformatics, Austria together with University of South Brittany, Mathematics, Computer Science, Statistics Department, France and Palacký University Olomouc, Department of Geoinformatics, Czech Republic.

Declaration

This thesis has been composed by Rodrigo Brust Santos for the Erasmus Mundus Joint Master's Degree Program in Copernicus Master in Digital Earth for the academic year 2023/2024 at the Department of Geoinformatics, Faculty of Natural Sciences, Paris Lodron University Salzburg, and Department of Mathematics, Computer Science and Statistics, Faculty of Science and Engineering, University of South Brittany

Hereby, I declare that this piece of work is entirely my own, the references cited have been acknowledged, and the thesis has not been previously submitted to the fulfillment of the higher degree.

Vannes, June 2024
Rodrigo Brust Santos

Acknowledgments

I want to express my gratitude to my friends that were by my side during the rainy year in Vannes, specially to Candela, who has supported me in the darkest moments. To my parents, that even though far away, gave me the knowledge to face the struggles. To Tom Allavaneda, who's helped me uncountable times with code implementation. To Jeuvaugh, who's helped me to review my thesis.

Last but not least, I want to thank my supervisors Dr. Sébastien Lefèvre and Dr. Zahra Dabiri for guiding me and providing me the necessary tools to be able to accomplish this work.

This thesis, was proudly developed with multiple open-source software products. Their recognition is fundamental in order to keep the FOSS movement. For data visualization and manipulation, QGIS (QGIS Development Team, 2009), Matplotlib (Hunter, 2007), Geopandas (Jordahl et al., 2020), Numpy (Harris et al., 2020), Rasterio (Gillies et al., 2013), SAP (Santana Maia et al., 2021) and Higras (Perret et al., 2019).

Vannes, June 2024
Rodrigo Brust Santos

Abstract

Society is currently experiencing an energy transition, where geological hydrogen could be a clean solution. Naturally, the Earth emits H₂, which can be observed in Sub-Circular Depressions (SCDs) already identified on several continents. To accelerate the identification and mapping process of these SCDs, freely available earth observation data and well-established methodologies from mathematical morphology, such as Pattern Spectra (PS), can be utilized. This research aims to map SCDs using PS, Attribute Filtering (AF), and Shape-Spaces techniques applied to hierarchical Digital Elevation Model (DEM) representations.

These hierarchical representations, referred to as trees, form the basis of various methodologies in mathematical morphology, including the established PS technique. This methodology serves as a benchmark for the innovations introduced in this work: AF in Shape-Spaces and PS in Shape-Spaces. Additionally, considering trees and their characterizing attributes, this research theoretically presents a new attribute termed Estimated Slope.

When the new methodologies were applied to SCD delimitation, they exhibited limitations, particularly in filtering medium-sized structures. The Intersection over Union (IoU) and False Positive Rate (FPR) metrics were relatively low, ranging between 11-31% and 68-93.9%, respectively, primarily due to the high number of false positives that were filtered out. The selection of attributes and the threshold algorithm did not achieve optimal values for histogram segmentation. Despite these limitations, the new methods demonstrated increased robustness, as the number of false positives was reduced and only rounded structures were filtered out compared to the base methodology.

In order to verify the effectiveness of the new methodologies, it would be worthwhile applying them to other image analysis domains, to map the SCD with more suitable attributes, to use a modern threshold approach, and to develop the Estimated Slope attribute algorithm.

Contents

1	Introduction	1
1.1	Motivation	3
1.1.1	Thematic Motivation	3
1.1.2	Methodological Motivation	3
1.2	Research Questions	4
1.3	Objectives	4
1.4	Study Area	5
1.5	Thesis Structure	7
2	Literature Review	9
2.1	Background on Hydrogen Seeps	9
2.2	Background on Trees, Attribute Profile and Pattern Spectra	11
2.2.1	Trees	11
2.2.2	Pattern Spectra	12
2.3	Background on Attribute Profile, Tree Filtering and Shape-Space	13
2.3.1	Attribute Profile	13
2.3.2	Tree Filtering	14
2.3.3	Tree Filtering on Space of Shapes	15
2.4	Related Works	16
3	Methodology	17
3.1	Dataset Description	18
3.2	Ground Truth Generation	18
3.3	Tree Generation from TanDEM 30m	19
3.4	Tree Filtering Methodologies	20
3.4.1	Pattern Spectra	20
3.4.2	Attribute Filtering within Space of Shape	20
3.4.3	Pattern Spectra Applied to Space of Shapes	21
3.4.4	Theoretical Proposition of Estimated Slope As a New Attribute	21
3.5	Evaluation Metrics	24
4	Results	25
4.1	Pattern Spectra Applied to DEM	25
4.2	Attribute Filtering within Space of Shape	28
4.3	Pattern Spectra within Space of Shape	31

5	Discussions Conclusions	35
5.1	Metrics Evolution	35
5.2	Attributes Performance and Threshold Setting	37
5.2.1	Attributes	37
5.2.2	Threshold	37
5.3	Limitations	38
5.4	Final Thoughts	38
5.5	Future Works	39
	Bibliography	41
	List of Figures	50
	List of Tables	51
	List of Listings	51
	List of Abbreviations	53

CHAPTER 1

Introduction

Currently, we are experiencing a sharp energy transition from polluting sources to renewable energy sources. Climate pressure, coupled with heavy investment, is causing countries to diversify their energy matrices beyond coal, gas, and oil, for example (Floristean, 2020; Moretti and Webber, 2020).

Hydrogen is one of the energy sources that has received a lot of attention in recent years. This is due to its high energy potential and small carbon footprint. There are several ways to obtain hydrogen, and its different origins are classified by color, resulting in what is called the hydrogen rainbow. Figure 1.1 shows the most common energy sources of H₂ and the production process used. In total, there are 10 types of hydrogen, classified by the production process, energy source and whether CO₂ is emitted or not (Incer-Valverde et al., 2023).

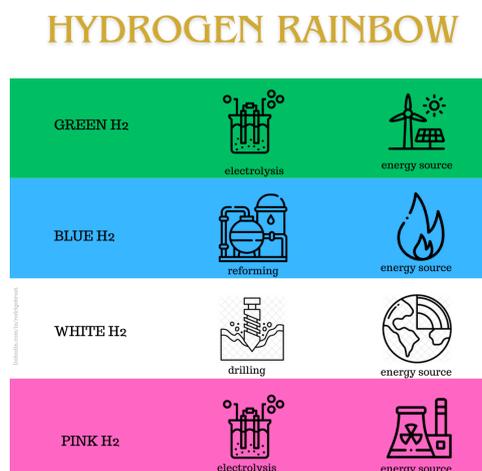


Figure 1.1: The main generation sources of H₂ and their classification: green hydrogen, generated from electrolysis using renewable energy and zero CO₂ emissions; blue hydrogen, generated from natural gas by reforming it, leading to CO₂ emission; white hydrogen, naturally generated; pink H₂, produced in nuclear power-plants by electrolysis and without CO₂ emissions. Source: Author, 2024.

Although Incer-Valverde et al. (2023) briefly mention that it has origin from natural sources

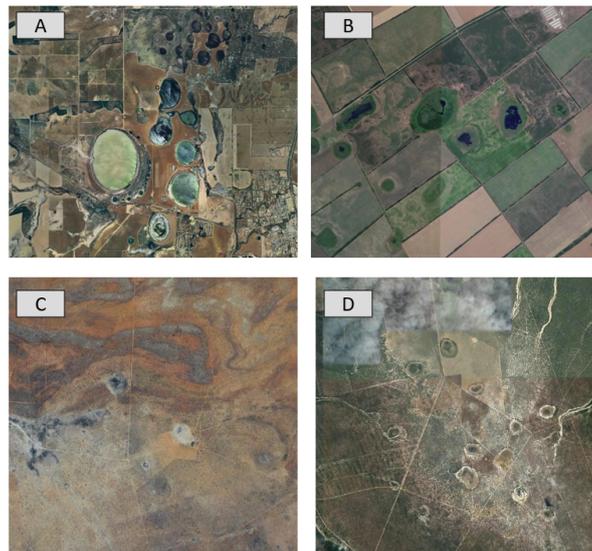


Figure 1.2: Example of studied fairy circles in several countries. A) Australia; B) Russia; C) Namibia; D) Brazil. All images at a scale of 1:50.000. Satellite source: Google Earth.

in Earth's subsurface. These deposits can be accessed via a well, as has been the case in Mali since 2012, producing energy with zero emissions and at a cost of less than 1 dollar per kilo of H₂ (Moretti and Webber, 2020; Willige, 2022). Despite this example, hydrogen system and setting must be further investigated in order to transform the raised hypothesis into factual concepts (Frery et al., 2021).

Some studies try to identify possible origins for white hydrogen, which depends on the rock types and chemical environment (Lévy et al., 2023; Moretti et al., 2022). Such subsurface deposits are subject to geological conditions, and H₂ can be exuded through conductive faults, generating rounded depressions without the presence of vegetation, where high concentrations of H₂ and methane are detectable in the soil when compared to the surrounding regions (Larin et al., 2015).

Many ellipsoid structures, known as Fairy Circles (FC) or Sub-Circular Depressions (SCD), were already identified within cratonic regions, such as in Namibia (Moretti et al., 2022), Australia (Frery et al., 2021), Turkey (Etiope, 2023), Russia (Larin et al., 2015) and Brazil (Prinzhofer et al., 2019), as shown in Figure 1.2.

1.1 Motivation

1.1.1 Thematic Motivation

Numerous researches have been carried out on all continents with the aim of obtaining more information about SCD and how they can contribute to the development of the energy industry. However, the use of satellite images and remote sensing has been little explored in the identification of hydrogen sources, except for a couple of examples Carrillo Ramirez et al. (2023); Lévy et al. (2023); Moretti et al. (2022); Mosquera-Rivera et al. (2024).

From a geological point of view, there are several round structures that can be formed during the evolution of a terrain. The SCD is one more ellipsoid/rounded formation that can be seen from space with the aid of Earth Observation (EO) products, such as freely available Digital Elevation Model (DEM). Moretti et al. (2021) conducted a comprehensive morphological profiling of SCD and compared them with other geological structures, such as dolines, which are also rounded formations. Their research revealed that SCDs typically have a gentle slope and are shallow, with depths up to 10 meters. In contrast, dolines are characterized by depths greater than 10 meters and slopes exceeding 50%.

Moretti et al. (2021) concluded that the primary goal of their study was to establish morphological criteria. Later, this knowledge about the SCD could be applied in remote sensing products, such as multispectral imagery and DEM. So far, Moretti et al. (2022) were able to use these products to analyze around the SCD, but not in an automated form. Considering this goal and the detailed profiling of SCD, questions arise regarding how to develop a more direct and automated workflow for identifying and delineating SCD structures using EO products. The development of such a procedure could significantly benefit researchers and companies by aiding in the early exploration phase of projects. This would ensure that regions of interest contain the desired SCD formations rather than other geological structures, such as dolines, leading to more accurate and efficient field mapping.

1.1.2 Methodological Motivation

To carry out this SCD scanning, a field of Computer Vision called Mathematical Morphology can be used. In this area there are several methodologies capable of performing operations such as image segmentation, image filtering, pixel classification, both on regular images and on images generated from remote sensing. Two widely established methodologies, such as Pattern Spectra and Attribute Filtering, may be tools capable of delineating SCDs. Despite the advances made in this area over the last decade, there are no researches combining these methodologies with H2 exploration, as it is a very new subject.

With the advancement in the field of image analysis, image processing and image classification, this whole workflow could be done utilizing the newest deep learning approaches. However, one of the goals here is to perform structure delineation without using methodologies that are computationally expensive. In this matter, PS and Attribute Filtering are strong candidates to find this equilibrium. In addition, the developments in this field in the last decade show that considering approaches other than deep learning are still valid (Santana Maia et al., 2021).

In this way, the present work has the motivation of uniting these two different fields of study, as well as proposing novelties using the Pattern Spectra and Attribute Filtering methodologies, with more recent topics, such as Shape-Spaces, to extract SCD from DEM. The importance of such novelties goes beyond the development of mathematical morphology methodologies, but also to demonstrate whether the combination of PS and Attribute Filtering with Shape-Spaces is a valuable approach or not. In this sense, the current research can help in two fronts: in the computer vision methodological advancements and in using such methods for the H2 mapping and exploration.

1.2 Research Questions

In order to achieve the desired implementation of the workflow to delineate and identify the SCD, we will be using computer vision techniques applied to the DEM, such as Pattern Spectra (PS) and Shape-Spaces, where from attributes (i.e., compactness, area, height), it is possible to filter regions in the image that correspond to a specific threshold. To analyze the application of computer vision tools on the extraction of SCDs, the following research questions are raised:

1. *Which accuracy can be achieved for fairy circle delineation using medium resolution DEM and Pattern Spectra and the proposed methodological novelties?*
2. *Does regular Pattern Spectra application performs better than the new methodologies proposed?*
3. *Can new pattern spectra attributes be generated based on prior knowledge of the inclination of structures from past morphological studies?*

1.3 Objectives

The research questions presented lead to the following objectives:

- To use regular pattern spectra attribute profile to delineate the desired structures;
- To identify possible strengths and limitations on regular attributes for extracting these structures;
- To apply PS and attribute filtering within Shape-Spaces as a new methodological approach;
- To develop a new attribute based on expert knowledge, such as presented by Moretti et al. (2021) about the slopes of the SCDs.

1.4 Study Area

The study area is located in the Moora-Pingarrega area, Western Australia (Figure 1.3). This area was chosen for the following reasons: the presence of studies in the same area with measurements proving the presence of H₂ in the soil of the structures (Frery et al., 2021); the existence of geological and geophysical data; the availability of DEMs for the region of interest; the high density of SCDs of different sizes and different roundness characteristics, thus allowing the different methodologies to be tested.

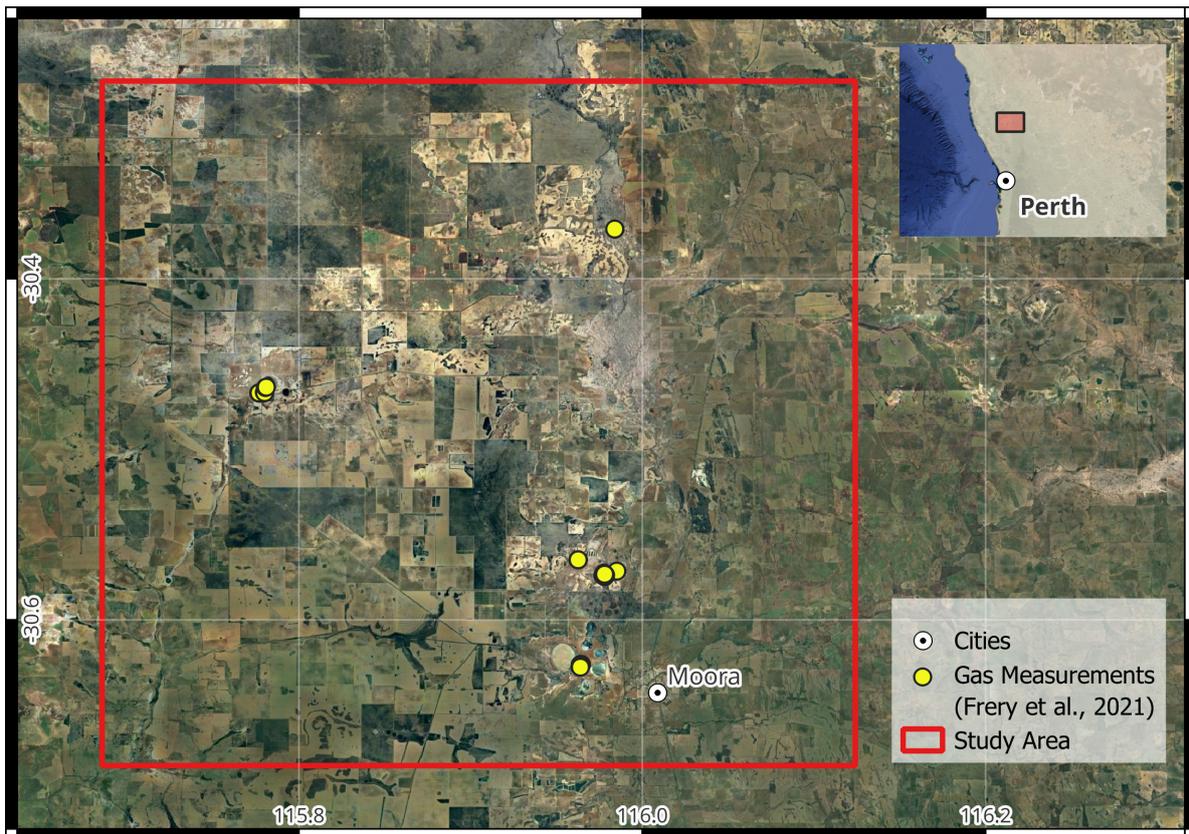


Figure 1.3: Location of the Study Area and H₂ Measurements done by Frery et al. (2021)

With the availability of expert knowledge and ground truth measurements makes it possible to assess the structures using EO products. With these types of data, it is feasible to identify, delineate the Fairy Circles and extract their features.

The Moora-Pingarrega region is within the North Parth Basin (Figure 1.4). Its terrain is mainly undulating, with some hills and valleys (Mory et al., 2005). The area was chosen by Frery et al. (2021) due to its geological setting. Located in this basin with complex structural geology, in the west part of the Yilgarn Craton, the region is controlled by the Darling Fault.

A geological model was developed by Frery et al. (2021), based on the analysis of regional geology, structural, geophysics and lithology. The hypotheses based on the model indicate that the hydrogen seeps are located along major Darling Fault zone, which makes its surroundings compartmentalized. This characteristic means that beneath the surface, there are a lot of

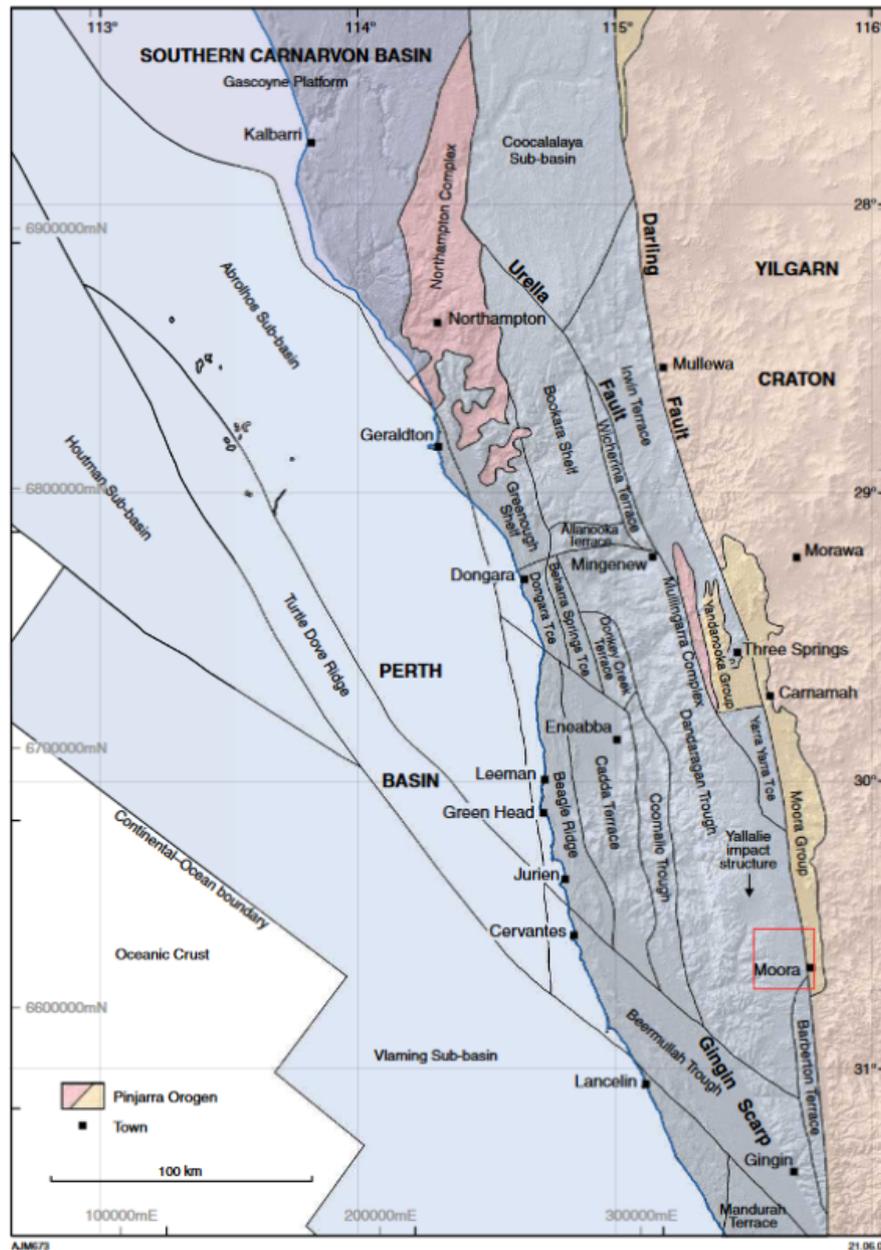


Figure 1.4: North Perth Basin, modified from Mory et al. (2005)

smaller fault families that are responsible for conducting the H₂ gas from the depths to the surface, forming the SCD. In addition to conducting, these faults can also work as traps, leading to an H₂ reservoir, however it needs further investigations in order to prove it (Frery et al., 2021).

The H₂ generation rocks in the region could be related to the Proterozoic Iron-Rich Granite or Pre-Cambrian ultramafic rocks near-by the Darling Fault. These possible sources mean that, according to Lévy et al. (2023), the H₂ in the region could be classified as Type I or Type

II, depending on which rocks the chemical reactions are occurring (see Table 2.1 for types explanation).

1.5 Thesis Structure

The structure of the thesis is as follows:

- **Chapter 1** introduces the basic concepts of hydrogen-emitting structures, study area, and the motivation for carrying out this research and the objectives.
- **Chapter 2** presents the literature review on hydrogen, related work and the mathematical morphology topics, such as trees, attribute profiles, pattern spectra and connected operators.
- In the **Chapter 3** the materials and methods used to develop the current thesis, besides the proposed methodologies.
- **Chapter 4** the results obtained are presented and,
- discussed in **Chapter 5**, followed by the limitations, conclusions and future works.

CHAPTER 2

Literature Review

2.1 Background on Hydrogen Seeps

Hydrogen seeps, also known as Fairy Circles and Carolina Bays, are SCD, with depths reaching up to 10m, gentle slopes, having their diameter ranging from a hundred meters up to a couple of kilometers (Moretti et al., 2021; Zgonnik, 2020). One can only affirm that it is a hydrogen seep when it leaks H₂, that's why it is necessary to monitor these structures during a couple of months in order to guarantee that it is emitting H₂, since the activity can be intermittent (Moretti et al., 2021; Prinzhofer et al., 2019). Some studies point that the emission of H₂ happens in pulses, having moments with higher and lower flow, being able to reach zero (Larin et al., 2015; Prinzhofer et al., 2019).

Hydrogen seeps with H₂ exudation has been mapped all around the world (Carrillo Ramirez et al., 2023; Etiope, 2023; Frery et al., 2021; Larin et al., 2015; Lefeuvre et al., 2021; Moretti et al., 2022; Prinzhofer et al., 2019), with one special case in Mali, where is actively producing energy for the last decade (Diallo et al., 2022).

Despite this single example, natural hydrogen exploration is in its infancy. Researches still have a lot to understand about the gas generation, geological setting for trapping and conducting the gas (Frery et al., 2021; Moretti et al., 2021; Zgonnik, 2020); and studying from its leakages on Earth's crust can point to large reservoirs in the subsurface (Zgonnik, 2020). Tracing a parallel, right now the hydrogen industry is in the same pace level as was the oil and gas industry back in the 19th century (Smith et al., 2005). Back then, the oil and gas exploration happened in shallow depths, and turned to be successful, which could be a cheap and fast way to also test for hydrogen (Moretti et al., 2021).

The Hydrogen Cycle needs to be better understood. Yet, there is a proposed classification of the three main processes that generate the natural hydrogen in the subsurface (Lévy et al., 2023):

- Redox of Fe⁺²
- H₂O Radiolysis
- Pyrolysis of Organic Matter

The rocks (Lévy et al., 2023) and geological context (Moretti et al., 2021) necessary for these reactions to happen are:

Table 2.1: Rock Types and Geologic Context necessary for H₂ reactions.

Rock Types	Geological Context
Basic and ultrabasic rocks from the mantle	Extension zones (e.g., Atlantic Ridge)
Iron bearing rocks	Compression zones with ophiolites (e.g., Oman)
Radioactive rocks	Stable cratonic basins with Archean to Proterozoic basement (e.g., São Francisco Craton)
Organic rocks, such as coal and shale	

Once the H₂ gas was generated in the depths, it needs a fault, so it's possible to migrate (Johnsgard (1988); and trap, to avoid it from leaking. According to Subbota and Sardonnikov (1968, see on page 9, apud Zgonnik (2020)), such traps could be a layer of clay, salt or an aquifer (Vacquand, 2011). Due to the fact the hydrogen gas has the least amount of molecules from the whole elements groups, it is the lightest, giving it a characteristic of rapid diffusing, in air and materials (Zgonnik, 2020). Summed up to that, it is a gas without color, odorless and non-toxic, which becomes even harder to identify some leakage. As a consequence of its characteristics, the hydrogen seeps weren't discovered until recently (Larin et al., 2015; Prinzhofer et al., 2019, 2018; Sukhanova et al., 2013). So far, for example, there are a couple of models and studies that indicates that these type of structures occurs near fault zones (Frery et al., 2021; Larin et al., 2015).

The hydrogen industry and researches are quickly escalating. As the time passes by, more institutions and governments are interested in this topic due to the possibility of generating clean energy with relatively low cost. A couple of example are: the Colombian government has created law to foment these types of discoveries (Carrillo Ramirez et al., 2023); and Germany put natural hydrogen in their energy strategy plan in 2020 Diallo et al. (2022).

This governmental fomentation and companies willingness across the world builds the perfect momentum to increase the knowledge over the H₂ system and the SCDs. Using EO products combined with mathematical morphology can be an interesting strategy applied to the H₂ emitting structures, seizing industry's momentum.

2.2 Background on Trees, Attribute Profile and Pattern Spectra

2.2.1 Trees

In digital image processing, connected pixels with similar values are grouped into objects, giving rise to the connected components. This concept, which was presented by Serra (1998), had such a positive effect that it led to other studies, like connected operators in binary and grayscale images (Heijmans, 1999; Salembier and Serra, 1995; Serra, 1998; Serra and Salembier, 1993).

It was then realized that this concept would be very relevant, since it could be implemented in image filtering and segmentation (Salembier and Serra, 1995), with the use of thresholds to select the desired elements (Perret et al., 2012). These components, which are selected from the values, can easily be represented hierarchically in a tree structure (Salembier et al., 1998).

Hierarchical trees, a type of connected operator, have two main groups (Bosilj et al., 2013): **inclusion trees and partitioning trees**, as exemplified in Figure 2.1.

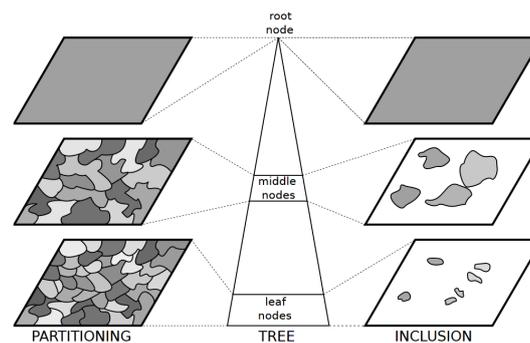


Figure 2.1: Differences from Partition and Inclusion Trees. Source: Bosilj et al. (2017).

Inclusion Trees are representations of partial partitions of an image, involving nested supports and components that are produced via the creation, expansion, and merging of image blocks (Ronse, 2014), for example, there are **max-tree**, **min-tree**, and **tree-of-shape**. On the other hand, **partitioning trees** are full partitions of the image. These trees begin with the finest partitions stored in the leaf nodes. Pixels are successively merged until one node is left per level, known as the root node, into which the whole image domain fits (Bosilj et al., 2018). **Binary partition tree**, **alpha tree** and **omega tree** are examples of partition trees.

It is worth noting that inclusion trees need a total order of pixel values, but partitioning trees are constructed based on the dissimilarity measure between neighboring pixels and therefore require a total order of the image's edges, but not of pixels themselves (Bosilj et al., 2017). Moreover, inclusion trees are designed to capture extrema-oriented information; such trees may contain small regions or points (e.g., local image maxima or minima), meanwhile partitioning trees are used to partition an image and leaves and cuts indeed represent different partitions of an image. For that reason, partitioning trees are well suited for processing

multivariate data, as they can hold information about an object at intermediate gray levels without necessitating a strict total order of pixel values (Bosilj et al., 2017, 2018).

The **min-tree** and **max-tree** are types of inclusion trees (Ballester et al., 2003). The **max-tree** has its local maxima (bright pixels) in the leaves, meanwhile the root is composed by dark pixels. The **min-tree** has the opposite behavior. These are the trees commonly used in pattern spectra analysis (Bosilj et al., 2016a; Guiotte et al., 2020), which is the next covered topic. **Tree of Shapes** (ToS) is a combination of **max-tree** and **min-tree** that seeks representing both dark and bright pixels. The local extrema's levels of ToS don't necessarily represent maximum and minimum intensities (Géraud et al., 2013).

Binary Partition Tree (BPT) (Salembier and Garrido, 2000), have pixels, flat areas or other thin partition in the leaves. As a consequence, it is the tree with most flexibility due to the fact that distinct structures can be generated by utilizing distinct definitions of the region model and the merge criteria (Merciol et al., 2014). The **alpha-tree** is generated based on the pixel's similarity. Its regions are defined by a fixed alpha distance between two pixels. Finally, the **omega-tree** tries to solve the alpha-trees' chaining effect. Meanwhile, alpha-tree sets an alpha distance for pixels, the **omega-tree** have an attribute called omega that limits the maximum difference between the lowest and highest grayscale in each connected component.

Since its proposal, the trees were responsible for the basis of mathematical morphology in many developed researches. There are uncountable works that used either inclusion trees or partition trees. The list is quite extensive, but as example of trees usage, it's possible to list image filtering and image segmentation (Jones, 1997), change detection (Tushabe and Wilkinson, 2008), classification (Urbach et al., 2007) and remote sensing (Boldt et al., 2014; Cavallaro et al., 2015; Ghamisi et al., 2014). In this research, the objective is to generate an inclusion tree, specifically a min-tree from a DEM and apply different mathematical morphology methods, such as PS.

2.2.2 Pattern Spectra

Pattern Spectra (PS), proposed by Maragos (1989), is an image descriptor based on mathematical morphology. It has become an essential tool for understanding the structural composition of images by analyzing the distribution of sizes and shapes of image components (Bosilj et al., 2016b).

PS features a structure based on histograms, displaying the distribution of elements within an image. Following the construction of trees, various statistical and geometrical attributes can be calculated for each tree node. PS then illustrates how these attributes behave in a histogram (Mirmahboub et al., 2021). By establishing a direct link between attribute values and corresponding image areas, PS can highlight significant regions based on selected attributes, making it a strong tool for image analysis (Guiotte et al., 2020).

Reading the histogram generated by the attributes is straightforward. In each axis there is an attribute. The sum of two one-dimensional histograms lead to a 2D Spectrum (Bosilj et al., 2016b), exhibiting a distinct pattern. This graphical representation is delineated by the selection of threshold values for each attribute. Consequently, only a subset of nodules is delineated, facilitating the identification of regions of interest within the analyzed image.

Since it can be efficiently applied to a max-tree and min-tree hierarchy, PS have received several extensions from the original concept throughout the years. In this manner, PS was effectively applied to image classification (Urbach et al., 2007) and image retrieval (Tushabe and Wilkinson, 2008). Bosilj et al. (2016b) demonstrated its efficacy in retrieving image patches instead of the whole image. Another extension was the application of descriptors based on local pattern spectra, introduced by Bosilj et al. (2016a). Such novelty brought speediness to the retrieval of satellite imagery. In addition, it reduced the dependency on feature vectors, improving the results related to dense STIF approach, proposed by Ozkan et al. (2014). Also, Lefèvre (2009), proposed the combinations of attributes that describe various object characteristics, such as orientation, color, spectral, spatial information and intensity, but they presented limitations because of their 1D properties.

More recently, Mirmahboub et al. (2021) employed the fast PS which aimed parallel processing by a sliding window over a large image and find the areas of interest. The results were positives, since it was capable of improving local pattern spectra. In the application field, Guiotte et al. (2020) utilized PS to extract terrain features from a high-resolution DTM generated from LiDAR. These examples underscore the versatility of PS in diverse image analysis tasks. For the current work, PS will serve as a baseline method, and also be implemented within Shape-Spaces. Finally, all methodologies will be compared in order to see whether these proposals are relevant or not.

2.3 Background on Attribute Profile, Tree Filtering and Shape-Space

2.3.1 Attribute Profile

Introduced by Dalla Mura et al. (2010), an Attribute Profile (AP) is a method in image analysis that involves the application of attribute filters based on specific criteria to analyze and decompose an image according to the chosen attributes. It allows for a detailed examination of the image's structural information by considering various characteristics or properties of the regions within the image.

Such novelty aimed to overcome the limitations with the Morphological Profile (MP) approaches, such as computation complexity, partial characterizations of objects leading to a partial analysis, and structure element constraint (Dalla Mura et al., 2009, 2010). Since AP and Attribute Filtering are correlated topics, a lot of the extensions and improvements done in the AP can be applied in the other. Specially when it comes to threshold setting and selection of attributes.

The first publications about AP presented only 4 attributes: area, moment of inertia, diagonal length of bounding box and standard deviation (Dalla Mura et al., 2010). With time, more attributes were introduced such as entropy, homogeneity (Dalla Mura et al., 2010), complexity (Das et al., 2018), perimeter and area of bounding box (Bhardwaj et al., 2019), solidity (Bhangale et al., 2017), and many others more.

Santana Maia et al. (2021) states that it is fundamental to choose a significant set of attributes

in order to obtain a good classification based on the interested structures. In case of not knowing the optimal combination, it is good practice to use the most common combined attributes.

Regarding the threshold of the attributes, it is known that the optimal selection of values have a direct relationship with the AP performance (Aptoula et al., 2016). In order to find the best threshold, many studies propose equations (Ghamisi et al., 2014) and supervised learning algorithms for attributes clustering (Mahmood et al., 2012). Besides these two examples, there are other recent develop that propose solutions for the topic, such as Cavallaro et al. (2017) proposal on using Granulometry Characteristics Functions (CGFs) for selecting thresholds automatically. With these approaches, it is possible to overcome the expert knowledge based threshold setting (Dalla Mura et al., 2011).

In the tree filtering part, there wasn't developments as in the other AP parts, as mentioned before. Next section brings more information about tree filtering.

2.3.2 Tree Filtering

Connected Operators, defined by Serra and Salembier (1993), are a filtering strategy in order to obtain flat zone regions (sets of connected pixels in an image that have uniform intensity values) by merging them. Connected operators have an outstanding performance on preserving contour properties, being capable of recognize high-level objects but also low-level filtering (Salembier and Wilkinson, 2009; Xu et al., 2016).

Filtering a tree's attributes with increasing behavior is elementary. It can be a straight process if the attribute has increasing criterion (Santana Maia et al., 2021). If the attribute's node does not respect the threshold, it is deleted. When an attribute increases, it's similar to cutting off a branch from the tree because if a node doesn't meet the threshold, none of its descendants do either (Xu et al., 2016). On the other hand, non-increasing attributes presents a complicated approach, as the descendants of a node to be removed may not have been removed, leading to oscillations in decisions (Salembier and Wilkinson, 2009; Santana Maia et al., 2021). As a workaround, there were two tree pruning and two threshold-based approaches, such as **Max Rule**, **Max Rule Direct Rule** (Salembier et al., 1998; Salembier and Wilkinson, 2009) and **Subtractive Rule** (Urbach et al., 2007).

The **Max Rule** involves pruning the branches from the leaves up to the first node that needs to be maintained. This strategy ensures that the nodes closer to the tree's root, which need to be preserved based on the non-increasing criterion, are retained while removing unnecessary branches. The **Min Rule** has the same idea as the Max Rule, but instead of pruning the leaves up to the first node to be preserved, it does the contrary, by pruning branches from the leaves up to the last node that needs to be removed (Salembier et al., 1998).

The **Direct Rule** involves directly applying the non-increasing criterion to each node in the tree and making pruning decisions based on this criterion. This strategy directly evaluates the criterion at each node without considering the relationships between nodes, allowing for a straightforward pruning process based on the criterion (Salembier et al., 1998).

Finally, the **Subtractive Rule** concerns subtracting the non-increasing criterion value of a parent node from the criterion values of its children nodes to determine the pruning

decisions. When using the subtractive rule, the pruning decisions are based on the difference in criterion values between parent and child nodes, allowing for a more nuanced approach to tree simplification with non-increasing criteria (Urbach et al., 2007).

Despite the workaround approaches to the difficulties of filtering a tree with non-increasing criterion, Xu et al. (2016) states that such rules have downsides, specially when trying to prune a tree where two objects within the same branch are the desired ones to be segmented.

2.3.3 Tree Filtering on Space of Shapes

As there was a methodological gap in this research field, Xu et al. (2016) proposed the Shape-Space. It is a derived filtering domain built upon the Image-Space. It allows for filtering operations on subtrees, leading to results that generalize existing tree-based connected operators systematically. The Figure 2.2, better illustrates how the methodology works.

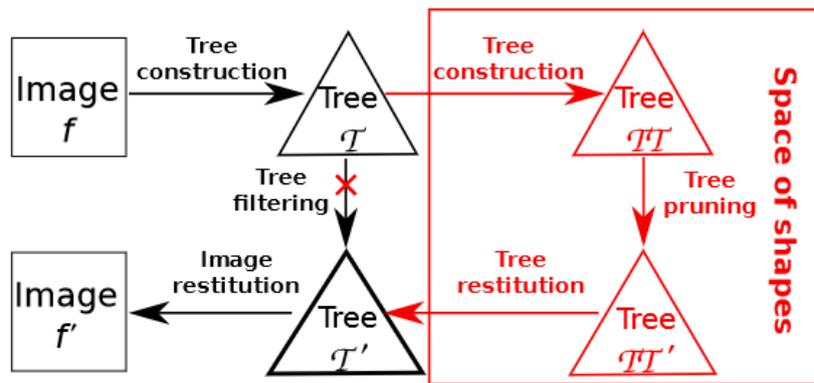


Figure 2.2: The black path represents the classical connected operators. The proposal made by Xu et al. (2016) is represented in the red and black path, which is shape-space filtering. Source: Xu et al. (2016).

It uses the connected components benefits of preserving image contours effectively during processing. By encoding the inclusion relationships between connected components, shape-space methods maintain the structural information of shapes in the image. This preservation of contours leads to more accurate shape analysis, segmentation, and feature extraction tasks. In addition, shape-space representations are robust to variations in image properties due to their invariance to contrast changes.

Shape-space techniques aim to improve shape analysis by providing a structured and organized environment for shape manipulation. By computing attributes that characterize each shape or connected component within the graph representation, shape-space enables the quantification of shape properties and facilitates shape-based operations.

The computation of Shape-Spaces involves transforming the image into a graph structure, computing shape attributes, and applying connected filters for shape manipulation. Attributes are derived to quantify shape properties such as area, gray level, or more advanced shape attributes.

All these concepts presented in the **2.3 Section** are the fundamental basis for the novelty that shall be presented in this research. From these ideas, it will be proposed the extension of Shape-Spaces by applying Attribute Filtering and PS within it, seeking to identify the strengths and weaknesses of these approaches.

2.4 Related Works

The work done by Moretti et al. (2021) which brought the statistical characterization of the SCDs leveraged a substantial potential for the development of works using remote sensing and EO products in this specific field. By then, the SCD didn't have a profiling about its shape and depth. Since the publication, this research field started utilizing remote sensing in the pre-field phase of the exploration campaigns.

Later, Moretti et al. (2022) proposed the utilization of normalized indexes from Landsat-8 images, such as Normalized Difference Vegetation Index (NDVI), Normalized Derived Built-up Index (NDBI), Soil-adjusted Vegetation Index (SAVI), DEM and costal aerosol bands. The goal was to delineate the vegetation ring around the SCD. When compared with the previous work, where it was mainly used Google Earth Images as data source, in this more recent publication presented a more structured workflow using EO data. Using the more robust approach, Lévy et al. (2023) analyzed different SCDs in distinct periods of time, aiming to present the relevant tools and workflows to identify and map potential emitting structures.

More recently, Mosquera-Rivera et al. (2024) presented a workflow using multiple methods to identify possible concentrations of SCDs. A combination of numerous datasets, such as PRISMA Hyperspectral Imagery, Landsat Level-2, Copernicus DEM and Geophysical data (Radiometric, Megnetometric and Gravimetric) were used along with multiple methodological strategies, like PCA, band ratio, water and vegetation indexes. So far, it is the most extensive work done applying remote sensing on the identification and mapping of Sub-Circular Depressions (SCD)s. The results presented were promising, since they were capable of identifying potential regions. For further works, Mosquera-Rivera et al. (2024) suggest the application of other spectral algorithms, machine learning for classification as well as the integration of morphology-based features using DEM.

Despite the recent evolution in the application of Earth Observation (EO) products in the mapping of SCDs, the contribution of the current research aims to provide one methodological approach applied to DEM in order to delineate the SCDs. It solves the gap in the literature by using applying mathematical morphology methodologies, such as PS and Attribute Profiling on DEMs. The objective is to help to map these structures over large areas, as well as give the necessary attributes to confirm whether it is a SCD or not by its slope characteristics, an expert knowledge generated in Moretti et al. (2021).

CHAPTER 3

Methodology

This chapter discusses the materials and methods used in this study. Figure 3.1 represents the workflow, providing a better visualization and understanding of the thesis' entire development. Below, each of the boxes will be addressed, with the assumptions and strategies pointed out. The boxes with a red border stand out, as they indicate the novelties presented by this work.

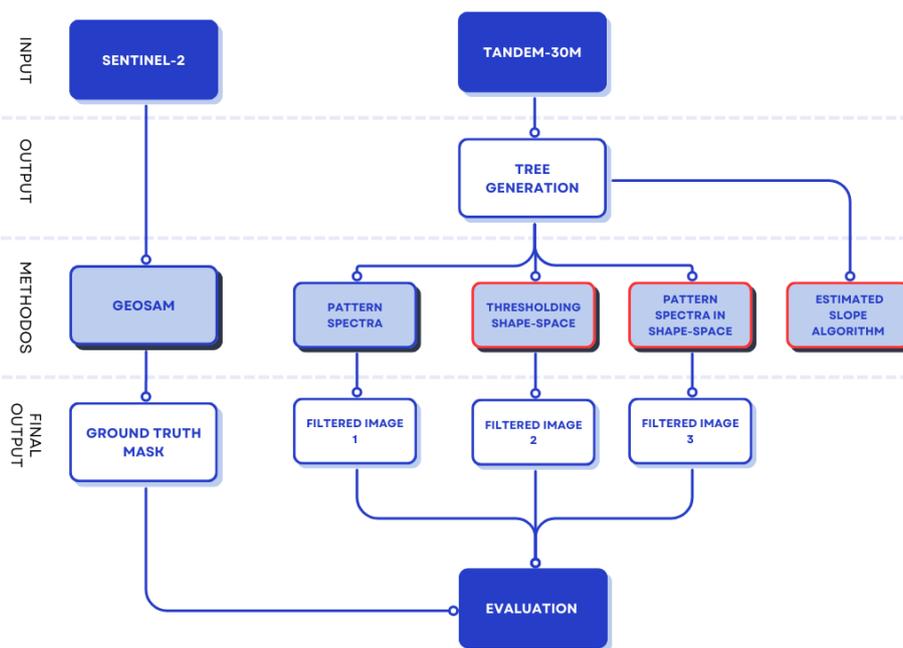


Figure 3.1: Thesis materials and methods. The boxes in red are the proposed methodologies.

3.1 Dataset Description

For all analysis the TanDEM-X 30m Edited DEM is used, with 30m resolution, provided by the German Aerospace Center (DLR). It is an edited product from the TanDEM-X Global DEM, which has 1 arc second. In fact, this product represents the surface, not the ground elevation, as described in its documentation, therefore being technically characterized as a Digital Surface Model (DSM). Yet, it's an accurate product with absolute horizontal and vertical accuracy below 10m; slopes below 20% have 2m accuracy, and above 20% have 4m, also with a 90% confidence level (Wessel, 2018).

3.2 Ground Truth Generation

As described in the previous section, the area of study was chosen due to previous studies in the area.

Despite developing the research in an area already surveyed, there aren't ground-truth masks available as a guide to compare with the outputs. Having that in mind, it was generated a ground-truth mask SAMGEO (Zhao et al., 2023), a QGIS plug-in based on Meta's Segment Anything Model (SAM) that allows the user to segment features automatically. The image used for segmentation was the RGB composition, with 10m spatial resolution, from Sentinel-2, imaged on March 21, 2024. The ground truth was generated around the areas where H₂ leakage was measured, as shown in Figure 3.2.

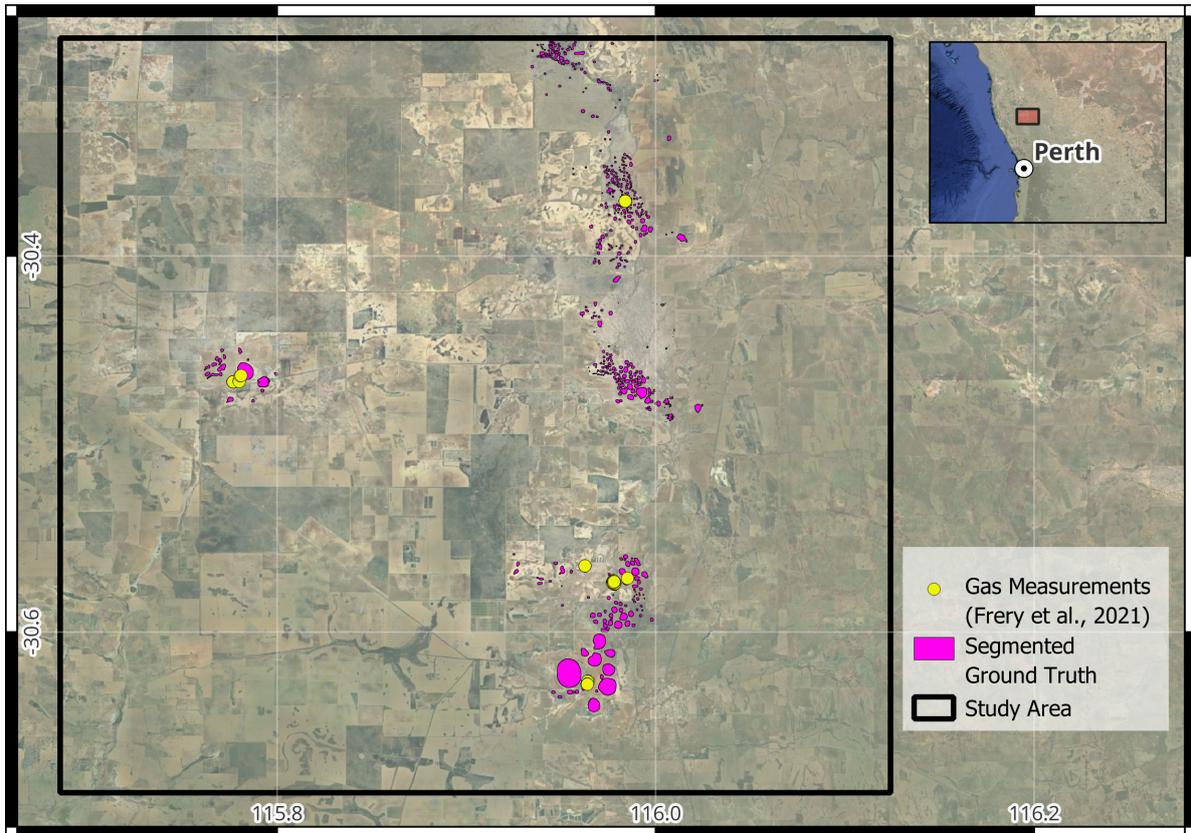


Figure 3.2: Distribution of Ground Truth measurements based on image segmentation.

3.3 Tree Generation from TanDEM 30m

The Higma library will be used to generate a min-tree from the DEM, although the same can be done with SAP. The idea of using Higma is because it has a higher control level than SAP, and also, the latter was built on the first. With its creation, it's possible to store the tree edges and altitude easily, which will be important to future processing steps. Also, the functions for characterizing, processing and iterating the tree have better compatibility with this specific library.

The tree generation with Higma is required to first generate a graph that will serve as the weights of the tree. For the generation of the graph, there are a lot of building options, but for this research, it was chosen to create trees based on implicit undirected 4 adjacency graphs of the image's shape.

Following that, the computation of the min-tree takes as input the recent generated graph and the DEM, and its generation have as output the tree and its node altitudes. From the tree output, it's possible to generate different attributes, filter the tree's node in order to perform different types of image analysis.

3.4 Tree Filtering Methodologies

3.4.1 Pattern Spectra

The implementation of Pattern Spectra is important to serve as a baseline and further be compared with other proposed methodologies, since it is a widely used approach. Based on Guiotte et al. (2020), the selected PS attributes selected were:

- **area**: the surface area on the node;
- **volume**: the volume of a node equals the area multiplied by the difference in elevation with its parent, plus the combined volumes of all its children;
- **height**: the difference in elevation between the node's parent and the lowest node in its subtree;
- **compactness**: compactness of the node's shape, calculated as 16 times the area divided by the perimeter squared. Compactness values range from 1 for compact shapes like circles to 0 for non-compact shapes.

First, a **min-tree** was generated from the DEM with the SAP library. Then, each one of the selected attribute is calculated. To obtain the thresholds aiming to highlight the structures, LaPSUs software (Avellaneda and Merciol, 2024) was utilized to select the most interesting values for each one of the attributes above (Table 3.1). Following that, 6 2D Spectrum maps were created in order to check which pair of filtered attributes performed the best. The maps are: area x compactness, area x height, area x volume, volume x height, volume x compactness and height x compactness. Finally, the evaluation part was done with IOU and FNR.

Table 3.1: Minimum and maximum values for different attributes.

Attribute	Min	Max
Area	$10^2 \times 1$	$10^3 \times 2$
Volume	$10^2 \times 4$	$10^3 \times 4$
Height	0	12
Compactness	0.05	0.6

3.4.2 Attribute Filtering within Space of Shape

One of the novelties of this work involves applying automatic thresholding within Shape-Spaces in order to reconstruct the image using Yen Threshold algorithm (Yen et al., 1995). It is a multilevel threshold algorithm that takes into consideration: the disparity between the thresholded and initial image; and the amount of bits necessary to represent the thresholded image. The aim of its utilization is to test whether this approach turns to be accurate or not when reconstructing the image using non-increasing attributes. Since the H2 seeps are mainly

circular, it is calculated the circularity index for each tree node, in order to identify which nodes fit into a seep characterization.

The steps are similar to what Xu et al. (2016) has presented. First, a tree is generated from the original DEM, followed by the removal of large area components. After, the circularity for each node is calculated and added to the tree as edges. Following that, the second tree is created based on the undirected graph for circularity values and the circularity values itself. Next, it is obtained the area, depth, volume, and height attribute values based on the second tree to perform further filtering.

Instead of filtering the tree in Shape-Spaces with an arbitrary value, let this part be done by the Yen Threshold. To implement such approach, one should apply the Yen threshold algorithm to the non-increasing criteria attributes that were obtained in the last step. Finally, reconstruct the attributes to the first tree and obtain the results.

The Yen Threshold method was chosen due to its capabilities of finding an optimal threshold value that separates the foreground from the background. In this case, it is expected that the segmentation happens on the edge between the round structures and image background.

3.4.3 Pattern Spectra Applied to Space of Shapes

The other innovation brought by this research is the addition of Pattern Spectra (PS) application inside the Space of Shapes. The idea, just like the proposal presented in the last section, is to use non-increasing criterion attributes to perform the filtering based on them, also using the Yen Threshold algorithm to select the boundaries.

With PS, it is possible to obtain the value of distinct attributes and better filter the regions. This approach allows obtaining more tree nodes that have a better representation of the structures, since only one attribute might not be the most suitable.

To apply PS within the Shape-Spaces, the same strategy as before is utilized. After obtaining the values of thresholding with Yen algorithm, apply the PS, followed by image reconstruction.

3.4.4 Theoretical Proposition of Estimated Slope As a New Attribute

Given the limitations of the previous attributes, summed to the geomorphological characterization of FC (Moretti et al., 2021) and the Higma software (Perret et al., 2019), this research proposes the Estimated Slope as a new attribute for the Pattern Spectra analysis, following then a characterization of the H2 seeps. Also, as a consequence of this novelty, the research can contribute to the development of Higma.

In studies using DEM, one of the alternatives for obtaining the slope of a given region is through the difference in height between the initial point Y_0 and the final point Y_1 divided by the horizontal distance L (Equation 3.1) (U.S. Geological Survey, 2022). The estimated slope can be calculated by having the attribute values of volume and area for each tree node, and also by obtaining the diameter of the structures.

$$\text{Slope} = \frac{Y_1 - Y_0}{L} \quad (3.1)$$

In order to implement the estimated slope algorithm, it is necessary to initially generate a tree. Then identify the center of the leaves from the average of the coordinates. Next, from the central point, calculate the diagonal of the smallest possible bounding box that surrounds the structure at each node-level, represented by the letter **L**. The minimum values of **L**, which represent the diameter of the leaves, need to be saved so that they can be used in the future for the difference between the diameter of a top node and the leaf.

To calculate the average height of a node, it is necessary to use Volume (**V**) and area (**A**), because dividing them gives the mean grayscale value $\bar{X}G$, which represents the height of a specific node. Once again, it is essential to keep the mean grayscale values of the leaf, for the same reason as the diameter.

Since **L** is the diameter of a structure, it's necessary to divide the Equation (3.2) by 2 to obtain the radius, which represents the distance between the center point and the edge. This, will guarantee that the estimated slope is in one direction only.

With these attributes, you can estimate the slope by substituting the variables into Equation (3.1). Thus, the estimated slope equation is:

$$\text{Slope} = \frac{\bar{X}G_{\text{node}} - G_{\text{min}}}{\frac{L_{\text{node}} - L_{\text{min}}}{2}} \quad (3.2)$$

For the purpose of simplification of the equation, we finally have Equation (3.3):

$$\text{Slope} = 2 \times \frac{\bar{X}G_{\text{node}} - G_{\text{min}}}{L_{\text{node}} - L_{\text{min}}} \quad (3.3)$$

Figure 3.3 brings an example for calculating the Estimated Slope, taking into consideration all the necessary variables to solve the equation. The example considers that **G3** is 15, and **Gmin** is 3, and their diagonals are respectively 10 and 1.5.

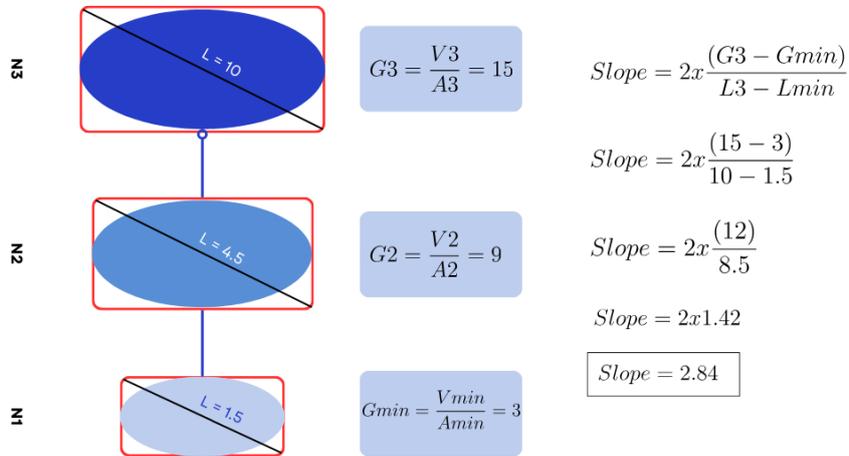


Figure 3.3: Estimated Slope conceptual example

Estimated Slope Algorithm Implementation

In the code block below, the logical implementation of the Estimated Slope algorithm:

```

1 Build the Min-Tree:
2   Construct a min-tree from the input image.
3
4 Identify Center Nodes:
5   For each leaf node in the tree:
6     Store the area (A) and volume (V) of each leaf node.
7     Compute the sum of X and Y coordinates of all pixels within the node
8     .
9     Calculate the center coordinates:
10    X_center = sumX / area
11    Y_center = sumY / area
12    centroid = X_center, Y_center
13 Find Edge Points:
14   For each node in the tree:
15     Generate a bounding box (BB) around the nodes pixels from the
16     centroid.
17     Store the minimum and maximum bounding box values for each node.
18     Calculate the diagonal (L) of the bounding box, representing the
19     diameter of the node structure.
20     Store the diagonal value that is minimum for further computations (

```

```

19     L_min).
20 Calculate Grayscale Values:
21   For each node in the tree:
22     Compute the average grayscale (G) value:
23     G = volume / area
24   Store the grayscale value that is minimum (G_min).
25
26 Estimate Slope:
27   For each node in the tree:
28     Calculate the estimated slope from each node:
29     estimated_slope = (Node Average G - G_min) / (L - L_min)
30     Divide by two to have one direction slope in percentage.

```

3.5 Evaluation Metrics

Considering the different methodologies that will be used, it is necessary to carry out the evaluation based on ground truth. In this way, Intersection Over Union (IOU) and False Positive Rate (FPR) were chosen.

The IOU, commonly used in binary classification tasks, indicates how much of the prediction intersects the ground truth (Shah, 2023). An IOU of 1 means that the prediction completely overlaps the ground truth, so the higher the index, the better the result. As these are binary predictions, the IOU is calculated as shown in Equation 3.4 :

$$IOU = \frac{TP}{TP + FN + FP} \quad (3.4)$$

where True Positive (TP) are the corrected predicted pixels; False Negative (FN) means that the classification indicates a condition that does not exist, characterizing a Type II error. FN indicates the pixels that are missed by the model. False Positive (FP) represents the pixels that are falsely identified by the model.

The FPR is a very common metric used in binary classification, and uses the same classes as IOU, as shown in Equation 3.5. Therefore, the False Positive Rate (FPR) indicates the probability that a true negative will be incorrectly classified as positive by the test. The higher the FPR, the poorer the performance of the classification algorithm.

$$FPR = \frac{FP}{FP + TN} \quad (3.5)$$

CHAPTER 4

Results

4.1 Pattern Spectra Applied to DEM

Pattern Spectra applied to DEM using arbitrary threshold values, as the baseline method, overall, as shown by the metrics in Table 4.1 and by the images in the figs. 4.2 to 4.7.

All attribute maps generated by the methodology presented low IOU values, varying between 15.5% and 20.8%, in addition to high FPR, between 84.9% and 89.8%. Negative results are highly influenced by structures that are very small, as the algorithm was not able to detect them. Furthermore, several structures present in the tree generated from the DEM were highlighted, even if they do not have rounded or ellipsoidal characteristics.

The methodology also presented an inability to filter the largest FCs in some cases. In other words, in certain maps, there was only delimitation of the medium structures, thus being another reason for the deterioration of the metrics.

To obtain these results, the values presented in Table 3.1 were used, and represented in an illustrated form in Figure 4.1. In each image, each axis represents a different attribute. The horizontal lines indicate maximum and minimum values for the Y axis, and the vertical lines represent the values for the X axis. Therefore, the filtering took into account only the tree nodes that are within the polygon generated from these limits.

Table 4.1: Metrics obtained from the PS methodological application

Attributes	IOU	FPR
Volume Compactness	0.208	0.849
Volume Height	0.189	0.869
Compactness Height	0.172	0.898
Area Compactness	0.159	0.892
Area Height	0.156	0.897
Volume Area	0.155	0.888

The attribute map with the best metrics was **Volume x Compactness**, with IOU of 20.8% and FPR of 84.9%. This map, presented in Figure 4.3, was able to highlight large and medium H2 emitting structures, but was unable to filter out the smaller ones, which have a high density

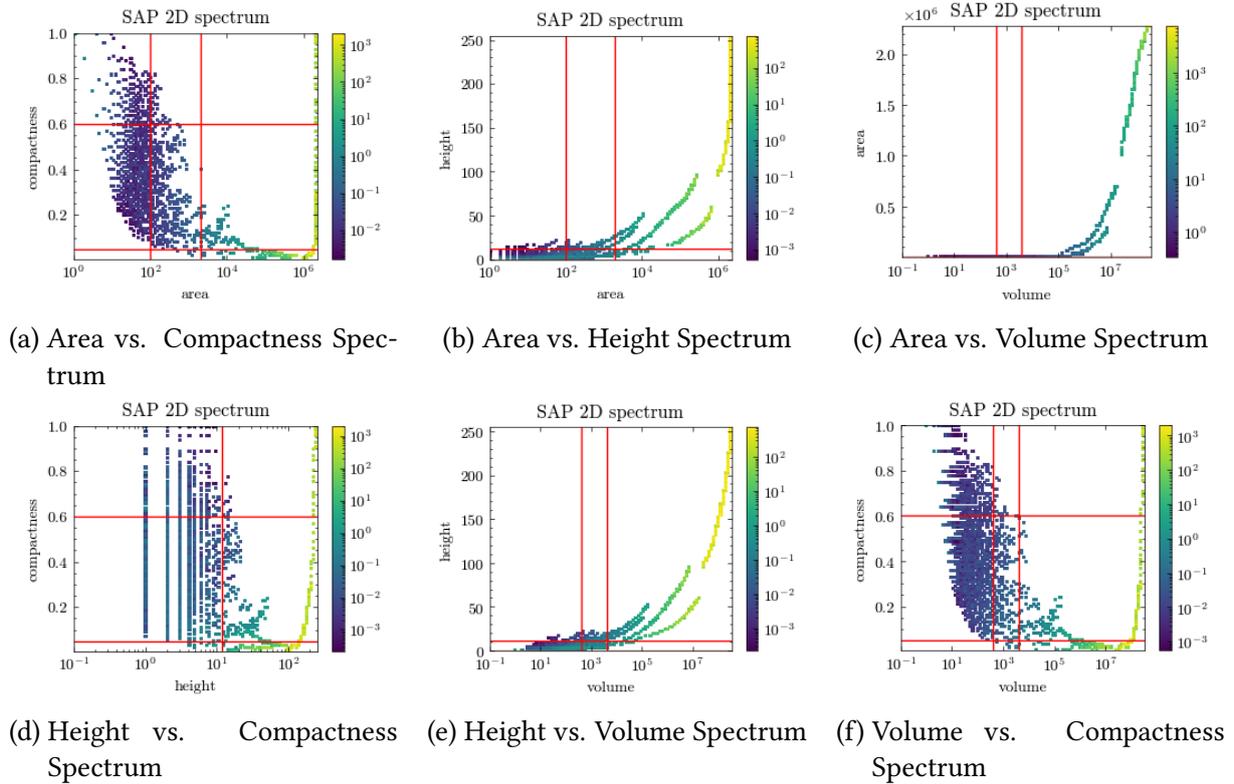


Figure 4.1: 2D Spectrum of 2 Attributes Variables

in the study area. Another point of attention is the presence of false positives with more elongated shapes, different from FCs. Despite this, it was the map that presented the greatest ability to filter FP, when compared to the others. The Volume x Height attribute map (Figure 4.2) presents a very similar pattern, but with more FP, which justifies the degradation of its metrics.

The third-best attribute map (Figure 4.4) shows a substantial difference in its visualization. Although it is still able to delineate medium, large and some small FCs, it presents a substantial amount of FP. As a consequence, it has the highest value for FPR, despite having the third-best IOU. Once again, like the previous ones, this map was not able to filter structures that are only rounded.

The other attribute maps (Figures 4.5, 4.6, 4.7) follow the pattern of filtering FP structures with non-rounded formats. Furthermore, they are unable to filter small and large structures, only medium ones, showing a greater level of inefficiency.

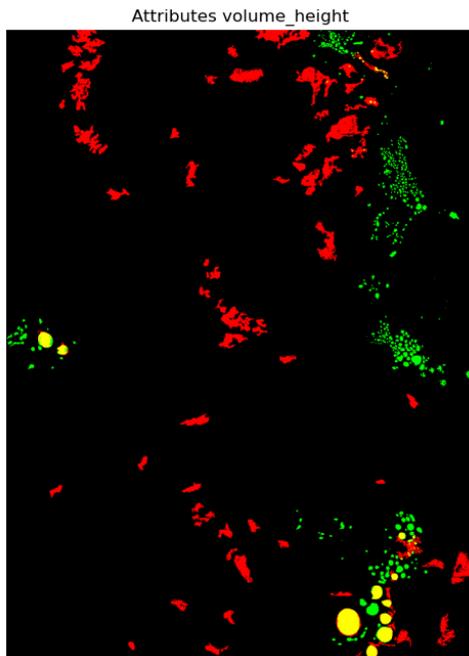


Figure 4.2: Volume x Height Attributes Map.

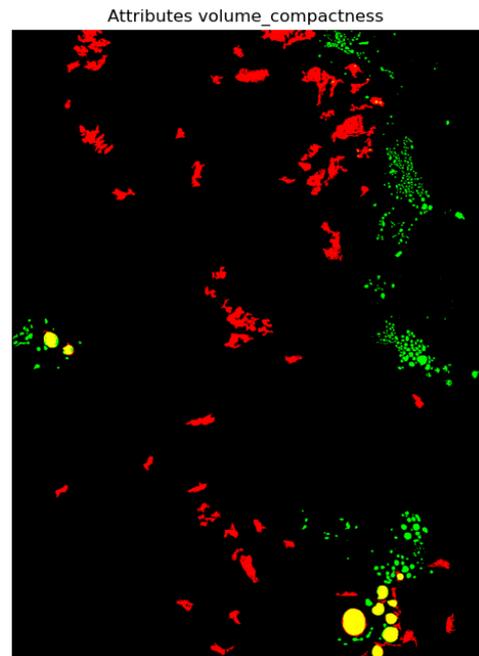


Figure 4.3: Volume x Compactness Attribute Map.

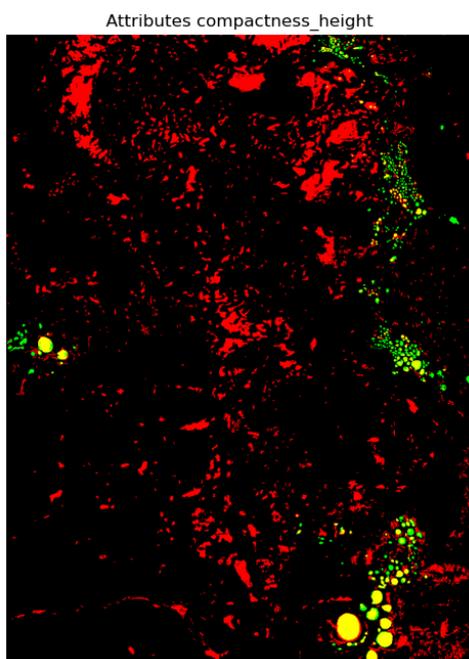


Figure 4.4: Compactness x Height Attributes Map.

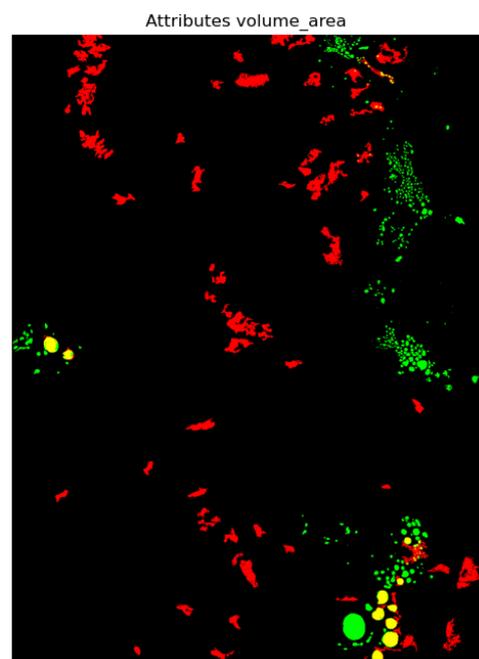


Figure 4.5: Volume x Area Attribute Map.

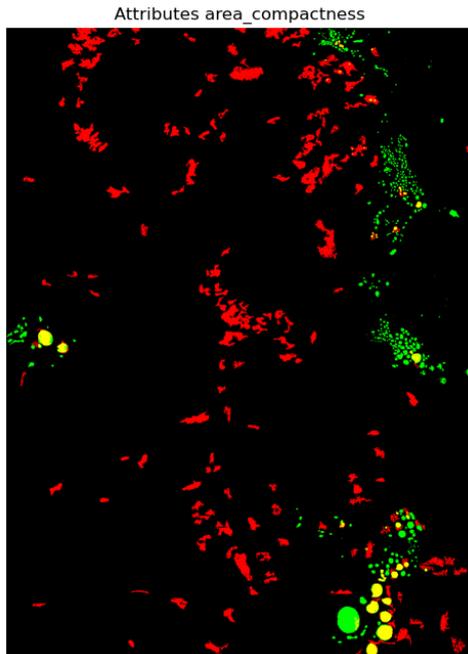


Figure 4.6: Area x Compactness Attributes Map.

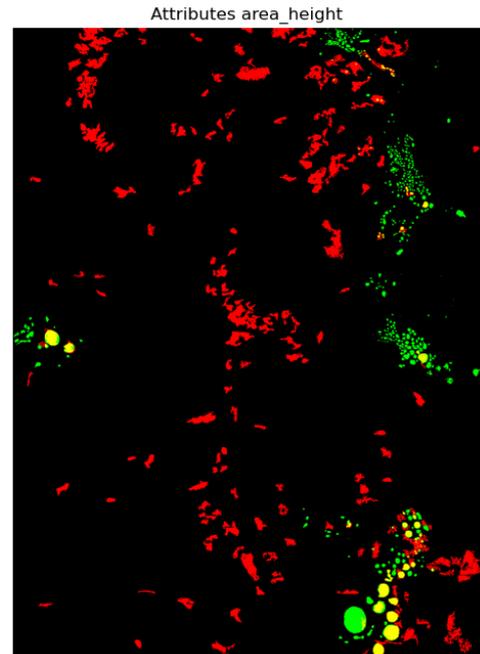


Figure 4.7: Height x Area Attribute Map.

4.2 Attribute Filtering within Space of Shape

The second methodology, Attribute Filtering in the Space of Shapes, presented an improvement when compared with the baseline methodology, which was the Pattern Spectra. Table 4.2 and figs. 4.9 to 4.12 show the results here obtained. The better IOU and reduction in FPR indicate that this proposed methodology can better delineate structures with fewer false positives. This result may be linked to some factors such as the strategy of applying circularity weight to the space of shapes tree and also the affinity of the methodology of filtering nodes in a non-increasing criteria system.

Table 4.2: Metrics for the Non-Increasing Criterion attribute filtering.

Attributes	IOU	FPR
Depth	0.113	0.939
Height	0.268	0.781
Volume	0.268	0.781
Area	0.270	0.778

In this methodological application, three of the four attributes showed better results than those previously obtained with PS. Except for the results obtained by filtering the depth attribute, the others showed IOU values between 26.8% and 27%, and a lower FPR rate. In both metrics, there was a positive variation of 10%, i.e. an improvement on the results obtained with the other methodology.

With IOU between 11% and 27%, FPR between 77.8% and 93.9%, the maps generated by the four attributes mostly filtered out rounded structures, unlike what was seen in results involving PS. However, this methodology proved incapable of delimiting the plurality of FC sizes. Some large and small structures were highlighted, but without a well-defined pattern. This analysis does not apply to the **Depth** attribute (Figure 4.9), since it delimited almost all FC, but also highlighted numerous structures that are not rounded, showing a very high FPR rate.

The histogram of nodes in the shape space tree was segmented using the Yen algorithm, marked by the red line in Figure 4.8. Looking at the histograms, we see that **Depth** has a different distribution of nodes to the other attributes. This behavior is probably related to filtering from the outer tree to the inner tree, within the space of shapes. The depth attribute computed from the tree in space of shapes has values for all nodes, which indicates the depth of that node within that tree. Differently from the other attributes, where they had part of their nodes removed due to the filtering from the first to the second tree.

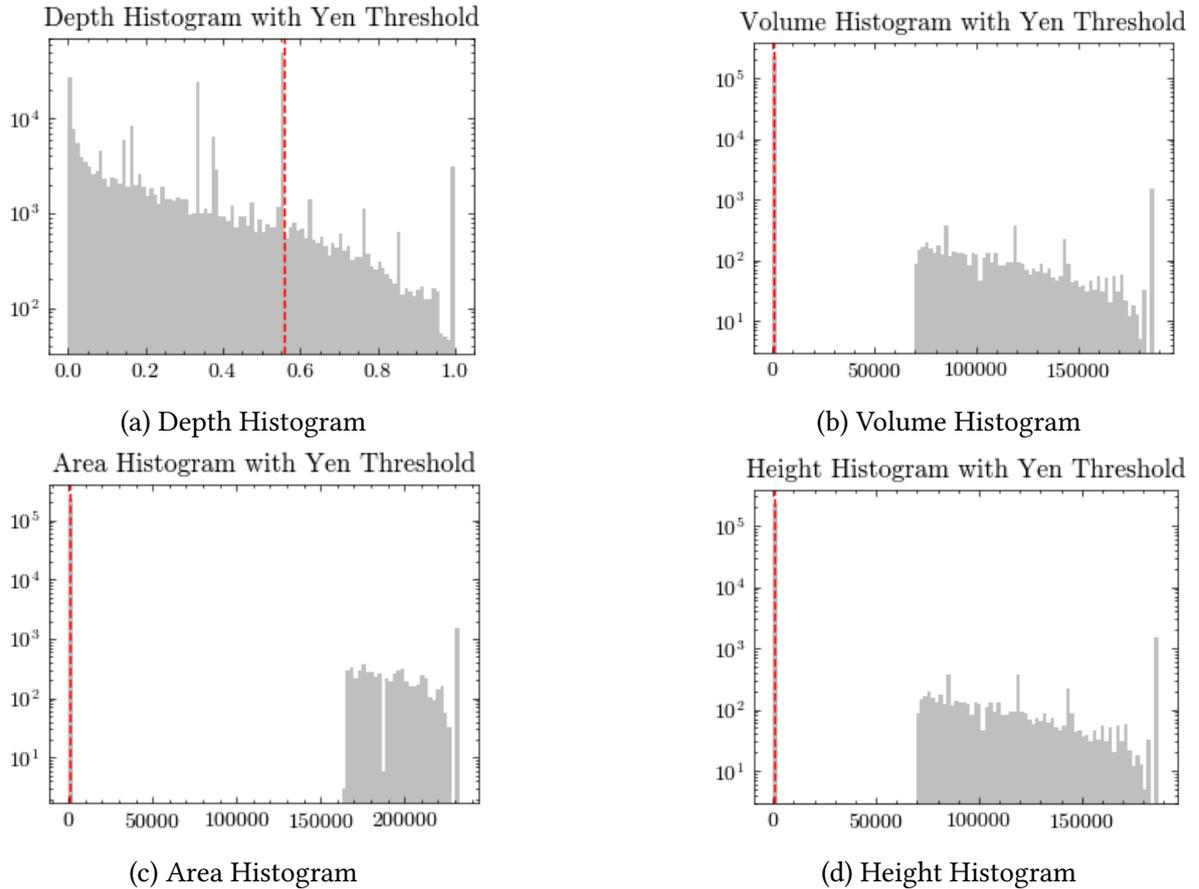


Figure 4.8: Attribute Filtering within Space of Shape Histograms

The attribute with the best metrics was **Area** (Figure 4.11), with IOU of 27% and FPR of 77.8%. When compared with the best map generated by PS, it shows an improvement of 7.2% in IOU and a reduction of 7.1% in FPR. Making the same comparison with the other attributes

that performed well in the same methodological application, **Height** and **Volume** (Figures 4.10, 4.12), it can be seen that there was a slight change in the output, but not significantly impacting the final result. The same filtering pattern for rounded structures such as FP, as well as highlighting practically the same FCs shows why the difference between IOU and FPR are 0.2% and 0.3% respectively.

In general, the results presented by filtering attributes in the space of shapes seem to be more robust to FP when compared to PS. Applying a circularity weight to the space of shapes tree helps in this regard. This prevents structures that meet the threshold but do not have a sufficient circularity index from being removed from the image.

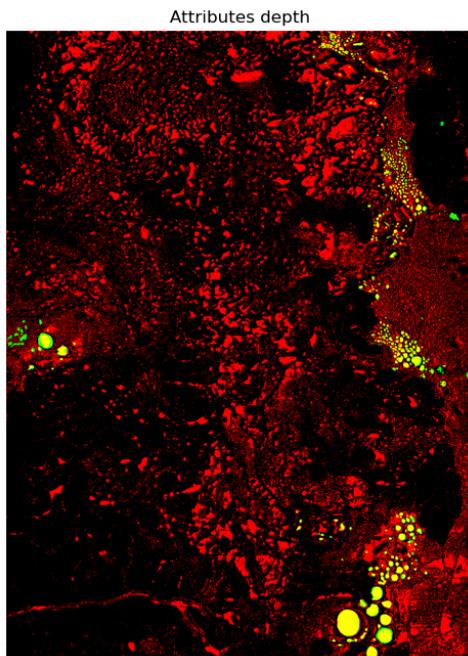


Figure 4.9: Depth Map.

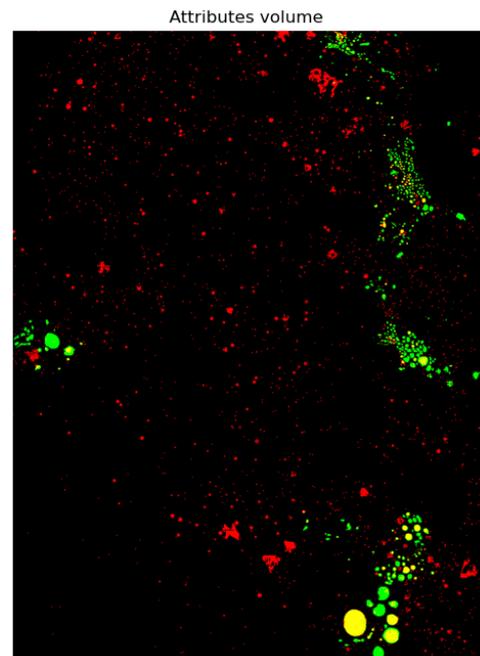


Figure 4.10: Volume Map.

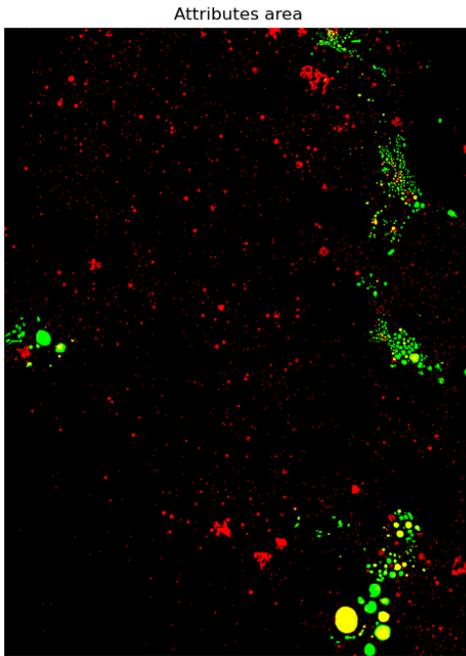


Figure 4.11: Area Map.

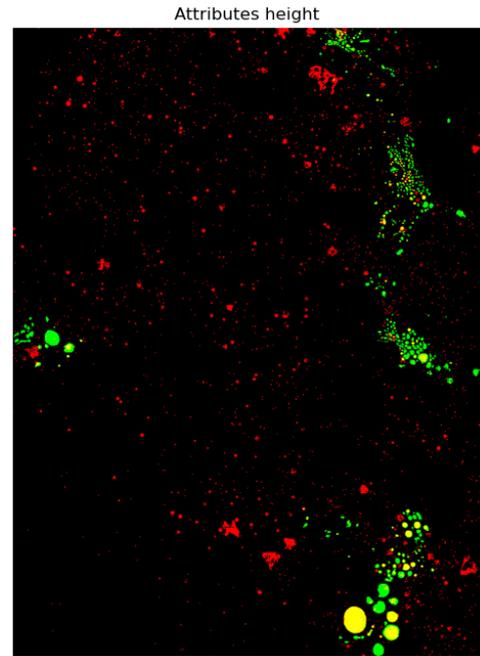


Figure 4.12: Height Map.

4.3 Pattern Spectra within Space of Shape

The last methodological approach, Pattern Spectra within Space of Shape, is the one that presents the best results so far when comparing with both PS and Attribute Filtering in Space of Shape. Table 4.3 shows how this strategy has achieved the best indicators so far, 31% and 68% respectively in IOU and FPR. Figures 4.14 to 4.19 displays the generated 6 maps.

Table 4.3: Metrics for the Pattern Spectra Applied in the Space of Shape

Attributes	IOU	FPR
Depth Volume	0.310	0.680
Depth Area	0.310	0.680
Height Area	0.280	0.781
Area Volume	0.280	0.781
Height Volume	0.280	0.781
Depth Height	0.115	0.938

Deficiencies of the previous methodologies can still be seen here, such as the lack of a pattern in the delimitation of FCs. Once again, the medium structures have not been highlighted, only the large and small ones. The structures that indicate FP also have the same rounded model, which is a consequence of the circularity index.

Another relevant fact is that filtering based on Pattern Spectra helps to reduce the percentage of FP, thus indicating that filtering an image with two attributes makes the process less susceptible to error. In the three best maps presented in the three methodologies, there was a

reduction in FPR from 84.9% in PS, to 77.8% in attribute filtering in space of shape to 68% in PS within Space of Shape. In total, a reduction of 16.9%. The combination of attributes brought better results and then the tree filtering with only one, as presented in the last subchapter.

Figure 4.13 shows the 2D spectra and how they were filtered according to the Yen threshold algorithm. From these histograms, the trees were filtered, and the results were obtained. The Depth x Volume and Depth x Area attribute maps showed the best results. The other three, Height x Area, Area x Volume and Height x Volume performed well, with IOU and FPR metrics above the other methodologies, being 28% and 78.1% respectively.

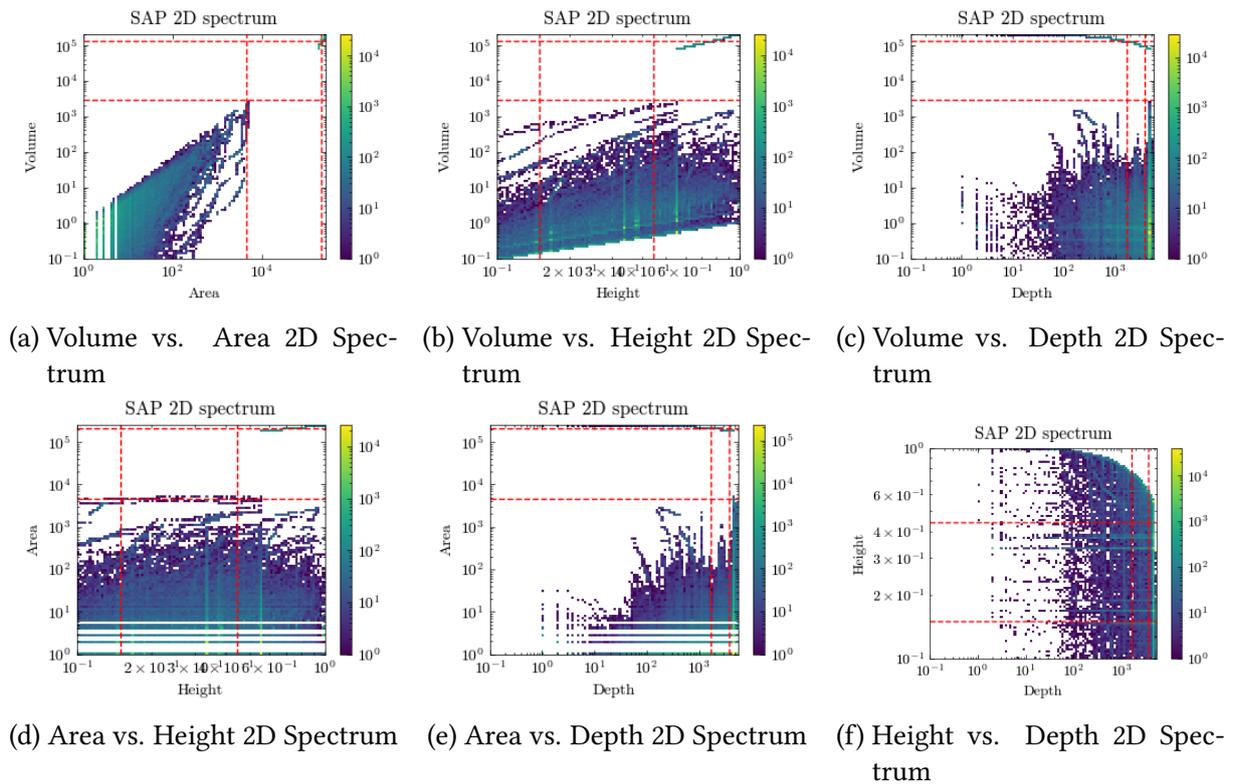


Figure 4.13: Attributes Spectrum of PS within Space of Shape.

The main difference between the two best maps and the other three was the amount of FP, since they were able to highlight similar FCs. The only attribute map that didn't perform well, following the same pattern presented in the previous methodology, was **Depth x Height**.

Thus, it can be seen that PS applied to the Space of Shapes is more robust to FP, which improves the metrics in general.

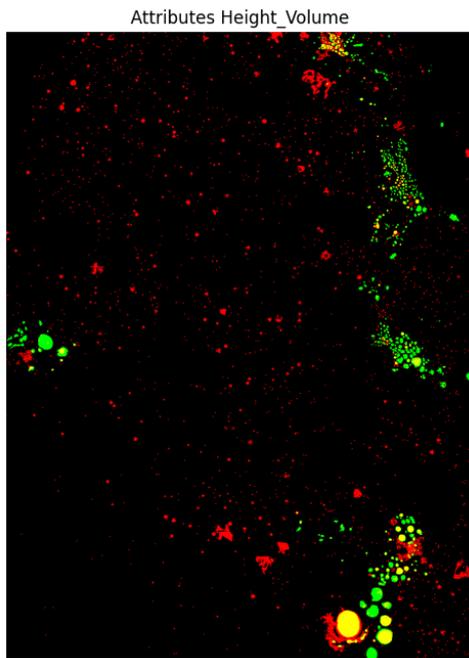


Figure 4.14: Volume x Height Attributes Map.

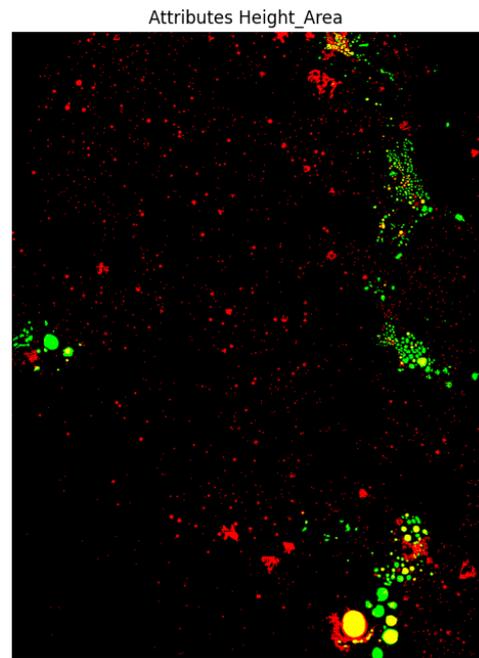


Figure 4.15: Area x Height Attribute Map.

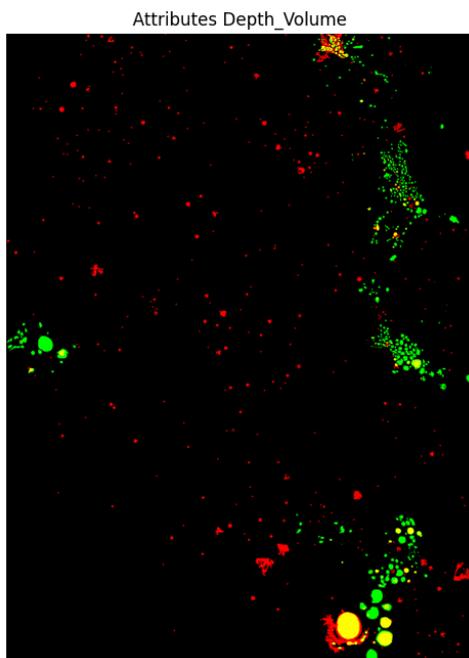


Figure 4.16: Depth x Volume Attributes Map.

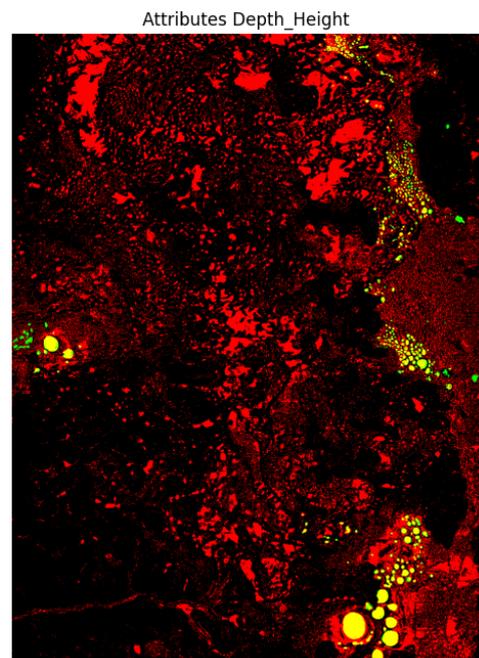


Figure 4.17: Depth x Height Attribute Map.

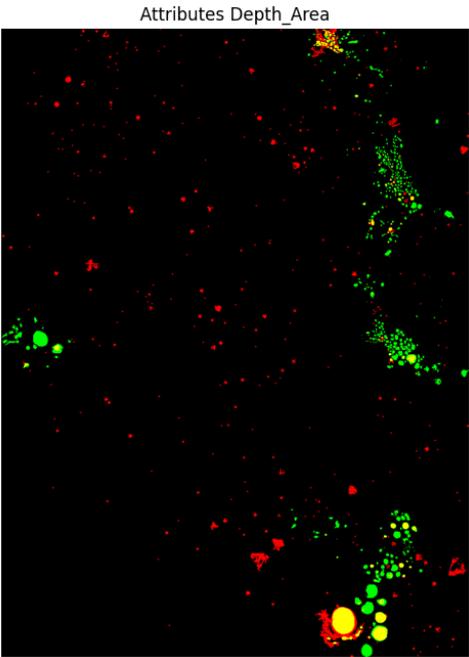


Figure 4.18: Area x Depth Attributes Map.

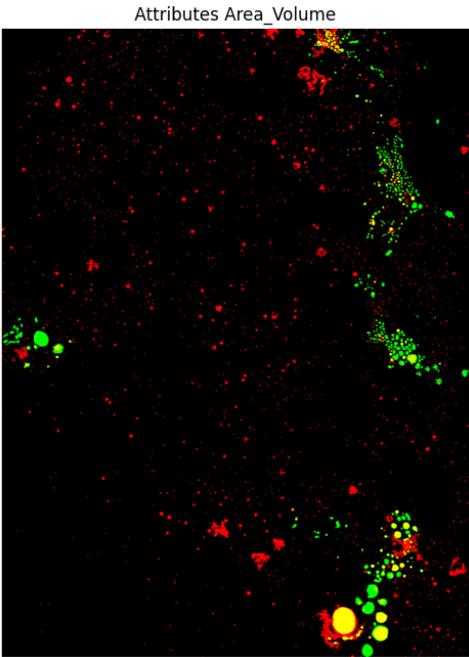


Figure 4.19: Volume x Area Attribute Map.

Discussions Conclusions

5.1 Metrics Evolution

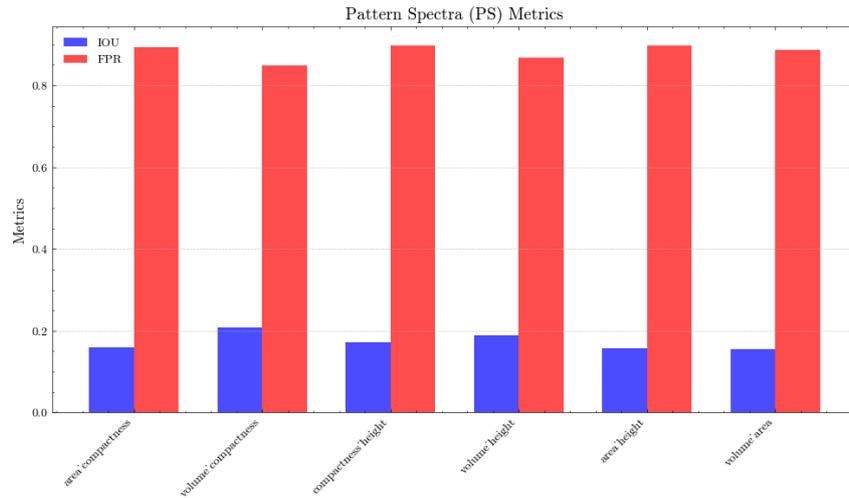
The results indicate that there were improvements in the metrics when comparing the baseline method (PS) with the two novel methods introduced: attributes filtering and PS applied within the Space of Shape (Figure 5.1).

The improvement in metrics is directly associated with a reduction in FP present in the images. The reduction of FPR is attributed to the construction of the tree in the Space of Shape based on circularity weight. The tree construction based on circularity is evident in the outcomes of the methodologies within Space of Shape. While the PS output exhibits a range of elongated structures that are neither rounded nor ellipsoidal, the other two methods predominantly filtered out nodes with rounded characteristics.

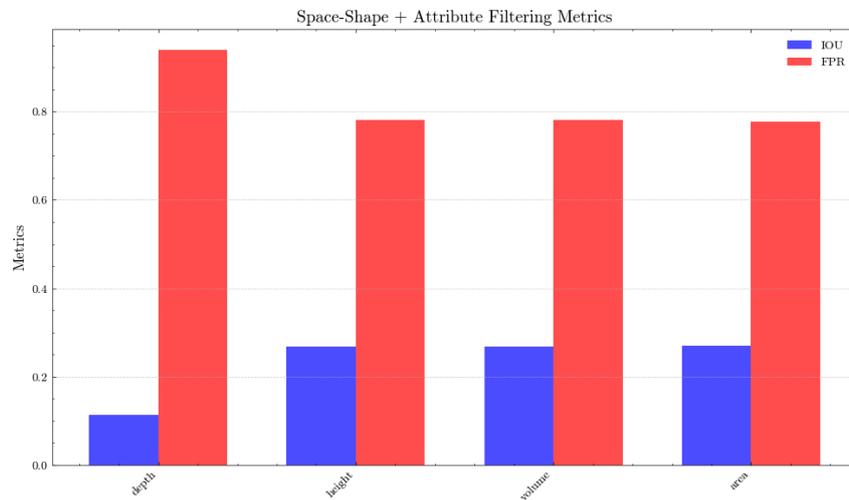
A difference in the filtering of structures was observed only in the outputs that had depth as an attribute. This attribute, utilized exclusively in the Space of Shape methodologies, produced both the most and least favorable results. This variability is likely linked to the generation of the second tree within the Space of Shape after the filtering of the first tree. As previously mentioned, the depth attribute within the Space of Shape includes all depth values, takes into consideration all node depths from root to leaves. In contrast, other attributes had some nodes removed during the transition from the first to the second tree. This behavior is illustrated in Figure 4.8.

The reason for changing from using compactness to depth from the first methodology to the last two was due to compactness returning several null values, which occurred during the calculation of circularity, resulting in divisions by zero. It was hypothesized that the depth of the nodules would be pertinent for filtering, particularly in the context of a non-increasing criteria scenario. This scenario implies that filtering is not straightforward; for instance, a parent node may not be part of the region of interest for filtering, yet its child node is.

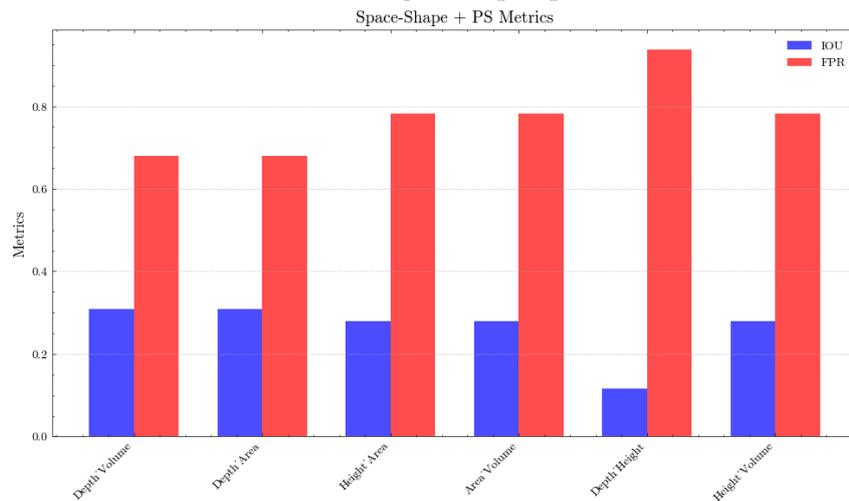
Contrary to initial expectations, this substitution of attributes appears to have influenced the results of the methodologies within the Space of Shape. As noted by Santana Maia et al. (2021), achieving accurate results through attribute filtering requires selecting the most relevant attributes for the analysis. The high variability observed suggests that depth may not have been the appropriate attribute.



(a) PS Metrics



(b) Attribute Filtering in Shape-Spaces Metrics



(c) PS in Shape-Spaces Metrics

Figure 5.1: Metrics for the three methodologies presented in this research.

5.2 Attributes Performance and Threshold Setting

5.2.1 Attributes

Among all the attributes employed in the three methodologies, none exhibited a significant standout effect. Specifically, no attribute was capable of substantially increasing the IOU rates and reducing the FPR while selectively filtering SCDs. An unexpected scenario emerged when analyzing the depth attribute, which was responsible for both the best and worst metrics.

In the Pattern Spectra application, the attribute that performed optimally was compactness, especially when combined with other attributes. In the Attribute Filtering in Shape-Spaces methodology, the attributes of area, volume, and height achieved IOU values that were relatively close to each other, indicating an improvement over the previous methodology. Lastly, in the Pattern Spectra applied to Shape-Spaces, depth combined with area and volume enhanced the results further but produced the worst outcomes when combined with height.

This high variability and imprecision in filtering the structures may be attributed to several factors. Firstly, the attributes chosen may not have been the most significant, as indicated by (Santana Maia et al., 2021) Secondly, the threshold algorithm may have failed to optimally delineate the regions in the histogram.

To address the issue of attribute selection, future research should evaluate the use of alternative attributes for image filtering. Another promising strategy could involve the use of the attribute proposed in this work, Estimated Slope. This attribute considers the calculation of the slope of the structures, which are known to have a slope up to 8% (Moretti et al., 2021). The integration of Estimated Slope, along with new attributes, may facilitate more precise delimitation of SDGs, thus demonstrating its effectiveness.

5.2.2 Threshold

Another factor contributing to the inaccurate delineation of structures might be the use of the Yen threshold algorithm. Threshold selection is crucial as it directly influences performance output (Aptoula et al., 2016). The algorithm was selected for its ability to separate the histogram such that the background is distinguished from the foreground. It offers an automatic method for value delimitation, enhancing reproducibility compared to manual methods.

The Yen algorithm is not ideal for histogram separation for this specific case. Yet, it surpasses manual selection methods. To achieve better delimitation of the SDGs and consequently improve IOU and reduce FPR, it is necessary to explore other threshold strategies.

In the last decade, numerous authors have proposed automatic threshold selection for Attribute Profiling, which can be applied to Attribute Filtering and Pattern Spectra. One such approach is the utilization of GCFs, leveraging tree construction to identify meaningful thresholds (Cavallaro et al., 2017). Another possible approach is the use of machine learning algorithms to determine the most significant values for threshold.

5.3 Limitations

In summary, the methodologies presented here have limited capacity to delimit SCDs, which is the thematic focus of this work. There has been a notable improvement in the proposed methodologies, especially with regard to the reduction in FPs, indicating better robustness. This increase in robustness can be interpreted as a positive point of the methodological proposals of this research.

The limitations are found especially when it comes to delimiting SCDs with medium sizes, since very small or very large structures can be highlighted. This makes it difficult to create a workflow capable of identifying and mapping structures in large regions.

For future work, updating the parameters identified here is necessary and important if the limitations are to be reduced as much as possible. In addition, the changes will help to understand whether the application of these methodologies in the identification and mapping of SCDs is relevant or not.

5.4 Final Thoughts

Energy transition is an ongoing global issue that is currently being tackled to ensure sustainable development. The emergence of H₂ as a non-polluting energy can play a key role in this transition. The geological hydrogen exploration industry is taking its first steps, and understanding the exudation structures and all the geology involved is a fundamental step for the best development of operations.

In this way, the SCD served as a thematic basis for proposing and testing methodologies for the theoretical basis, which is morphological mathematics. Using methodologies such as Pattern Spectra, Attribute Filtering and Shape-Space, the aim was to identify structures from trees generated from a DEM, test and propose attributes, and check the behavior of Yen as a threshold algorithm.

More specifically, regarding the methodological novelties presented here, the idea is to combine strategies that are already well-established in the field of morphological mathematics, and to create an extension by applying them to the Space of Shape. This emerges as a solution for filtering trees that are characterized by non-increasing criteria, which makes filtering trees more complicated.

The application of the methodologies to the thematic base presented a few limitations, such as the failure to delimit medium-sized structures. This behavior was directly reflected in the low IOU and FPR values. Despite this drawback, there was a substantial increase in the metrics when compared to the PS methodology, used as a baseline. This increase in metrics have a relationship with the decrease in FP, indicating that applying PS and Attribute Filtering within Shape-Spaces can increase the robustness of the filtering.

From the methodological development, it became clear that using the Yen algorithm is not an interesting choice for finding the optimal threshold values. Nevertheless, it is a better alternative than manually selecting the values. Also, the introduction of the Estimated Slope attribute, although not yet algorithmically implemented, can potentially enhance the structure detection by incorporating expert knowledge. Future work focusing on algorithmic

development and testing this attribute could validate its effectiveness and relevance.

The use of freely available 30-meter resolution DEM data underscores the scalability and practicality of these methodologies. They can be adapted and applied to different regions and datasets, making them versatile tools for geological hydrogen exploration.

5.5 Future Works

In order to continue the methodological development using Shape-Spaces, it is necessary to carry out the following in future work:

- Apply more advanced methods for selecting values from histograms. Having a well-optimized algorithm is fundamental to obtaining good results;
- Try to use other attributes, since these are also directly related to good filtering results. The attributes presented here showed great variability, indicating that they may not be ideal.
- Algorithmically develop, test and evaluate the proposed Estimated Slope attribute. This could have a positive impact on the mapping of vast regions.
- Apply the methodologies proposed in Shape-Spaces to other areas of image processing and classification in order to validate and understand whether they can really improve the robustness of the final results.

Bibliography

- Aptoula, E., Dalla Mura, M., and Lefevre, S. (2016). Vector Attribute Profiles for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3208–3220. → [p14], [p37]
- Avellaneda, T. and Merciol, F. (2024). Lapsus (landmark pattern-spectrum understanding). OBELIX. [Computer Software], version: alpha. Not published yet. → [p20]
- Ballester, C., Caselles, V., and Monasse, P. (2003). The tree of shapes of an image. *ESAIM: Control, Optimisation and Calculus of Variations*, 9:1–18. → [p12]
- Bhangale, U., Durbha, S. S., King, R. L., Younan, N. H., and Vatsavai, R. (2017). High performance GPU computing based approaches for oil spill detection from multi-temporal remote sensing data. *Remote Sensing of Environment*, 202:28–44. → [p13]
- Bhardwaj, K., Patra, S., and Bruzzone, L. (2019). Threshold-Free Attribute Profile for Classification of Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):7731–7742. → [p13]
- Boldt, M., Thiele, A., Schulz, K., and Hinz, S. (2014). SAR Image Segmentation Using Morphological Attribute Profiles. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-3:39–44. → [p12]
- Bosilj, P., Aptoula, E., Lefèvre, S., and Kijak, E. (2016a). Retrieval of Remote Sensing Images with Pattern Spectra Descriptors. *ISPRS International Journal of Geo-Information*, 5(12):228. → [p12], [p13]
- Bosilj, P., Damodaran, B. B., Aptoula, E., Mura, M. D., and Lefèvre, S. (2017). Attribute Profiles from Partitioning Trees. pages 381–392. MAG ID: 2606115022. → [p11], [p12], [p49]
- Bosilj, P., Kijak, E., and Lefèvre, S. (2018). Partition and Inclusion Hierarchies of Images: A Comprehensive Survey. *Journal of Imaging*, 4(2):33. → [p11], [p12]
- Bosilj, P., Lefèvre, S., and Kijak, E. (2013). Hierarchical Image Representation Simplification Driven by Region Complexity. In Petrosino, A., editor, *Image Analysis and Processing – ICIAP 2013*, Lecture Notes in Computer Science, pages 562–571, Berlin, Heidelberg. Springer. → [p11]
- Bosilj, P., Wilkinson, M. H. F., Kijak, E., and Lefèvre, S. (2016b). Local 2D Pattern Spectra as Connected Region Descriptors. *Mathematical Morphology - Theory and Applications*, 1(1). Publisher: De Gruyter Open Access. → [p12], [p13]

- Carrillo Ramirez, A., Gonzalez Penagos, F., Rodriguez, G., and Moretti, I. (2023). Natural H₂ Emissions in Colombian Ophiolites: First Findings. *Geosciences*, 13(12):358. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute. → [p3], [p9], [p10]
- Cavallaro, G., Dalla Mura, M., Benediktsson, J. A., and Bruzzone, L. (2015). Extended Self-Dual Attribute Profiles for the Classification of Hyperspectral Images. *IEEE Geoscience and Remote Sensing Letters*, 12(8):1690–1694. → [p12]
- Cavallaro, G., Falco, N., Dalla Mura, M., and Benediktsson, J. A. (2017). Automatic Attribute Profiles. *IEEE Transactions on Image Processing*, 26(4):1859–1872. → [p14], [p37]
- Dalla Mura, M., Benediktsson, J. A., and Bruzzone, L. (2009). Modeling structural information for building extraction with morphological attribute filters. page 747703. → [p13]
- Dalla Mura, M., Benediktsson, J. A., Waske, B., and Bruzzone, L. (2010). Morphological Attribute Profiles for the Analysis of Very High Resolution Images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3747–3762. → [p13]
- Dalla Mura, M., Villa, A., Benediktsson, J. A., Chanussot, J., and Bruzzone, L. (2011). Classification of Hyperspectral Images by Using Extended Morphological Attribute Profiles and Independent Component Analysis. *IEEE Geoscience and Remote Sensing Letters*, 8(3):542–546. → [p14]
- Das, A., Bhardwaj, K., and Patra, S. (2018). Morphological complexity profile for the analysis of hyperspectral images. In *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, pages 1–6, Dhanbad. IEEE. → [p13]
- Diallo, A., Cissé, C. S. T., Lemay, J., and Brière, D. J. (2022). La découverte de l’hydrogène naturel par Hydroma, un « Game Changer » pour la transition énergétique. *Annales des Mines - Réalités industrielles*, vembre 2022(4):154–160. Publisher: Institut Mines-Télécom. → [p9], [p10]
- Etiopie, G. (2023). Massive release of natural hydrogen from a geological seep (Chimaera, Turkey): Gas advection as a proxy of subsurface gas migration and pressurised accumulations. *International Journal of Hydrogen Energy*, 48(25):9172–9184. → [p2], [p9]
- Floristean, A. (2020). Hydrogen Future. *Renewable Matter*, 34. → [p1]
- Frery, E., Langhi, L., Maison, M., and Moretti, I. (2021). Natural hydrogen seeps identified in the North Perth Basin, Western Australia. *International Journal of Hydrogen Energy*, 46(61):31158–31173. Publisher: Pergamon. → [p2], [p5], [p6], [p9], [p10], [p49]
- Ghamisi, P., Benediktsson, J. A., and Sveinsson, J. R. (2014). Automatic Spectral–Spatial Classification Framework Based on Attribute Profiles and Supervised Feature Extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5771–5782. → [p12], [p14]
- Gillies, S. et al. (2013). Rasterio: geospatial raster i/o for Python programmers. [Computer Software]. → [pv]

-
- Guiotte, F., Etaix, G., Lefèvre, S., and Corpetti, T. (2020). Interactive digital terrain model analysis in attribute space. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020:1203–1209. Conference Name: XXIV ISPRS Congress, Commission II (Volume XLIII-B2-2020) - 2020 edition Publisher: Copernicus GmbH. → [p12], [p13], [p20]
- Géraud, T., Carlinet, E., Crozet, S., and Najman, L. (2013). A quasi-linear algorithm to compute the tree of shapes of nd images. volume 7883, pages 98–110. Springer Berlin Heidelberg. → [p12]
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Rio, J. F., Wiebe, M., Peterson, P., Gerard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362. → [pv]
- Heijmans, H. J. A. M. (1999). Connected Morphological Operators for Binary Images. *Computer Vision and Image Understanding*, 73(1):99–120. → [p11]
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95. → [pv]
- Incer-Valverde, J., Korayem, A., Tsatsaronis, G., and Morosuk, T. (2023). “Colors” of hydrogen: Definitions and carbon intensity. *Energy Conversion and Management*, 291:117294. → [p1]
- Johnsgard, S. (1988). *Fracture pattern of north-central Kansas, hydrogen soil gas anomalies over the Midcontinent Rift System*. PhD thesis, University of Kansas. → [p10]
- Jones, R. (1997). Component trees for image filtering and segmentation. In Coyle, E., editor, *Proceedings of IEEE Workshop on Nonlinear Signal and Image Processing (NISP)*, Mackinac Island. → [p12]
- Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, and Leblanc, F. (2020). geopandas/geopandas: v0.8.1. [Computer Software], version: 0.8.1. → [pv]
- Larin, N., Zgonnik, V., Rodina, S., Deville, E., Prinzhofer, A., and Larin, V. N. (2015). Natural Molecular Hydrogen Seepage Associated with Surficial, Rounded Depressions on the European Craton in Russia. *Natural Resources Research*, 24(3):369–383. → [p2], [p9], [p10]

- Lefeuvre, N., Truche, L., Donzé, F.-V., Ducoux, M., Barré, G., Fakoury, R.-A., Calasou, S., and Gaucher, E. C. (2021). Native H₂ Exploration in the Western Pyrenean Foothills. *Geochemistry, Geophysics, Geosystems*, 22(8):e2021GC009917. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021GC009917](https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021GC009917). → [p9]
- Lefèvre, S. (2009). Beyond morphological size distribution. *Journal of Electronic Imaging*, 18(1):013010. → [p13]
- Lévy, D., Roche, V., Pasquet, G., Combaudon, V., Geymond, U., Loiseau, K., and Moretti, I. (2023). Natural H₂ exploration: tools and workflows to characterize a play. *Science and Technology for Energy Transition*, 78:27. Publisher: IFP Énergies nouvelles (IFPEN), Commissariat à l'Énergie Atomique et aux Energies Alternatives (CEA). → [p2], [p3], [p6], [p9], [p10], [p16]
- Mahmood, Z., Thoonen, G., and Scheunders, P. (2012). Automatic threshold selection for morphological attribute profiles. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 4946–4949, Munich, Germany. IEEE. → [p14]
- Maragos, P. (1989). Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):701–716. → [p12]
- Merciol, F., Chapel, L., and Lefèvre, S. (2014). Hyperspectral image representation through alpha-trees. pages 37–40. → [p12]
- Mirmahboub, B., Moré, J., Youssefi, D., Giros, A., Merciol, F., and Lefèvre, S. (2021). Fast Pattern Spectra Using Tree Representation of the Image for Patch Retrieval. In Lindblad, J., Malmberg, F., and Sladoje, N., editors, *Discrete Geometry and Mathematical Morphology*, volume 12708, pages 107–119. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science. → [p12], [p13]
- Moretti, I., Brouilly, E., Loiseau, K., Prinzhofer, A., and Deville, E. (2021). Hydrogen Emanations in Intracratonic Areas: New Guide Lines for Early Exploration Basin Screening. *Geosciences*, 11(3):145. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute. → [p3], [p4], [p9], [p10], [p16], [p21], [p37]
- Moretti, I., Geymond, U., Pasquet, G., Aimar, L., and Rabaute, A. (2022). Natural hydrogen emanations in Namibia: Field acquisition and vegetation indexes from multispectral satellite image analysis. *International Journal of Hydrogen Energy*, 47(84):35588–35607. → [p2], [p3], [p9], [p16]
- Moretti, I. and Webber, M. E. (2020). Natural hydrogen: a geological curiosity or the primary energy source for a low-carbon future? 34. → [p1], [p2]
- Mory, A., Haig, J., D., W., McLoughlin, S., and Hocking, R. M. (2005). *Geology of the northern Perth Basin, Western Australia — a field guide*. → [p5], [p6], [p49]

-
- Mosquera-Rivera, J. E., Jiménez-Vergara, J. M., Vargas-Jiménez, Alberto, C., Philip, B., and Morales, H. (2024). Preliminary Remote Spatial Analysis of Fairy Circles: an Approximation of Hypersectral and Geophysical Data from Hydrogen Seeps. *First Break*, 42(6):65–78. Publisher: European Association of Geoscientists & Engineers Type: Journal Article. → [p3], [p16]
- Ozkan, S., Ates, T., Tola, E., Soysal, M., and Esen, E. (2014). Performance Analysis of State-of-the-Art Representation Methods for Geographical Image Retrieval and Categorization. *IEEE Geoscience and Remote Sensing Letters*, 11(11):1996–2000. → [p13]
- Perret, B., Chierchia, G., Cousty, J., F. Guimarães, S. J., Kenmochi, Y., and Najman, L. (2019). Higr: Hierarchical Graph Analysis. *SoftwareX*, 10:100335. → [pv], [p21]
- Perret, B., Lefevre, S., Collet, C., and Slezak, (2012). Hyperconnections and Hierarchical Representations for Grayscale and Multiband Image Processing. *IEEE Transactions on Image Processing*, 21(1):14–27. → [p11]
- Prinzhofer, A., Moretti, I., Françolin, J., Pacheco, C., D’Agostino, A., Werly, J., and Rupin, F. (2019). Natural hydrogen continuous emission from sedimentary basins: The example of a Brazilian H₂-emitting structure. *International Journal of Hydrogen Energy*, 44(12):5676–5685. → [p2], [p9], [p10]
- Prinzhofer, A., Tahara Cissé, C. S., and Diallo, A. B. (2018). Discovery of a large accumulation of natural hydrogen in Bourakebougou (Mali). *International Journal of Hydrogen Energy*, 43(42):19315–19326. → [p10]
- QGIS Development Team (2009). *QGIS Geographic Information System*. Open Source Geospatial Foundation. → [pv]
- Ronse, C. (2014). Ordering Partial Partitions for Image Segmentation and Filtering: Merging, Creating and Inflating Blocks. *Journal of Mathematical Imaging and Vision*, 49(1):202–233. → [p11]
- Salembier, P. and Garrido, L. (2000). Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *Image Processing, IEEE Transactions on*, 9:561–576. → [p12]
- Salembier, P., Oliveras, A., and Garrido, L. (1998). Antiextensive connected operators for image and sequence processing. *IEEE Transactions on Image Processing*, 7(4):555–570. → [p11], [p14]
- Salembier, P. and Serra, J. (1995). Flat zones filtering, connected operators, and filters by reconstruction. *IEEE Transactions on Image Processing*, 4(8):1153–1160. → [p11]
- Salembier, P. and Wilkinson, M. H. (2009). Connected operators. *IEEE Signal Processing Magazine*, 26(6):136–157. → [p14]

- Santana Maia, D., Pham, M.-T., Aptoula, E., Guiotte, F., and Lefevre, S. (2021). Classification of Remote Sensing Data With Morphological Attribute Profiles: A decade of advances. *IEEE Geoscience and Remote Sensing Magazine*, 9(3):43–71. → [pv], [p3], [p13], [p14], [p35], [p37]
- Serra, J. (1998). Connectivity on Complete Lattices. *Journal of Mathematical Imaging and Vision*, 9(3):231–251. → [p11]
- Serra, J. C. and Salembier, P. (1993). Connected operators and pyramids. In *Image Algebra and Morphological Image Processing IV*, volume 2030, pages 65–76. SPIE. → [p11], [p14]
- Shah, D. (2023). Intersection over Union (IoU): Definition, Calculation, Code. Accessed on May 9th, 2024. <https://www.usgs.gov/educational-resources/determine-percent-slope-and-angle-slope>. → [p24]
- Smith, N. J. P., Shepherd, T. J., Styles, M. T., and Williams, G. M. (2005). Hydrogen exploration: a review of global hydrogen accumulations and implications for prospective areas in NW Europe. *Geological Society, London, Petroleum Geology Conference Series*, 6(1):349–358. → [p9]
- Sukhanova, N. I., Trofimov, S. Y., Polyanskaya, L. M., Larin, N. V., and Larin, V. N. (2013). Changes in the humus status and the structure of the microbial biomass in hydrogen exhalation places. *Eurasian Soil Science*, 46(2):135–144. → [p10]
- Tushabe, F. and Wilkinson, M. H. F. (2008). Content-Based Image Retrieval Using Combined 2D Attribute Pattern Spectra. In *Advances in Multilingual and Multimodal Information Retrieval*, volume 5152, pages 554–561. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science. → [p12], [p13]
- Urbach, E. R., Roerdink, J. B., and Wilkinson, M. H. (2007). Connected Shape-Size Pattern Spectra for Rotation and Scale-Invariant Classification of Gray-Scale Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):272–285. → [p12], [p13], [p14], [p15]
- U.S. Geological Survey (2022). To Determine Percent of Slope and Angle of Slope. Accessed on April 16th, 2024. <https://www.usgs.gov/educational-resources/determine-percent-slope-and-angle-slope>. → [p21]
- Vacquand, C. (2011). Genesis and mobility of natural hydrogen: energy source or storable energy carrier? Technical report, France. FRNC-TH-8654 INIS Reference Number: 44098391. → [p10]
- Wessel, B. (2018). Tandem-x ground segment – dem products specification document. Technical report, EOC, DLR, Oberpfaffenhofen, Germany. Public Document TD-GS-PS-0021, Issue 3.2. → [p18]
- Willige, A. (2022). The colors of hydrogen: Expanding ways of decarbonization. Accessed on December 16th, 2023. <https://spectra.mhi.com/the-colors-of-hydrogen-expanding-ways-of-decarbonization>. → [p2]

- Xu, Y., Géraud, T., and Najman, L. (2016). Connected filtering on tree-based shape-spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6):1126–1140. → [p14], [p15], [p21], [p49]
- Yen, J.-C., Chang, F.-J., and Chang, S. (1995). A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing*, 4(3):370–378. → [p20]
- Zgonnik, V. (2020). The occurrence and geoscience of natural hydrogen: A comprehensive review. *Earth-Science Reviews*, 203:103140. → [p9], [p10]
- Zhao, Z., Fan, C., and Liu, L. (2023). Geo SAM: A QGIS plugin using Segment Anything Model (SAM) to accelerate geospatial image segmentation. → [p18]

List of Figures

1.1	The main generation sources of H ₂ and their classification: green hydrogen, generated from electrolysis using renewable energy and zero CO ₂ emissions; blue hydrogen, generated from natural gas by reforming it, leading to CO ₂ emission; white hydrogen, naturally generated; pink H ₂ , produced in nuclear power-plants by electrolysis and without CO ₂ emissions. Source: Author, 2024.	1
1.2	Example of studied fairy circles in several countries. A) Australia; B) Russia; C) Namibia; D) Brazil. All images at a scale of 1:50.000. Satellite source: Google Earth.	2
1.3	Location of the Study Area and H ₂ Measurements done by Frery et al. (2021) .	5
1.4	North Perth Basin, modified from Mory et al. (2005)	6
2.1	Differences from Partition and Inclusion Trees. Source: Bosilj et al. (2017). . .	11
2.2	The black path represents the classical connected operators. The proposal made by Xu et al. (2016) is represented in the red and black path, which is shape-space filtering. Source: Xu et al. (2016).	15
3.1	Thesis materials and methods. The boxes in red are the proposed methodologies.	17
3.2	Distribution of Ground Truth measurements based on image segmentation. .	19
3.3	Estimated Slope conceptual example	23
4.1	2D Spectrum of 2 Attributes Variables	26
4.2	Volume x Height Attributes Map.	27
4.3	Volume x Compactness Attribute Map.	27
4.4	Compactness x Height Attributes Map.	27
4.5	Volume x Area Attribute Map.	27
4.6	Area x Compactness Attributes Map.	28
4.7	Height x Area Attribute Map.	28
4.8	Attribute Filtering within Space of Shape Histograms	29
4.9	Depth Map.	30
4.10	Volume Map.	30
4.11	Area Map.	31
4.12	Height Map.	31
4.13	Attributes Spectrum of PS within Space of Shape.	32
4.14	Volume x Height Attributes Map.	33
4.15	Area x Height Attribute Map.	33

List of Figures

4.16 Depth x Volume Attributes Map. 33
4.17 Depth x Height Attribute Map. 33
4.18 Area x Depth Attributes Map. 34
4.19 Volume x Area Attribute Map. 34

5.1 Metrics for the three methodologies presented in this research. 36

List of Tables

2.1	Rock Types and Geologic Context necessary for H ₂ reactions.	10
3.1	Minimum and maximum values for different attributes.	20
4.1	Metrics obtained from the PS methodological application	25
4.2	Metrics for the Non-Increasing Criterion attribute filtering.	28
4.3	Metrics for the Pattern Spectra Applied in the Space of Shape	31

List of Abbreviations

AP	Attribute Profile
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EO	Earth Observation
FC	Fairy Circles
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GCF	Granulometry Characteristics Functions
IOU	Intersection Over Union
MP	Morphological Profile
NDBI	Normalized Derived Built-up Index
NDVI	Normalized Difference Vegetation Index
PCA	Principal Component Analysis
PS	Pattern Spectra
SAM	Segment Anything Model
SAVI	Soil-adjusted Vegetation Index
SCD	Sub-Circular Depressions
TP	True Positive