M 2022

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Biofilm Structure and Properties Characterization through Optical Coherence Tomography Image Processing

MANUEL MONTEIRO
MASTER THESIS IN CHEMICAL ENGINEERING
PRESENTED TO THE FACULTY OF ENGINEERING
OF THE UNIVERSITY OF PORTO

Master in Chemical Engineering

Biofilm Structure and Property Characterization through Optical Coherence Tomography Image Processing

Master dissertation

of

Manuel Monteiro

Developed within the course of dissertation

held in

FEUP

Logo of company/R&D unit

Supervisor at FEUP: **Prof. Fernando Martins**

Joint supervisors at FEUP: **Dr. Diogo Narciso**

Dr. Ana Alexandra Pereira

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

DEPARTAMENTO DE ENGENHARIA QUÍMICA

July 2022

Acknowledgment

As the writing of this document draws to a close, I would like to extend my thanks to a series of important people who collaborated, directly or indirectly and allowed for its existence.

Initially, I would like to thank my closest family members. I would like to thank my mother for always believing in me, especially when I did not, and always supporting me regardless of my decisions or eventual missteps. I would like to thank my father, for instilling in me the drive and work ethic required to reach this step of my higher education. I would like to thank my grandmother for always being there to lend a helping hand and listen to my rants, even when they must have sounded incomprehensible. A special thanks is also in order to my brother, for always reminding me that there is more to life than work.

I would like to thank my friends, without which I never would be close to graduating, and whose contributions extend far beyond what any few words can say. Thank you for keeping me grounded, and showing me that the struggle was communal, and never a singular, solitary effort. A special thanks to all those who were there from the first day in college, helped me and were present through all the years of my study up to this point.

As it pertains to this work specifically, a special word of thanks is in order to my long-time friend Rui Ribeiro, for proofreading this document multiple times and helping me understand which parts were suitably written, and which needed special attention or further development.

As a final note on this more personal section, I would be remiss to not mention my girlfriend, Beatriz Silva, without who I would not be writing this document at all. The impact she had on my evolution, not only as a person, but especially as a student, a learner, and an educator, will always be understated when put down in words. For all the nights where I was close to giving up and she was there beside me, no mere acknowledgement will ever suffice. Additionally, the invaluable study and dedication she has shown to her own personal work, have and will continue to inspire me to strive for a better me.

On a professional level, I must extend my acknowledgments and most profound thanks to FEUP, as it has prepared me for the adversities I will undoubtedly face in the real world.

As it pertains to this work specifically, I extend special thanks to my supervisors, Ana Pereira, Fernando Martins, and Diogo Narciso, for providing extensive feedback, support, and always being available and ready to clarify any doubts and explain what needed improvements.

This work was financially supported by LA/P/0045/2020 (ALiCE), UIDB/00511/2020 and UIDP/00511/2020 (LEPABE), funded by national funds through FCT/MCTES (PIDDAC).

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

REPÚBLICA PORTUGUESA

Abstract

Biofilms can be described as aggregates of microorganisms which clump together and are often attached to a surface. As mesoscale systems, imaging techniques are required to obtain information about their properties and structure. One such technique is Optical Coherence Tomography. The images obtained through this technique are complex and contain large amounts of valuable information that can be used to characterize a given biofilm. Therefore, it is in the best interest of anyone studying them to utilize automatic processing. In the case of Optical Coherence Tomography, a software tool has been recently developed to analyse two-dimensional biofilm images. However, no such tool exists that allowed for the analysis of three-dimensional images.

The main work developed and described in this document is the expansion of the existing two-dimensional software tool, into a three-dimensional perspective. This corresponds to the updating of the previously existing algorithms so as to allow for the analysis of three-dimensional biofilm images. In order to achieve this, a thorough look and rewriting of each function presented in previous work is contained within this document.

As the result of the work described in this document, several algorithms are available for the processing of three-dimensional biofilm images. Several improvements have been identified for future work to reduce processing times and allow for higher convenience.

Keywords: Biofilm, Imaging, Automatic Processing, Software

Resumo

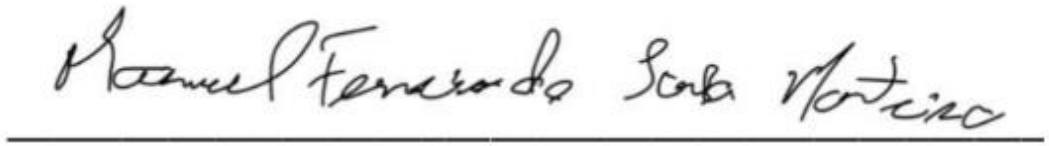
Os biofilmes podem ser descritos como agregados de microrganismos que se aglomeram e estão frequentemente ligados a uma superfície. Como sistemas em mesoscala, são necessárias técnicas de imagiologia para obter informações sobre as suas propriedades e estrutura. Uma dessas técnicas é a Tomografia de Coerência Óptica. As imagens obtidas através destas técnicas são complexas e contêm informação valiosa que pode ser utilizada para caracterizar um dado biofilme. Por conseguinte, é do interesse de qualquer utilizador que as estude recorrer ao seu processamento automático. No caso da Tomografia de Coerência Óptica, foi recentemente desenvolvido uma aplicação informática de análise de imagens de biofilmes bidimensionais. No entanto, não existia ainda uma aplicação informática deste tipo que permita a análise de imagens tridimensionais.

O principal trabalho desenvolvido e descrito neste documento é a expansão da aplicação bidimensional existente, para uma perspectiva tridimensional. Isto corresponde à actualização de algoritmos anteriormente desemvolvidos para processamento unicamente a 2 dimensões de modo a permitir a análise de imagens de biofilmes tridimensionais. Para o conseguir, o presente documento contém uma visão completa e reescrita de cada função apresentada em trabalhos anteriores.

Como resultado do trabalho descrito neste documento está disponível um conjunto de algoritmos que permitem o processamento de imagens tridimensionais de biofilmes. Ainda permanece como trabalho futuro o desenvolvimento de alguns códigos que permitam reduzir os tempos de processamento e uma maior comodidade na obtenção de resultados.

Declaration

I hereby declare, under word of honour, that this work is original and that all non-original contributions is indicated and due reference is given to the author and source



(Manuel Ferreira de Sousa Monteiro)

04 of July of 2022

INDEX

List of Figures

List of Tables

Notation and Glossary

C_F	Compaction parameter	
I	Individual pixel/voxel Intensity	
i	Grayscale pixel intensity (i=0,…,255)	
i^{thresh}	Threshold intensity for pixel binarization	
i^{max}	Maximum Pixel intensity for a given z	
$\bar{L}_F$	Average biofilm thickness	μm
L_F (x,y)	Biofilm thickness at given (x,y)	μm
m	Multiplication parameter for threshold intensity calculation (m > 1)	
nP(x,y)	Number of biofilm pixels in *z at given (x,y)*	
N_x	Number of pixels in horizontal axis	
N_y	Number of pixels in depth axis	
N_z	Number of pixels in vertical axis	
p	Percentile parameter for threshold intensity calculation (0≤p≤100)	
p_{len}	Pixel length	μm
R	Biofilm Roughness	
S^{bin}	Pixel status matrix (for binarization)	
S^{int}	Pixel status matrix (for Structure Detection 2)	
S^{str}	Pixel status matrix (for Structure Detection 1)	
x	Pixel position in horizontal axis (x=0,1, …,N_x-1)	
x_{bands}	Number of parallel processing bands in x direction	
y	Pixel position in depth axis (y=0,1, …,N_y-1)	
y_{bands}	Number of parallel processing bands in y direction	
z	Pixel position in vertical axis (z=0,1, …,N_z-1)	
z^{imax}(x,y)	Pixel position for maximum pixel intensity	
$z_{diff(x,y)}$	Difference between biofilm thickness at a given (x,y)	
z_{lim}	*z* value corresponding to lower bound for substratum detection	
z_{max}	*z* value corresponding to lower bound for image cut	
z_{min}	*z* value corresponding to upper bound for image cut	
z^{bot} (x)	Complete set of coordinates at biofilm bottom interface	
z^{top} (x)	Complete set of coordinates at biofilm top interface	

List of Acronyms

BISCAP	Biofilm Imaging and Structure Classification Processor
CLSM	Confocal Laser Scanning Microscopy
EPS	Extracellular Polymeric Substances
OCT	Optical Coherence Tomography
SEM	Scanning Electron Microscopy

1 Introduction

1.1 Framing and presentation of the work

Biofilms can be described as aggregates of microorganisms which clump together and are often attached to a surface (Hall-Stoodley et al., 2004). There is considerable variety within their properties and applications, as they can be comprised of any number or association of cells. Their use in industry has been documented, such as in Microbial Fuel Cells (Wang et al., 2013), as filters on sewage treatment plants (Lewandowski et al., 2011), as well as in oil degradation from freshwater systems (Golyshin et al., 2008).

As mesoscale systems, imaging techniques are required to obtain information about their properties and structure. These may include Scanning Electron Microscopy (SEM), (Relucenti et al., 2021) laboratory x-ray (Carrel et al., 2017), confocal microscopy (Schlafer et al., 2017 as well as Optical Coherence Tomography (OCT) (Hou et al., 2019), (Wagner et al., 2010), which is the main focus of this work. This technique allows for the acquisition of 3D images of a given biofilm, through the use of light refraction. A beam of light is directed at the tissue of interest, and the scattered light is collected and treated through interferometry. It has applications in medicine and surface characterization, as it is able to obtain results without damaging the sample (Aquino et al. 2019).

The images obtained from these techniques are then processed and return valuable information on morphology, structure, and physical properties, as well as a clear visual representation of the studied biofilm. These are then often used to ascertain if a biofilm is appropriate for a given system (Drexler et al., 1999).

The analysis of these images through manual processing is inefficient, complex, and down to personal interpretation as they require pixel-by-pixel analysis. Therefore, automation is in the best interest of anyone analysing these images. Automatic processing of these images is available for confocal microscopy, as reported, for example in (Heydorn et al., 2000) or (Mueller et al., 2006) and has recently been developed for 2D OCT images in (Narciso et al., 2022) and forms the basis on which this work was developed.

With this in mind, the main objective of this work was to expand on the algorithms devised for 2D methodology, so as to adapt them for the 3D case. As three-dimensional structures, they can only be characterized by three-dimensional processing. Two-dimensional analysis is naturally limited, as there is variability in depth along a given biofilm, with a notable difference generally observed between the first and last 2D slices. Additionally, due to how image processing depends on continuity detection, a given section of an image may be incorrectly

characterized as non-biofilm if the analysis is done in two dimensions only, given that it may have continuity in the non-measured direction. Properties such as roughness, porosity and average thickness depend on all three spatial dimensions and as such 2D analysis can paint an incomplete picture and return information that may not truly characterize the biofilm. Therefore, for analysis to be complete would require information that is only available through the use of 3D imaging.

However, for 3D, the amount of information to process is considerably larger. As the three-dimensional images are stacks of two-dimensional ones, a significative increase of the number of pixels to analyse occurs. Additionally, as each pixel requires more operations to be processed, the computational complexity of these issues is considerably raised by the extension to 3D. Processing times may be elevated from seconds in 2D processing (image sizes around 10MB) to hours or days in 3D processing (image sizes around 1500 MB). Therefore, a potential solution is parallel processing.

Parallel processing is the division of a given task into multiple threads, which can be independently worked on by a given processor and is often applied to previously existing algorithms to improve their efficiency (Alamsi et al., 1989). In order to obtain results in a reasonable timescale, it was in our best interest to develop parallel processing functions that would be able to shorten the processing time.

1.2 Contribution of the author to the work

Taking the two-dimensional algorithms developed by Narciso et al. (2022) as a reference, a new version of the code was devised to analyse 3D images. I studied each of the original algorithms presented in this article and adapted them to meet the demands of three-dimensional processing.

These tasks consisted of the addition and writing of a series of lines of code, so as to adapt the previously existing functionality. Most of the work focused on the main processing functions as well as the two continuity tests used to obtain the biofilm's structure. This consisted of the bulk of the work, as it required the adaptation of the algorithms to contemplate a considerably larger amount of points and calculations. Besides these two continuity tests, every auxiliary function was adapted in order to enable 3D image processing. This includes the reading of images as well as their pre-processing. Changes were required and applied to the output functions, as well as the calculation of parameters.

This work was done in close collaboration with my supervising team and can be divided into three separate stages:

(I) Discussion of necessary code changes and implementations;

(II) Code writing, debugging, troubleshooting, and testing;

(III) Code review and functionality confirmation.

This workflow was undertaken for the update of all necessary algorithms. All of the functions were updated by me, with additional support from the supervising team, especially in the parallel algorithms.

Upon completion of code writing, I performed a series of performance tests so as to provide initial insight into potential improvements to processing times through manipulation of parallel parameters. This provides preliminary information on how to achieve faster processing of images.

1.3 Organization of the dissertation

Chapter 1 presents the introduction, with a general overview of the principles required, as well as the motivations for this work.

Chapter 2 corresponds to the state of the art, where a series of basic principles and analysis of current developments in OCT treatment of biofilm images will be presented.

Chapter 3 presents a description of the materials and algorithms used. Pseudocode is presented and explained for all the algorithms, and schematics are shown to represent the study of the problem at hand.

Chapter 4 presents three case studies. These cases report the results of the use of the code structure developed on the analysis of separate images, as well as a direct comparison to previously existing work, and a preliminary study on the optimization of the key parallel processing parameters to achieve faster computing times.

In the conclusion presented in Chapter 5, a retrospective look on the work as a whole is shown, as well as suggestions for future work in this field and the next steps to be taken in development.

In Chapter 6, an informal assessment of the work done is also shown, as well as the author's personal thoughts of the work developed.

2 Context and State of the art

2.1 Biofilm Imaging

The study of biofilms is a vast and varied area of study, with potential applications in many different fields, as in the medical, food, and energy industries.

A Biofilm can be described as an aggregate of microorganisms, where there is direct contact between each microorganism's cells. Besides these microorganisms, there is also an extracellular matrix of polymeric substances (EPS), which are produced by the cells contained within the biofilm. They have highly hydrated three-dimensional structure, and form on any type of surface, upon exposure to environmental factors. After a given microorganism is exposed to the necessary conditions, it undergoes a dynamic change to its genes, causing it to grow in a way that is beneficial to the formation of biofilm (An et al., 2007).

Biofilms' general properties are affected in large part by their morphology, as demonstrated in (Bishop et al.,1995). In this work, the biofilm structure was measured and shown to cause variations in mass transportation efficiency between different elements of the biofilm's matrix. This structure can be characterized by parameters such as average thickness and roughness. Their application and efficiency for a given problem can vary with these properties. As an example, in (Maria Culell et al., 2015)'s work, the nitrification levels achieved with different thickness of ammonia oxidizing bacteria biofilms within a moving bed reactor were measured. Although biofilm thickness was not found to affect the ammonium production, nitrate production was higher for higher thickness biofilms. As another example, in Yun shen et al., (2012) adhesion of pathogenic Legionella pneumophila biofilms to water surfaces were measured for different roughness biofilms. By changing the flow velocities, it was demonstrated that smoother biofilm surfaces lead to easier detachment of pathogens. It can then be said that the determination of biofilm properties is of key importance for the biofilm function (how it behaves).

When unattended, a given biofilm will grow on a given liquid or solid substrate up to a macroscopic level. However, in many cases of interest, there is a focus on the microscopical aspects of biofilms, including individual cell growth. There has also been a growing interest in a mesoscale approach, to ascertain mechanical properties. Therefore, imaging techniques are required to adequately characterize them. A summary and review of most current techniques is available in (Azeredo et al., 2016).

Light absorption is one important asset in biofilm microscopy. For example, the effects studied by Carvalho and da Fonseca (2007), through optical microscopy showed that that there is a linear relation between the intensity of light absorbed in a given area of biofilm is related to the thickness of biofilm at that point. This corresponds to different brightness levels of the corresponding image. Although optical microscopy is cheap, easy to perform, and only requires simple preparation of a given sample, it has severe limitations when it comes to the amount of information extractable. The level of magnification achieved does not allow for in depth characterization of specific sections of the biofilm, and there is a maximum value of saturation detectable, after which no further information can be gained.

In contrast, Confocal Laser Scanning Microscopy (CLSM) is widely accepted to be the most versatile and powerful technique available to characterize biofilms in their microscopic scale. It works based on fluorescency. While in typical fluoroscency microscopy the sample is evenly illuminated by the light source, in CLSM, the beam of light is passed through a pinhole, and onto the sample allowing for a specific point of a given system to be illuminated. Using an oscillating mirror arm, multiple parallel points can be analysed at once, and a two or three-dimensional structure can be imaged. This technique has been applied to many different types of biofilm, such as those described in (Guilbaud et al., 2015; Sun et al., 2015; Villacorte et al., 2015), and allows for the calculation of volumes, thicknesses and roughnesses, as detailed in Bridier et al. (2010)

Through the use of different dyes (Neu and Lawrence, 2014a, b) or genetic modification (Klausen et al., 2003b; Sanchez et al., 2015; Tolker-Nielsen & Sternberg, 2014) biofilms fluorescency can be achieved to allow for the application of CLSM. However, this is also the biggest limitation to the use of CLSM, as any system that cannot be made fluorescent is invalid for analysis through this method.

Scanning Electron Microscopy (SEM) has also been applied to biofilms. In this method, a beam of electrons is used to scan the surface of a material, which then scatters and absorbs these electrons and returns information to characterize the surface. It is, however, unable to return information in depth, as the electrons do not penetrate the surface. Although SEM has been used as a comparative method (Di Bonaventura et al., 2003, 2006; Hasan et al., 2015; Li et al., 2015;) for its positive properties, such as the ability to visualize a large amount of information at once when compared to other imaging techniques (from 50 to 100 nm), SEM's sample preparation is time consuming and tedious, and may destroy the structure or cause artifacting, as reported in (Hannig et al., 2010). This limits the use of SEM, as the biofilm samples may be valuable or unique, or as the loss of information from artifacting may be inacceptable if precision is required.

Optical coherence tomography (OCT) is currently undergoing study and interest as a suitable imaging tool for description of biofilms on a mesoscale.

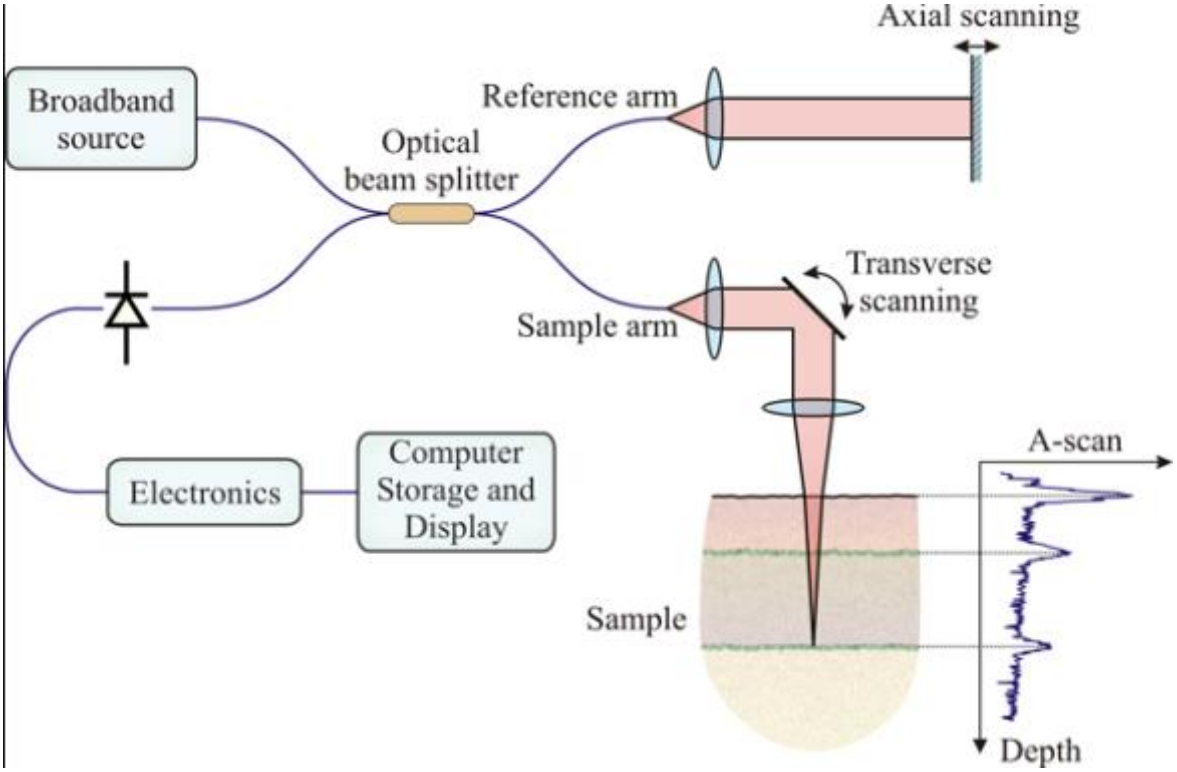


Figure 1: Schematic of OCT functioning (Taken from:(OBEL, n. d)

As shown in Figure 1, OCT images are obtained from the use of an interferometer, a light source, a beam splitter and two arms, one for reference and one for sampling. The sampling arm consists of a lens pointing at the object of interest (in this particular case, biofilm). The light from the light source is reflected and scattered by the sample, which causes interference with the beam splitter. This interference is what provides the information that will be recorded and analysed. It is then returned as information in the form of an image, a dataset in greyscale, which returns higher (brighter) values for higher biofilm concentrations, and lower (darker) values to indicate the absence of biofilm in a given region. In the particular case of biofilm imaging, the association of depth information is called an "A-scan". The sum of consecutive parcels of this information is called a "B-scan". Although a name for consecutive B-scans has not been formally attributed yet, it is referred to in literature as a C-scan. As OCT detects scattering and reflection, it is well suited to describe the general structure of a given biofilm. It also avoids the main problem of techniques such as CLSM, as samples do not need to be fluorescent.

OCT showed to be promising in several biofilm studies, like for example the one from Xi et al. (2006) who monitored the growth of a *Pseudomonas aeruginosa* biofilm within a glass capillary, non-invasively and in-situ. In 2010, Wagner et al. developed a study where they first visualized the complete structure of a three-dimensional biofilm and demonstrated that its compaction (normalized measure of how strongly compacted biofilm structure is) depended on the Reynolds number of its fluid. In this study, single OCT scans were obtained within two minutes with no sample pre-treatment. A series of following studies, such as those found in Fortunato et al. (2016) and Rasmussen et al. (2016) cemented OCT as a versatile tool for the characterization of biofilm structures at the mesoscale. As an example, result, Figure 2 shows typical raw data in image format obtained from OCT in the study of these systems:

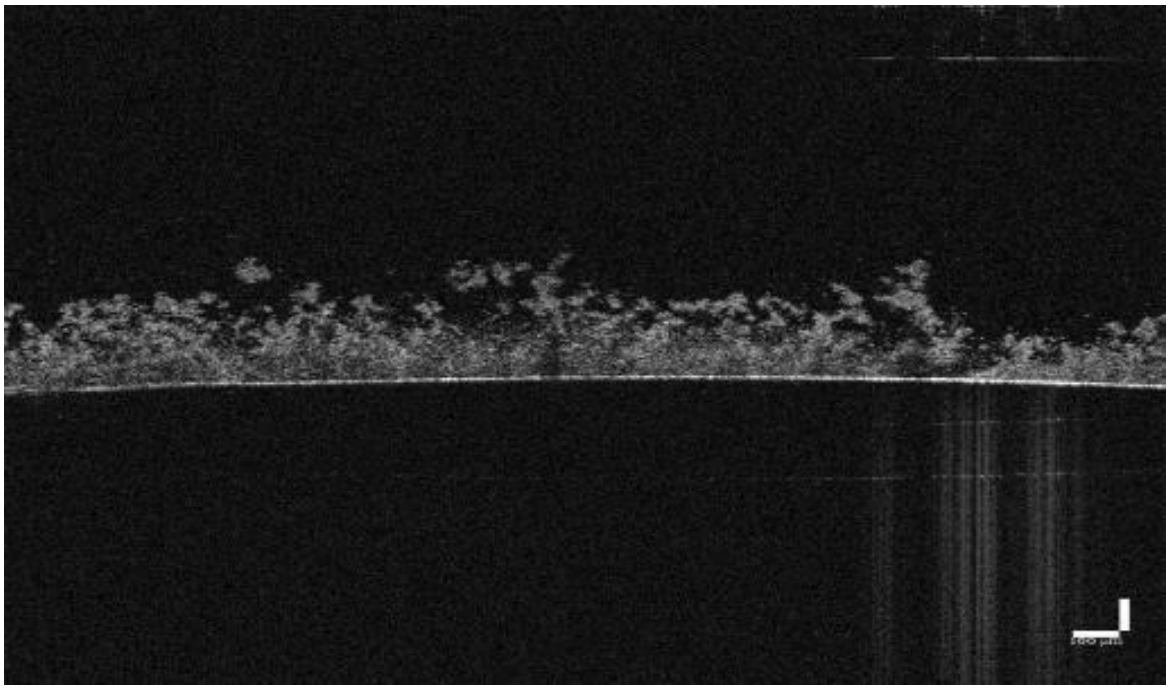


Figure 2: Image obtained from OCT analysis of biofilm

As can be seen, the data is available in a greyscale image, where brighter whites correspond to the existence of biofilm. Some visual artifacting can be seen in the bottom right of the image.

2.2 Image processing

While the aforementioned studies focused mainly on OCT's ability to accurately portray and represent biofilm's general morphology or even growth over a given timeframe, the interest lies in the complete characterization, with the calculation of parameters such as thickness and roughness. In order to derive these parameters from OCT imaging, a series of image processing treatments should be applied to the datasets. As shown in Figure 2, OCT returns a greyscale

image, which can then be read by a computer as a three-dimensional matrix. Each pixel having a different value for X and Z to describe position, as well as a matching greyscale intensity. It is these three-dimensional matrices that are then manipulated through a series of operations. This is the starting point for this work, and the area to which it corresponds.

Initially, the datasets are binarized, and each individual pixel is separated into biofilm or void. This requires the separation based on a threshold. Though multiple thresholding algorithms are available to determine this value, the Otsu method (Otsu, 1973 ; Yerly et al., 2007) and the triangle methods (Derlon et al., 2013) are commonly used. This thresholding can also be applied manually such as in Blauert et al. (2015). After the information is binarized, it may be further processed. For example, in Wagner et al. (2010), downscaling was performed to lower image size prior to binarization, and following it a white line was drawn to separate the biofilm base from the attached structure after which the height of biofilm in each scan was calculated. In contrast, in Dreszer et al. (2014), custom MATLAB scripts were used to initially locate and separate the base from the structure, after which edge detection was used to obtain an outline of the biofilm, filtering was applied to separate real voids from those caused by visual artifacting, as well as removal of spurious nose pixels prior to height calculation. More recently, in (Narciso et al., 2022), an independent, fully automatic and novel software was presented, entitled Biofilm Imaging and Structure Classification Automatic Processor, shortened by the authors to BISCAP, constitutes a simple way to automatically process 2D OCT images.

This software functions based on a novel thresholding and substratum detection strategy. As detailed by the authors, pre-processing is non automatic and consists of the indication of a maximum (z_{max}) and minimum (z_{min}) value of z, which will correspond to the top and bottom interfaces of the biofilm in study. Afterwards, the image is cut along these two lines. A representative schematic is shown in Figure 3:

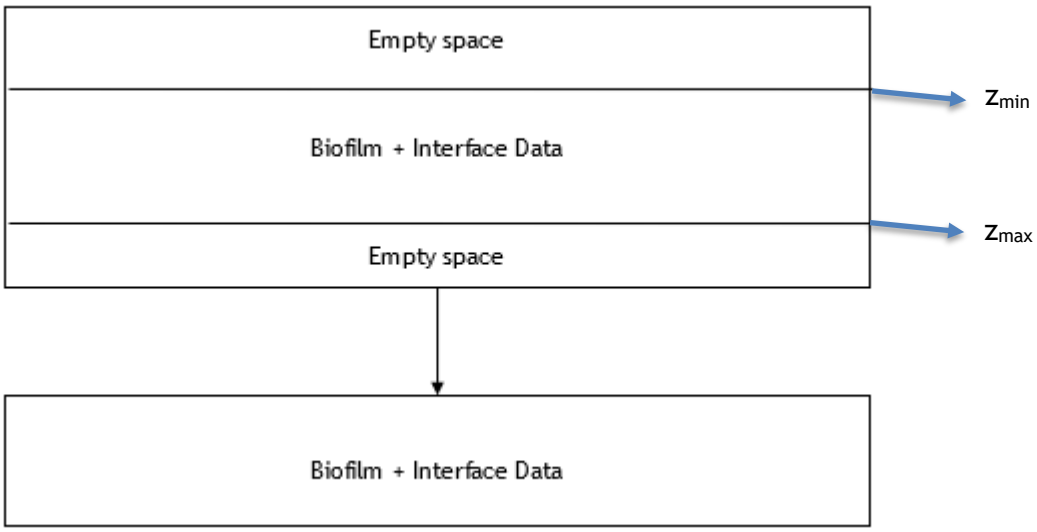


Figure 3: Schematization of pre-processing execution

This is however, a completely optional step that can be disregarded. However, the gains in computation time vastly outweigh the relatively lower investment of time corresponding to the marking of the lower and upper bounds. Additionally, during this step, a lower bound is given for the determination of the base of the biofilm interface. This is referred to as the limit value *for* z (z_{lim}) and will provide an initial value for the next algorithm to search. These values are provided via manual pre-processing.

There is then a series of automatic processing steps, which are divided by the authors into four steps.

(I) Initially, the bottom interface's position is calculated. This interface is often identified by a series of very bright pixels occurring in the lower bounds of the image. This observation results from the analysis of a large number of OCT images and is not a necessary condition. The authors report this algorithm as being able to accurately detect the position of the bottom interface for the majority of OCT datasets.

In a first step, all points with the highest intensity in *x* are marked, and then filtered. These values are then estimated as a series of straight lines between the filtered points, resulting in an accurate depiction of biofilm interface. The flowchart depicting the steps required is shown below in Figure 4.

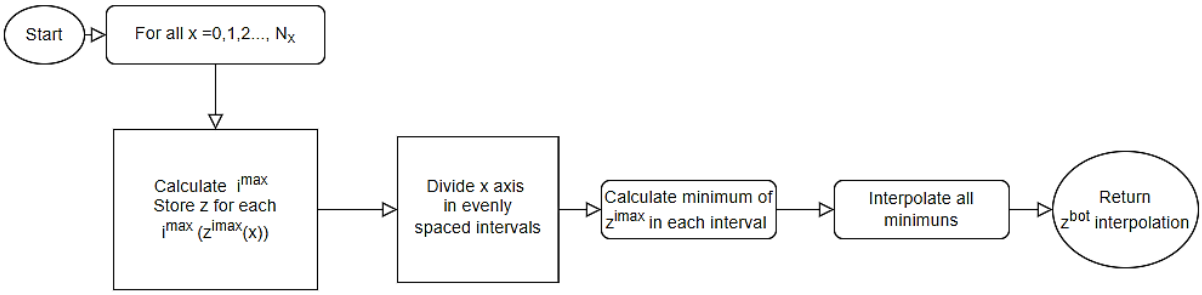


Figure 4: Flowchart for biofilm interface

(II) In order to accurately portray the data and separate biofilm from background, binarization is applied. This is the second step, and corresponds to the calculation of a threshold intensity, and subsequent classification of every pixel as biofilm (above threshold) or void (below threshold) respectively.

The threshold is calculated by taking the average of pixel intensity for a certain percentage of the pixels present in the top sections of the image to obtain a value that accurately estimates the average background pixel intensity and is defined by two user parameters. These parameters are *p* (corresponding to the p-percentile of highest pixel intensities) and *m*, which is a multiplication factor greater than 1. As the threshold value is subjective and varies from image to image, the combination of these two factors ensures a decent estimate is used.

Upon thresholding, pixel binarization occurs. In this step, every pixel is classified as potential biofilm or as void. Its corresponding schematic representation is shown in Figure 5:

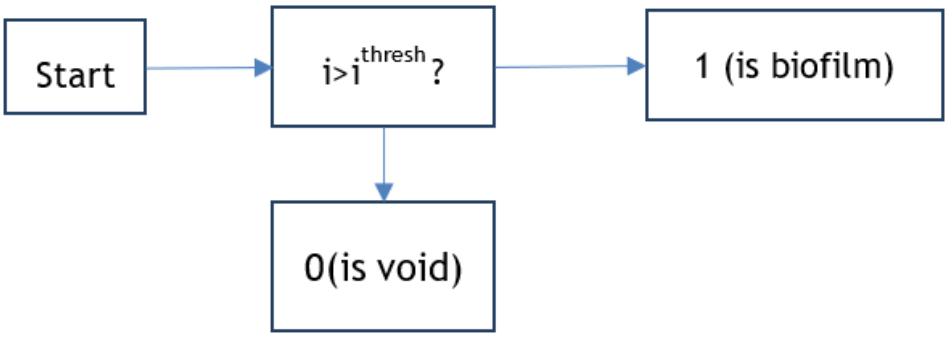
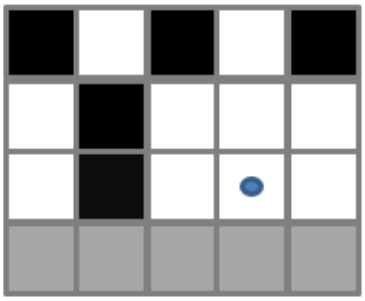


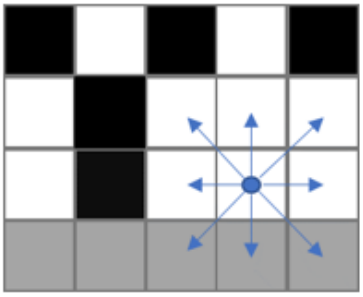
Figure 5:Schematic representation of binarization

(III)As a third step, the biofilm's structure is then discovered using a volume filtration strategy, whereupon only the pixels connected, whether directly or indirectly to the substratum are

biofilm. This leads to a continuity search in all 8 spatial directions, as explained in Figure 6's schematic, where white represents biofilm, black represents void and grey is the interface.

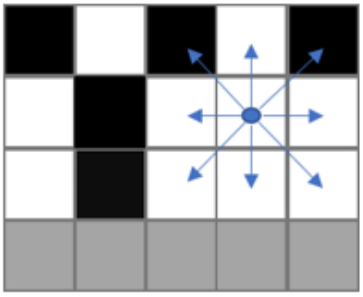


1: The initial point (blue dot) is located directly above the biofilm interface(grey).



2: Continuity test

All 8 points surrounding the original point are tested. Any white blocks in direct contact with the original point are validated as biofilm.



3: Continuity Test (continued)

From all points previously validated as biofilm, a new continuity check is executed. Any white points connected to this new point are also validated as biofilm. This is repeated sequentially until there are no more possible connections.

Figure 6: Structure Detection 1 schematization

Additionally, as a fourth and final step, the top interface is determined through the detection of the initial biofilm pixels, starting from the top of the image. A similar method to the previous algorithm is adopted, although detection starts from the top of the image and finishes upon

the detection of the first biofilm pixel within a given slice. The schematic is demonstrated below, in Figure 7.

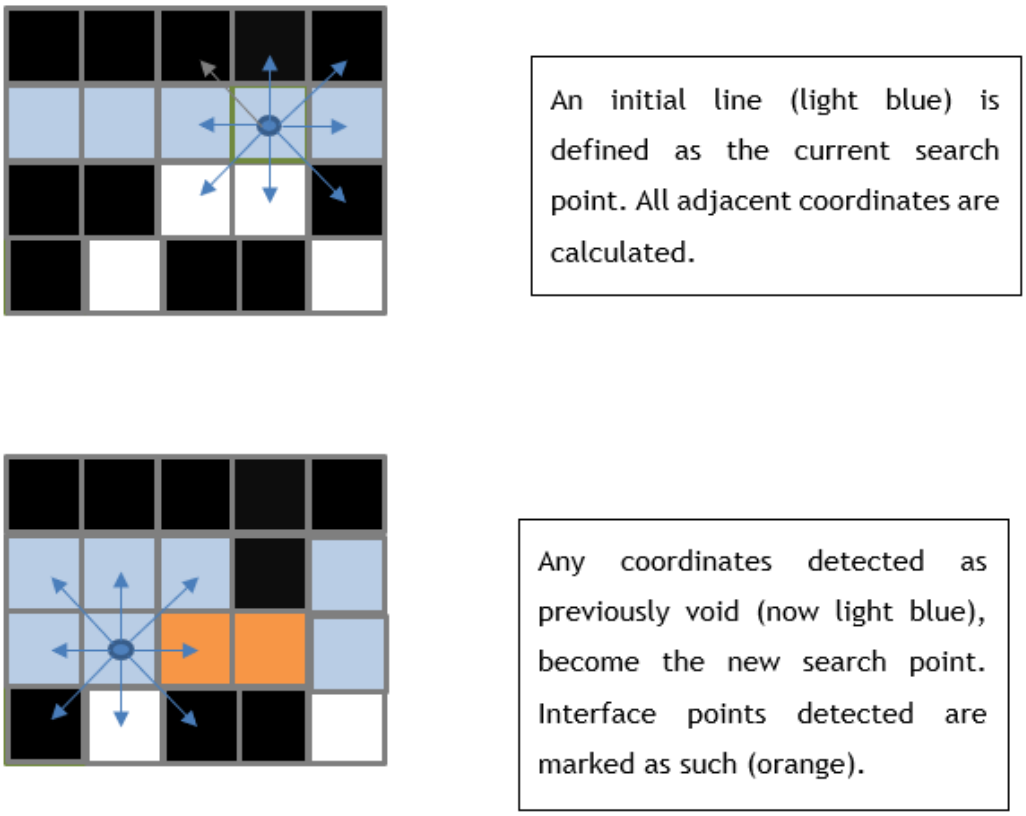


Figure 7:Structure Detection 2 schematization

Upon completion of these processing steps, there is complete information on biofilm structure, which then allows for the calculation and return of a series of user defined outputs, including values for parameters that have shown their relevance, such as thickness and roughness, as well as a novel “compaction parameter”, which presents a measure for how much of a given biofilm’s main structure is actually comprised of voids.

This software is therefore able to return parameters such as thickness and roughness, accurately and completely characterizing a given B-scan for a system. However, to the authors best knowledge, no similar software to BISCAP exists for the treatment of three-dimensionally stacked images, or C-scans. As previously stated, biofilms are a three-dimensional structure and, to obtain accurate and complete information, a similar analysis would have to be

employed, but contemplating the complete set of information returned by a thorough OCT imaging of a system.

This dissertation then sets out to expand and improve on the software developed in (Narciso et al., 2022), so as to allow for the analysis of three-dimensional cases. In order for this, it is expected that changes will be required to each of the previously described algorithms, so as to take into consideration the additional spatial dimension.

The main output results from this work is a functional software that adequately and accurately describes any biofilm system into a 'clean image' with no artifacting and that returns values for biofilm structural parameters, such as thickness and roughness. These parameters will have been redefined and recalculated to accommodate the addition of the third spatial dimension.

Additionally, problems with computational efficiency and time to process are expected to arise, as the number of calculations is exponentially higher, not only from the increased amount of information to process, but also from the additional continuity testing in the volume filtration method. Therefore, it is also expected that some advancements will be made in the sense to lower the calculation time required to obtain results.

3 Materials and Methods

To advance within the work done, a series of algorithms and methods were developed, as can be roughly summarized in the Flowchart shown in Figure 8.

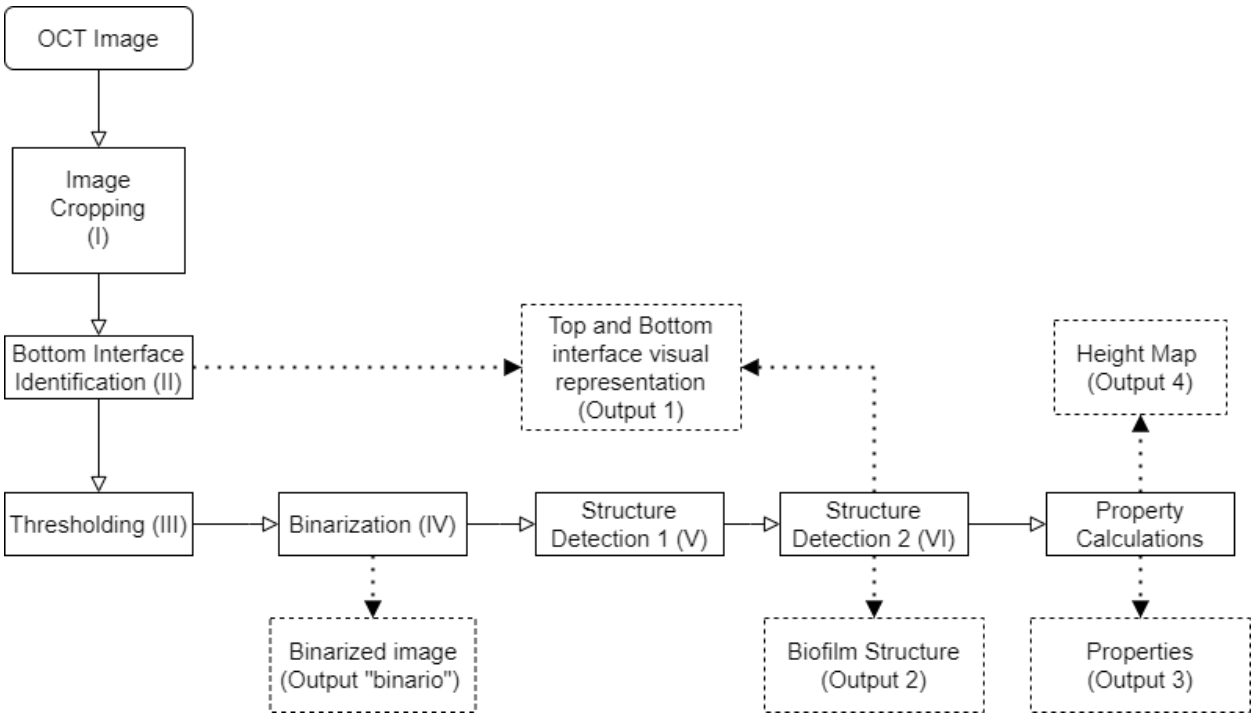


Figure 8: High level readability of general processing

Step I corresponds to image pre-processing, while steps II to VI correspond to processing. The dotted arrows point to the outputs obtained through each stage. The steps taken for their updating will be detailed in this chapter. To prove their functioning as well as to help develop these algorithms, they were implemented in Python. The materials used for this implementation and research are also described within this chapter.

3.1 Methods

The base material for this work are images obtained from OCT imaging. For consistent and correct analysis, an axis system was devised to work as a consistent referential. These axes are shown in Figure 9.

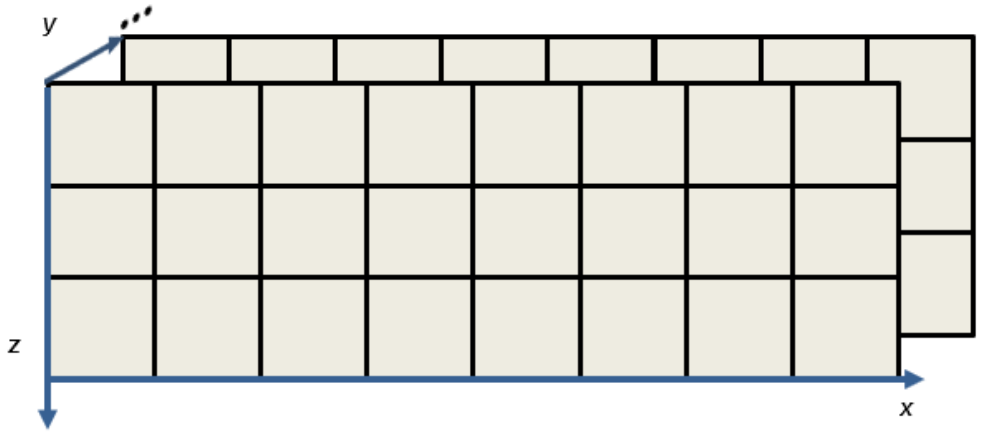


Figure 9: Axes system for three-dimensional imaging

As can be seen in Figure 9, *z* axis corresponds to height information, *x* axis to horizontal information, and y to depth information. In a two-dimensional case, the axis system is the same, but there is no y dimension. It is also important to note that higher *z* values correspond to the lower parts of the image.

Initially, the raw greyscale images are imported as a 3D matrix and created as *NumPy* arrays. As the treatment of this data progresses, differing 3D matrices are created and processed in order to guarantee adequate information gathering. This treatment can be divided into 3 sections, which are: pre-processing, processing, and property calculations. The updates required for pre-processing, processing and post-processing functions algorithms are summarized, schematized, and outlined in the following section. Additionally, pseudocode is included as a demonstration of the logic behind the more complex algorithms.

3.1.1 Pre-Processing

Image cropping (I)

Initially, the image is loaded as a three-dimensional matrix. Information is provided to crop this initial matrix, to remove unnecessary information, namely the top and bottom parts of each z/x slice, which contain only background information, consistently with the principles highlighted in Figure 3 for the 2D case.

3.1.2 Processing

Bottom interface identification (II):

From the cropped image obtained in step I, the brightest voxels present within the image are identified, and used as an initial basis for the bottom interface searching algorithm.

In order to function for three-dimensional imaging, this algorithm required adaptation. It had added onto it a repetition loop, in order to call it as many times as there are slices in *y*. This results in a two-dimensional matrix of pairs of *x*-*y* values, which each include the z coordinate that corresponds to the height of the biofilm's surface for the subsequent calculations.

Thresholding (III)

This step follows the same principles described in Section 2.2 for thresholding. The main difference is the use of points from every y slice, to accurately represent the complete image.

Voxel binarization (IV):

With the threshold obtained from step III, all voxels are binarized, and identified as either biofilm or background.

This algorithm functions based on the same principles described in Section 2.2, and simply attributes the value to each three-dimensional voxel instead of each two-dimensional pixel. This results in the creation of three-dimensional matrix S^{bin}.

Structure Detection 1 (V):

The structure detection functions consisted of the bulk of the work. They are complex, highly precise algorithms who must be functioning correctly in order to ensure that the biofilm's structure is completely characterized.

This algorithm determines if there is continuity between biofilm voxels, both in the vertical (*z*) as well as the depth (*y*) and horizontal (*x*) directions. As there is now a cube like structure, search starting from an interior voxel leads to a direct check of all 26 points that are directly adjacent to the initial point considered, in direct contrast to two-dimensional problems, where only 8 adjacent pixels are checked. This is well represented by the schematic shown in Figure 10.

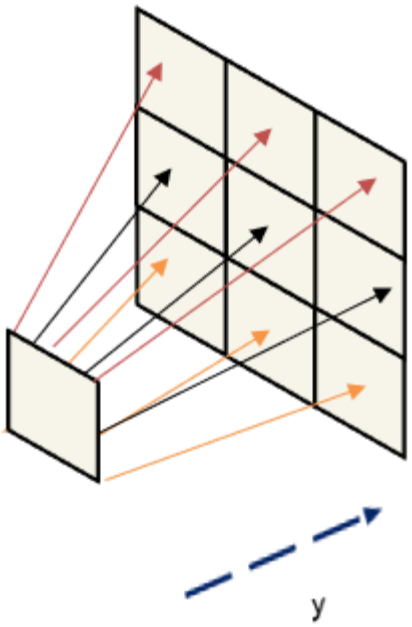


Figure 10: Schematic representation of continuity detection in y direction.

In the schematic, different z values are colour coded, red for lower values, black for the same z value as the coordinate of the point being checked, and orange for higher values. Voxels pertaining to lower y or equal y values are not shown to allow for visual clarity, similarly to Figure 9. This structure is therefore repeated in a mirrored fashion for y values lower than the one of the voxels currently being checked, and checking for the coordinates with the same y value occurs identically to the structure detection described in section 2.2. This then leads to a total of 26 voxels to check for any given interior voxel (as when y is the same, the point itself is not checked). This corresponds to 9 voxels in the y-1 slice, 9 voxels in the y+1 slice, and the 8 adjacent voxels in current y slice.

The pseudocode for this updated function is shown below:

1.1 CREATE Matrix S^{str} with same size as Image.

1.2 INITIALIZE all voxels in S^{str} as “non-biofilm”, except the biomass voxels (from S^{bin}) as “unexplored”.

1.3 FOR all (y, x) pairs from 0 to N_{y-1} and N_{x-1} respectively, directly above (in y,z-1,x) to bottom interface voxels, label every biomass voxel as “biofilm”. This step corresponds to the initial iteration and is the source of continuity.

1.4 FOR all voxels identified from previous step or step 1.6, calculate every adjacent voxel’s coordinates in all directions (26 voxels total).

1.5 For all voxels from step 1.4, if its status is “unexplored” update it to biofilm (it is adjacent to at least one biofilm voxel as calculated in an earlier iteration) and add it to a list of voxels

to be processed in the next iteration. End current iteration. If there are additional voxels to be processed, begin a new iteration and go to step 1.4. Else, terminate.

1.6 When step "1.4" is completed, update any remaining "unexplored" voxels to "Non biofilm". (While they are biomass, no continuity exists with bottom interface). Return final S^{str}

This was achieved through the updating of the previously existing two-dimensional functions. These updates correspond to the addition of the third dimension to the continuity checking. This implied a logic structure, so as to ensure no repeat calculations are done.

The search for new voxels must occur in a way as to not include coordinates outside the image limits. In order to implement this, the logic structure divides the processing area into 9 different sections, as described in Table 1.

Table 1: Logic Structure sections

x coordinates	y coordinates	Directions (x)	Directions (y)	Number of points
$0<x<N_x-1$	$0<y<N_y-1$	Left,Right	Back,Front	26
$0<x<N_x-1$	$y=0$	Left,Right	Back	17
$0<x<N_x-1$	$y=N_y-1$	Left,Right	Front	17
$x=N_x-1$	$0<y<N_y-1$	Left	Back,Front	17
$x=0$	$0<y<N_y-1$	Right	Back,Front	17
$x=N_x-1$	$y=N_y-1$	Left	Back	8
$x=0,$	$y=N_y-1$	Right	Back	8
$x=N_x-1$	$y=0$	Left	Front	8
$x=0$	$y=0$	Right	Front	8

For each of the sections presented in Table 1, the relevant coordinates are then added to the list of points to check. This logic structure is essential for the functioning of the algorithm, as otherwise points outside of the matrix will attempt to be checked, resulting in failure. This is the most complex part of the algorithm, as the correct coordinates must be added in order to not double-check unnecessary points, and the coordinates are calculated individually for each new iteration.

As can be seen, the main difference is the added continuity dimension. This can be likened to voxel search occurring in a cube shape instead of a square. The original algorithm is unable to detect depth continuity, even if called for each slice individually. This difference is what makes it three-dimensional analysis.

Structure Detection 2 (VI):

An updated version of this algorithm was also devised and implemented for this work, which will take into consideration that any of the 26 neighbouring points could be the initial interface point, whose pseudocode is contained below.

2.1CREATE Matrix S^{int} with same size as Image.

2.2 Label all voxels as "unexplored", except the voxels below z^{bot} *and above* z^{max} , which are labelled as "background".

2.3 DEFINE all voxels at z= z^{max-1} as the starting voxels for the first iteration,

2.4 FOR all voxels to process (from step 2.3 or step 2.6), check the matching voxel in S^{str}.

2.5 IF voxel in S^{str} =" non biofilm"; check every adjacent voxel.

2.6 IF any voxel checked in 2.4 is marked as "biofilm", mark iteration's voxel as "interface (water side)".

2.7 ELSE mark it as "background", and add its coordinates to the next iteration.

2.8ELSE mark voxel as "interface (biofilm side)".

2.9 Stop current iteration. If there are additional voxels from step 2.6 begin a new iteration and go to step 2.3. Else terminate.

In the case of this algorithm, the logic structure is highly similar to the one present previously.

The nine different sections present in Table 1 are once again considered, so as to avoid checking of unnecessary or inexistent points. However, testing is now for the presence of any single biofilm voxel in the vicinity. Upon detection of any voxel corresponding to biofilm, no further search is conducted from that voxel in particular. To ensure this, each voxel has every adjacent voxel checked in all three directions.

The main difference between the new version and the previously existing algorithm is, as in Structure Detection 1, the number of directions that must be checked for new points. Each individual point will have, for most cases, 26 adjacent voxels to check.

These two algorithms correspond to the largest portion of the work and calculations and are the functions that require the most processing power. The number of voxels to analyse in each of them rise exponentially with the size of the image.

3.1.3 Property Calculations and Outputs

As the program runs and the matrix is updated, visual outputs are created to return valuable information, and allow for visual confirmation of the software's functioning.

The first output, named binario, corresponds to the result of the binarization step. It is therefore an image of the same size as the original, where every voxel marked from thresholding as potential biofilm is shown with maximum intensity ("pure" white), while everything else is black, signalling voids. This is achieved by multiplying each value within the binarized matrix by the maximum intensity (0 ->0 (black), 1 ->255 (white)) and is useful to confirm if the threshold parameters used were appropriate, as it visually shows if things that are clearly not biomass (such as minor visual artifacts) are being considered as such.

The second output, named Output1, corresponds to the result of the Structure Detection steps. Every point of the biofilm is shown in its original intensity, while all voxels corresponding to either the top and bottom interfaces are marked as a user defined colour. This allows for an identification of the biofilm bounds.

The third output, named Output2, corresponds to the final result of all steps. It returns an image where everything that is not considered biofilm is considered as background. This is equivalent to showing only the biofilm structure and is a good representation of the biofilm in each slice.

Upon completion of the structure detection steps, the matrix with all the information required for property calculations is available. These are calculated according to the set of equations below, and are compiled in the excel file corresponding to Output3:

Equation 1 refers to the calculation of biofilm thickness,

$$L_F(x,y) = (z^{bot}(x,y) - z^{top}(x,y)) * p_{len} \quad (1)$$

where $L_F(x,y)$ is the biofilm's thickness (z) for each x,y pair. $z^{bot}(x,y)$ and $z^{top}(x,y)$ are positions in z for the bottom and top interfaces, and p_{len} corresponds to each voxel's real life equivalent length.

Equations 2, 3 and 4 refer to the calculation of biofilm roughness,

$$(\bar{L}_F) = \frac{\sum L_F(x,y)}{N_x} \quad (2)$$

$$z_{diff(x,y)} = |(L_F(x,y) - \bar{L}_F)| \quad (3)$$

$$R = \frac{1}{(N_x\, N_y)} \sum z_{diff} \quad (4)$$

where, $\bar{L}_F$ corresponds to average biofilm thickness, $z_{diff(x,y)}$corresponds to the difference between the biofilm thickness in a given *x,y* pair, *R* refers to the roughness, and N_x and N_y correspond to the image's pixel size in the *x* and *y* directions respectively.

Equation 5 refers to the calculation of the compactness parameter,

$$C_F = \frac{nP}{\bar{L}_F} \frac{pL}{(N_x\ N_{y)}} \tag{5}$$

where C_F is the compaction factor, and *nP(x,y)* is the number of biofilm pixels in the *z* direction for a single *x,y* pair. It is therefore a measure of the number of pixels that actually correspond to biofilm (shown by the biofilm thickness' average) and the total number of pixels from top to bottom interface.

The final output, Output4, is a height map of the biofilm. It is obtained by normalizing the value of thickness to greyscale, and then returning a two-dimensional image. It therefore corresponds to a x/y slice, where the brightness of each pixel is directly proportional to its thickness.

3.1.4 Parallel Processing

In order to reduce the processing time and obtain results and characterization within a workable timeframe, additional parallel processing versions of the algorithms were devised. They allow for the processing of different sections of each image in a different process, allowing for the complete use of multi-core processors, as are common within modern computers. These can be activated or deactivated through the change of a parameter within the code.

This is achieved through the division of the image into equally spaced "sections", within the x and/or y direction. z sectioning is not feasible due to the way continuity is detected, as contact with the interfaces would be impossible to guarantee. The number of sections the image is to be divided in is defined by the user. These are divided as equally as possible, so as to ensure processing is efficient. The division meticulously separates the sections into bands, which are individual three-dimensional matrices, maintaining potential continuity information. A visualization of this process is shown in Figure 11.

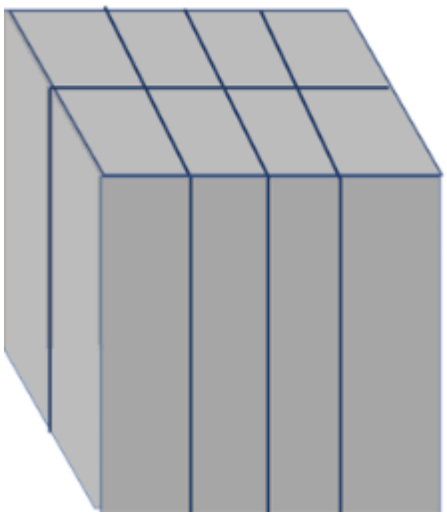


Figure 11: Sectioning of biofilm in x and y directions (4 x slices, 2 y slices)

After this sectioning, a series of parallel threads are created, in order to run the original continuity algorithms on each section. This returns information on each section independently and is lacking information on the lines through which the "cut" was executed. Therefore, updating has to be undertaken to consider the complete image, as otherwise information would be irrevocably lost, and incorrect information returned.

These updating functions are precise and complex, and required a different approach to every other function. They are based off of "incomplete" information derived from the sectioning, and therefore had to be carefully planned. They initially check every point along the individual directions of X and Y and update accordingly. This means defining an upper and lower bound for each section in each direction, so as to ensure no checks are attempted beyond the limit of each band. Additionally, the zones where there is intersection between bands are carefully checked.

This updating is slightly different for each of the Structure Detection algorithms defined in the previous section. For the second test, there is need of an update to the voxel status in order to ensure that the status of each voxel is the correct one. This is due to the different possibilities in Structure Detection 2, where voxels can be defined as interface or simply biofilm. This means the addition of a final step, where the voxels along the z direction are updated to guarantee the top interface is correctly calculated.

The schematization for parallel algorithm one is displayed below in Figure 12.

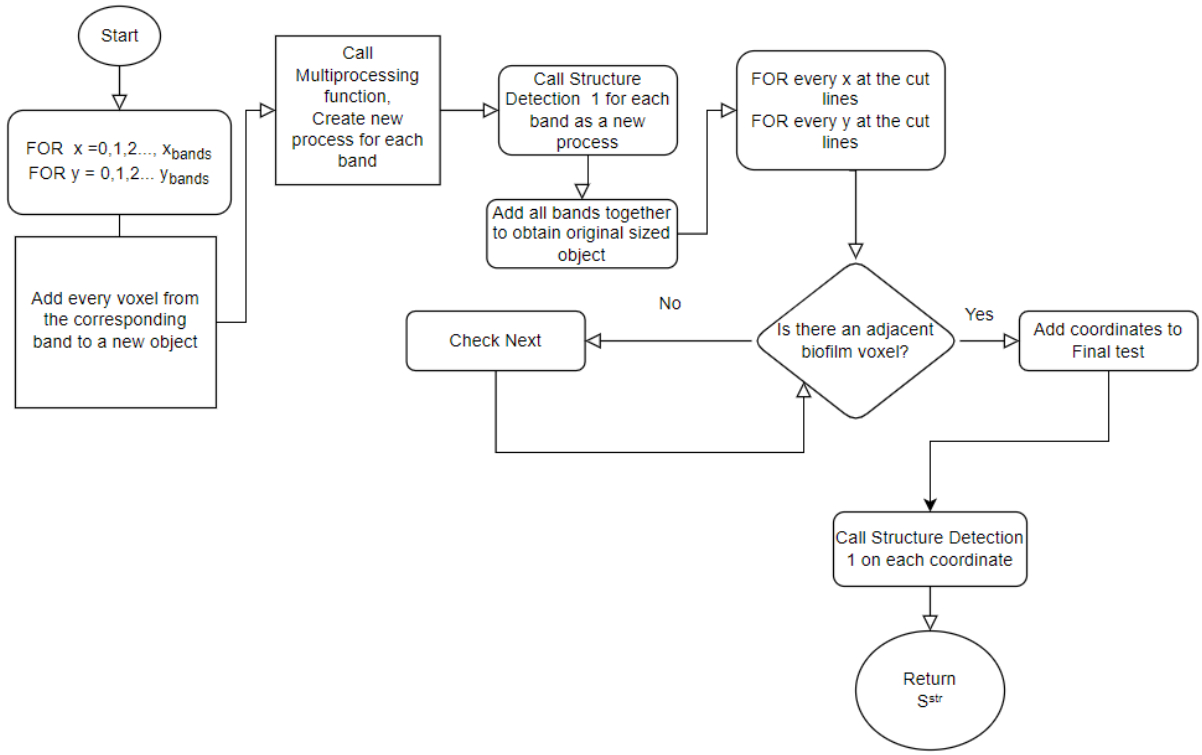


Figure 12: Parallel Processing schematization and explanation

To summarize, parallel processing takes place through the application of five distinct steps:

1. First, all coordinates for a given band must be calculated, and all bands are made into a distinct matrix.
2. All bands are processed with the previously described corresponding algorithm.
3. All individual results are added, and the original matrix is mostly restored.
4. Every coordinate where an addition occurred, has every point checked for biofilm existence.
5. For every point where biofilm is found, the original algorithm is run for final continuity testing.

Figure 13 Illustrates these 5 steps schematically.

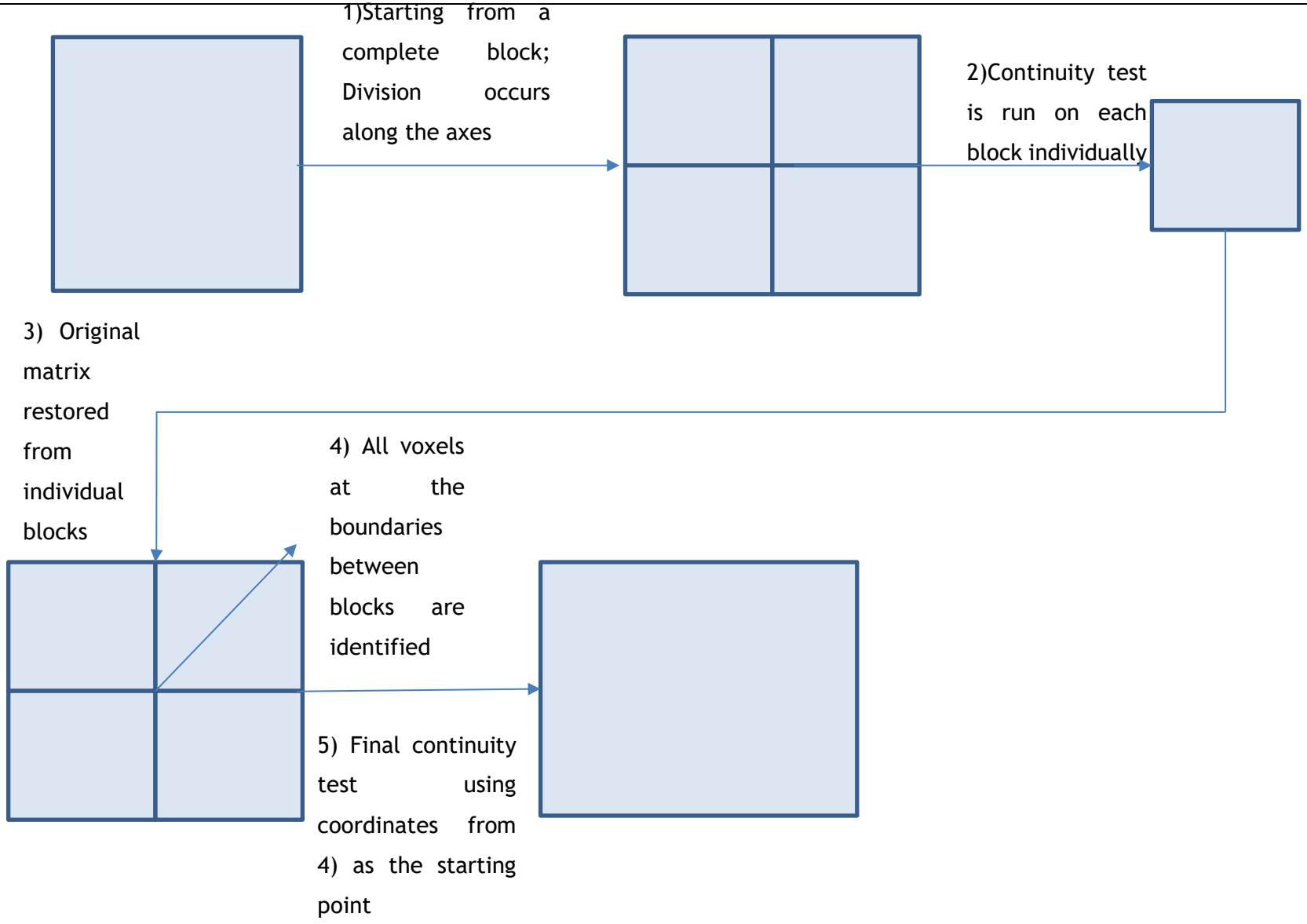


Figure 13: High Level description of parallel steps

This series of steps ensures that the parallel implementation gives the same results as a completely sequential approach.

At the end of the execution of the parallelized tests, an object identical to the one that would be obtained through the non-parallelized, previously described algorithm is achieved.

3.2 Materials

As a way to confirm their functioning by implementing the algorithms presented in the previous section, code was written in the Python programming language.

Python is a high-level, general-purpose programming language. It's widely known for its ease of use and code readability. It's highly customizable through the use of modules, which allow it to be used in a variety of fields. There is a very large amount of libraries available, with a wide range of functionalities, from Data analytics to graphical user interfaces (Kuhlman, 2012). These characteristics made it ideal for the development of this work, as there were libraries available for the main challenges of this work, which were matrix manipulation and image reading and writing.

A brief description of the libraries used is found in the following section:

NumPy

NumPy stands for numerical python, and is a library used to allow python to work with traditional arrays, instead of its default lists. It consists of a series of pre-compiled C code and allows for efficient and fast processing of large amounts of data (Harris et al., 2020). It is the basis for other libraries which require the array structure to function properly.

OpenCV (cv2)

OpenCV, which stands for open Computer Vision, is an open-source library used for machine learning and image processing. It allows, among other functionalities, the conversion of images into a 3D array or matrix, for further processing. It allows for this work to be developed, as it reads the biofilm images and turns them into an array which is then manipulated to identify the biofilm properties.

Pandas

Pandas allows for the creation of DataFrames, an object type that can be imported or exported to excel. It is mostly used to allow the exportation of the processed results and properties (McKinney, 2011), (Severance, 2009).

It should be noted that NumPy holds special significance to this work, as it is what allows for all the critical matrix operations which make up most of the tasks. All other libraries were used for the simplification of auxiliary work.

Collaboration platform GitHub was also used as a means to allow for cooperative code-writing and peer reviewing as well as for code storage during the intermediate stages.

All the results were obtained through the use of a remote workstation running a UNIX System, on an Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz with 128 GB of RAM as there is a large sum of operations, depending on the size of the image.

Images were acquired with a Ganymede Spectral Domain (SD) OCT-Optical Coherence Tomography system (Thorlabs GmbH, Germany) with a wavelength of 930 nm. A visualization field of 3.66 mm_2.98 mm was used for the X-Z section. A refractive index of 1.40 was used.

All the examples shown in this work are from biofilm formed with a *Pseudomonas fluorescens* ATCC 13525 inoculum, supplied with a nutritional medium. Biofilm growth occurred in a CDC biofilm reactor, under a 2800 Reynolds number in water. The detailed experimental procedure and materials for these growth conditions are described in Pereira et al. (2008).

4 Results and discussion

In this chapter, three case studies will be shown. Initially, a complete analysis of a three-dimensional image will be presented, as well as the results obtained from the execution of the previously mentioned code on a complete OCT Image. Comparisons will be made with typical values found in literature for the system in study. In the second case, a comparison between two dimensional results and three-dimensional results for the same section of a given biofilm image will be shown, so as to better display the differences between both cases, as to show that information is lost. In the third case, analysis performed on the parallel processing parameters will be displayed, as an attempt to provide initial insights into the best way to utilize parallelism to reduce the processing time.

4.1 Complete three-dimensional analysis

Starting from a sampled three-dimensional image, of dimensions 720(x)/1024(z)/509(y) voxels, the series of processing steps can be shown and analyzed from the program outputs, which consist of a series of image files and an excel file. All these results were obtained from the same image and required a runtime of 178343 seconds (approximately 49 hours, 32 minutes).

4.1.1 Pre-processing Results

As is shown in Figure 14, a large section of the image is not actually biofilm.

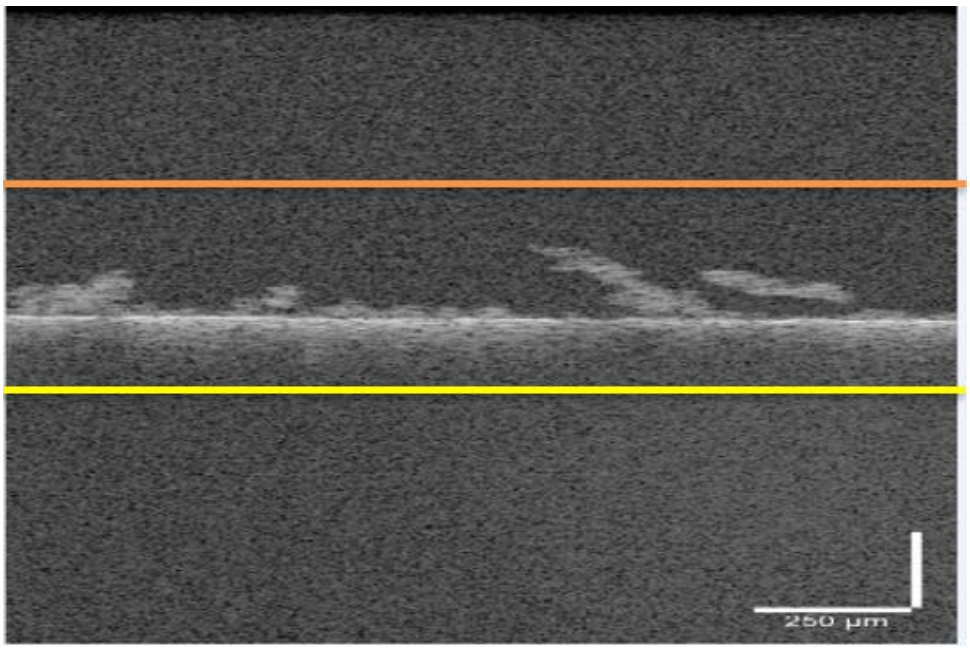


Figure 14: Initial two-dimensional slice of the image analysed.

Everything demonstrated above the orange line, and below the yellow one, is unnecessary information, as there is no biofilm to analyze.

Therefore, information is provided in an excel file to cut unnecessary z clutter to reduce processing time. For the image in question, the values provided are shown in Table 1 below.

Table 2: Limit values for z

	z_{min}	z_{max}	z_{lim}
"sample3D"	300	526	500

Upon pre-processing, the image is now equal to the one presented below in Figure 15.

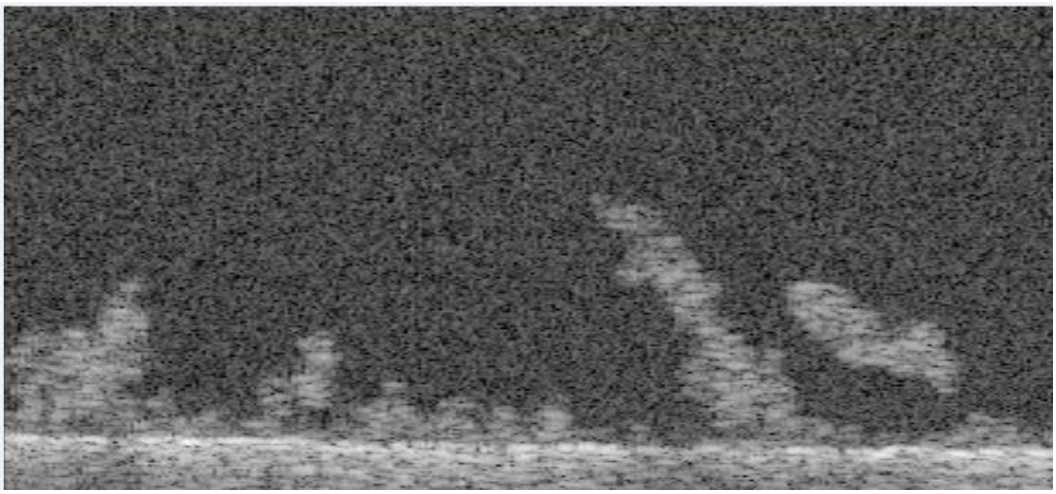


Figure 15: Pre-processed, two-dimensional slice of the analysed image

Care was taken to ensure no sections of biofilm were missing. For example, in y slice 209 (shown below in Figure 16), a series of pixels than clearly be identified as biofilm are found in a lower value of than in every other previous frame. The cut has to be executed above this value.

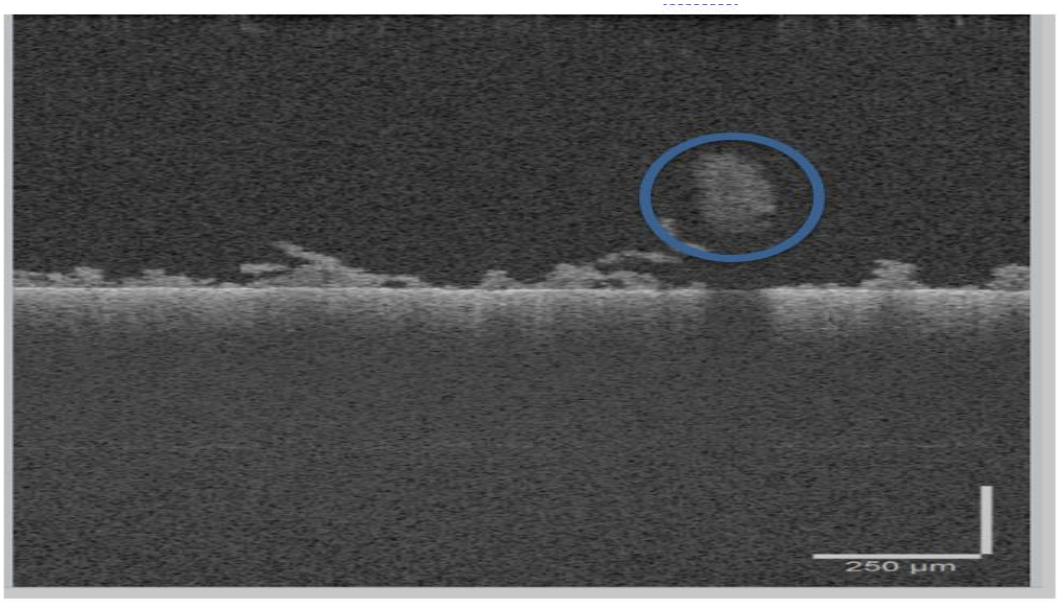


Figure 16: Slice for y = 209 of image showing highest biofilm point in all stacks.

4.1.2 Processing Results

The matrix created by binarization is then calculated and represented in the slices of output "binario" present in Figures 17 and 18.

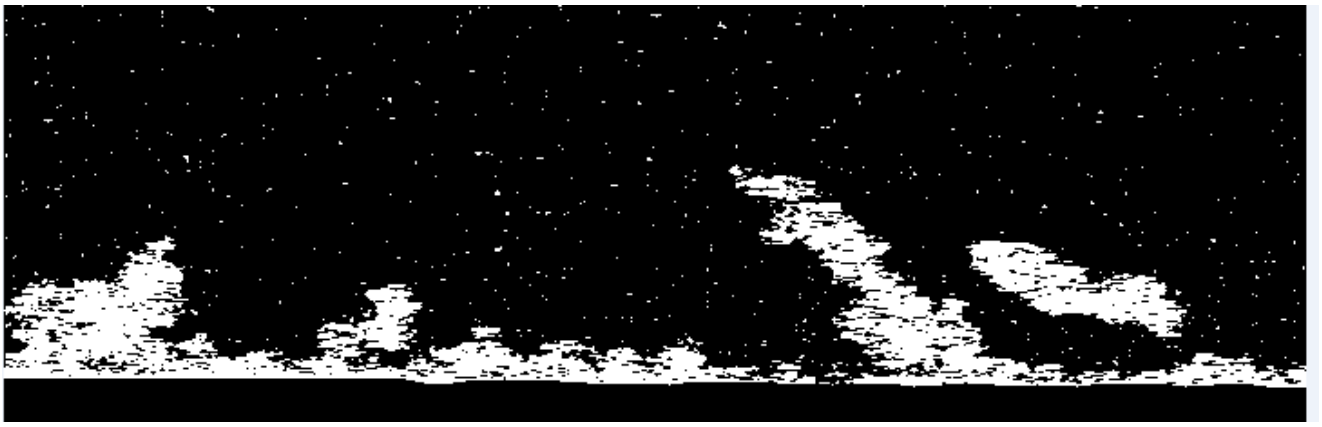


Figure 17: Slice corresponding to y = 1 of the binarized, processed image.

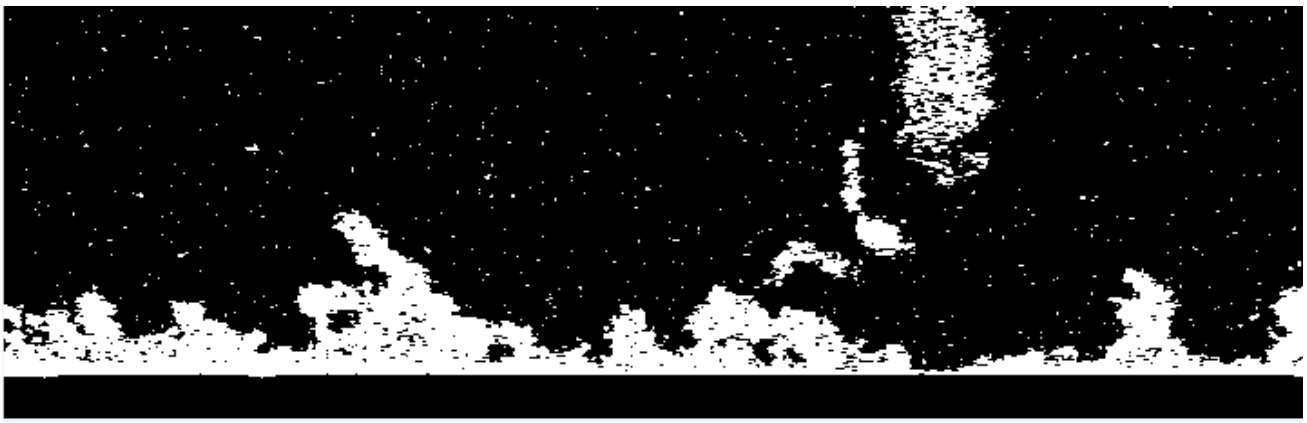


Figure 18: Slice corresponding to y= 409 of the binarized, processed image.

As can be seen, thresholding initially allows for all pixels in a stack to be classified as potential biofilm or background. Floating biofilm pixels may, however, still correspond to continuity in the y direction or visual artifacting.

After this, structure detection occurs, and is well represented in output 1, as shown in Figure 19.

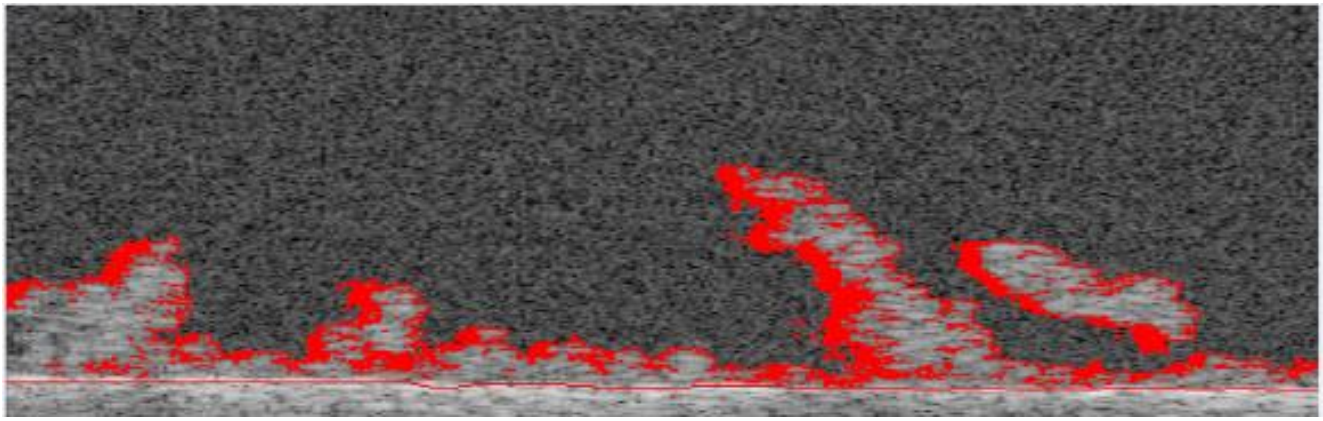


Figure 19: Biofilm and interface structure for first two-dimensional slice.

This demonstrates the functioning of structure detection quite well, as the interfaces are clearly visible.

In addition, Figure 20 shows directly adjacent slices in terms of y continuity, and it can be demonstrated that something that would initially, through two-dimensional analysis, be seen as floating, is actually connected to the biofilm as a whole through depth continuity. The relevant section is highlighted with a blue circle. All images were cropped and then added together and are identified.

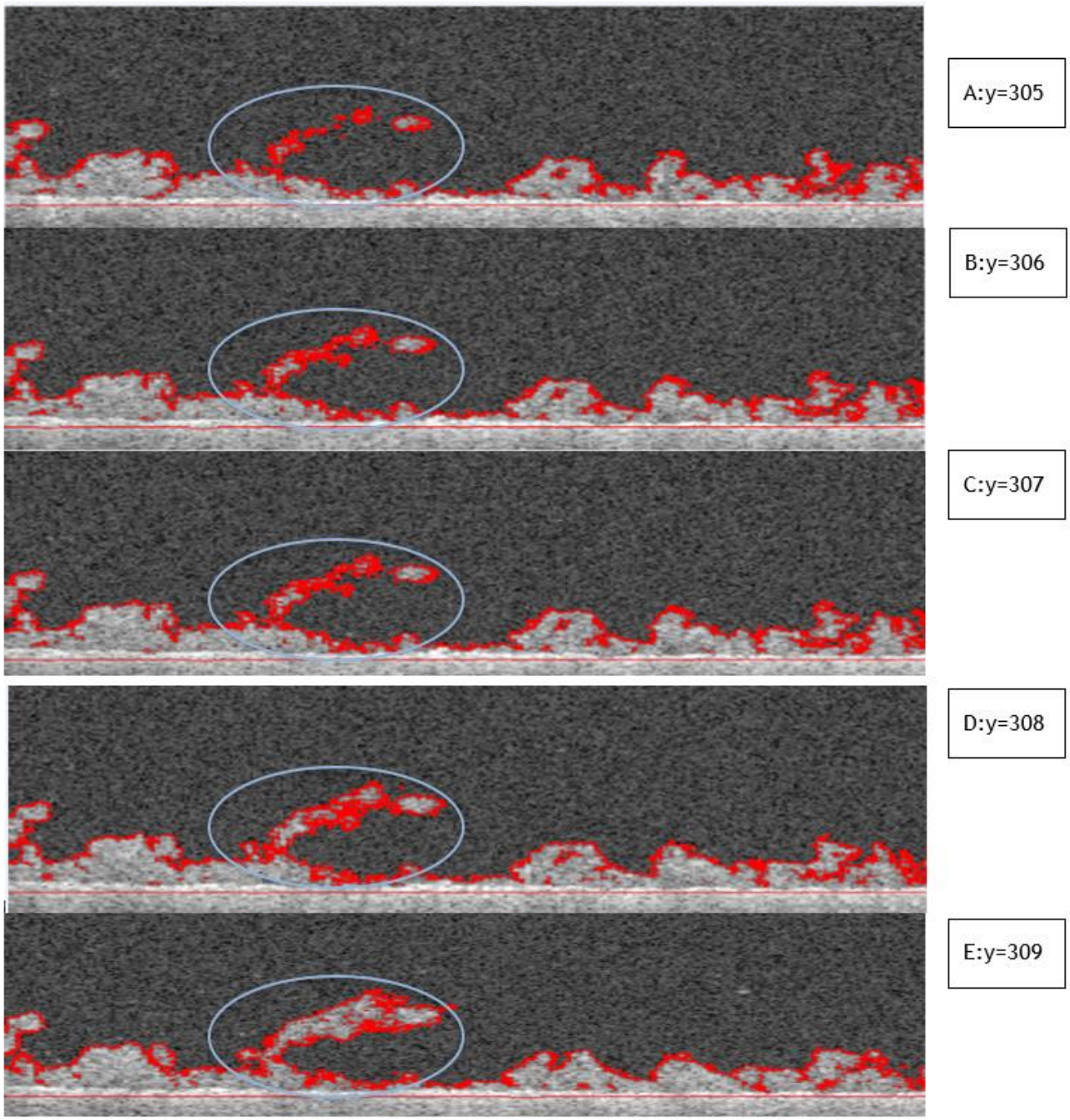


Figure 20: Adjacent biofilm stacks from y = 305 to y= 309

As the image proceeds further along in depth, it is clear that the section marked with the blue circle, initially perceived as floating biofilm, is actually part of the complete structure.

The same can be seen in many other examples throughout the image. This gives certainty in the correct functioning of the algorithm and is a good indicator that it's working as intended.

After the structure detection algorithms have run, every value required for the calculation of Output3 is available and is then exported as shown in Table 3.

Table 3: Output 3

	value
Average thickness	80.78 (μm)
Median thickness	60.76 (μm)
St. dev. thickness	63.01 (μm)
Max. thickness	451.38 (μm)
Roughness	44.74
Roughness coefficient	0.58
Compaction parameter	0.74
Threshold	105.39

The thickness values are consistent with the image analyzed. There is high variability in the biofilm, as it is very "wavy" in the y direction. The maximum thickness corresponds to the area attached to the base, where the biofilm is consistently present.

Upon completion of all the analysis, output 2 can be created, which corresponds to biofilm structure, with every other pixel colored as complete black. All greyscale intensities are also normalized. This is shown, for example, in Figure 21:

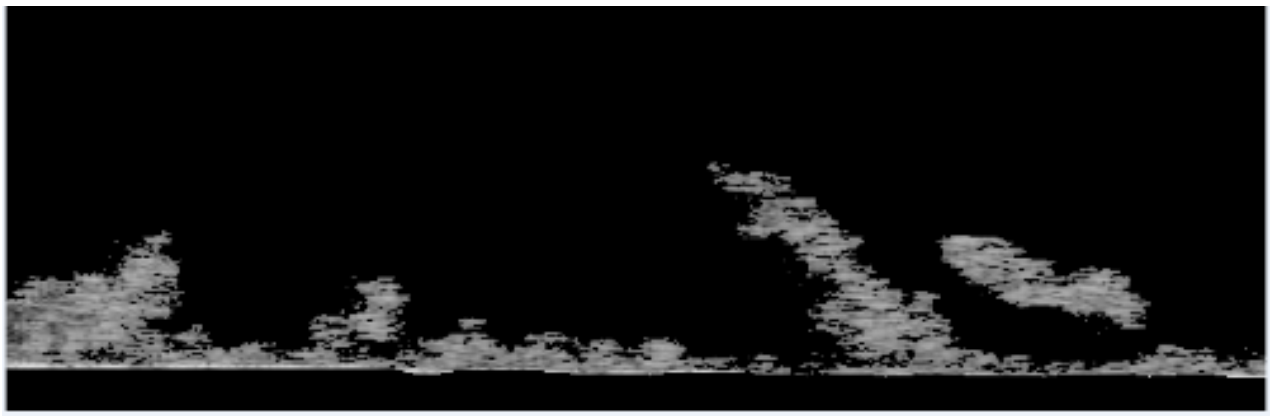


Figure 21: Structure of biofilm in y=0.

Output 4 is the final one and corresponds to a density map for the biofilm. It is a two-dimensional representation of biofilm density for the x and y directions. It is shown in Figure 22.

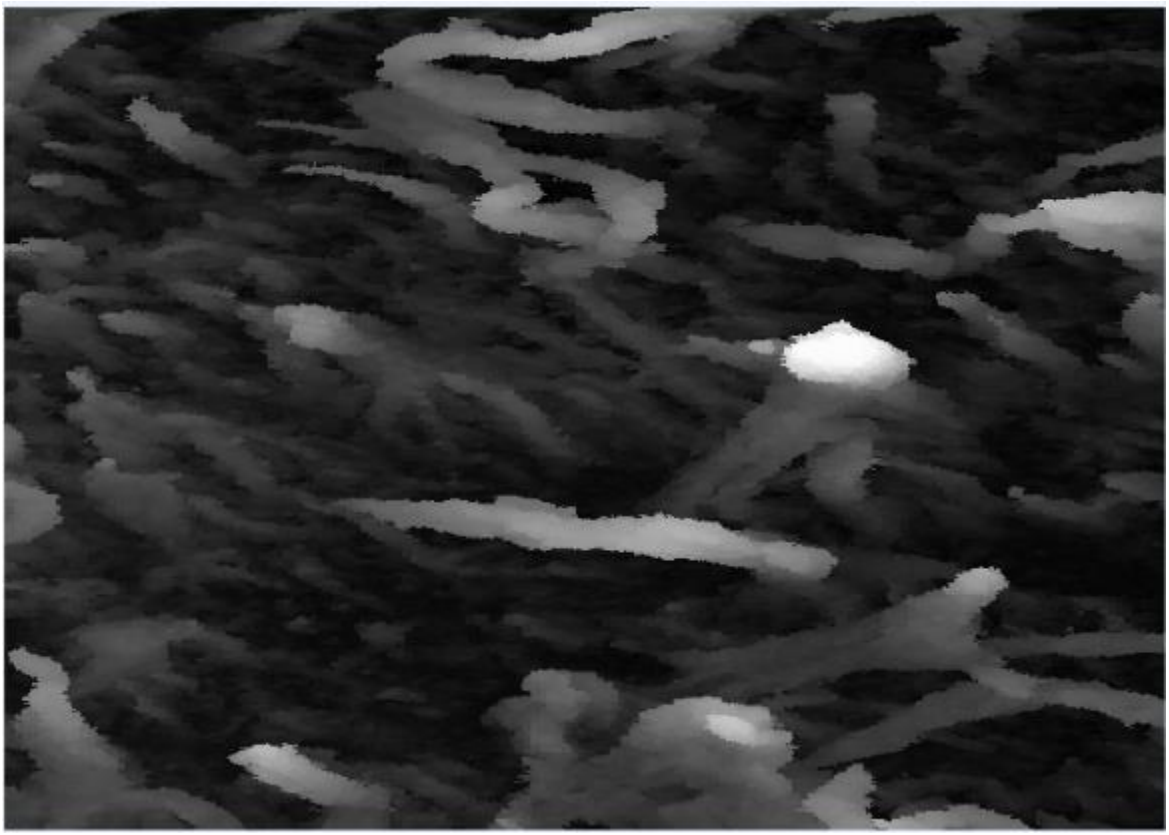


Figure 22: Thickness map of biofilm, in the x/y plane.

Brighter whites correspond to points where there is more often biofilm in the z dimension, indicating that the biofilm is thicker at that point.

4.2 Two-Dimensional comparison

In this section, a series of two-dimensional sections from the same three-dimensional image analyzed in the previous case are checked and compared one by one. This will allow for better comparison of the functioning of the developed software.

In Figure 23, the comparative results for the first stack are shown.

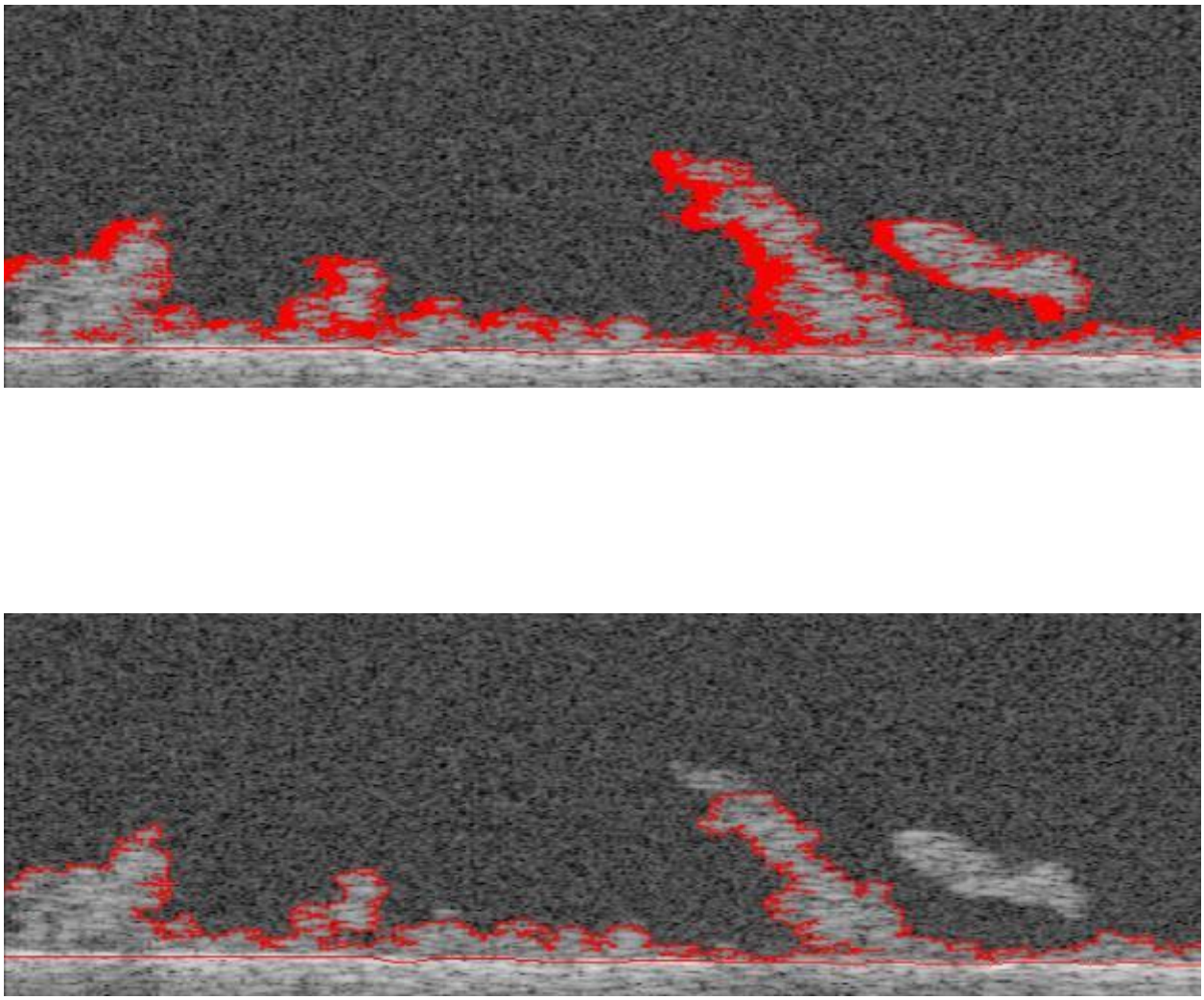


Figure 23: Comparison between first stack for three-dimensional and two-dimensional analysis. The top image and bottom images correspond to three-dimensional and two-dimensional analysis respectively.

As can be seen, on the right-hand side of the image, a series of points are not counted due to lack of direct continuity.

In Figure 24, comparative results for stack number 209 are shown.

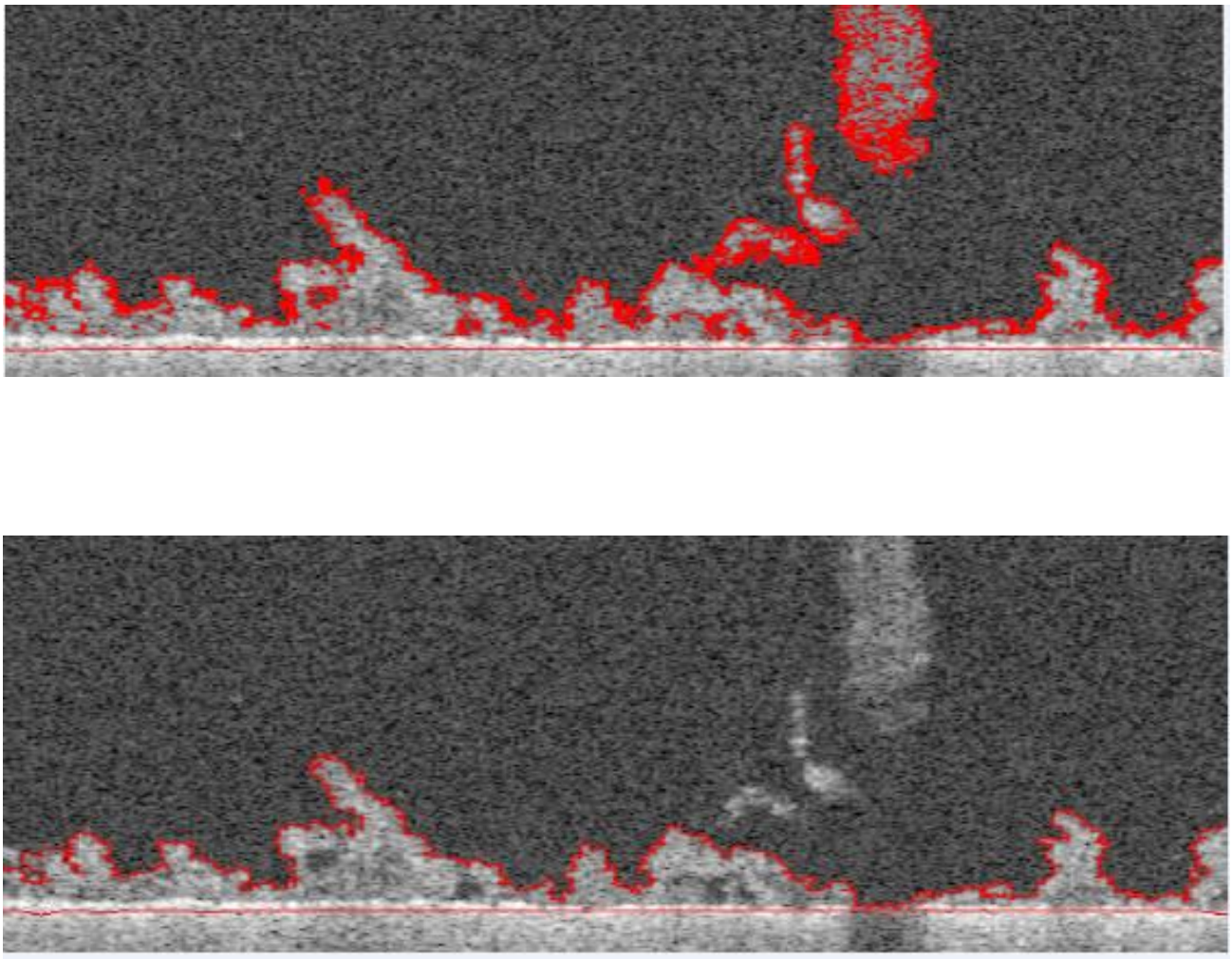


Figure 24: Comparative results for y =209. As in the previous example, the top image corresponds to three-dimensional analysis and bottom image to two-dimensional analysis.

4.3 Parametric optimization of parallel processing

A small amount of parametric optimization was done for the available code. To this end, the same image was run multiple times with different x and y cuts. This was done for the image presented in Figure 25. This image has an original size of 750(x)/400(z)/149(y) voxels, considerably smaller than the full-scale example shown in previous cases. This was done as the processing time for the full-scale image with the presently available algorithms was unfeasible for suitable testing within the available timeframe.

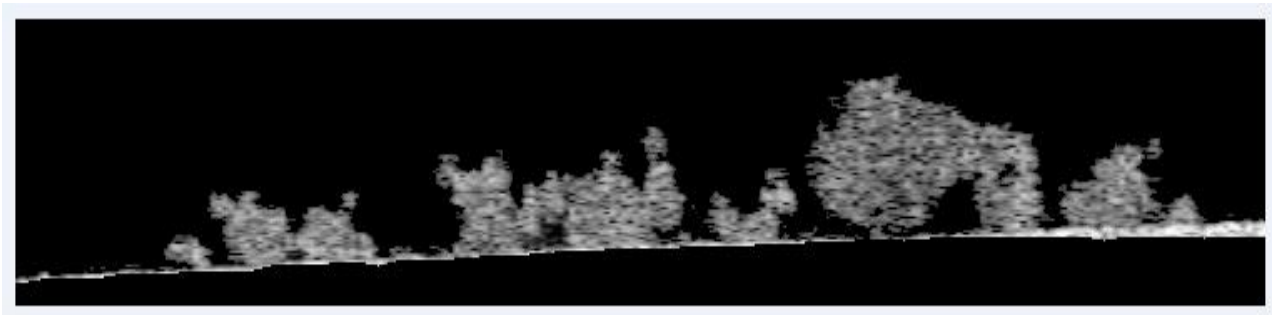


Figure 25: Output 2 for the y = 0 stack in the example image used to test parallel parameters

The runtime for each of these tests is shown in the table depicted in Table 4:

Table 4: Parallel processing cuts and processing time

	x_{bands}	y_{bands}	**Time (s)**
Case 1	1	15	13223
Case 2	3	10	36250
Case 3	3	3	2312
Case 4	5	5	3728
Case 5	7	7	19774
Case 6	10	3	5477
Case 7	15	1	2522

The parameters analysed were the number of x and y divisions. This is then correlated with the size of each band. More divisions will lead to smaller band sizes, and more parallel processes to calculate each band.

As can be seen, runtime appears to grow considerably for this test with y distance. This can be theorized to be due to the image's complexity in the y direction. As shown in Figure 26, which represents the last stack in the y direction for this image, there is a very large difference between the first stack and the last one.

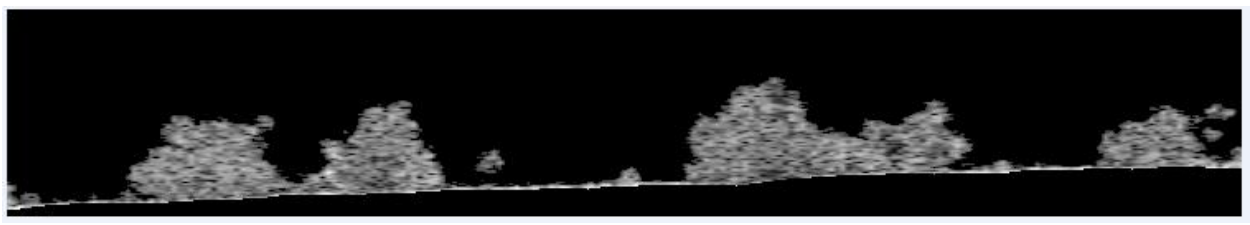


Figure 26: Output 2 for the y = 149 stack in the test image.

As this complexity grows, the weight of the processing for each final re-processing grows. This seems to indicate that the sectioning should occur in the direction where continuity is smoother.

The value for non-parallel processing is not shown in the chart and is equal to 86564 seconds. This corresponds to over double the highest value achieved with parallel processing.

These results, while preliminary unequivocally show 2 things. The first is that parallel processing was correctly implemented and works to reduce processing times considerably. The second is that a series of parametric studies are warranted. While this test image had relatively manageable runtimes, it would be in the best interest of efficiency to conduct further study into what the most effective parameters to use would be in different types of images.

5 Conclusions and future work

5.1 Conclusions

In this work, a series of algorithms were created to allow the complete processing of three-dimensional images of biofilm obtained from Optical Coherence Tomography. The algorithms were successfully implemented in Python and are able to return results for every biofilm image tested so far. These algorithms have not yet been translated to a graphic user interface or similar system, and are currently only available through direct Python scripting, requiring basic programming knowledge to operate.

All available algorithms are based on well-known principles and have been expanded in a functional way to encompass the three-dimensionality of the task at hand. This included the complete expansion of binarization, thresholding, and Structure Detection algorithms. All of these tasks can be considered complete.

Additionally to the main algorithms, parallel processing algorithms were devised and successfully implemented to allow for the reduction of processing time, with initial tests showing at least a 50% reduction in processing time when compared to completely sequential image treatment.

The combination of the developed sequential algorithms with the parallel processing ones allows for the complete characterization of any given biofilm image, within a reasonable timeframe. This corresponds to a complete assessment of the biofilm's structure and properties, resulting in the return of values such as biofilm thickness and roughness. This process is done automatically, with minimal user input. These algorithms are of use to any individual or research group who desires to adequately describe a biofilm system at the mesoscale.

After completion of algorithm writing and implementation, a series of examples were tested to ensure the functioning of the developed scripts. The software was able to adequately characterize and show biofilm morphology, as well as calculate biofilm parameters. Multiple examples are shown in this work of this correct functioning, and no problems were detected with the detection of any type of biofilm structure.

Direct comparison between previous existing algorithm's results and the new developed ones' was also performed, to unequivocally show the correct implementation of the three-dimensional principles required for biofilm characterization.

Upon initial testing, the parallel processing algorithms were also iterated upon, and initial testing was performed to obtain a clearer picture of the best implementation and use of the

developed algorithms so as to reduce processing time. This included a series of tests where different parameters were considered, and processing times measured. Preliminary results were obtained for the best values of these parameters to use, considering images of similar nature to those tested.

Overall, a major step forward in the study of biofilms through OCT can be said to have been taken, as the means to adequately characterize biofilms through three-dimensional OCT analysis is now available after the completion of this work.

5.2 Future Work

As previously mentioned, a series of improvements are possible:

Initially, the bottom interface detection algorithm could be improved upon. It currently relies on each biofilm section independently in two dimensions, which may lead to mistakes when the surface is highly uneven. Implementation of extra detection in the Y direction could help avoid such errors with the estimate.

While two algorithms were initially developed for Structure detection 1 and 2, only one was extended to the three-dimensional case. Additional tests and analysis could be done to ensure the most efficient implementation possible.

While the algorithms have proven to be functional, they are implemented as a series of python scripts, which are not very user-friendly. Implementation of a graphic user interface would be a valuable addition and help with pre-processing visualization.

Parallel processing is implemented from a basic standpoint but could be particularly scrutinized to increase efficiency and decrease processing times. While an initial search for parallel processing parameter's optima is described in this work, a lot more tests must be executed before conclusions can be taken about the best way to section each image.

Additionally, the base algorithms could be replaced with faster versions designed from the ground-up to function with parallel processing, such as Swensen-Wang or Wolff. These algorithms rely on random clustering and could effectively detect continuity in a less intensive fashion.

All outputs are currently shown in the original form for an OCT image, which is a stack of two-dimensional images. A three-dimensional model image shown from perspective would be a relevant output to be able to easily visualize three-dimensional structure.

6 Assessment of the work done

6.1 Objectives Achieved

By the end of the writing of this document, a fully functional version of the intended work is available. It adequately characterizes biofilms through the analysis of three-dimensional OCT images and calculates every parameter besides porosity. The main objective of the work can therefore be said to be complete.

All algorithms are functional and correctly updated to encompass three-dimensionality, and preliminary results show high levels of promise as to their correct implementation. The series of principles applied ensure the correct detection of biofilm structure, as initially intended. Additional, improvements have been identified such as development of a graphic user interface to allow for easier use.

Parallelization was successfully implemented and considerably reduced processing times for the tested cases and is a major step forward in the accessibility of this analysis' usage. Additionally, initial tests to confirm parallelization parameters are available to provide guidelines for future work and research.

Overall, the work can be said to have been a success, as the main objective was completed, and a major step forward was taken in allowing the characterization of biofilms through OCT analysis.

6.2 Other Work Carried Out

Before any other work was developed, initial timing for two algorithms available for two-dimensional processing was checked, so as to ascertain which would be more viable.

All the functions in the main algorithms were timed for multiple images, so as to obtain information as to the most optimizable sections.

6.3 Final Assessment

As of the writing of this final section, the main objective of the work was completed, as the image processing algorithms are available for use.

The precise nature of the work undertook was a challenge for me and has definitely been a learning experience which positively impacted my work ethic and experience.

7 References

Almasi, G. S., & Gottlieb, A. (1989). *Highly Parallel Computing*. Benjamin-Cummings Publishing Co., Inc.

An, D., & Parsek, M. R. (2007). The promise and peril of transcriptional profiling in biofilm communities. *Current Opinion in Microbiology*, *10*(3), 292–296. https://doi.org/10.1016/j.mib.2007.05.011

Aquino, M. M., Mota, C. C. B. O., Santos, J. P. C. L., Nascimento, P. L. A., Campello, S. L., & Gomes, A. S. L. (2019). Optical coherence tomography as a tool to visualize biofilm formation over removable prosthesis. *Clinical and Preclinical Optical Diagnostics II*, 11078_69. https://doi.org/10.1117/12.2527013

Azeredo, J., Azevedo, N. F., Briandet, R., Cerca, N., Coenye, T., Costa, A. R., Desvaux, M., di Bonaventura, G., Hébraud, M., Jaglic, Z., Kačániová, M., Knøchel, S., Lourenço, A., Mergulhão, F., Meyer, R. L., Nychas, G., Simões, M., Tresse, O., & Sternberg, C. (2017). Critical review on biofilm methods. *Critical Reviews in Microbiology*, *43*(3), 313–351. https://doi.org/10.1080/1040841X.2016.1208146

Bishop, P. L., Zhang, T. C., & Fu, Y.-C. (1995). Effects of biofilm structure, microbial distributions and mass transport on biodegradation processes. *Water Science and Technology*, *31*(1), 143–152. https://doi.org/10.1016/0273-1223(95)00162-G

Blauert, F., Horn, H., & Wagner, M. (2015). Time-resolved biofilm deformation measurements using optical coherence tomography. *Biotechnology and Bioengineering*, *112*(9), 1893–1905. https://doi.org/10.1002/bit.25590

Bridier, A., Dubois-Brissonnet, F., Boubetra, A., Thomas, V., & Briandet, R. (2010). The biofilm architecture of sixty opportunistic pathogens deciphered using a high throughput CLSM method. *Journal of Microbiological Methods*, *82*(1), 64–70. https://doi.org/10.1016/j.mimet.2010.04.006

Carrel, M., Beltran, M. A., Morales, V. L., Derlon, N., Morgenroth, E., Kaufmann, R., & Holzner, M. (2017). Biofilm imaging in porous media by laboratory X-Ray tomography: Combining a non-destructive contrast agent with propagation-based phase-contrast imaging tools. *PLOS ONE*, *12*(7), e0180374-. https://doi.org/10.1371/journal.pone.0180374

de Carvalho, C. C. C. R., & da Fonseca, M. M. R. (2007). Assessment of three-dimensional biofilm structure using an optical microscope. *BioTechniques*, *42*(5), 616–620. https://doi.org/10.2144/000112403

Derlon, N., Koch, N., Eugster, B., Posch, T., Pernthaler, J., Pronk, W., & Morgenroth, E. (2013). Activity of metazoa governs biofilm structure formation and enhances permeate flux during Gravity-Driven Membrane (GDM) filtration. *Water Research*, *47*(6), 2085–2095. https://doi.org/10.1016/j.watres.2013.01.033

di Bonaventura, G., Pompilio, A., Picciani, C., Iezzi, M., D'Antonio, D., & Piccolomini, R. (2006). Biofilm formation by the emerging fungal pathogen Trichosporon asahii: development, architecture, and antifungal resistance. *Antimicrobial Agents and Chemotherapy*, *50*(10), 3269–3276. https://doi.org/10.1128/AAC.00556-06

di Bonaventura, G., Spedicato, I., D'Antonio, D., Robuffo, I., & Piccolomini, R. (2004). Biofilm formation by Stenotrophomonas maltophilia: modulation by quinolones, trimethoprim-sulfamethoxazole, and ceftazidime. *Antimicrobial Agents and Chemotherapy*, *48*(1), 151–160. https://doi.org/10.1128/AAC.48.1.151-160.2004

Dreszer, C., Wexler, A. D., Drusová, S., Overdijk, T., Zwijnenburg, A., Flemming, H.-C., Kruithof, J. C., & Vrouwenvelder, J. S. (2014). In-situ biofilm characterization in membrane systems using Optical Coherence Tomography: Formation, structure, detachment and impact of flux change. *Water Research*, *67*, 243–254. https://doi.org/10.1016/j.watres.2014.09.006

Drexler, W., Morgner, U., Kärtner, F. X., Pitris, C., Boppart, S. A., Li, X. D., Ippen, E. P., & Fujimoto, J. G. (1999). In vivo ultrahigh-resolution optical coherence tomography. *Opt. Lett.*, *24*(17), 1221–1223. https://doi.org/10.1364/OL.24.001221

Fortunato, L., Bucs, S., Linares, R. V., Cali, C., Vrouwenvelder, J. S., & Leiknes, T. (2017). Spatially-resolved in-situ quantification of biofouling using optical coherence tomography (OCT) and 3D image analysis in a spacer filled channel. *Journal of Membrane Science*, *524*, 673–681. https://doi.org/10.1016/j.memsci.2016.11.052

Golyshin, P. N., dos Santos, V. A. P. M., Yakimov, M. M., Timmis, K. N., Golyshin, P. N., & Díaz, E. (2008). *Genomic insights into oil biodegradation in marine systems.*

Guilbaud, M., Piveteau, P., Desvaux, M., Brisse, S., & Briandet, R. (2014). Exploring the Diversity of Listeria monocytogenes Biofilm Architecture by High-Throughput Confocal Laser Scanning Microscopy and the Predominance of the Honeycomb-Like Morphotype. *Applied and Environmental Microbiology*, *81*. https://doi.org/10.1128/AEM.03173-14

Hall-Stoodley, L., Costerton, J. W., & Stoodley, P. (2004). Bacterial biofilms: from the Natural environment to infectious diseases. *Nature Reviews Microbiology*, *2*(2), 95–108. https://doi.org/10.1038/nrmicro821

Hannig, C., Follo, M., Hellwig, E., & Al-Ahmad, A. (2010). Visualization of adherent micro-organisms using different techniques. *Journal of Medical Microbiology*, *59*(1), 1–7. https://doi.org/10.1099/jmm.0.015420-0

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 Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hasan, S., Danishuddin, M., & Khan, A. U. (2015). Inhibitory effect of zingiber officinale towards Streptococcus mutans virulence and caries development: in vitro and in vivo studies. *BMC Microbiology*, *15*(1), 1. https://doi.org/10.1186/s12866-014-0320-5

Heydorn, A., Nielsen, A. T., Hentzer, M., Sternberg, C., Givskov, M., Ersbøll, B. K., & Molin, S. (2000). Quantification of biofilm structures by the novel computer program COMSTAT. In *Microbiology* (Vol. 146). www.im.dtu.dk

Hou, J., Wang, C., Rozenbaum, R. T., Gusnaniar, N., de Jong, E. D., Woudstra, W., Geertsema-Doornbusch, G. I., Atema-Smit, J., Sjollema, J., Ren, Y., Busscher, H. J., & van der Mei, H. C. (2019). Bacterial Density and Biofilm Structure Determined by Optical Coherence Tomography. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-46196-7

Klausen, M., Aaes-Jørgensen, A., Molin, S., & Tolker-Nielsen, T. (2003). Involvement of bacterial migration in the development of complex multicellular structures in *Pseudomonas aeruginosa* biofilms. *Molecular Microbiology*, *50*, 61–68. https://doi.org/10.1046/j.1365-2958.2003.03677.x

Kuhlman, D. (2011). *A Python Book: Beginning Python, Advanced Python, and Python Exercises*. Platypus Global Media. https://books.google.pt/books?id=1FL-ygAACAAJ

Lewandowski, Z., & Boltz, J. (2011). Biofilms in Water and Wastewater Treatment. In *Treatise on Water Science* (Vol. 4, pp. 529–570). https://doi.org/10.1016/B978-0-444-53199-5.00095-6

McKinney, W. (2011). *pandas: a Foundational Python Library for Data Analysis and Statistics*.

Mueller, L. N., de Brouwer, J. F., Almeida, J. S., Stal, L. J., & Xavier, J. B. (2006). *Analysis of a marine phototrophic biofilm by confocal laser scanning microscopy using the new image quantification software PHLIP*. https://doi.org/10.1186/1472-6785-6-1

Narciso, D. A. C., Pereira, A., Dias, N. O., Melo, L. F., & Martins, F. G. (2022). Characterization of biofilm structure and properties via processing of 2D optical coherence tomography images in BISCAP. *Bioinformatics*, *38*(6), 1708–1715. https://doi.org/10.1093/bioinformatics/btac002

Neu, T., & Lawrence, J. (2014). Advanced Techniques for In Situ Analysis of the Biofilm Matrix (Structure, Composition, Dynamics) by Means of Laser Scanning Microscopy. *Methods in Molecular Biology (Clifton, N.J.)*, *1147*, 43–64. https://doi.org/10.1007/978-1-4939-0467-9_4

OBEL. (n.d.). *Introduction to Optical Coherence Tomography (OCT)*. Retrieved June 6, 2022, from https://sites.google.com/site/obeluwa/research/oct/introduction-to-oct

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Piculell, M., Welander, P., Jonsson, K., & Welander, T. (2015). Evaluating the Effect of Biofilm Thickness on Nitrification in Moving Bed Biofilm Reactors. *Environmental Technology*, *37*, 1–28. https://doi.org/10.1080/09593330.2015.1080308

Pons, L., Délia, M.-L., & Bergel, A. (2011). Effect of surface roughness, biofilm coverage and biofilm structure on the electrochemical efficiency of microbial cathodes. *Bioresource Technology*, *102*(3), 2678–2683. https://doi.org/10.1016/j.biortech.2010.10.138

Rasmussen, K., Reilly, C., Li, Y., & Jones, R. S. (2016). Real-time imaging of anti-biofilm effects using CP-OCT. *Biotechnology and Bioengineering*, *113*(1), 198–205. https://doi.org/10.1002/bit.25701

Relucenti, M., Familiari, G., Donfrancesco, O., Taurino, M., Li, X., Chen, R., Artini, M., Papa, R., & Selan, L. (2021). Microscopy Methods for Biofilm Imaging: Focus on SEM and VP-SEM Pros and Cons. *Biology*, *10*(1), 51. https://doi.org/10.3390/biology10010051

Sanchez-Vizuete, P., le Coq, D., Bridier, A., Herry, J.-M., Aymerich, S., & Briandet, R. (2015). Identification of ypqP as a New Bacillus subtilis biofilm determinant that mediates the protection of Staphylococcus aureus against antimicrobial agents in mixed-species communities. *Applied and Environmental Microbiology*, *81*(1), 109–118. https://doi.org/10.1128/AEM.02473-14

Schlafer, S., & Meyer, R. L. (2017). Confocal microscopy imaging of the biofilm matrix. *Journal of Microbiological Methods*, *138*, 50–59. https://doi.org/10.1016/j.mimet.2016.03.002

Severance, C. R., Andrion, A., Blumenberg, S., & Hauser, E. (2016). *Python for Everybody: Exploring Data Using Python 3*. Charles Severance. https://books.google.pt/books?id=zjqzDAEACAAJ

Shen, Y., Monroy, G. L., Derlon, N., Janjaroen, D., Huang, C., Morgenroth, E., Boppart, S. A., Ashbolt, N. J., Liu, W.-T., & Nguyen, T. H. (2015). Role of Biofilm Roughness and Hydrodynamic Conditions in Legionella pneumophila Adhesion to and Detachment from Simulated Drinking Water Biofilms. *Environmental Science & Technology*, *49*(7), 4274–4282. https://doi.org/10.1021/es505842v

Sun, L., Liao, K., & Wang, D. (2015). Effects of magnolol and honokiol on adhesion, yeast-hyphal transition, and formation of biofilm by Candida albicans. *PloS One*, *10*(2), e0117695–e0117695. https://doi.org/10.1371/journal.pone.0117695

Tolker-Nielsen, T., & Sternberg, C. (2014). Methods for Studying Biofilm Formation: Flow Cells and Confocal Laser Scanning Microscopy. *Methods in Molecular Biology (Clifton, N.J.)*, *1149*, 615–629. https://doi.org/10.1007/978-1-4939-0473-0_47

Villacorte, L., Ekowati, Y., Neu, T., Kleijn, M., Winters, H., Amy, G., Schippers, J., & Kennedy, M. (2015). Characterisation of algal organic matter produced by bloom-forming marine and freshwater algae. *Water Research*, *73*, 216–230. https://doi.org/10.1016/j.watres.2015.01.028

Wagner, M., & Horn, H. (2017). Optical coherence tomography in biofilm research: A comprehensive review. *Biotechnology and Bioengineering*, *114*(7), 1386–1402. https://doi.org/10.1002/bit.26283

Wagner, M., Taherzadeh, D., Haisch, C., & Horn, H. (2010a). Investigation of the mesoscale structure and volumetric features of biofilms using optical coherence tomography. *Biotechnology and Bioengineering*, *107*(5), 844–853. https://doi.org/10.1002/bit.22864

Wagner, M., Taherzadeh, D., Haisch, C., & Horn, H. (2010b). Investigation of the mesoscale structure and volumetric features of biofilms using optical coherence tomography. *Biotechnology and Bioengineering*, *107*(5), 844–853. https://doi.org/10.1002/bit.22864

Wang, V. B., Chua, S.-L., Cao, B., Seviour, T., Nesatyy, V. J., Marsili, E., Kjelleberg, S., Givskov, M., Tolker-Nielsen, T., Song, H., Loo, J. S. C., & Yang, L. (2013). Engineering PQS Biosynthesis Pathway for Enhancement of Bioelectricity Production in *Pseudomonas aeruginosa* Microbial Fuel Cells. *PLOS ONE*, *8*(5), e63129-. https://doi.org/10.1371/journal.pone.0063129

West, S., Wagner, M., Engelke, C., & Horn, H. (2016). Optical coherence tomography for the in situ three-dimensional visualization and quantification of feed spacer channel fouling in

reverse osmosis membrane modules. *Journal of Membrane Science*, *498*, 345–352. https://doi.org/10.1016/j.memsci.2015.09.047

Xi, C., Marks, D. L., Schlachter, S., Luo, W., & D, S. A. B. M. (2006). High-resolution three-dimensional imaging of biofilm development using optical coherence tomography. *Journal of Biomedical Optics*, *11*(3), 1–6. https://doi.org/10.1117/1.2209962

Yerly, J., Hu, Y., Jones, S. M., & Martinuzzi, R. J. (2007). A two-step procedure for automatic and accurate segmentation of volumetric CLSM biofilm images. *Journal of Microbiological Methods*, *70*(3), 424–433. https://doi.org/10.1016/j.mimet.2007.05.022