



# DEVELOPMENT OF A TOMOGRAPHIC ATMOSPHERIC MONITORING SYSTEM BASED ON DIFFERENTIAL OPTICAL ABSORPTION SPECTROSCOPY

RUI FILIPE CANTARINO VALENTE DE ALMEIDA

Master in Biomedical Engineering

BIOMEDICAL ENGINEERING

NOVA University Lisbon  
September, 2021



**NOVA**

NOVA SCHOOL OF  
SCIENCE & TECHNOLOGY

DEPARTMENT  
OF PHYSICS

---

# DEVELOPMENT OF A TOMOGRAPHIC ATMOSPHERIC MONITORING SYSTEM BASED ON DIFFERENTIAL OPTICAL ABSORPTION SPECTROSCOPY

**RUI FILIPE CANTARINO VALENTE DE ALMEIDA**

Master in Biomedical Engineering

**Adviser:** Prof. Dr. Pedro Vieira  
*Auxilliary Professor, NOVA University Lisbon*

BIOMEDICAL ENGINEERING

NOVA University Lisbon  
September, 2021

## **Development of a Tomographic Atmospheric Monitoring System based on Differential Optical Absorption Spectroscopy**

Copyright © Rui Filipe Cantarino Valente de Almeida, NOVA School of Science and Technology, NOVA University Lisbon.

The NOVA School of Science and Technology and the NOVA University Lisbon have the right, perpetual and without geographical boundaries, to file and publish this dissertation through printed copies reproduced on paper or on digital form, or by any other means known or that may be invented, and to disseminate through scientific repositories and admit its copying and distribution for non-commercial, educational or research purposes, as long as credit is given to the author and editor.

*To my parents.*

## ACKNOWLEDGEMENTS

First and foremost, I need to thank FCT NOVA and FutureCompta, for having me and letting me work with them and develop this thesis. In addition, I want to acknowledge the contribution from the Portuguese Foundation for Science and Technology, who funded this work through the NOVA I4H doctoral programme, grant PDE/B-DE/114549/2016.

I would also like to thank my advisor, Prof. Pedro Vieira, for all the support he has given me, on many levels and throughout the years.

To my friends, old and new, I extend my utmost gratitude for helping me tilt my head back and laugh once in a while.

Finally, to my family, without whom I would have never been able to do anything, much less something as significant as what I present in this dissertation.

*“I know the pieces fit, 'cause I watched them fall away. ”*  
*(TOOL - Schism)*

*“Silenzio, Bruno! ” (Luca Paguro)*

## ABSTRACT

The aim of this thesis is to describe the design and development of a proof of concept for a commercially viable large area atmospheric analysis tool, for use in trace gas concentration mapping and quantification.

Atmospheric monitoring is a very well researched field, with dozens of available analytical systems and subsystems. However, current systems require a very important compromise between spatial and operational complexity. We address this issue asking how we could integrate the [Differential Optical Absorption Spectroscopy \(DOAS\)](#) atmospheric analysis technique in a [Unmanned Aerial Vehicle \(UAV\)](#) with tomographic capabilities.

Using a two-part methodology, I proposed two hypothesis for proving the possibility of a miniaturised tomographic system, both related to how the spectroscopic data is acquired. The first hypothesis addresses the projection forming aspect of the acquisition, its matrix assembly and the resolution of the consequent equations. This hypothesis was confirmed theoretically by the development of a simulation platform for the reconstruction of a trace gas concentration mapping.

The second hypothesis deals with the way in which data is collected in spectroscopic terms. I proposed that with currently available equipment, it should be possible to leverage a consequence of the Beer-Lambert law to produce molecular density fields for trace gases using passive [DOAS](#). This hypothesis was partially confirmed, with definite conclusions being possible only through the use of complex autonomous systems for improved accuracy.

This work has been a very important first step in the establishment of [DOAS](#) tomography as a commercially viable solution for atmospheric monitoring, although further studies are required for definite results. Moreover, this thesis has conducted to the development of a [DOAS](#) software library for Python that is currently being used in a production environment. Finally, it is important to mention that two journal articles were published from pursuing this work, both in important journals with Impact Factors over 3.0.

**Keywords:** [DOAS](#), tomography, [UAV](#), drones

## RESUMO

Era o objectivo deste trabalho descrever o processo de desenho e implementação de uma prova de conceito para um sistema de avaliação atmosférica comercialmente viável, para uso no mapeamento das concentrações de compostos traço na atmosfera.

A avaliação atmosférica é um campo muito estudado, estando no presente momento disponíveis para instalação diversos sistemas e subsistemas com estas capacidades. No entanto, é marcante o compromisso que se verifica entre a resolução espacial e a complexidade operacional destes equipamentos. Nesta tese, desafio este problema e levanto a questão sobre como se poderia desenvolver um sistema com os mesmos fins, mas sem este premente compromisso.

Usando uma metodologia a duas partes, proponho duas hipóteses para comprovar a exequibilidade deste sistema. A primeira diz respeito à formação da matriz tomográfica e à resolução das equações que dela derivam e que formam a imagem que se pretende. Confirmei esta hipótese teoricamente através do desenvolvimento de uma plataforma de simulação para a reconstrução tomográfica de um campo de concentrações fantoma.

A segunda é dirigida a aquisição de dados espectroscópicos. Proponho que com o material presentemente disponível comercialmente, deverá ser possível aproveitar uma consequência da lei de Beer-Lambert para retirar os valores de concentração molecular de gases traço na atmosfera. Foi apenas possível validar esta hipótese parcialmente, sendo que resultados mais conclusivos necessitariam de equipamentos automatizados dos quais não foi possível dispôr.

No final, este trabalho constitui um importante primeiro passo no estabelecimento da técnica de [DOAS](#) tomográfico como uma alternativa comercialmente viável para a análise atmosférica. Ademais, o desenvolvimento desta tese levou à escrita de uma biblioteca em Python para análise de dados [DOAS](#) actualmente usada em ambiente de produção. Por fim, importa realçar que dos trabalhos realizados no decorrer da tese foram publicados dois artigos em revistas científicas com *Impact Factor* acima de 3.

**Palavras-chave:** [DOAS](#), tomografia, drones

# CONTENTS

|   |             |
|---|-------------|
| <b>List of Figures</b>  | <b>xi</b>   |
| <b>List of Tables</b>   | <b>xiii</b> |
| <b>List of Listings</b>   | <b>xiv</b>  |
| <b>Acronyms</b>   | <b>xv</b>   |
| <b>1 Methods</b>  | <b>1</b>    |
| <b>1.1 First Hypothesis</b>   | <b>3</b>    |
| <b>1.1.1 Unmanned Aerial Vehicle</b>  | <b>3</b>    |
| <b>1.1.2 Collection System</b>  | <b>4</b>    |
| <b>1.1.3 Ground Station</b>   | <b>8</b>    |
| <b>1.1.4 Tomosim</b>  | <b>9</b>    |
| <b>1.2 Second Hypothesis</b>  | <b>17</b>   |
| <b>1.2.1 The Lambertian Hypothesis</b>  | <b>17</b>   |
| <b>1.2.2 Testing the Lambertian Hypothesis</b>                                    | <b>19</b>   |
| <b>2 Discussion</b>   | <b>22</b>   |
| <b>2.1 First Hypothesis</b>   | <b>22</b>   |
| <b>2.1.1 The UAV</b>  | <b>22</b>   |
| <b>2.1.2 Collection System</b>  | <b>23</b>   |
| <b>2.1.3 TomoSim</b>  | <b>23</b>   |
| <b>2.2 Second Hypothesis</b>  | <b>23</b>   |
| <b>2.2.1 Experiment - First Run</b>   | <b>23</b>   |
| <b>2.2.2 Experiment - Second Run</b>  | <b>23</b>   |
| <b>Bibliography</b>   | <b>24</b>   |
| <b>Appendices</b>   |             |
| <b>A Forest Fire Finder: DOAS application to long-range forest fire detection</b> |             |
|   | <b>26</b>   |

## CONTENTS

---

|   |           |
|---|-----------|
| <b>B TomoSim: a tomographic simulator for Differential Optical Absorption Spectroscopy</b>  | <b>40</b> |
| <b>C Systematic Review of DOAS Tomography</b>   | <b>61</b> |
| <b>D Testing Lambertian Absorption - handout protocol for the experiment of Section 1.2</b> | <b>84</b> |
| <b>E Telescope - Spectrometer Coupling</b>  | <b>94</b> |
| <b>F Refractor Telescope Design</b>   | <b>96</b> |

## Annexes

## LIST OF FIGURES

|      |   |    |
|------|---|----|
| 1.1  | General diagram of the atmospheric monitoring system, in an overview style. . . . .   | 2  |
| 1.2  | Schematic representation of the intended pollution measurement geometry. The second hypothesis of the Research Question (RQ) tells us that to get the Region Of Interest (ROI) we can subtract the concentrations obtained on the left of the ROI from the ones obtained to the right of the ROI. . . . .   | 2  |
| 1.3  | A schematic for the custom-designed drone system on which the pollution monitoring system would be mounted. . . . .   | 3  |
| 1.4  | Conducted experiment to evaluate advantages of using a direct light connection instead of fibre optics. The spectrometer support (that can be seen in red) was designed in Autodesk Inventor and 3D printed. . . . .  | 5  |
| 1.5  | Plots obtained through the collection of light managed by the assemblies depicted in Figure 1.4 . . . . .   | 5  |
| 1.6  | Telescope - Spectrometer coupling. 3D design rendered through Autodesk Inventor. . . . .  | 6  |
| 1.7  | The spectrum captured on a cloudy but luminous day using the ET90 telescope and the Avantes Mini spectrometer. . . . .  | 7  |
| 1.8  | UML diagram for the DOAS library. The Object Oriented Programming (OOP) approach that was followed allows for an instrument-oriented experiment parametrisation, which is not available in any other software . . . . .   | 9  |
| 1.9  | A flowchart describing the beginning stages of the Tomosim simulation. . . . .  | 10 |
| 1.10 | Flowchart representing the Tomosim operation during the reconstruction phase. . . . .   | 10 |
| 1.11 | Illustration of the projection gathering algorithm based on the assembly of groups of acquisitions as fanbeam projections. The lines presented on the drawing come from the right to the left. Each is a ray in fan, corresponding to one stop of the drone. In the second measurement moment, the drone moves to the exit point on the left and repeats the first moment collection. Here, both fans ans rays are separated by 5 degrees and there are 6 rays within each fan. A "real life"acquisition would feature a lot more rays, but that would be graphically complicated for the reader. Both ray and fan angular distance are customisable at runtime. Map taken from Google Maps in 2019. ©Google. . . . . | 11 |

## LIST OF FIGURES

---

|      |  |    |
|------|--|----|
| 1.12 | <i>P<sub>2</sub></i> Calculation. The software uses the position of the drone, as defined by the vertical angle, $\beta$ , and the distance between the drone and the centre of the trajectory, $D$ , to determine $P_2$ through the solution of a second degree equation. . . . . | 12 |
| 1.13 | A graphical representation of the new spectral phantom, custom built for the TomoSim application. . . . .  | 14 |
| 1.14 | Error estimation graphical representation. Note errors are extremely exaggerated for visualisation purposes. . . . .   | 15 |
| 1.15 | First draft of the Unified Modelling Language (UML) diagram for the new architectural model of the Tomosim library. Note the usage of several classes from the DOASlibrary described in Section 1.1.3. . . . .   | 17 |
| 1.16 | Location of observer points for the physical experiment. . . . .   | 18 |
| 1.17 | Published spectrum of the CREE XHP50.2 LED light. The LED that was used in this experiment corresponds to the blue line [2]. . . . .   | 19 |
| 1.18 | Spectrum measured with the experiment spectrometers, at a distance of approximately 50 m from the torchlight. . . . .  | 19 |
| E.1  | Mechanical drawing for the telescope-spectrometer coupling device. . . . .   | 95 |
| F.1  | Refractor telescope design: basic schematic . . . . .  | 96 |
| F.2  | 3D schematic for the refractor telescope design. . . . .   | 97 |
| F.3  | Spot diagram for the refractor telescope. . . . .  | 97 |
| F.4  | 3D rendering of the optical system . . . . .   | 98 |

## LIST OF TABLES

|     |  |    |
|-----|--|----|
| 1.1 | Specifications table for the Omegon MightyMak 60 telescope [10]. . . . .   | 7  |
| 1.2 | Table summarising the new phantom's construction details, as a sum of 5 Gaussian profiles and an ellipse designed using TomoPhantom. In this table, Co is the object's amplitude, Xo and Yo are its center coordinates, and a and b are the objects half-widths. The table is constructed using TomoPhantom's particular syntax and more information can be obtained at [8]. . . . . | 13 |
| 1.3 | Summary table for the two experiment assemblies. Note the difference in terms of material, due to the two different roles both assemblies play during the experiment. This is translated into not having the need of an artificial light source in the West Bank's assembly. . . . .   | 20 |
| 1.4 | Actions are the indivisible unit upon which each capture is built. The prescribed actions for this experiment are described in this table. . . . .   | 20 |
| 1.5 | Captures are particular sets of actions that are conducted according to a specific order, depending on the time of day on which the capture is run. This table describes the prescribed captures on which this experiment consisted.   | 20 |
| 2.1 | Component list for the hexacopter that was designed as a part of this thesis.  | 22 |

## **LIST OF LISTINGS**

## ACRONYMS

|                      |   |
|----------------------|---|
| <b>API</b>           | Application Programming Interface <a href="#">13</a>  |
| <b>ASCII</b>         | American Standard Code for Information Interchange <a href="#">20</a>                             |
| <b>DOAS</b>          | Differential Optical Absorption Spectroscopy <a href="#">vii, viii, xi, xii, 8, 9, 16, 17, 18</a> |
| <b>FBP</b>           | Filtered BackProjection <a href="#">13</a>  |
| <b>FFF</b>           | Forest Fire Finder <a href="#">6</a>  |
| <b>IEP</b>           | InfraEstruturas de Portugal <a href="#">19</a>  |
| <b>MLEM</b>          | Maximized Likelihood Expectation Maximization <a href="#">13, 14</a>                              |
| <b>MSE</b>           | Mean Squared Error <a href="#">16</a>   |
| <b>OOP</b>           | Object Oriented Programming <a href="#">xi, 8, 9, 16</a>  |
| <b>Project ATMOS</b> | ATmosphere MOnitoring System Project <a href="#">3</a>  |
| <b>ROI</b>           | Region Of Interest <a href="#">xi, 1, 2, 11, 12, 16</a>   |
| <b>RPi</b>           | Raspberry Pi <a href="#">4</a>  |
| <b>RQ</b>            | Research Question <a href="#">xi, 1, 2</a>  |
| <b>RTK GPS</b>       | Real Time Kinematic Global Positioning System <a href="#">15</a>                                  |
| <b>RTK-GPS</b>       | Real Time Kinematic GPS <a href="#">4, 8</a>  |
| <b>SART</b>          | Simultaneous Algebraic Reconstruction Technique <a href="#">13</a>                                |
| <b>SMA</b>           | SubMiniature version A <a href="#">4</a>  |
| <b>UAV</b>           | Unmanned Aerial Vehicle <a href="#">vii, ix, 3, 8, 22</a>   |
| <b>UML</b>           | Unified Modelling Language <a href="#">xii, 8, 17</a>   |
| <b>USB</b>           | Universal Serial Bus <a href="#">4, 20</a>  |

## TODO LIST

|  |    |
|--|----|
| link to theory . . . . .                   | 13 |
| Link to theory . . . . .                   | 13 |
| link to theory . . . . .                   | 13 |
| link to theory . . . . .                   | 14 |
| link to theory . . . . .                   | 16 |
| Figure: This is actually a table . . . . . | 22 |

## METHODS

Macroscopically, the approach to the RQ was conducted by working with two hypothesis:

**First Hypothesis:** The definition of a particular set of algorithmically defined projections in such a manner that they might be used for tomographic reconstruction of column densities of trace gases in the atmosphere, in a given ROI;

**Second Hypothesis:** We can retrieve the column density for a given trace gas (or set of trace gases) between two points by performing a spectral measurement in both of these points in the same direction and subtracting them one from the other.

Testing the first hypothesis implies the design of a whole system and requires a very diverse and multidisciplinary approach. I have already covered that the system should be drone-mounted. But which drone? And what is the collection system? And then again, what trajectory shall this drone adopt? Do we know it works?

Figure 1.1 is a general diagram of the envisaged solution. Section 1.1 aims to provide a complete description of how I have arrived at the system that I propose, and of the subsystems that comprise it. Subsection 1.1.1 is dedicated to the description of the drone and its subsystems, the blue box in the figure, Subsection 1.1.2 is about the particularities of the collection system and corresponds to the red box in the figure. The ground station is addressed in Subsection 1.1.3, which is the orange box in the figure. Finally, the trajectory simulation platform (green box) is the matter of Subsection 1.1.4.

## CHAPTER 1. METHODS

---

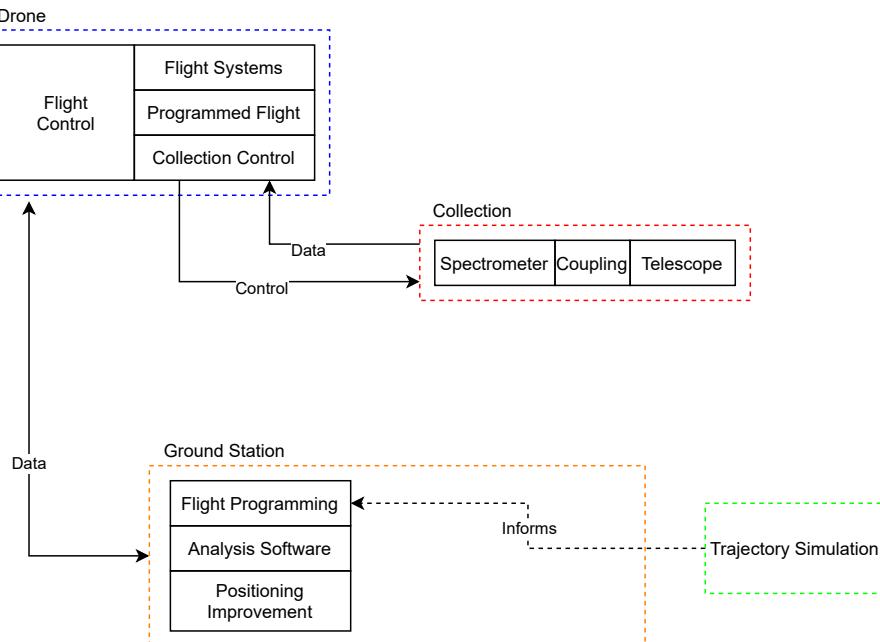


Figure 1.1: General diagram of the atmospheric monitoring system, in an overview style.

On Section 1.2, this chapter moves to a different subject. It is aimed at the testing of the second hypothesis. Figure 1.2 is a pictorial representation of what the proposed pollution monitoring system is designed to (ideally) do. Our second hypothesis is based on Lambertian theory, and tells us that one way of getting the density values in the ROI is to subtract whatever density is obtained to the left of the ROI from the density obtained by measuring on the right of the ROI. It is clear that, from a mathematical point of view, this is the case. But can I measure it with current-day off-the-shelf equipment? Subsection 1.2.1 addresses the theoretical perspective and reasoning for this idea. Subsection 1.2.2 presents the experiment that I have designed to test it.

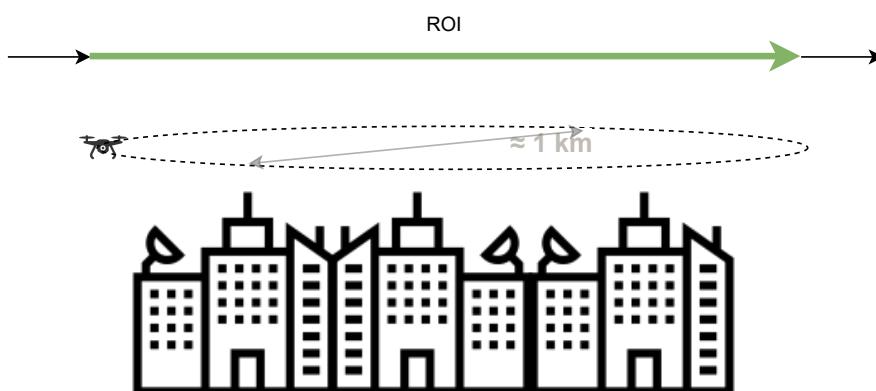


Figure 1.2: Schematic representation of the intended pollution measurement geometry. The second hypothesis of the RQ tells us that to get the ROI we can subtract the concentrations obtained on the left of the ROI from the ones obtained to the right of the ROI.

## 1.1 First Hypothesis

### 1.1.1 Unmanned Aerial Vehicle

Although it was not possible to assemble the final physical system onto an actual drone, I have managed to specify it completely in a custom design for which Figure 1.3 is a basic schematic <sup>1</sup>.

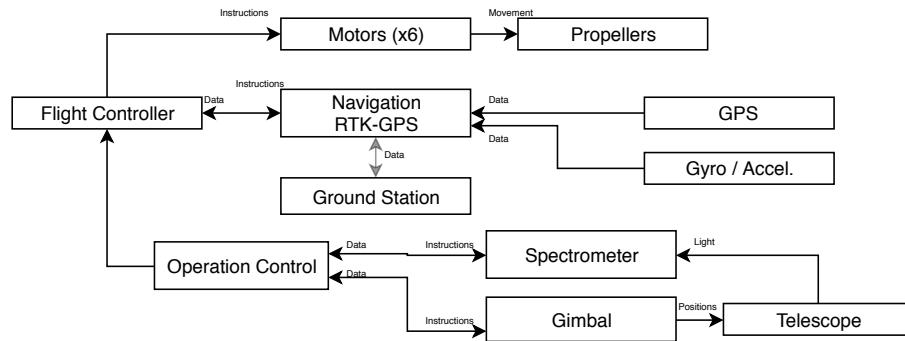


Figure 1.3: A schematic for the custom-designed drone system on which the pollution monitoring system would be mounted.

Our flight time requirements are between 30 to 40 minutes for each measurement cycle. During this time, the drone would have to carry itself and the collection equipment while still being able to land in a safe position before recharging or replacing the batteries. Sufficiently capable commercial systems were prohibitively expensive, thus I had to design my own drone.

I began by selecting the type of drone I would use. Hexacopter models are known around enthusiasts to be more efficient and stable, especially if there is a high-payload requirement. They have several advantages over the more "traditional" quadcopter designs. The most important is certainly the fact that they can fly and land safely in case of one or even two motors malfunctioning. Moreover, they are much stabler than their 4 motor counterpart, and they are for the most part immune to (up to) moderate winds. Finally, they are able to climb higher, since they are less affected by thin air. They also have disadvantages, such as their steeper cost, their size and their inherently higher complexity. These drawbacks were easily manageable though, through the European funds of project [ATmosphere MOnitoring System Project \(Project ATMOS\)](#) and the technical abilities of the team in which the project was being developed.

I have therefore settled on a DJI S900 frame. This was, at the time of selection, being discontinued by the Chinese brand. However, this model remained popular among custom builders because of the familiarity people already had with it and the fact that it was still commonly available in specialised retailers. I was advised by two specialists to increase the device's breadth through the addition of custom made 368mm arms. This inclusion served two purposes: on the one hand, it reduced weight, since the original arms were made of aluminium and the new ones were carbon fibre; on the other hand, they granted the room needed for wider propellers, with 17" blades. To

<sup>1</sup>To the reader: although this is still a fairly advanced design, recent advances in drone technology might render it a bit obsolete in some aspects. Please bear in mind that this design was finished in March 2019

accommodate the change in propeller size, it was also advisable to replace the original motors by more powerful units. In this case, I chose the DJI E1200. Larger propellers, lighter body and more powerful motors led to significant improvements in both flight time and carrying capacity. According to the manufacturer, this configuration is able to handle a maximum weight per rotor of 3900g, or almost 24 kg in total. Assuming our drone to weight 6 kg, this gives us a very comfortable weight margin for our payload, which in any case should be slightly below 2 kg [6].

A Pixhawk flight controller takes care of the aerial dynamics, movement and positioning. This flight controller comes with all the needed sensors (gyroscopes, barometers, magnetometers, accelerometers, etc.), requiring only an external navigation unit. The unit natively supports [Real Time Kinematic GPS \(RTK-GPS\)](#), a combination of inertial sensors and traditional satellite navigation data that achieves positioning precisions of a few tens of centimetres [15, 14].

### 1.1.2 Collection System

The flight controller unit is in permanent communication with the collection system, through the spectrometer controlling computer. This is an [Raspberry Pi \(RPi\)](#)-based single board device. This equipment controls the spectral acquisition process, which is conducted by an Avantes Mini spectrometer with 2048 channels and a [Universal Serial Bus \(USB\) 3.0](#) connection. The Avantes spectrometer can be interfaced with both Unix-based and Microsoft Windows operating systems through a set of software libraries that are available for download at this manufacturer's website. There are several physical interfaces available for both types of operating system. Nowadays, the most common and expedient way to run this connection is through the [USB](#) port. This is also the connection that allows a higher data throughput (allowing smaller integration times to be used). In the end, the proposed system will only run in Unix-based computers, but the Windows version was very important for initial experiments, which at the time were programmed in C# (code included in Appendix ??). The spectrometer control flow is very different in the two approaches. The Unix version works by continuously polling the spectrometer for the spectral data, which is the somewhat obvious strategy. The Windows version uses a sophisticated technique: Windows messages. This is a set of fixed-value event flags that are fired at the Operating System level and can be intercepted by running programs.

This type of spectrometer is usually shipped with an [SubMiniature version A \(SMA\)](#) connector, for direct connection of a fibre optics cable. While this is ideal as a bench-top solution and when size and weight restrictions are looser, we found it not to be the best fit for our particular case. We tend to think of fibre optics as an almost lossless medium for data transmission. This is, of course, true, but it does not hold for its connectors. They represent one of the most significant sources for signal loss in an optics fibre line. Avantes does not mention any value for [SMA](#) connector losses, but the traditional figure that appears in most manufacturer's catalogues is around 1dB [5, 9, 1]. To test how a direct connection fares when compared to a fibre connection, I designed a spectrometer support that can be attached to the telescope via its red-dot rail and fabricated it using a 3D printer. Technical drawings available in Appendix E.

The assembly is displayed in Figure 1.4. In this experiment, I have found 75%

energy difference between the fibre optics cable and a direct measurement in which the light goes directly from the telescope into the spectrometer's connector. These results are plotted in Figure 1.5.



(a) This assembly uses a traditional fibre optics connection to transport light between the telescope and the spectrometer.



(b) This assembly allows light to pass directly from the telescope to the spectrometer.

Figure 1.4: Conducted experiment to evaluate advantages of using a direct light connection instead of fibre optics. The spectrometer support (that can be seen in red) was designed in Autodesk Inventor and 3D printed.

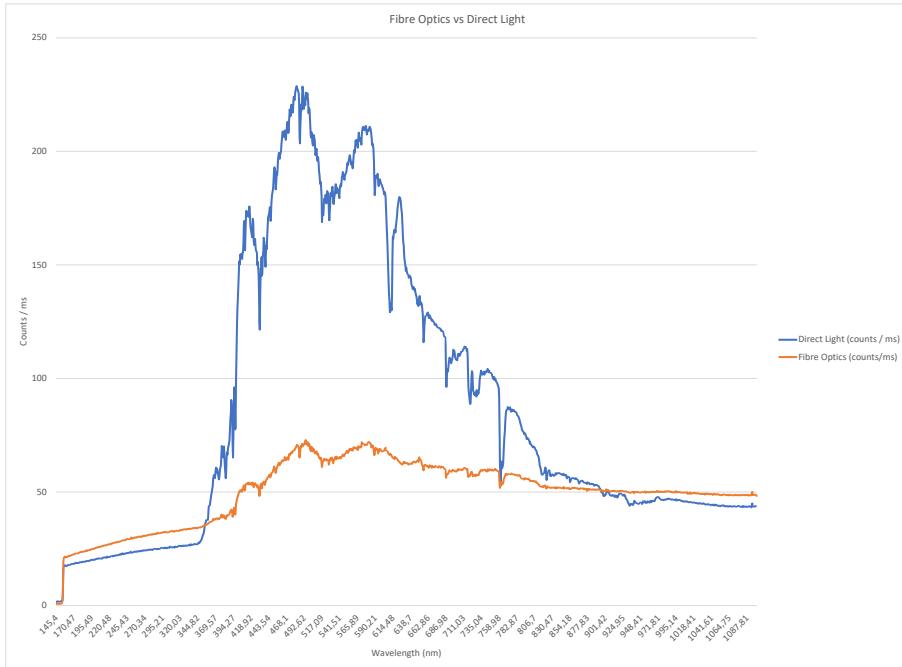


Figure 1.5: Plots obtained through the collection of light managed by the assemblies depicted in Figure 1.4

Circumventing the optical loss of signal is normally feasible just by selecting a more powerful light source. In our case, this would equate to the selection of a larger telescope, which would go against our size and weight constraints. Therefore, we had to solve the problem through a direct light connector that coupled the telescope to the spectrometer's entrance. This design was made with Autodesk's Inventor CAD

## CHAPTER 1. METHODS

---

software suite <sup>2</sup> and some renderings of the designs can be seen in Figure 1.6. Technical drawings are also available in Appendix E.

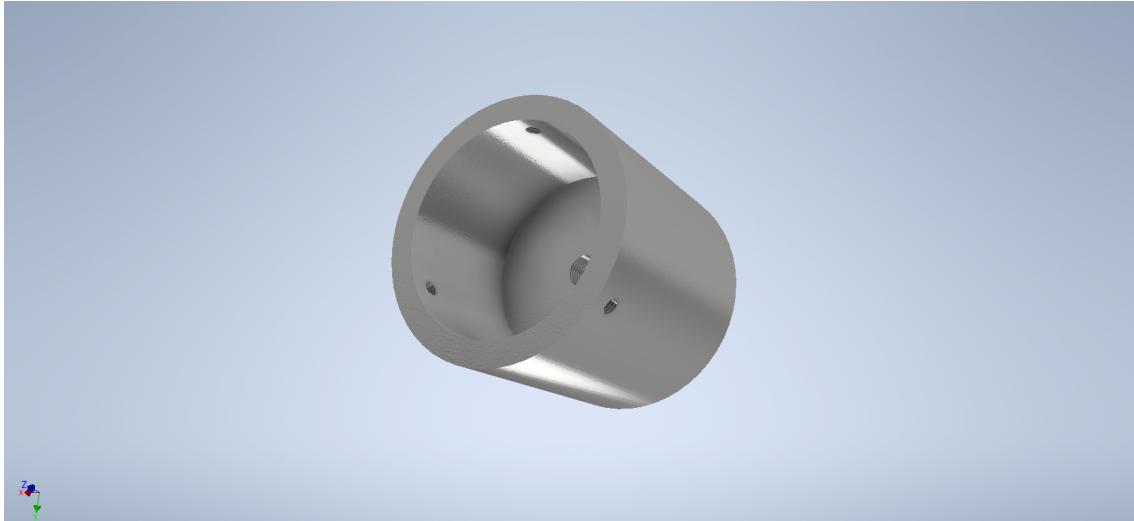


Figure 1.6: Telescope - Spectrometer coupling. 3D design rendered through Autodesk Inventor.

The centrepiece of this collection system is of course the telescope. From the beginning of the project, I understood that this was going to be one of the most important aspects of the endeavour. It has gone through some iterations before reaching the final design solution. The telescopes with which I began my experiments were some of the ones used in the [Forest Fire Finder \(FFF\)](#) project, which were normally Meade ET90 or similar (90mm diameter Maksutov-Cassegrain telescopes). These seemed too bulky to mount on a drone. They would fit, of course, but movement would become a very probable problem. Looking for alternatives at the time rendered no results. The 90mm diameter telescope market has a lot of competition, but smaller tubed reflecting telescopes are somewhat rare. On the refractors side of things, the problem was otherwise. Telescopes are usually built to look at the skies. Manufacturers try to offer the best balance between magnification, image quality and ease of use in each price range. Therefore, commercial refractor telescopes were too long to include in our drones, as they reached for larger focus lengths to get more magnification power. Our needs were in fact quite different. We wanted the light collection capabilities of a telescope tube, but since our idea was to connect it to a spectrometer, optical quality was not at all a priority, and neither was magnification. With this in mind, I started working on a custom made tube design.

I started by choosing the appropriate tube diameter. On a cloudy but luminous day, I used an ET90 to capture the spectrum that I show in Figure 1.7. This spectrum was collected with the Avantes Mini spectrometer, with an integration time of 50 ms. The Poissonian nature of light entering a telescope allows me to assume that its relationship with time is linear. Moreover, we know that the quantity of light that enters the telescope varies with the square of its diameter. Thus we can write the calculations in Equation 1.1.

---

<sup>2</sup>Thanks to a very generous protocol between Autodesk and many universities around the world that

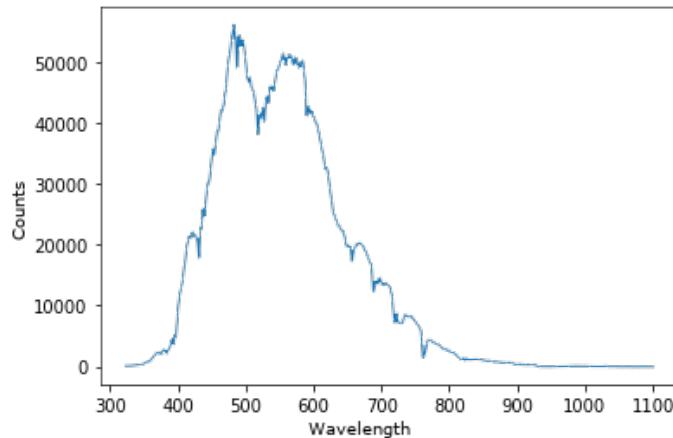


Figure 1.7: The spectrum captured on a cloudy but luminous day using the ET90 telescope and the Avantes Mini spectrometer.

Table 1.1: Specifications table for the Omegon MightyMak 60 telescope [10].

| Feature                             | Value | Unit |
|-------------------------------------|-------|------|
| <b>Aperture</b>                     | 60    | mm   |
| <b>Focal Length</b>                 | 700   | mm   |
| <b>f/</b>                           | 11,7  | N/A  |
| <b>Maximum Useful Magnification</b> | 120   | x    |
| <b>Weight</b>                       | 0,6   | kg   |

$$E \propto D^2 \Leftrightarrow \frac{E_{90}}{E_{25}} = \frac{90^2}{25^2} \approx 13 \quad (1.1)$$

The same spectrum of Figure 1.7 would take 13 times longer to collect with a telescope that had 25mm instead of 90mm diameter. In other words, 650 ms. This is a reasonable compromise given our application.

The first design attempt was a Maksutov-Cassegrain telescope. However, this was quickly abandoned due to 1) calculations indicated an impossible length for this kind of telescope and a 25mm diameter; and 2) the inherent complexity of these telescopes made assembling one prohibitively costly in terms of time.

The second design was much simpler. It was a refracting tube with the same diameter (25mm) and a focal length of 300mm. Designs for this telescope and its optical simulation were produced using OSLO-EDU software and are included in Appendix F. This ended up not being the telescope used in the final design, because as I was working in this part of the project, Omegon released their MightyMak 60, a 60mm diameter Maksutov-Cassegrain telescope that is small enough to be fitted onto a drone assembly. This was a perfect timing because the commercial availability of a suitable telescope allowed significant time savings on the whole ordering, assembling and testing process. The specifications of the new telescope are included in Table 1.1 [10].

makes this software free for all FCT NOVA's students.

### 1.1.3 Ground Station

As illustrated by Figure 1.1, the ground station is the device responsible for:

1. Improving positioning precision;
2. Programming the UAV's trajectory;
3. Analysing the incoming data.

Now, the first point is handled in automatic fashion by the RTK-GPS devices that are included in the assembly, so there is no intervention required on my behalf; the second point, flight programming, is achieved via Arducopter's programming API's. These are software libraries that allows one to program a drone's flight pattern using regular programming languages, such as Python. Unfortunately, there was not enough time for me to learn how this is done. Finally, the ground station is supposed to process the data that come from the drone.

The DOAS software library is a Python package developed specifically for this thesis data processing operations. It was, as other components, designed using an OOP approach and following the SOLID principles of Object Oriented Programming. This piece of software was written in response to the initial research that I undertook and that returned no usable results in terms of modular, compact Python libraries for DOAS applications, that I could use in my work. It is, as far as I know, the only DOAS solving application with this kind of structure.

The library (UML diagram presented in Figure 1.8) models a DOAS application through the instrumentation lens. A DOAS application is always parametrised through its spectrometer's physical features and limitations, which in turn determine the structure of the analysed spectral data, and even the differential cross sections of the trace gases that are to be studied. Of course, this library is much more limited in its capabilities than some specific programs that have become commonplace in this kind of application, such as QDOAS [3], but the fact that it can be operated through a Python program and that one can manipulate the data through such tools as Pandas DataFrames more than make up for this lack. Moreover, since it is in effect a software library, it is also as flexible as one is willing to expand it.

An important side note that attests to this library's relevance is that it has been fully integrated in FutureCompta's Bee2Fire software, with further developments being conducted through this team's efforts. This is the first commercially applicable result provided by the work in this thesis.

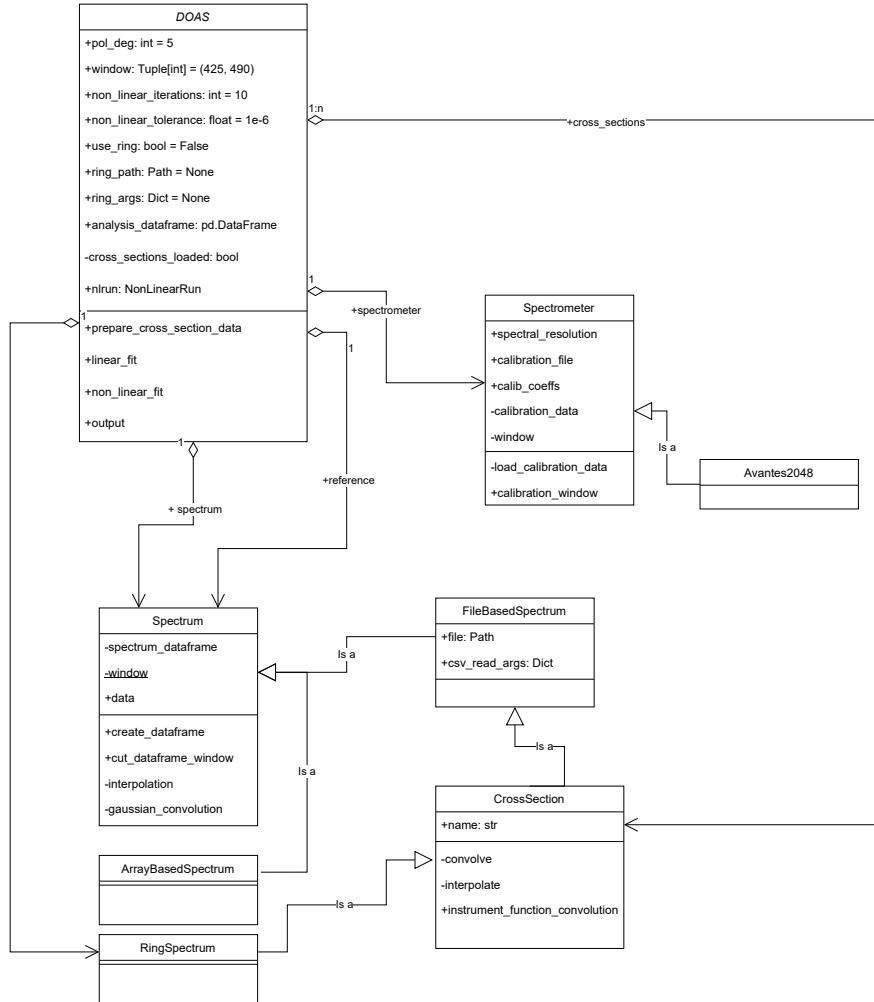


Figure 1.8: UML diagram for the **DOAS** library. The **OOP** approach that was followed allows for an instrument-oriented experiment parametrisation, which is not available in any other software

#### 1.1.4 Tomosim

Tomosim is the software program designed for the simulation of the image reconstruction process in the proposed atmospheric monitoring system of this dissertation. It corresponds to the last box of the schematic presented in Figure 1.1. The general workflow of this piece of software can be viewed in Figure 1.9 and Figure 1.10.

One of the main goals of this piece of software is the evaluation and characterisation of possible drone trajectories that allow the capture of sufficient projection data to perform the tomographic inversion. The trajectory is a key element of the system, as it determines basically everything in the experiment, and most prominently, its tomographic geometry. Currently, Tomosim can only use one type of geometry. This is the fanbeam geometry, as described in Section ???. It was chosen because it seemed to be more promising taking into account the several important restrictions that this system imposes, namely in terms of flight time reduction. In essence, the drone's trajectory (illustrated in Figure 1.11) is a horizontal circle which is parametrised to be at a certain height and to have a certain diameter. Both of these dimensions are

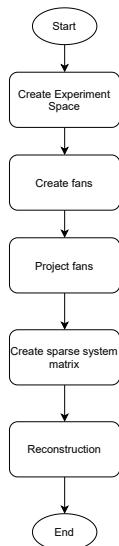


Figure 1.9: A flowchart describing the beginning stages of the Tomosim simulation.

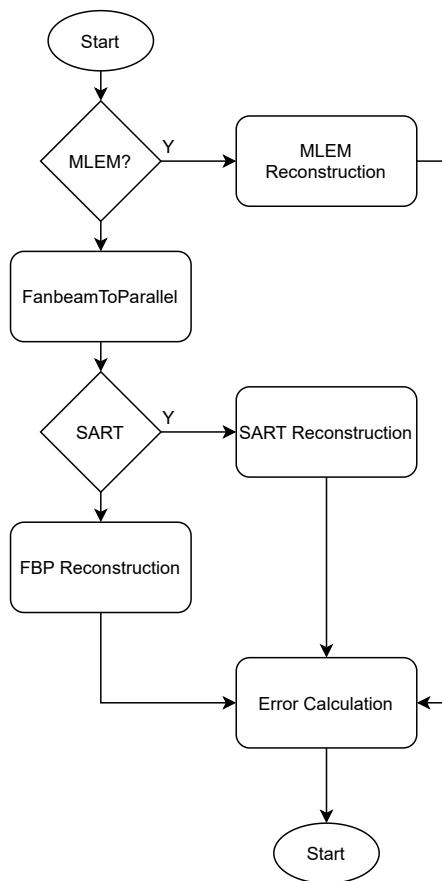


Figure 1.10: Flowchart representing the Tomosim operation during the reconstruction phase.

set at experiment / measurement time. The drone stops on this circle at regular angular intervals, say  $\alpha$  degrees. Each one of these stops ( $360 / \alpha$  stops) will generate a fanbeam projection, by pointing the optical system inwards (with respect to the circular macro-trajectory) and performing a series of spectroscopic measurements in different directions and also at regular intervals, say  $\gamma$  degrees. The particular case in which  $\alpha = \gamma$  is very interesting, because it then opens the possibility for resorting the fanbeams into parallel virtual-projections that are much easier to reconstruct tomographically, as introduced in Section ??.

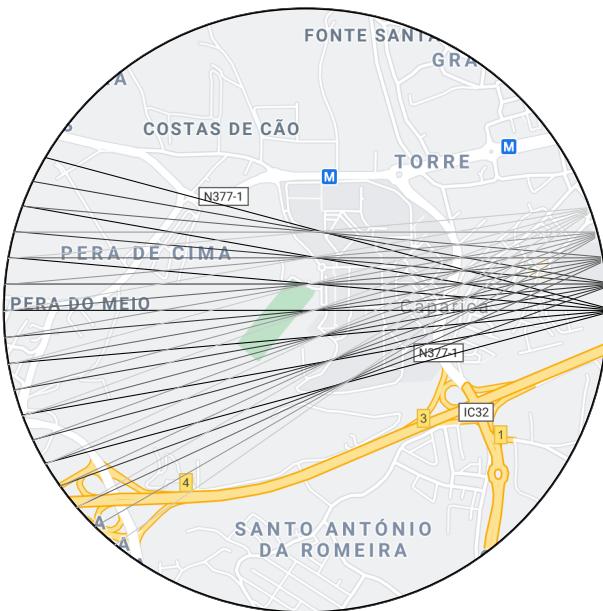


Figure 1.11: Illustration of the projection gathering algorithm based on the assembly of groups of acquisitions as fanbeam projections. The lines presented on the drawing come from the right to the left. Each is a ray in fan, corresponding to one stop of the drone. In the second measurement moment, the drone moves to the exit point on the left and repeats the first moment collection. Here, both fans and rays are separated by 5 degrees and there are 6 rays within each fan. A "real life" acquisition would feature a lot more rays, but that would be graphically complicated for the reader. Both ray and fan angular distance are customisable at runtime. Map taken from Google Maps in 2019. ©Google.

Of course, since we want to calculate molecular density in a give geographic region, we need to delimit our measurements to this space. The system is able to calculate the point where a given ray of light exits our **ROI** from the point of entry and its direction. Figure 1.12 is a schematic snapshot of a point in which the drone is taking a spectrum in one of its stops. Here, the drone's position ( $P_1$ ) is given by the distance  $D$  and the angle  $\beta$ . The gimbal is pointing at a direction at an angular distance of  $\gamma$  form line  $OP_1$ . Point  $P_2$ , which is not known, is at the intersection between the trajectory's circumference and line  $P_1P_2$ . Now, any point on this line can be expressed parametrically, with the sum of a point and a vector; while to say a point is on a circumference is the same as saying the distance between that point and the centre of this circumference is equal to its radius. The situation can be described by the expressions in Equation 1.2.

$$X = P_1 + t \cdot (P_2 - P_1) \quad (1.2)$$

$$|P_2| = D^2$$

Unravelling these expressions, and making use of the algebraic property that says that  $|A|^2 = A \cdot A$ , the expression becomes a second degree equation, as stated in Equation 1.3, writing  $P_2 - P_1$  as V.

$$t^2V^2 + 2 \cdot V \cdot P_1 \cdot t + P_1^2 - D^2 = 0 \quad (1.3)$$

If line  $P_1P_2$  non-tangentially intersects the circumference, solving Equation 1.3 renders two values for  $t$  (which correspond to  $P_1$  and  $P_2$ ). Selection is made by determining the returned value of  $t$  which maximises the euclidean distance between the produced point and  $P_1$ .

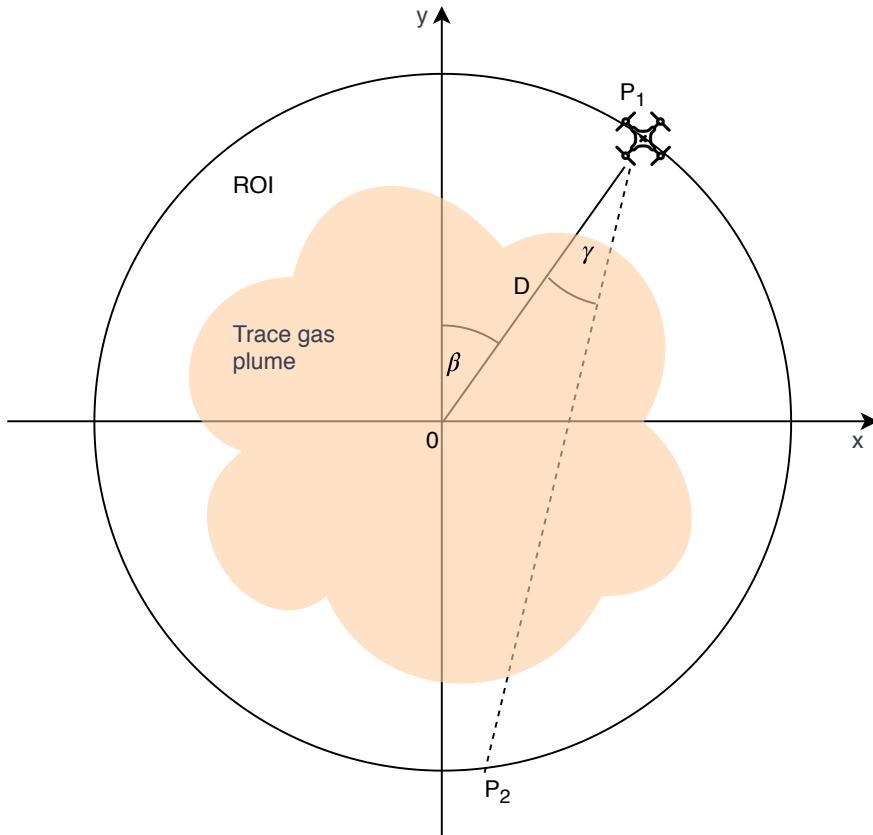


Figure 1.12:  $P_2$  Calculation. The software uses the position of the drone, as defined by the vertical angle,  $\beta$ , and the distance between the drone and the centre of the trajectory,  $D$ , to determine  $P_2$  through the solution of a second degree equation.

At this point, I have described the drone's trajectory and how the simulated drone calculates where the light enters and exits the ROI. But what exactly is in it? What will the simulation reconstruct?

A phantom is a device that represents the human body or some of its parts. They have been used in medical physics since the beginning of the field. In medical imaging, for instance, phantoms started being used in the late nineteenth century and early

twentieth century. At the time, it was very difficult to find volunteers for any kind of experiment that involved radiation, due to the common effects that were rapidly reported by the first people subject to this kind of intervention [4]. In spite of this difficulty, scientists and researchers still had to determine the dosimetry properties and physical limitations of their radiative devices, so medical physicists had to develop their own test models, or phantoms, for this effect. More recently, phantoms have been designed to develop computed tomography applications and algorithms. These phantoms mimic the body's attenuation properties in the X-Ray section of the electromagnetic spectrum, for instance.

Although the system that I propose does not aim at measuring or using the human body (or any other animal's), the concept still stands. To evaluate our reconstruction methods and the validity of our data gathering strategies, I needed an atmospheric phantom.

The distribution of gases in the atmosphere is completely different from biological tissue. Therefore, medical imaging phantoms were not adequate. The design that I have created is based on the premise that a two-dimensional Gaussian peak is more appropriate to describe the smoother nature of gaseous distribution [12]. This in contrast with the crisply defined edges of a medical tomography phantom such as Shepp-Logan's head phantom [11].

To design the phantom itself, I used a library called TomoPhantom [8], a tomographic phantom generator that provides a Python [Application Programming Interface \(API\)](#), making it trivial to include in the Tomosim simulator. The new phantom is comprised of 5 Gaussian profiles, depicting a static gas mixture. An ellipse is also in the phantom, near one of the corners. This serves mainly as a reference point for reconstruction, given its more solid and crisp nature. The new phantom can be seen in Figure 1.13 and its features are stated in Table 1.2.

Table 1.2: Table summarising the new phantom's construction details, as a sum of 5 Gaussian profiles and an ellipse designed using TomoPhantom. In this table, Co is the object's amplitude, Xo and Yo are its center coordinates, and a and b are the objects half-widths. The table is constructed using TomoPhantom's particular syntax and more information can be obtained at [8].

| Type     | Co | Xo   | Yo   | a    | b    | Angle |
|----------|----|------|------|------|------|-------|
| Gaussian | 1  | -0,1 | -0,1 | 0,25 | 0,5  | -45   |
| Gaussian | 1  | 0,6  | 0    | 0,65 | 0,45 | -45   |
| Gaussian | 1  | -0,6 | -0,4 | 0,8  | 0,8  | 0     |
| Gaussian | 1  | -0,4 | 0,8  | 0,7  | 0,7  | 0     |
| Ellipse  | 1  | 0,4  | -0,8 | 0,3  | 0,15 | 0     |

The phantom is then discretised, using Siddon's algorithm [before being reconstructed](#) using one of the algorithms described in Section ?? : [Filtered BackProjection \(FBP\)](#), [Maximized Likelihood Expectation Maximization \(MLEM\)](#) or [Simultaneous Algebraic Reconstruction Technique \(SART\)](#). Here, there is an additional step if opting for [FBP](#) or [SART](#). Both algorithms come directly from a dedicated tomography library and expect to receive parallel projection data as their input. Tomosim resorts the collected fanbeam projections onto a virtual parallel geometry [by using the procedure](#)

[link to theory](#)

[Link to theory](#)

[link to theory](#)

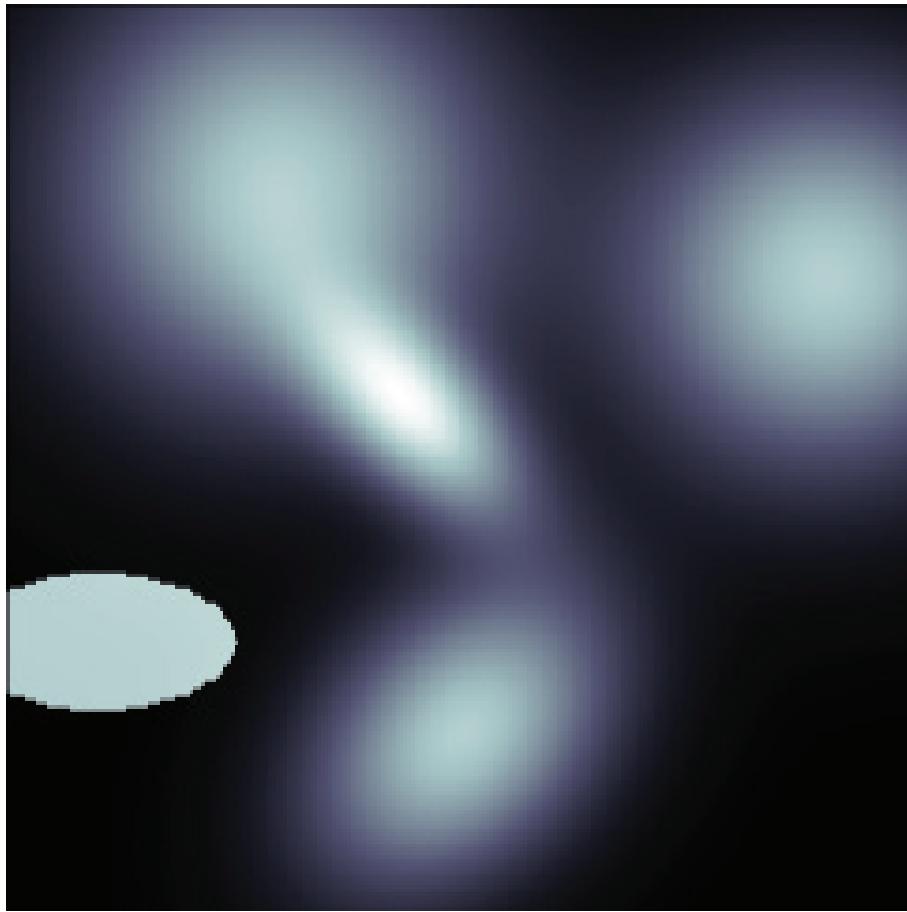


Figure 1.13: A graphical representation of the new spectral phantom, custom built for the TomoSim application.

described in Section ?? . I could not find any library that already applied this resorting algorithm, so I wrote my own, inspired by Matlab's `fanspara` routine.

In contrast, the `MLEM` algorithm was hand coded especially for Tomosim, and relies heavily on the fact that (as described in Section ?? ), this algorithm is a series of algebraic operations between matrices .

Error estimation is an important part of every tomography method and especially in simulations, as it allows one to approach the results with much more confidence of their similarity to the real world. Error sources for the Tomosim simulator come in four different natures: time errors, geometric errors, spectroscopic errors and reconstruction errors.

Time errors come from the fact that there are two moments of measurement. In a dynamic system, the time that passes between the two is enough for concentrations to change significantly. Tomosim does not address these errors, because they can be eliminated by the introduction of a second drone carrying the same type of equipment, which would eliminate said time difference.

Geometric errors exist due to the drone not being able to situate itself perfectly. There is always a positioning error, no matter how sophisticated the onboard equipment is. This type of error is addressed in the simulation through a Monte Carlo like approach. Positioning and pointing errors are assumed to have normal distributions.

link to theory

Each time a point is calculated by the drone, it generates a normally distributed random number, with a mean of 0 and a standard deviation equal to the nominal error of the positioning system. This number is then added to the theoretical point. Figure 1.14 is a graphical representation of the reasoning behind the calculation of the geometric error. The image deals with two types of error. One comes from the [Real Time Kinematic Global Positioning System \(RTK GPS\)](#) positioning system (the positioning error,  $\epsilon_p$ ); and the other that comes from the gimbal (the pointing error,  $\epsilon_\gamma$ ). The two  $\epsilon$  values are the nominal error for the positioning and the pointing devices. The error is introduced in the simulation through the values of  $\beta$  and  $D$  (see Figure 1.12) while calculating  $P_2$ . Given the very low nominal error for the gimbal, the small angle approximation is valid ( $\sin \theta = \theta$ ). This is used to determine the theoretical value of  $P_2$ , located on the device's circular trajectory. Finally, the software adds the positioning error, using the same process as in  $P_1$ 's case. The error depiction in Figure 1.14 is extremely exaggerated for visibility.

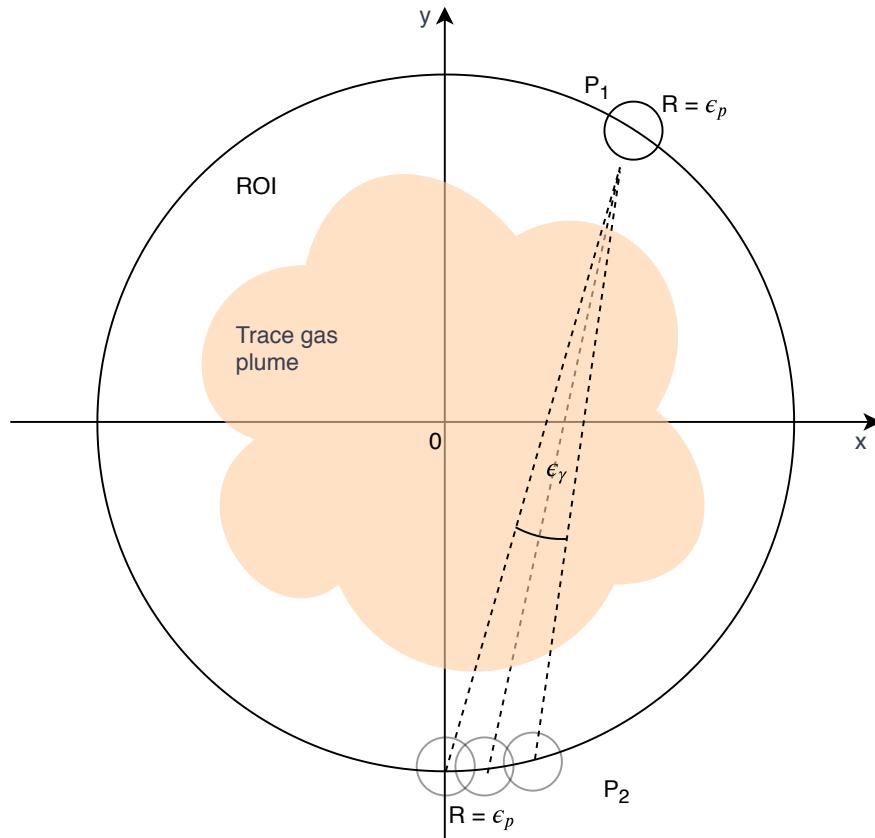


Figure 1.14: Error estimation graphical representation. Note errors are extremely exaggerated for visualisation purposes.

The third type of error are the spectroscopic errors. These come from the spectroscopic equipment that is used to gather projections. To take this noise into account, the simulator adds a Gaussian noise spectrum to each measurement, which is configurable through its standard deviation, as was previously done in [13]. This is a valid approach, insofar as the captured spectra are perfectly calibrated regarding spectral shift and squeeze. Since this is a simulation software, this is an acceptable assumption.

The final type of error that the system needs to contend with is the reconstruction error. In tomographic inversion problems, it is common to use techniques such as the **Mean Squared Error (MSE)** as a metric for an algorithm's performance. Tomosim was also evaluated in this light and in two separate ways. The first was to calculate the **MSE** for each pixel of the whole image. This information can still be viewed as an image (it is a two-dimensional grid of values) and paints an immediate picture of the general behaviour of the reconstruction algorithm. Moreover, it can tell the viewer if there are any types of shapes or areas in which the algorithm has more difficulties. The second way of using **MSE** to evaluate the reconstruction is to calculate a score through Equation 1.4. In this equation, and with respect to this simulator,  $f$  is the original image and  $g$  the one reconstructed from projections.

$$E = \sqrt{\frac{\sum |g(x, y) - f(x, y)|^2}{\sum |f(x, y)|^2}} \quad (1.4)$$

Just as the **DOAS** library, which is described in Section 1.1.3, the Tomosim simulation was programmed using the **OOP** paradigm. And also like the **DOAS** library, the SOLID principles of **OOP** were generally observed. The global flowcharts of the software's operation can be viewed in Figure 1.9 and Figure 1.10. As the application grew and I continued to work on it, it became quite clear that the original architecture was compromising the tool's performance and the code would need to be heavily overhauled. Although the time-frame of this work did not allow this, I was able to devise a basic new architecture for this software.

Ideally, Tomosim would be an extension of the more generic **DOAS** library. The only reason why it is not is purely historical: Tomosim's development was started before the former library library. Therefore, the first part of any refactoring exercise should be to mend this error. The new architecture, represented in Figure 1.15, is characterised by an increase in modularity. This is achieved through the introduction of the new classes "Reconstructor" and "RecEvaluator", which are composed into an also new class called "Experiment". The introduction of the two first classes greatly increases Tomosim's flexibility, as one would now be allowed to introduce any new reconstruction technique or evaluation method, without ever having to touch code already in place. To ensure this functionality, it is highly advisable that these two classes are built using the Template Design Pattern, with structural enforcing in the form of key abstract methods. This new design also favours the implementation of a mix between the factory design pattern and the strategy design pattern. This would result in factory methods being used to create Experiment objects based on some input parameters that would (for instance) determine the types of reconstruction and / or evaluation performed by said object.

Improving the system's architecture is a necessary and very important development, but the refactoring effort must also comprehend the performance side of the application. Currently, Tomosim's spends most of its running time discretising the **ROI**. This is expected. However, since Siddon's algorithm is not implemented using arrays, but instead makes heavy use of Python lists and for-loop-based iterations, it is absolutely imperative that this component is optimised

[link to theory](#)

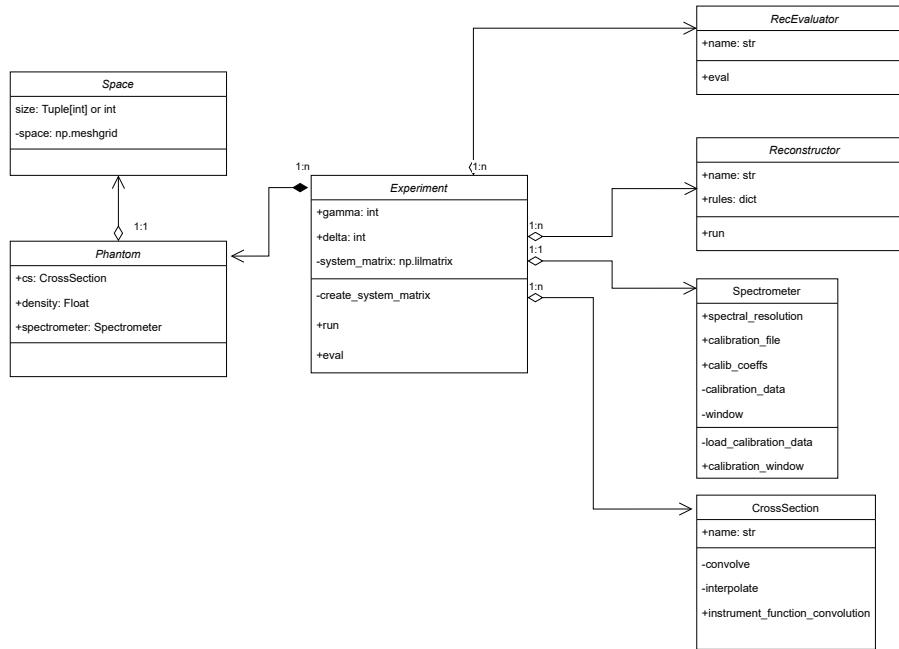


Figure 1.15: First draft of the [UML](#) diagram for the new architectural model of the Tomosim library. Note the usage of several classes from the [DOASlibrary](#) described in Section 1.1.3.

## 1.2 Second Hypothesis

### 1.2.1 The Lambertian Hypothesis

As stated in this chapter's introductory notes, this thesis main body of work revolves around two base assumptions, our hypothesis. The first one, about the information capturing by the idealised system, was addressed in Subsection 1.1.2. The second, more physical in nature, is the subject matter of this section. Our hypothesis states that the light absorption between points  $A$  and  $B$  (let's call it  $A_{AB}$ ) should be equal to the difference of the absorptions in  $A$  and  $B$ . We can write this, in a *Lambertian* manner as in Equation 1.5.

$$I_B = I_A \cdot \exp \left[ -AB \cdot \sum_i \sigma_{ABi} \cdot c_{ABi} \right] \quad (1.5)$$

This is to say that the light intensity reaching point  $B$  is given by the intensity reaching  $A$ , exponentially decreased by the absorbers at interval  $AB$ . The intensities at  $A$  and  $B$  are written as in Equation 1.6.

$$\begin{aligned} I_B &= I_0 \cdot \exp \left[ -L_B \cdot \sum_i \sigma_{Bi} \cdot c_{Bi} \right] \\ I_A &= I_0 \cdot \exp \left[ -L_A \cdot \sum_i \sigma_{Ai} \cdot c_{Ai} \right] \end{aligned} \quad (1.6)$$

If we join all this information in the same expression, the equation is transformed into its final form, presented in Equation 1.7.

$$I_0 \cdot \exp \left[ -L_B \cdot \sum_i \sigma_{Bi} \cdot c_{Bi} \right] = I_0 \cdot \exp \left[ -L_A \cdot \sum_i \sigma_{Ai} \cdot c_{Ai} \right] \cdot \left[ -AB \cdot \sum_i \sigma_{ABi} \cdot c_{ABi} \right] \quad (1.7)$$

Equation 1.7 can be greatly simplified: we take the natural logarithm of both sides and we state that  $\sum_i \sigma_{Xi} \cdot c_{Xi} = S_i$ . These operations result in the simplified form of Equation 1.8.

$$L_B \cdot S_B = L_A \cdot S_A + L_{AB} \cdot S_{AB} \quad (1.8)$$

Now,  $L_X \cdot S_X$  can be thought of as the wavelength dependent light absorption in path  $X$ . In this case, the wavelength interval is always the same. We can therefore conclude that, theoretically, our hypothesis is valid: light absorption between points A and B can be expressed in terms of the absorption on both these points and corresponds to their difference.

Although mathematically this seems clear-cut, in the real world things can become more problematic, since we have to deal with the imperfections that characterise a real physical system. Noise, instrumental limitations, adverse environmental effects, etc.. The experiment we describe in the next few paragraphs aimed at determining target trace gas concentration in a set analysis field. This field is dimension-wise compatible with those that would be employed in the final working system. This experiment is represented in Figure 1.16.

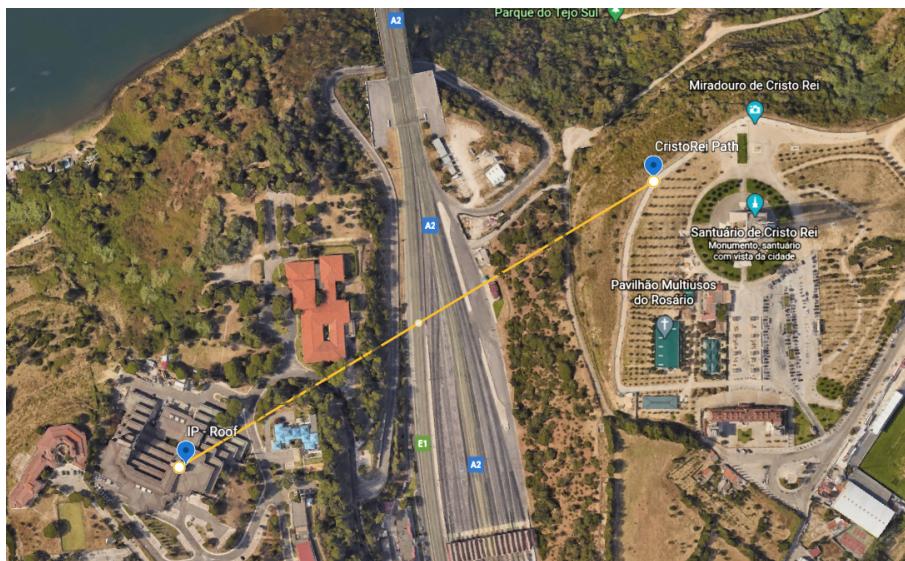


Figure 1.16: Location of observer points for the physical experiment.

The goal of the experiment was to compare passive and active DOAS measurements performed with a very short time difference between them. The passive measurement would employ the same acquisition strategy as the drone is expected to use. This comparison will be used to test our second hypothesis.

Finding two appropriate experiment sites proved to be the first difficulty: both telescopes should see sky on the back of the other telescope. Otherwise, the contribution from the terrain's reflection would have to be taken into account and the experiment conditions would be very different from the ones the drone will have. There are not

many site pairs that provide this, and most of the ones that exist are private and authorisations are not easy to obtain. In the end, we managed to run the experiment in the facilities of *InfraEstruturas de Portugal (IEP)* and the *Cristo-Rei* sanctuary, near our own base.

### 1.2.2 Testing the Lambertian Hypothesis

The experiment involved two different optical assemblies, which are summarised in Table 1.3. Both assemblies play two roles, which reflect the comparison between active and passive that is the entire aim of the test. To simplify, we will address the two as West Bank and East Bank. The West Bank is, as the name implies, the assembly that is placed further West, i.e., the one that is installed on IEP's roof. By exclusion, the East Bank assembly is the one placed on the sanctuary. The West Bank assembly is comprised of a telescope and tripod, a spectrometer (with the necessary fibre optics attached) and a laptop.

The East Bank assembly has exactly the same parts, but in addition to them, it features a hand-held torch that sports an XHP50.2 CREE LED. The manufacturer states that this torch is capable of illuminating by itself up to a distance of 300 m and produces luminous flux of at least 1500 lm. By fitting this torch on the telescope's eyepiece channel, we are able to further collimate the light that it produces, making it reach much further distances than originally stated, and being easily picked up by the other telescope. This is plain to see in Figure ???. The light spectrum that pertains to the CREE LED in use is published in this device's datasheet, and presented in Figure 1.17, which largely corroborates the spectrum in Figure 1.18, taken by the same spectrometers that were used in the experiment, at a distance of approximately 50 m from the torchlight.

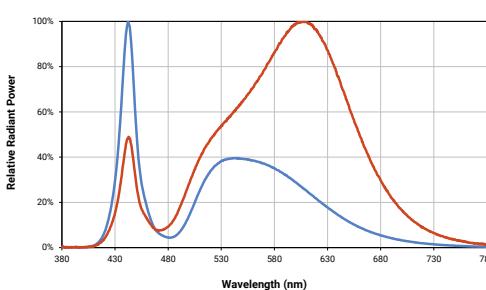


Figure 1.17: Published spectrum of the CREE XHP50.2 LED light. The LED that was used in this experiment corresponds to the blue line [2].

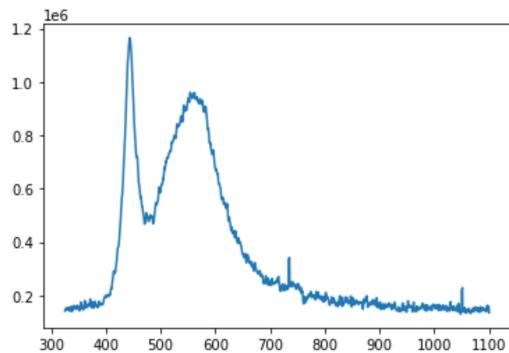


Figure 1.18: Spectrum measured with the experiment spectrometers, at a distance of approximately 50 m from the torchlight.

The experiment itself is scheduled to start at around 06:00 A.M.. It consists in capturing spectral measurements in both modes (active and passive) periodically, with the least amount of time possible between measurements in the same capture. In this case, I am calling capture to a particular group of actions that are defined in Table 1.4. Captures are defined according to the time at which they are run, and are summarised in Table 1.5. Closing time for this experiment was set on 11:00 A.M.. This time window

## CHAPTER 1. METHODS

---

Table 1.3: Summary table for the two experiment assemblies. Note the difference in terms of material, due to the two different roles both assemblies play during the experiment. This is translated into not having the need of an artificial light source in the West Bank's assembly.

|                                | <b>West Bank</b>          | <b>East Bank</b>          |
|--------------------------------|---------------------------|---------------------------|
| <b>Spectrometer</b>            | Avantes USB 2048 channels | Avantes USB 2048 channels |
| <b>Telescope</b>               | Meade ET90                | Meade ET90                |
| <b>Artificial Light Source</b> | N/A                       | Goobay CREE XHP50.2 torch |
| <b>Laptop</b>                  | Windows 10 laptop         | Windows 10 laptop         |
| <b>Software</b>                | AvaSoft 8.11              | AvaSoft 8.11              |

ensures measurements are taken during sunrise and until after the morning rush hour is over.

Table 1.4: Actions are the indivisible unit upon which each capture is built. The prescribed actions for this experiment are described in this table.

| Action ID | Action  | Description   |
|-----------|---|---|
| A         | Active trace gas concentration determination  | With the two telescopes facing each other, we collect spectra for two minutes with the light source turned off and another 2 min with the light source turned on. |
| B         | Passive trace gas concentration determination | With the two telescopes aligned and approximately facing West, we collect spectra for 2 minutes.  |
| C         | Passive reference collection                  | The West telescope points upwards and collects data for 2 minutes.  |

Table 1.5: Captures are particular sets of actions that are conducted according to a specific order, depending on the time of day on which the capture is run. This table describes the prescribed captures on which this experiment consisted.

| Time Frame      | Period     | Action  |
|-----------------|------------|---------|
| 05:00 - Sunrise | 15 minutes | A       |
| Sunrise         | Once       | C       |
| Sunrise - 11:00 | 15 minutes | A and B |

As displayed in Table 1.3, the spectrometers are both the same model, manufactured by Avantes and with 2048 channels, powered through the same [USB](#) cable that is used for data transfer. The spectra are acquired through Avantes' own collection software, AvaSoft 8. The spectrometer are configured to have an integration time of 20ms and immediately store every measurement on an [American Standard Code for Information Interchange \(ASCII\)](#) file. With the kind of lighting conditions that we are dealing with, this integration time allows us not to worry about saturation. However, to build usable spectra we need to sum the collected files. This is valid because given

the very little time it takes to make a measurement (2 minutes), the sun can be considered a constant light source, and therefore we can consider the photons to have a Poissonian statistic distribution [7].

# DISCUSSION

In Chapter 1, I introduced the methods with which I conducted the research and work in this thesis. In addition, I also introduced the notion that this project is divided naturally in two main parts, according to the research questions that I aim to answer.

This chapter maintains this division. In Section 2.1, I present and discuss what has come from the work conducted with reference to the technological problem of obtaining sufficient tomographic data for the retrieval of an atmospheric trace gas column density map for a given geographic area, through optical spectroscopy techniques.

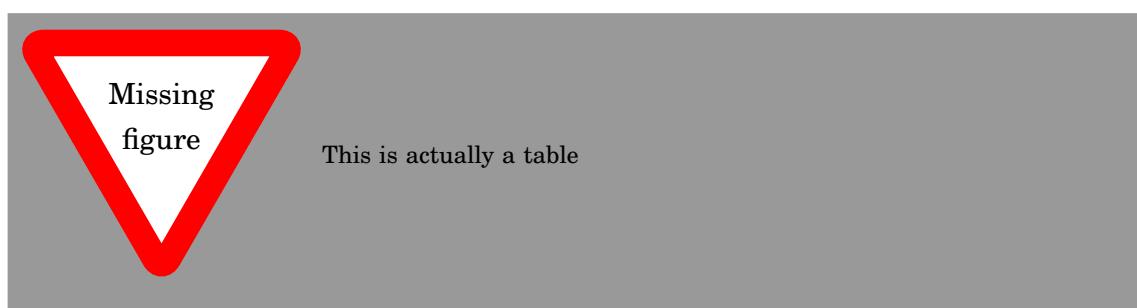
Section 2.2 is, naturally, dedicated to the results coming from the second hypothesis presented in Chapter 1. This is to say the *Lambertian* hypothesis that was tested through a practical experiment designed especially to this end.

## 2.1 First Hypothesis

### 2.1.1 The UAV

The component list presented in Table ?? follows the drone concept explained in Section 1.1.1. The idea was to build a hexacopter based on the very flexible and commonly used DJI S900 frame. This drone had to be capable of maintaining flight for at least 30 minutes while carrying a payload of 2kg (which is actually more than the collection system weighs). Note the inclusion of custom-made carbon fibre arms and 17" propellers.

Table 2.1: Component list for the hexacopter that was designed as a part of this thesis.



**2.1.2 Collection System**

The collection system is composed by the parts listed in Table ???. The collection system involved the design of several parts using specialised CAD software.

**2.1.3 TomoSim**

**2.2 Second Hypothesis**

**2.2.1 Experiment - First Run**

**2.2.2 Experiment - Second Run**

## BIBLIOGRAPHY

- [1] Amphenol Optics Fiber. *SMA Connectors , Adapters & Receptacles*. 2002 (cit. on p. 4).
- [2] CREE. *XLamp ® XHP50.2 LEDs Datasheet*. 2021 (cit. on p. 19).
- [3] T. Danckaert et al. *QDOAS*. 2015. URL: <http://uv-vis.aeronomie.be/software/QDOAS/> (cit. on p. 8).
- [4] L. A. Dewerd and M. Kissick. *The Phantoms of Medical and Health Physics*. Ed. by L. A. DeWerd and M. Kissick. Biological and Medical Physics, Biomedical Engineering. New York, NY: Springer New York, 2014, p. 286. ISBN: 978-1-4614-8303-8. DOI: [10.1007/978-1-4614-8304-5](https://doi.org/10.1007/978-1-4614-8304-5). URL: [http://www.springer.com/10.1007/978-1-4614-8304-5](http://www.springer.com/series/3740%20http://link.springer.com/10.1007/978-1-4614-8304-5) (cit. on p. 13).
- [5] Diamond FIber Optics. *F-SMA Standard Datasheet* (cit. on p. 4).
- [6] DJI. *DJI E1200 Pro Tuned Propulsion System*. 2015. URL: <https://www.dji.com/pt/e1200> (visited on 12/28/2020) (cit. on p. 4).
- [7] M. Fox. *Quantum optics*. Oxford, UK: Oxford University Press, 2006. ISBN: 9780198566724 (cit. on p. 21).
- [8] D. Kazantsev et al. “TomoPhantom, a software package to generate 2D–4D analytical phantoms for CT image reconstruction algorithm benchmarks”. In: *SoftwareX* 7 (Jan. 2018), pp. 150–155. ISSN: 2352-7110. DOI: [10.1016/J.SOFTX.2018.05.003](https://doi.org/10.1016/J.SOFTX.2018.05.003). URL: <https://www.sciencedirect.com/science/article/pii/S2352711018300335?via%7B%5C%7D3Dihub> (cit. on p. 13).
- [9] Newport Fibre Optics. *Fiber Optics : Fiber Preparation and Fiber Connectors* (cit. on p. 4).
- [10] Omegon. *Omegon Mighty Mak 60*. 2021 (cit. on p. 7).
- [11] L. A. Shepp and B. F. Logan. “The Fourier reconstruction of a head section”. In: *IEEE Trans. Nucl. Sci.* 21.3 (June 1974), pp. 21–43. ISSN: 0018-9499. DOI: [10.1109/TNS.1974.6499235](https://doi.org/10.1109/TNS.1974.6499235). URL: <http://ieeexplore.ieee.org/document/6499235> (cit. on p. 13).

- [12] C. Stachniss, C. Plagemann, and A. J. Lilienthal. “Learning gas distribution models using sparse Gaussian process mixtures”. In: *Auton. Robots* 26.2-3 (Apr. 2009), pp. 187–202. ISSN: 0929-5593. DOI: 10.1007/s10514-009-9111-5. URL: <http://link.springer.com/10.1007/s10514-009-9111-5> (cit. on p. 13).
- [13] J. Stutz and U. Platt. “Numerical analysis and estimation of the statistical error of differential optical absorption spectroscopy measurements with least-squares methods”. In: *Appl. Opt.* 35.30 (Oct. 1996), p. 6041. ISSN: 0003-6935. DOI: 10.1364/AO.35.006041. URL: <https://www.osapublishing.org/abstract.cfm?URI=ao-35-30-6041> (cit. on p. 15).
- [14] D. Triantafyllou and N. A. Aspragathos. *Intelligent Robotics and Applications*. Ed. by S. Jeschke, H. Liu, and D. Schilberg. Vol. 7101. Lecture Notes in Computer Science PART 1. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 509–519. ISBN: 978-3-642-25485-7. DOI: 10.1007/978-3-642-25486-4. URL: <http://link.springer.com/10.1007/978-3-642-25486-4> (cit. on p. 4).
- [15] F. Zimmermann et al. “PRECISE POSITIONING of UAVS - DEALING with CHALLENGING RTK-GPS MEASUREMENT CONDITIONS during AUTOMATED UAV FLIGHTS”. In: *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 4.2W3 (2017), pp. 95–102. ISSN: 21949050. DOI: 10.5194/isprs-annals-IV-2-W3-95-2017 (cit. on p. 4).

A

**FOREST FIRE FINDER: DOAS  
APPLICATION TO LONG-RANGE FOREST  
FIRE DETECTION**



# Forest Fire Finder – DOAS application to long-range forest fire detection

Rui Valente de Almeida<sup>1,2</sup> and Pedro Vieira<sup>1,2</sup>

<sup>1</sup>NGNS, Ingenious Solutions, Rua Cidade de Évora, 11, 2660-022 Loures, Portugal

<sup>2</sup>Faculty of Science and Technology, NOVA University of Lisbon, Caparica Campus, 2829-516 Caparica, Portugal

*Correspondence to:* Pedro Vieira (pedro.vieira@ngns-is.com)

Received: 23 September 2016 – Discussion started: 14 November 2016

Revised: 11 April 2017 – Accepted: 8 May 2017 – Published: 22 June 2017

**Abstract.** Fires are an important factor in shaping Earth's ecosystems. Plant and animal life, in almost every land habitat, are at least partially dependent on the effects of fire. However, their destructive force, which has often proven uncontrollable, is one of our greatest concerns, effectively resulting in several policies in the most important industrialised regions of the globe.

This paper aims to comprehensively characterise the Forest Fire Finder (FFF), a forest fire detection system based mainly upon a spectroscopic technique called differential optical absorption spectroscopy (DOAS). The system is designed and configured with the goal of detecting higher-than-the-horizon smoke columns by measuring and comparing scattered sunlight spectra. The article covers hardware and software, as well as their interactions and specific algorithms for day mode operation. An analysis of data retrieved from several installations deployed in the course of the last 5 years is also presented.

Finally, this paper features a discussion on the most prominent future improvements planned for the system, as well as its ramifications and adaptations, such as a thermal imaging system for short-range fire seeking or environmental quality control.

## 1 Introduction

Fire is a process by which elements chemically combine with oxygen, releasing energy (as heat and light) and smoke into the surrounding environment. Fires are an important factor in shaping Earth's ecosystems. Plant and animal life, in many

land habitats, are at least partially dependent on the effects of fire (Food and Agriculture Organisation , FAO).

The use of fire by hominids predates civilisation by thousands of years and, in today's society, there are almost no areas of technology or scientific knowledge that do not involve fire in one way or another. However, fire's destructive power is undeniable.

Forest fires are among the great concerns of the present day in industrialised countries. Research regarding wildfires has been targeted by many countries and unions worldwide in an effort to minimise the negative impact these events imply.

According to the Intergovernmental Panel on Climate Change (IPCC), climate change is expected to increase global temperatures and change rainfall patterns, leading to an increased risk of fire (IPCC, 2012). This means that the number of registered fires throughout the world is expected to increase, a phenomenon the world must be ready to address.

In the European Union, the Horizon2020 research programme states that there must be a union-wide investment in research concerning forest protection and recovery from fires. In the past, the FP7 programme had sponsored the development of an automatic forest fire detection system called FireSense, an investment of over EUR 2.5 million (European Comission, 2012).

The United States Forest Service acknowledge the importance of understanding wildland fire dynamics, running a network of research centres solely dedicated to the study of this subject. Research endeavours take 6 % of the service's annual budget, which is currently directed primarily towards fire suppression (United States Forest Department, 2015).

Australia is another geographic region where wildfires have had a great impact. As a response, its government has created the Bushfire and Natural Hazards Cooperative Research Centre. The institution builds upon more than 10 years of experience dealing with Australian bushfires and aims to produce internationally recognised research regarding the study and modelling of wildfires in Australia and New Zealand (BNHCRC, 2016).

In spite of this global investigation effort regarding fires and their behaviour, every year, material losses as a result of fires ascend to billions of dollars and thousands of lives are lost in the same way. This leads to a strong increase in the size of the fire protection market, including passive and active detection platforms, which is expected to grow at a cumulative aggregate growth rate of 11.53 % from 2014 to 2020 (Research and Markets, 2016).

## 2 State of the art

In recent years, several methods have been developed in an attempt to automatically and reliably detect forest fires. These systems differ primarily in their strategic approach to the issue at hand, creating three main categories:

- Satellite monitoring techniques: satellite data have been used for fire monitoring purposes since the late 20th century. The MODIS (MODerate resolution Imaging Spectroradiometer) and AVHRR (Advanced Very High Resolution Radiometer) sensors, deployed respectively in the Aqua/Terra and NOAA satellites, have had extensive use in this regard. However, their low temporal resolution (2 and 4 times per 24 h, respectively) make them poor candidates for fire detection uses. Geostationary satellites overcome this difficulty by continuously scanning a single, very large geographic region. They have, nevertheless, a low spatial resolution of 1 km, which means that small fires are difficult for them to detect (Manyangadze, 2009).
- Wireless network sensing: the wireless sensor network approach to fire detection is completely different from the other two categories. Instead of having a single device patrolling the target region, these systems are designed on the capabilities of a high number of extremely small battery-operated sensor boards that can communicate among themselves (Alkhatib, 2014; Liyang et al., 2005).

The sensor boards are equipped with several sensors, from temperature and humidity to luminance detectors. In spite of their great fire detection capabilities, these networks present various drawbacks, such as their very limited individual range of detection and their 2-year lifetime or the fact that their remains might imply an environmental issue (Alkhatib, 2014).

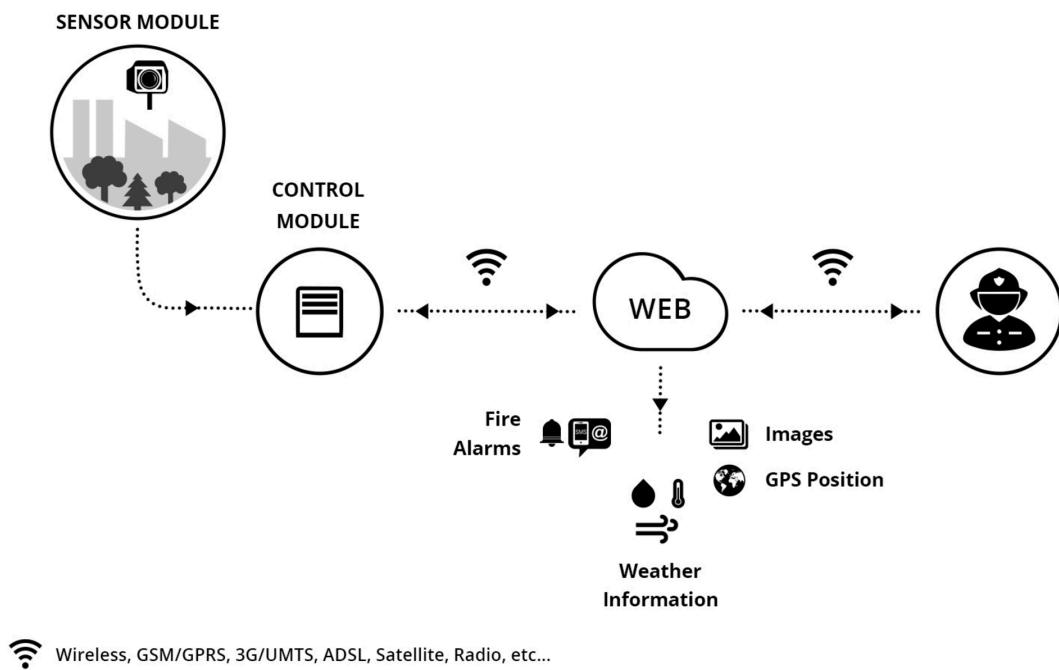
- Large-area remote sensing: this family of systems is designed with the goal of minimising the number of deployed devices in a given target region. Their architecture implies the use of an optical principle in order to detect smoke or flames, whether optical cameras or spectrometers.

There are already several commercially available systems, such as the Forest Fire Finder (FFF; the main subject of this paper), FireWatch, ForestWatch, AlarmEYE or EyeFi SPARC. Although these are commercial products, the available information is sparse and many times outdated, so a true comparison not only is beyond the scope of the article but also would require more research efforts. Nevertheless, it is important to briefly describe the operating principles of the more prominent systems.

- ForestWatch, developed in South Africa by EnviroVision Solutions, uses optical object recognition software, coupled to a very specific camera system. It detects smoke during the day and the flame glow during the night, at a maximum distance of 24 km in every direction, in a semi-automatic fashion. It is probably the most commercially successful system, with more than 300 currently operating towers (Envirovision Solutions, 2015; Hough, 2007).
- FireWatch is a commercial system operated and sold by IQ Wireless GmbH, in Germany. The system uses optical sensors and object recognition algorithms to detect smoke at a maximum distance of 15 km. It is important to mention that the FireWatch system is not a fully automatic fire detection platform, requiring a control room to operate correctly (IQ-Wireless, 2016).
- The FFF was developed in Lisbon, in a partnership between the NOVA University of Lisbon and NGNS-IS, Ltd., in 2006. This patented system uses a spectroscopic technique to assess the atmosphere and detect smoke columns (NGNS-IS, 2016). During the night, the system changes its operation mode and relies solely on image processing to detect a fire's glow. Its maximum rate detection range is of 15 km, and it acts with complete autonomy, requiring minimal human intervention (see Fig. 1).

The FFF's most significant advantage over its rivals is its low number of false alarms (typically one or two per week). This comes from the fact that the system's smoke-detection capabilities do not rely on image processing. This in turn means that reliable detections can be achieved by a smaller number of deployed devices (only one, three for triangulation). However, more reliable alarms imply less human intervention, which translates into less financial expenditure over time.

The FFF system is the only one to use an optical spectroscopy technique to detect fire through smoke presence in



**Figure 1.** The Forest Fire Finder system (NGNS-IS, 2016).

real time. Since the analysis is carried out in an outdoor scenario, the process is not as straightforward as in laboratory experiments. This article addresses only the spectroscopic techniques used in the system's daytime operation mode.

### 3 The technique

The FFF system makes use of a spectroscopic technique called differential optical absorption spectroscopy (DOAS). This is a well-established and widely used technique in the field of atmospheric studies (Platt and Stutz, 2007).

There are two main categories of DOAS experiment assemblies, with different goals and capabilities:

- Active systems, of which a simple illustration is presented in Fig. 2, are characterised by relying on an artificial light source for their measurements. A spectrometer at the end of the light path performs spectroscopic detection. Active DOAS techniques are very similar to traditional in-lab absorption spectroscopy techniques (Platt and Stutz, 2007);
- Passive DOAS techniques, illustrated in Fig. 3, use natural light sources, such as the Sun and the moon, in their measurement process. An optical system is pointed in certain elevation and azimuth angles and sends the captured light into a spectrometer, connected to a computer. The system returns the total value of the light absorption in its path (Platt and Stutz, 2007; Merlaud, 2013). Since the FFF system is basically a passive DOAS sys-

tem, we will centre our discussion on this category from this point forward.

DOAS itself is based on Lambert–Beer's law, which can be written as (Platt and Stutz, 2007)

$$I(\lambda) = I_0(\lambda) \cdot \exp(-\sigma(\lambda) \cdot c \cdot L), \quad (1)$$

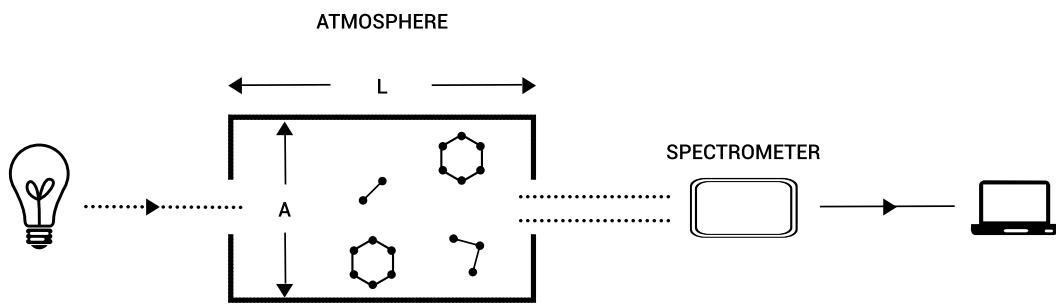
where  $\lambda$  is the wavelength of the emitted light;  $I(\lambda)$  is the light intensity as measured by the system;  $I_0(\lambda)$  is the intensity of the light as emitted by the source; and  $\sigma(\lambda)$  is the absorption cross section of absorber, which is wavelength dependent;  $c$  is the concentration of the absorber we want to measure.

This law allows the definition of optical thickness ( $\tau$ ) (Platt and Stutz, 2007):

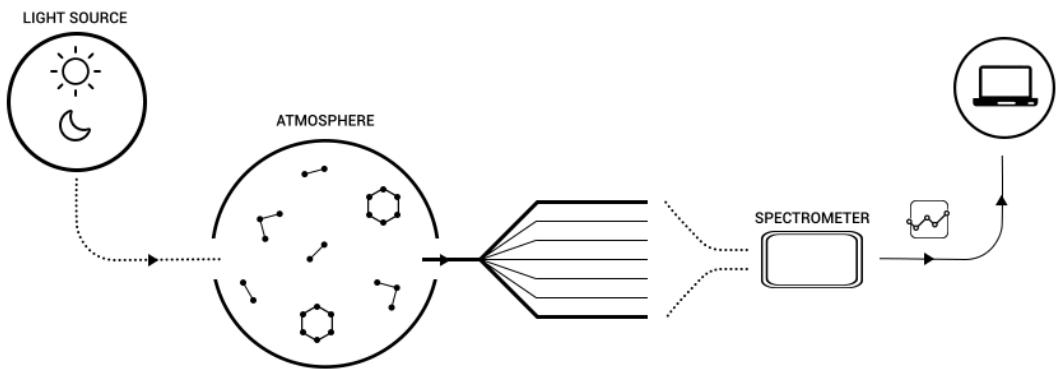
$$\tau(\lambda) = \ln\left(\frac{I_0(\lambda)}{I(\lambda)}\right) = \sigma(\lambda) \cdot c \cdot L. \quad (2)$$

In a laboratory setting, Eq. (1) or (2) can be used to directly calculate an absorber's concentration, provided there is knowledge of its cross section. In the open atmosphere, however, absorption spectroscopy techniques are far more complex. On one hand,  $I_0(\lambda)$  is not accessible since we measure from inside the medium we want to measure. On the other hand, there are several environmental and instrumental effects that influence measurement results. These effects include the following (Platt and Stutz, 2007).

- Rayleigh scattering is due to small molecules present in the atmosphere and is heavily influenced by wavelength (hence the blue colour of the sky).



**Figure 2.** Active DOAS schematic.



**Figure 3.** Passive DOAS schematic.

- Mie scattering is caused by particles and larger molecules suspended in the atmosphere and is not very dependent on the wavelength (hence the white colour of clouds).
- Instrumental and turbulence effects are the instrument's transmissivity and atmospheric turbulence in the optical path also limit light intensity.

In addition, we also have to take into account that, in the atmosphere, there are a number of trace gases that interfere with passing light.

Another aspect worth mentioning is that our device is never pointed directly at the light source (the Sun) but always processes light that has been scattered at some unknown point in the optical path. This means that the light that reaches our detector is only the scattered fraction of the sunlight, depending on the system's position and geometry, as well as wavelength.

The expansion of Lambert–Beer's equation to include all these effects results in Eq. (3).

$$I(\lambda) = I_0(\lambda) \cdot A(\lambda, \dots) \cdot S(\lambda) \cdot \exp \left[ - \int \left[ \left( \sum_i \sigma_i(\lambda, s) \cdot c_i(s) \right) + \epsilon_M(\lambda, s) \right] ds \right],$$

$$+ \epsilon_R(\lambda, s) \right] ds \Big], \quad (3)$$

where  $A(\lambda, \dots)$  is the fraction of scattered light that reaches the device,  $S(\lambda)$  represents instrumental and turbulence effects,  $\sigma_i(\lambda, s)$  is the absorption cross section of absorber  $i$ ,  $c_i$  is the concentration of absorber  $i$ ,  $\epsilon_R(\lambda)$  represents Rayleigh's extinction coefficient and  $\epsilon_M(\lambda)$  represents Mie's extinction coefficient.

The interest of this equation lies within the retrieval of  $c_i$ , a given absorber's concentration. Since the integral is taken along the total atmospheric path of the measured photons, and considering that their cross sections do not vary significantly in atmospheric conditions, it is possible to define the concept of slant column, which is of great importance (Merlaud, 2013).

$$SC_i = \int c_i(s) ds \quad (4)$$

This quantity, as Eq. (4) shows, equals the integral of an individual absorber's concentration along the atmospheric optical path of relevance.

Now, without knowledge of  $I_0(\lambda)$ , these equations cannot give us absolute concentration values. We can, however, use another scattered light spectrum as reference in Eq. (2). Instead of absolute densities, this will yield relative changes in the atmosphere. We thus arrive at Eq. (5).

$$\begin{aligned} \ln\left(\frac{I_{\text{ref}}}{I}(\lambda)\right) &= \ln\left(\frac{A_{\text{ref}}}{A}(\lambda, \dots)\right) + \ln\left(\frac{S_{\text{ref}}}{S}(\lambda)\right) \\ &+ \sum_i (\sigma_i(\lambda) \cdot \Delta SC_i(\lambda)) + \Delta \tau_M(\lambda) \\ &+ \Delta \tau_R(\lambda), \end{aligned} \quad (5)$$

where  $\Delta SC_i$  is the relative slant column of absorber  $i$ ;  $\Delta \tau_M$  is the relative Mie scattering term, integrated to its optical thickness; and  $\Delta \tau_R$  is the relative Rayleigh scattering term, integrated to its optical thickness.

This is where the principle of DOAS is applied. Instrument features, scattering and other atmospheric effects have broad absorption spectral profiles, which vary slowly with wavelength. Several trace absorbers have narrow and rapidly varying spectral signatures in at least a small section of the spectrum. By using Eq. (6), we can separate these contributions (Danckaert et al., 2015).

$$\sigma(\lambda) = \sigma'(\lambda) + \sigma_0(\lambda) \quad (6)$$

Here, the broad part of the optical thickness ( $\sigma_0(\lambda)$ ) can be separated from the narrow part ( $\sigma'(\lambda)$  – differential) by approximating it by a low-order polynomial, resulting in Eq. (7).

$$\ln\left(\frac{I_{\text{ref}}}{I}(\lambda)\right) = \sum_{i=1}^n \sigma'_i(\lambda) \cdot \Delta SC_i + \sum_{j=0}^m a_j \cdot \lambda^j, \quad (7)$$

where  $\sum_{i=1}^n \sigma'_i(\lambda) \cdot \Delta SC_i$  is the differential part (narrow-band, rapidly varying with wavelength) and  $\sum_{j=0}^m a_j \cdot \lambda^j$  is a low-order polynomial, used to remove the broadband spectral features resulting from atmospheric and instrumental phenomena.

In practice, the mathematical solving of Eq. (7) is not enough since it does not account for the Ring effect or the non-linearities that result from stray light and wavelength shift in measured and cross-section spectra.

The Ring effect is a consequence of rotational Raman scattering: molecules in the atmosphere do not absorb photons in a purely elastic (Rayleigh scattering) fashion. A small portion of the light–matter interaction is in fact inelastic (Brinkmann, 1968; Merlaud, 2013). This changes the light source frequencies as seen from the detector. This phenomenon was first noticed by Grainger and Ring in 1962. At the time, they noticed that the well-known Fraunhofer lines would slightly change when one observed them by using moonlight instead of scattered daylight (Grainger and Ring, 1962).

From the occurrence of these phenomena, it results that the mathematical procedure for DOAS measurements consists in solving a linear and a non-linear problem. The linear problem is solved by writing Eq. (7) in its matrix form:

$$\boldsymbol{\tau} = \mathbf{A} \cdot \mathbf{X}. \quad (8)$$

$\mathbf{A}$  is an  $m \times n$  matrix, with its columns being the differential cross sections  $\sigma'_i(\lambda)$  and the wavelength powers taking the polynomial  $P(\lambda) = \sum_{j=0}^m a_j \cdot \lambda^j$  into account. Since the number of lines in  $\mathbf{A}$  is much larger than the number of columns, the system is overdetermined and, in this case, we must use methods to numerically approximate a solution. It is common to use the least-squares approach, in which the best solution is the one that minimises  $\chi^2 = [\boldsymbol{\tau} - \mathbf{A} \cdot \mathbf{X}] \cdot [\boldsymbol{\tau} - \mathbf{A} \cdot \mathbf{X}]^T$ .

While the Ring effect is treated as a pseudo-absorber, a synthetically produced (Chance and Spurr, 1997) cross section that is fitted just like any other absorber, non-linearities are addressed by applying Levenberg–Marquardt's approach to non-linear fitting problems to Eq. (9) (Merlaud, 2013; Bevington and Robinson, 2003):

$$\begin{aligned} \ln\left(\frac{I_{\text{ref}}(\lambda)}{I(\lambda + \text{shift}) + \text{offset}}\right) &= \sum_{i=1}^n \sigma'_i(\lambda) \cdot \Delta SC_i \\ &+ \sum_{j=0}^m a_j \cdot \lambda^j, \end{aligned} \quad (9)$$

where shift and offset, which represent spectral wavelength shifts and stray light offsets, respectively, are responsible for the non-linear character of the problem.

The FFF system and its algorithm are based on the Passive DOAS technique by making scattered sunlight spectral measurements. Our algorithm differs from the original method mainly because of its very particular goal – fire detection through smoke. This objective is very different from the precise quantification of a certain trace gas concentration levels and means that both hardware and software must be adapted to the task at hand, as described in Sect. 4.

#### 4 The device

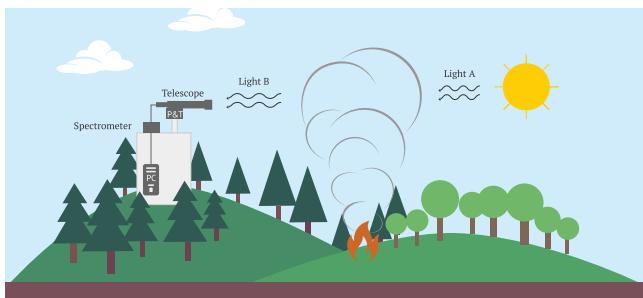
The Forest Fire Finder (see Fig. 4) is a remote sensing system that has the goal of detecting forest fires. It is a sophisticated piece of equipment with many features and customisation possibilities. Its complexity and the fact that it is meant to operate 24 h per day create a need for control electronics and instrumentation. These devices are out of scope for this paper and will be revisited for another article that will include a detailed description of the FFF control software. This section aims to give a brief and basic hardware/software presentation for daytime spectroscopic operation and fire detection.

The FFF scans the horizon for the presence of a column of smoke by performing sequential spectroscopic measurements of its surrounding environment using only the Sun as a light source, as illustrated by Fig. 5. Sunlight is captured with a Maksutov–Cassegrain telescope and guided through an optical fiber cable into a spectrometer, which will transform it into an electric signal.

The system has to cover wide areas, which is why the telescope is mounted on an ENEO VPT-501 pan and tilt unit



**Figure 4.** The Forest Fire Finder system in one of its deployment locations in the north of Portugal.



**Figure 5.** FFF illustration. The system continuously scans the horizon, in search of a smoke column.

that ensures the device's movement. The pan and tilt head unit assembly also includes a full-HD camera, which is used primarily for the optical alignment of the system and for human validation and supervision. During the night, this camera is also used for fire detection purposes; however, that is not within the scope of this paper and will be approached in another publication.

The Maksutov–Cassegrain telescope design uses the folded tube of the Cassegrain types and the spherical shape of primary mirror, secondary mirror and corrector lens of the Maksutov. In the FFF case, the chosen 90 mm aperture and 13.8 *f* ratio telescope, with a field of view of 1.4°, represents the best compromise between size, magnification and amount of captured light. In addition, it is also a cost-effective solution for the task at hand.

The AvaSpec 2048 is a popular 2048-pixel CCD photo array spectrometer. It can be customised with several slit sizes and gratings in order to suit the application it is intended for. In the case of the FFF system, a 50 µm slit is used in conjunction with a 300 lines mm<sup>-1</sup> grating, which ensures a wavelength range of 800 nm, from 300 to 1100 nm at a spectral resolution of 2.4 nm.

The spectrometer is connected to a computer, which is responsible for data processing and fire detection. It runs a custom-made software, developed in MathWorks' MATLAB development suite and C#. This software is deployed as a Microsoft Windows Service, as part of the FFF software suite.

## 5 Automatic smoke detection

The Forest Fire Finder is an electronic device that performs a spectroscopic analysis of the sky above the horizon, with the aim of detecting the presence of a smoke column. Smoke detection depends on the fire's emissions, which influence the composition of the atmosphere and on the system's spectroscopic algorithms, which allow those changes to be detected.

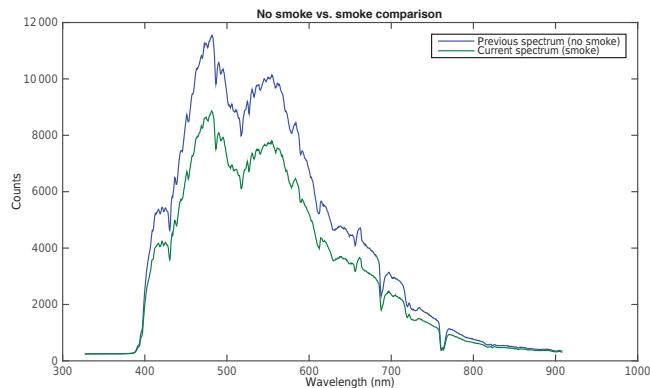
### 5.1 Forest fire emissions and DOAS

Forest fire smoke is a complex mixture of gases and aerosols that considerably changes the atmosphere (Urbanski et al., 2008). Among its key components are carbon oxides (CO and CO<sub>2</sub>), methane (CH<sub>4</sub>), non-methane hydrocarbons, volatile organic components, nitrous oxides (NO<sub>x</sub>) and particulate matter (Van Der Werf et al., 2010; Ward and Hardy, 1991; Spichtinger et al., 2004). Trace gases in smoke have a definite impact on the atmosphere's optical properties since some absorb light in the visible region of the electromagnetic spectrum. In addition to this, and depending on the combustion process, fire gives rise to the formation and emission of solid particles (Ward and Hardy, 1991). Given their size, these particles become aerosols, which influence light in all wavelengths due to Mie's scattering.

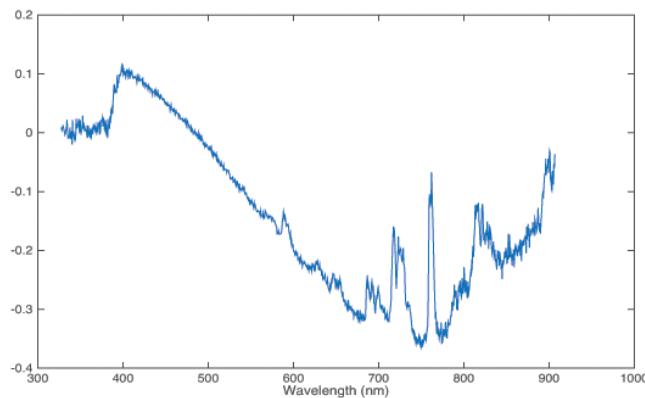
In contrast, fire emissions also alter the balance between the perceived column densities of water (H<sub>2</sub>O), oxygen (O<sub>2</sub>), ozone (O<sub>3</sub>) and the oxygen dimer (O<sub>4</sub>). All of these chemicals' cross sections are significant in the visible part of the spectrum.

Passive DOAS measurements are commonly used to retrieve the atmospheric column densities of several chemical compounds. Smoke columns, however, present themselves as sudden and localised changes in atmospheric concentrations. If one were to use this technique and analyse their absolute concentration values per se, it would be very difficult to infer the presence of smoke.

This does not mean passive DOAS cannot be used in this context. In fact, this method is very effective in detecting smoke if we put a "smoky" spectrum as  $I$  and a "normal" spectrum as  $I_0$  in Eq. (2), resulting in a signal as displayed in Fig. (7). Thus, by continuously acquiring spectra in a set of fixed azimuths and comparing the retrieved DOAS signals in pairs by azimuth, a narrow change such as the one produced by a forest fire becomes discernible in time. The FFF algorithm, presented in Sect. 5.2, does precisely this. These alterations are often difficult for the human eye to see, but there are some artificial intelligence algorithms that



**Figure 6.** This plot shows how a smoke column can influence a spectral measurement. Both these spectra were acquired on 29 December 2014, with a time difference of approximately 5 min and no azimuth difference.



**Figure 7.** Example of a signal obtained by dividing two consecutive spectra of the same azimuth and calculating the logarithm of this division, the differential optical thickness. It is this signal that is fitted through Eq. (9).

have been shown to be effective in separating the sky from a smoke column event and which we will discuss in Sect. 5.2.2.

## 5.2 The FFF algorithm

The FFF algorithm uses the mathematical ingenuity behind DOAS measurements to attempt the detection of forest fire smoke columns. It is important to bear in mind that this device is meant for real-time automatic detection of a forest fire. This creates strict time and memory constraints that the algorithm must adapt to in order to accomplish its function.

The system relies on its continuous movement at constant speed to provide spatially accurate detections. The spectrometer acquires  $2 \text{ spectra s}^{-1}$ , which are all analysed by the computer. Spectral integration time varies from 60 ms to 450 ms and is typically 210 ms. DOAS calculations are currently taking between 250 ms and 350 ms in the industrial computers powered by Intel i5 we use on the system. This

means that the software is almost always lagging behind the hardware. The system is designed to cope with this delay, and does so with great robustness, but in larger scans this lag becomes sufficiently significant as to compromise real-time fire detection.

These limits are reflected in several steps in the DOAS calculation process, such as the non-inclusion of stretch effects in the non-linear DOAS problem or the consideration that there are no atmospheric temperature fluctuations over the optical path. This approach would not be valid if we wanted to make a precise quantification of a certain compound's atmospheric column, but it works given the fact that we only want to distinguish between a smoky spectrum and non-smoky spectrum.

The algorithm is divided into two separate phases. The first phase, which is run in real time as soon as the spectrum arrives, is the chemical phase, described in Sect. 5.2.1. Results from this stage are stored in memory and accessed at the end of each scan by the second phase, which performs the classification of the analysis as a detection or a non-detection. This phase is described in Sect. 5.2.2.

### 5.2.1 The chemical phase

This algorithm section happens immediately after spectral acquisition. It corresponds to a passive DOAS analysis (see Sect. 3) of the spectrum in two different conceptual levels, as illustrated in Fig. 10. The first level uses the same azimuth spectrum of the previous scan as a reference spectrum in the DOAS calculations. The second level uses the mean of the ten spectra immediately to the left and to the right of the analysed spectrum for the same purpose. This processing level was created in an empirical way, after observing that in the presence of strong winds, smoke columns move horizontally. Both levels are calculated using literature spectra, compiled in Table 1. In practice, these two processing levels represents three possible ways of applying Eq. (2): considering  $I_0$  as the previous spectra acquired with the same azimuth; considering  $I_0$  as the average of the ten immediate spectra to the left of current azimuth; and considering  $I_0$  as the average of the ten immediate spectra to the right of current azimuth (this last method implies delayed processing of the current spectrum).

In this stage, the algorithm proceeds to retrieve column density values for  $\text{NO}_2$ ,  $\text{H}_2\text{O}$ ,  $\text{O}_2$ ,  $\text{O}_3$  and  $\text{O}_4$ , using Eq. (9) and the Levenberg–Marquardt algorithm. Figures 8 and 9 show actual retrieved column densities for one of the Peneda-Gerês FFF systems, taken on 26 December 2014. The charts were produced using the QDOAS software (Danckaert et al., 2015). These plots are all relative to the first level of processing, as described above. In parallel to the column density retrieval, short- and long-wavelength energy contribution and signal-to-noise ratio (SNR) are also computed for the analysed spectrum optical density (with relation to the previous spectrum of the same azimuth). All the processed data are

stored in a single matrix, which will be used by the second stage of the FFF algorithm, described in Sect. 5.2.2.

As stated, in the first level of processing the FFF uses the same azimuth spectrum of the previous scan as a reference. This means that between the spectrum of interest and the reference there is only an approximately 5 min difference. This in turn implies that, when plotted, the fitting signal is almost always negligible for non-smoky spectra (see the top of Figs. 8 and 9). However, smoky spectra have higher column density differences, which produce fitting plots with somewhat more-pronounced signals (see bottom of Figs. 8 and 9).

At this point, it is important to note two things. First and foremost, this kind of measurements produce results that are near the limit of the FFF's detection capabilities, which can still be numerically used to train a detection algorithm. Second, it should be highlighted that smoke-detection patterns were not manually defined. It would be extremely hard for a human to consider all the subtle changes in the millions of spectra acquired by the FFF systems and find a discerning pattern with that information. Instead, and as will be shown in Sect. 5.2.2, a machine learning algorithm was used to perform this task.

### 5.2.2 The classification phase

The classification stage runs at the end of each scan and goes through the chemical data gathered and stored in the previous algorithmic stage. It is divided into two levels, which are sequentially run. The first level acts as a filter for the second level, which is comprised of a support vector machine (SVM) classifier, an artificial intelligence supervised learning algorithm.

The first level starts by assembling two signals from the gathered data: one is comprised of the SNR value for each spectrum in the scan and the other is assembled by calculating an average signal energy per pixel, by dividing the sum of the square of each spectrum and by the number of its pixels and sequentially storing this result in an array.

The algorithm then applies a peak detection routine to the second artificial signal. For each detected peak, the system evaluates the SNR of the corresponding spectrum and compares it to a fixed threshold value, which is set in a configuration file. If this value is higher than said threshold, the spectrum is marked to be further analysed.

The next step involves the calculation of column density ratios between O<sub>2</sub>, O<sub>4</sub>, NO<sub>2</sub>, H<sub>2</sub>O and O<sub>3</sub> for the marked spectra. These ratios are then fed to the SVM, which returns a binary classification result (1 or 0).

An SVM is an algorithmic approach to the problem of classification in the context of supervised learning (Press et al., 2007). Introduced in 1992, by Boser et al. (1992), this method has since proved itself of great usefulness by providing relatively straightforward solutions to previously complicated classification applications. SVMs are generally easier to implement and understand, and this has also contributed

**Table 1.** Literature spectra used for the FFF's passive DOAS calculations. All cross sections are downloaded from the MPI-Mainz UV/VIS Spectral Atlas of Gaseous Molecules of Atmospheric Interest (Keller-Rudek et al., 2013).

| Compound                            | Interval used (nm) | Reference key         | Year |
|-------------------------------------|--------------------|-----------------------|------|
| Oxygen (O <sub>2</sub> )            | 600–800            | Bogumil et al. (2003) | 2003 |
| Ozone (O <sub>3</sub> )             | 500–650            | Bogumil et al. (2003) | 2003 |
| Oxygen dimmer (O <sub>4</sub> )     | 400–800            | Hermans (2011)        | 2011 |
| Water vapour (H <sub>2</sub> O)     | 400–500            | Coheur et al. (2002)  | 2002 |
| Nitrogen dioxide (NO <sub>2</sub> ) | 400–500            | Vandaele (2002)       | 2002 |

for their fast spread. The general concept behind the SVM methodology is to find and define the hyperplane that better separates data into two classes (Press et al., 2007).

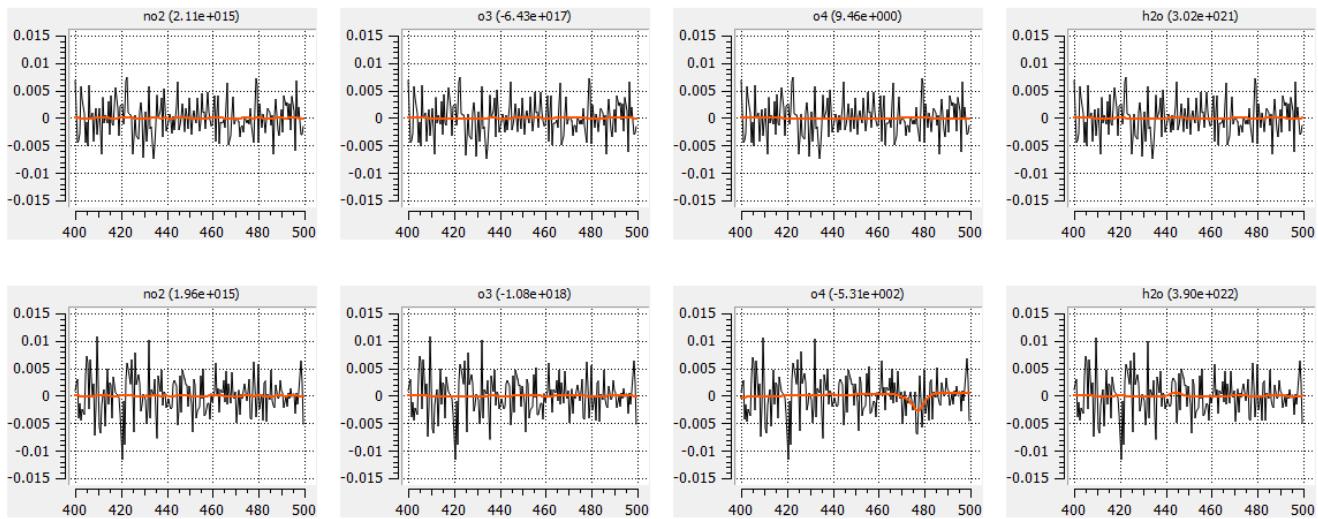
Like all supervised learning techniques, SVMs need to be trained prior to being used. In the case of the FFF classification algorithm, SVMs were trained in successive generations, with each generation built upon the results of the previous.

The first FFF SVM generation was built using data from 60 different moments in 2014 (fire and non-fire), in 13 different locations in the Peneda-Gerês National Park (PNPG). A specially designed software tool was used to manually search and store fire data from 30 different events and 30 non-fire moments in said year and an SVM was trained with the resulting information. Another custom-made software was used to classify data of several months with the created SVM. The second-generation SVM was created by repeating the process, taking the results of the first-generation SVM into account and testing against a new set of data.

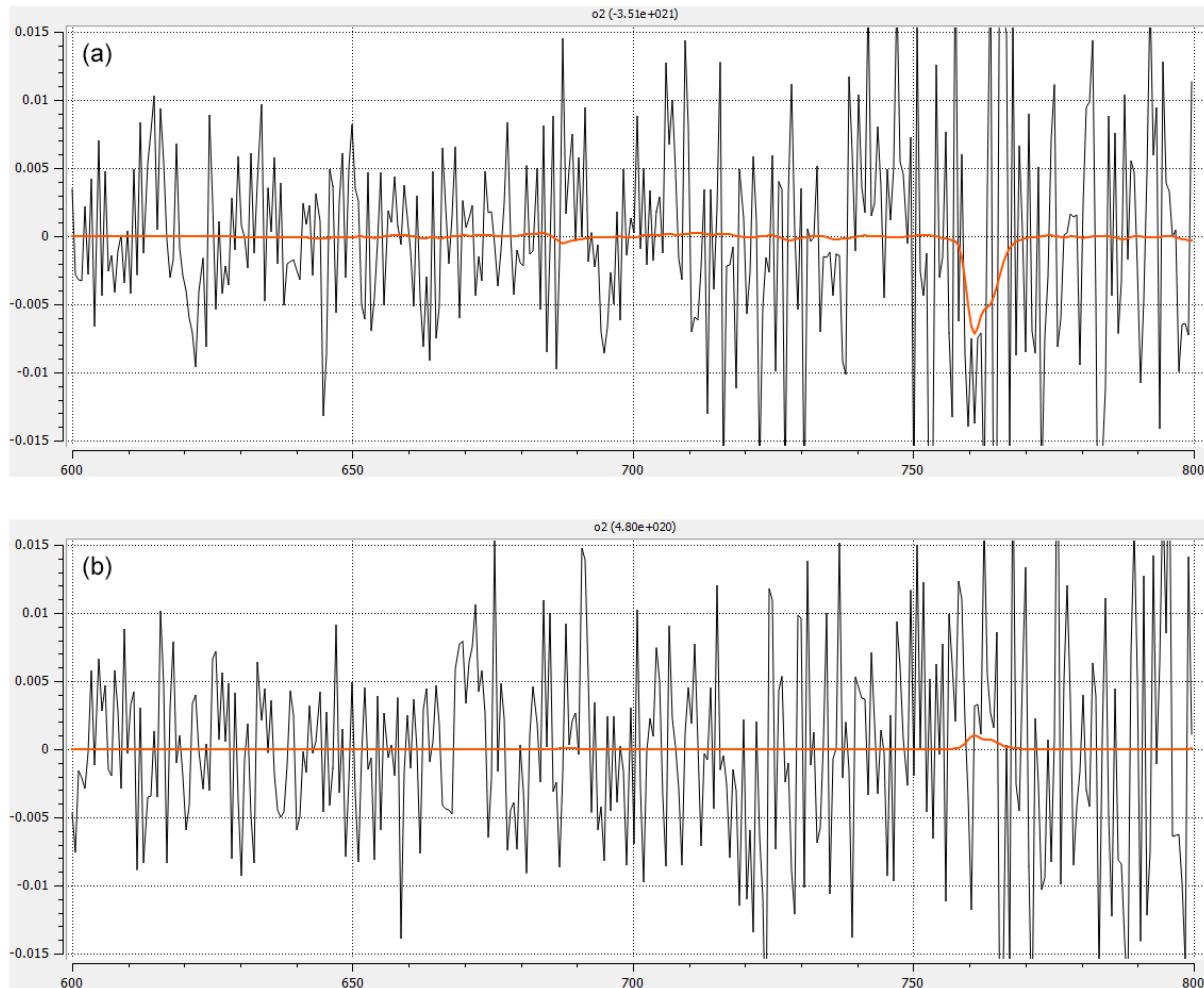
One downside of using this kind of classifier is that, after training and application, the classification becomes opaque. It is not feasible to understand what caused a false detection. This raises the problem of how to improve the detection capabilities of the FFF system. There are two fundamental approaches to this issue: focusing on the SVM or adding more information to the decision process.

The current line of thought concerning this situation is that, in normal circumstances, the classifier's performance is quite satisfactory. Still, some peculiar events cause the system to issue wrong detections. These events are many times of systematic nature, such as particular kind of cloud that only appears in the early morning or emissions from local industrial facilities. These interferences can be detected themselves, without altering the smoke column detection algorithm, which is where the majority of improvement efforts will be made in the near future (see Sect. 7).

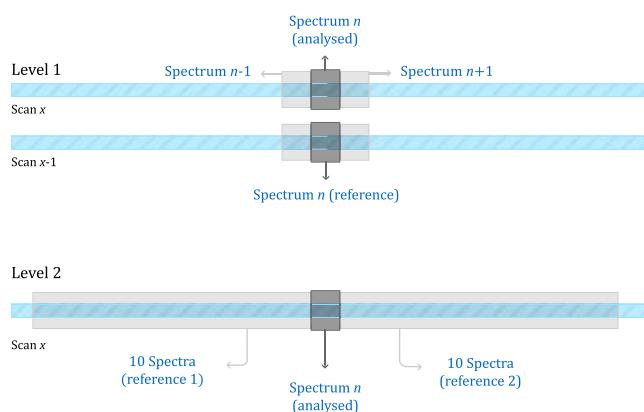
Finally the algorithm reaches the final point, at which it has to say whether the data that were run correspond to a smoke column in the horizon. The spectral information entered to the SVM includes the results from the two chemical processing levels debated in Sect. 5.2.1 and originates three classification possibilities:



**Figure 8.** Column densities for NO<sub>2</sub>, O<sub>3</sub>, O<sub>4</sub> and H<sub>2</sub>O, retrieved between 400 and 500 nm. The plots below correspond to a smoke column detection. In these plots, the horizontal axis corresponds to wavelength in nm and the vertical axis to the differential optical thickness.



**Figure 9.** Column density for O<sub>2</sub>, retrieved between 600 and 800 nm. The plot below corresponds to a smoke column detection. In these plots, the horizontal axis corresponds to wavelength in nm and the vertical axis to the differential optical thickness.



**Figure 10.** Illustration of FFF's two processing levels in the chemical stage of the algorithm.

- smoke column detected between previous and current scan;
- smoke column detected on the left of analysed spectrum;
- smoke column detected on the right of analysed spectrum.

If two of these classification results are positive, the system issues an alarm.

## 6 Results and discussion

In 2013, 13 FFF devices were deployed in the PNPG, in the north of Portugal. Their placement reflected topography, local accessibility and fire protection needs.

In 2015, FFF data were gathered and compared to official data from the Portuguese National Authority of Civil Protection (ANPC), the country's institution responsible for forest fire protection and management.

Table 2 shows said data and comparison. In it, a confirmed detection occurs when a smoke column is sensed by an FFF device and the detection is validated by a human operator. This is different from a registered fire event (RFE), which is a fire that was inserted into ANPC's database.

Official statistics count 132 fire events in 2015 within PNPG. During the same period, the FFF network issued 578 detections, of which 369 were false detections, and 209 confirmed events, of which 53 were coincident with RFEs.

Although the false detections may seem to be high in comparison to confirmed detection, it is important to bear in mind that each system has an average working period of 12 h per day. At two spectra per second, this means an average of 86 400 analysed spectra per system per day. Since each and every one of these analysis can trigger an alarm by itself, false detections reach only 0.0000833 % in all systems. In addition, there have been some events that were wrongly

**Table 2.** FFF statistics for 2015 in the Peneda-Gerês National Park.

|                             |             |
|-----------------------------|-------------|
| Registered fire events      | 132         |
| Total fire detections       | 578         |
| False events                | 369         |
| Confirmed detections        | 209         |
| Estimated network           | 56 940      |
| Operation time (h)          |             |
| Estimated analysis          | 409 968 000 |
| False detection %           | 0.0000900   |
| False alarms per system/day | 0.07776607  |

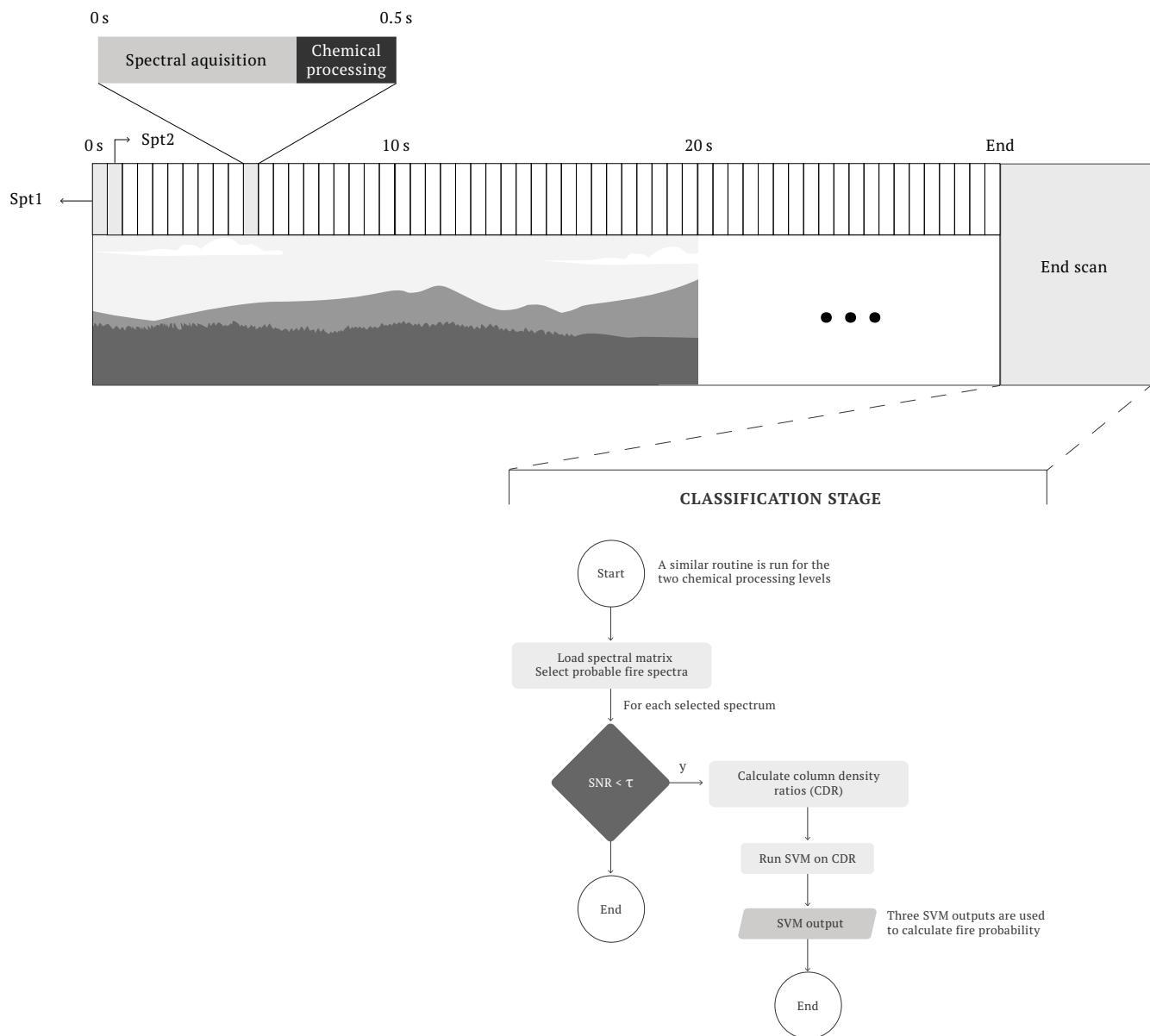
marked as false detections due to misunderstandings on how the system is to be handled by humans. These events correspond mainly to small fires and prescribed burns, large enough to be detected. Security issues, regarding the Portuguese Civil Protection Authority, prevent the exact quantification of human errors, but they amount to a significant percentage of false alarms.

Although the presented numbers are enough to paint a general picture of the FFF system's behaviour, the available data do not allow a thorough quantification of the system's performance, since there is no correct gold standard regarding forest fires due to fire registry procedures not being clearly established. This becomes exceedingly apparent when comparing the number of RFEs and the number of confirmed detections: every confirmed detection was deemed relevant by a human operator, yet there are only 132 RFEs for the 209 confirmed detections.

Another important result that becomes noticeable in Table 2 is the fact that false positives and true detections vary in similar ways. This can be explained by the fact that the FFF is a spectroscopic system at its heart. Fire releases chemical components into the atmosphere, which in turn are detected by the system. If there are many fires in a small geographic region, such as the PNPG, it is possible that an FFF is able to sense it, without the presence of a visible smoke column in its patrol path. We have also noticed a trend for false alarms on specific cloudy days. We believe this is due to pollutant particles carried by the clouds. Light scattering by these particular clouds sometimes breaches through the system's energy and SNR filters (see Sect. 5.2.1) and is incorrectly classified as a smoke column.

## 7 Future developments

The FFF has been an ongoing development for NGNS-IS, Ltd. Since 2006, the device has undergone two complete redesign processes, motivated by hardware improvements. The current version is without a doubt the most robust and reliable design ever, achieving uptimes of 99 %. There will always be room for improvement regarding hardware, but given the operating level of the current version these will not be a priority in the near future.



**Figure 11.** FFF algorithm simplified work-flow diagram.

The software architecture selected for the system allows total freedom for future development needs, with minimal integration efforts. As stated in Sect. 4, this was a requirement because most customers need some level of customisation.

As mentioned in Sect. 5, the FFF is subject to a number of strict time constraints, mostly related to the real-time nature of the system. As a consequence, one of our ever-present goals is algorithm optimisation. Currently, the spectral algorithms are being rewritten, with the inclusion of more features in less processing time. It is also true that as computers become more powerful, it is possible to add more details to the algorithms.

Another line of research currently being pursued is the study of false positives and their relation to the presence of aerosols in the atmosphere. These developments are, however, entirely dependent on the optimisation, since it is currently not possible to add this functionality without sacrificing real-time operation.

DOAS (see Sect. 3) is a widely used atmospheric analysis technique, with much broader uses than just fire sensing. In fact, the technique is used for urban air pollution monitoring (OPSIS, 2016), almost-real-time volcanic plumes monitoring for aviation control (Brenot et al., 2014) and quantification of volcanic gases (Galle et al., 2010). The experience attained while developing FFF allowed the creation of another

project, called Project ATMOS, designed to monitor and control air pollution, crop maturity, water stress and agricultural plagues. This project started in April 2016 and is expected to conclude in September 2019.

## 8 Conclusions

Life on Earth is greatly influenced and shaped by fire events. Humans in particular depend on fire to maintain their technology and way of life. However, forest fires are a global menace that cause concern all over the world. Several industrialised countries have allocated a great deal of resources to researching wildfires and their behavioural dynamics.

It is this concern that generates a very large market for remote sensing equipment for early forest fire detection, a market which is expected to grow 11.53 % by 2020.

This article addresses one of such equipments. The FFF was developed in Lisbon in 2006 by NGNS-IS, Ltd. and is the only one that is based on optical spectroscopy, particularly differential optical absorption spectroscopy.

In 2013, a 14-element network of FFF devices was installed in the Peneda-Gerês National Park. In 2015, this network was able to detect a confirmed 209 fire events, a number significantly higher than the officially registered 132 fire occurrences in the same region.

For the same period the system has issued 369 false detections, but it is worth considering that confirmed and false detections have similar trends. This is due to the fact that fires pollute the atmosphere with the chemicals that the system aims to detect and is in agreement with the device's operating physical principle, optical spectroscopy.

Evaluation of a fire detection system is a very difficult task. There is no formal definition of how large a forest fire must be to be considered an event and this means there is no perfect classification model (a gold standard) to compare the system's performance to.

The FFF is an automatic forest fire detection system which has proven itself to be effective and detected a great number of forest fires (see Table 2 in Sect. 6). In addition, the system's current hardware and software configuration has resulted in extremely high uptime levels, contributing for an adequate fire detection coverage and, consequently, optimal levels of fire protection.

*Data availability.* The data used in this paper are not publicly available nor are we authorised to make them so.

*Competing interests.* The authors are affiliated with NGNS, which developed the Forest Fire Finder system and may be affected by the research reported in the enclosed paper.

*Acknowledgements.* The authors would like to thank the NGNS team for their support, especially Hélia Pinto, for the numerous drawings and diagrams. This work was funded in part by the FCT PhD grant PD/BDE/114549/2016 and by the FEDER-3491 project, through Lisbon's Operational Program LISBOA2020.

Edited by: M. Van Roozendael  
Reviewed by: two anonymous referees

## References

- Alkhatib, A. A. A.: A review on forest fire detection techniques, International Journal of Distributed Sensor Networks, 2014, 597368, <https://doi.org/10.1155/2014/597368>, 2014.
- Bevington, P. R. and Robinson, D. K.: Data Reduction and Error Analysis for the Physical Sciences, 2003.
- BNHCRC: Bushfire & Natural Hazards CRC, available at: <http://www.bnhcrc.com.au/home>, last access: 10 May 2016.
- Bogumil, K., Orphal, J., Homann, T., Voigt, S., Spietz, P., Fleischmann, O., Vogel, A., Hartmann, M., Kromminga, H., Bovensmann, H., Frerick, J., and Burrows, J.: Measurements of molecular absorption spectra with the SCIAMACHY pre-flight model: instrument characterization and reference data for atmospheric remote-sensing in the 230–2380 nm region, J. Photoch. Photobio. A, 157, 167–184, [https://doi.org/10.1016/S1010-6030\(03\)00062-5](https://doi.org/10.1016/S1010-6030(03)00062-5), 2003.
- Boser, B. E., Guyon, I. M., and Vapnik, V. N.: A training algorithm for optimal margin classifiers, in: Proceedings of the fifth annual workshop on Computational learning theory – COLT '92, ACM Press, New York, New York, USA, 144–152, <https://doi.org/10.1145/130385.130401>, 1992.
- Brenot, H., Theys, N., Clarisse, L., van Geffen, J., van Gent, J., Van Roozendael, M., van der A, R., Hurtmans, D., Coheur, P.-F., Clerbaux, C., Valks, P., Hedelt, P., Prata, F., Rasson, O., Sievers, K., and Zehner, C.: Support to Aviation Control Service (SACS): an online service for near-real-time satellite monitoring of volcanic plumes, Nat. Hazards Earth Syst. Sci., 14, 1099–1123, <https://doi.org/10.5194/nhess-14-1099-2014>, 2014.
- Brinkmann, R. T.: Rotational Raman scattering in planetary atmospheres, Astrophys. J., 154, 1087–1093, <https://doi.org/10.1086/149827>, 1968.
- Chance, K. V. and Spurr, R. J. D.: Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, Appl. Opt., 36, 5224, <https://doi.org/10.1364/AO.36.005224>, 1997.
- Coheur, P.-F., Fally, S., Carleer, M., Clerbaux, C., Colin, R., Jenouvrier, A., Mérienne, M.-F., Hermans, C., and Vandaele, A. C.: New water vapor line parameters in the 26000-region, J. Quant. Spectrosc. Ra., 74, 493–510, [https://doi.org/10.1016/S0022-4073\(01\)00269-2](https://doi.org/10.1016/S0022-4073(01)00269-2), 2002.
- Danckaert, T., Fayt, C., Van Roozendael, M., de Smedt, I., Letocart, V., Merlaud, A., and Pinardi, G.: QDOAS, available at: <http://uv-vis.aeronomie.be/software/QDOAS/> (last access: 12 June 2017), 2015.
- Envirovision Solutions: Forestwatch by EnviroVision Solutions, EVS, 4th Wildfire Detection Workshop, 11–12 March 2015, Edmonton, Alberta, Canada, 2015.

- European Comission: Catalogue of EU funded projects in Environmental research, 0, 2012.
- Food and Agriculture Organisation (FAO): Fire management – global assessment 2006, FAO Forestry Paper 151, 135 pp., available at: <http://www.fao.org/docrep/009/a0969e/a0969e00.htm> (last access: 12 June 2017), 2007.
- Galle, B., Johansson, M., Rivera, C., Zhang, Y., Kihlman, M., Kern, C., Lehmann, T., Platt, U., Arellano, S., and Hidalgo, S.: Network for Observation of Volcanic and Atmospheric Change (NOVAC) – A global network for volcanic gas monitoring: Network layout and instrument description, *J. Geophys. Res.-Atmos.*, 115, 1–19, <https://doi.org/10.1029/2009JD011823>, 2010.
- Grainger, J. F. and Ring, J.: Anomalous Fraunhofer Line Profiles, *Nature*, 193, 762–762, <https://doi.org/10.1038/193762a0>, 1962.
- Hermans, C.: Collisionally induced absorption spectrum of oxygen attributed to the oxygen dimer (O<sub>2</sub>)<sub>2</sub>, Tech. rep., BIRA-IASB Spectroscopy Lab, available at: <http://spectrolab.aeronomie.be/index.htm> (last access: 12 June 2017), 2011.
- Hough, G.: Vision Systems for Wide Area Surveillance: ForestWatch – a long-range outdoor wildfire detection system, in: TASSIE FIRE Conference Proceedings, <https://doi.org/10.1017/CBO9781107415324.004>, 2007.
- IPCC: Managing the risks of extreme events and disasters to advance climate change adaptation, <https://doi.org/10.1596/978-0-8213-8845-7>, 2012.
- IQ-Wireless: FireWatch, available at: <http://www.fire-watch.de/>, last access: 19 May 2016.
- Keller-Rudek, H., Moortgat, G. K., Sander, R., and S'renzen, R.: The MPI-Mainz UV/VIS Spectral Atlas of Gaseous Molecules of Atmospheric Interest, *Earth Syst. Sci. Data*, 5, 365–373, <https://doi.org/10.5194/essd-5-365-2013>, 2013.
- Liyang, Y., Wang, N., and Meng, X.: Real-time forest fire detection with wireless sensor networks, in: Proceedings, 2005 International Conference on Wireless Communications, Networking and Mobile Computing, 2, 1214–1217, <https://doi.org/10.1109/WCNM.2005.1544272>, 2005.
- Manyangadze, T.: Forest fire detection for near real-timemonitoring using geostationary satellites, International Institute for Geo-Information Science and Earth Observation, MSc, 80, 2009.
- Merlaud, A.: Development and use of compact instruments for tropospheric investigations based on optical spectroscopy from mobile platforms, Books.Google.Com, 2013.
- NGNS-IS, L.: Long range fire detection exclusive solution, available at: <http://www.ngns-is.com/products/ngns-all-over/forest-fire-finder>, last access: 10 May 2016.
- OPSIS: Opsiis, availabel at: <http://www.opsis.se/Aboutus/tabid/996/Default.aspx>, last access: 19 May 2016.
- Platt, U. and Stutz, J.: Differential Optical Absorption Spectroscopy, Springer, Heidelberg, Germany, 2007.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes: The Art of Scientific Computing, 3rd Edn., 2007.
- Research and Markets: Fire Protection Systems Market by Technology (Active, Passive), Product (Fire Detection (Sensors & Detectors, RFID), Management (Fire Sprinklers, Extinguishers), Fire Analysis & Response), Service, Vertical, and Geography – Global Forecast to 2022, Tech. rep., Research And Markets, avaible at: <http://www.researchandmarkets.com/reports/3696148/fire-protection-systems-market-by-technology> (last access: 12 June 2017), 2016.
- Spichtinger, N., Damoah, R., Eckhardt, S., Forster, C., James, P., Beirle, S., Marbach, T., Wagner, T., Novelli, P. C., and Stohl, A.: Boreal forest fires in 1997 and 1998: a seasonal comparison using transport model simulations and measurement data, *Atmos. Chem. Phys.*, 4, 1857–1868, <https://doi.org/10.5194/acp-4-1857-2004>, 2004.
- United States Forest Department: The Rising Cost of Fire Operations: Effects on the Forest Service's Non-Fire Work, United States Department of Agriculture, 13 pp., available at: <http://www.fs.fed.us/sites/default/files/media/2014/34/nr-firecostimpact-082014.pdf> (last access: 12 June 2017), 2015.
- Urbanski, S. P., Hao, W. M., and Baker, S.: Chapter 4 Chemical Composition of Wildland Fire Emissions, *Develop. Environ. Sci.*, 8, 79–107, [https://doi.org/10.1016/S1474-8177\(08\)00004-1](https://doi.org/10.1016/S1474-8177(08)00004-1), 2008.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, 10, 11707–11735, <https://doi.org/10.5194/acp-10-11707-2010>, 2010.
- Vandaele, A. C.: High-resolution Fourier transform measurement of the NO<sub>2</sub> visible and near-infrared absorption cross sections: Temperature and pressure effects, *J. Geophys. Res.*, 107, 4348, <https://doi.org/10.1029/2001JD000971>, 2002.
- Ward, D. E. and Hardy, C. C.: Smoke emissions from wildland fires, *Environment International*, 17, 117–134, [https://doi.org/10.1016/0160-4120\(91\)90095-8](https://doi.org/10.1016/0160-4120(91)90095-8), 1991.

B

**TOMOSIM: A TOMOGRAPHIC SIMULATOR  
FOR DIFFERENTIAL OPTICAL ABSORPTION  
SPECTROSCOPY**

## Article

# TomoSim: A Tomographic Simulator for Differential Optical Absorption Spectroscopy

Rui Valente de Almeida <sup>1,2,\*</sup>, Nuno Matela <sup>3</sup> and Pedro Vieira <sup>1</sup>

<sup>1</sup> Physics Department, FCT NOVA, NOVA University of Lisbon, 2829-516 Caparica, Portugal; pmv@fct.unl.pt

<sup>2</sup> FutureCompta SA, 1495-190 Algés, Portugal

<sup>3</sup> Biomedical Engineering and Biophysics Institute, University of Lisbon, 1749-016 Lisbon, Portugal; nmatela@fc.ul.pt

\* Correspondence: rf.almeida@campus.fct.unl.pt

**Abstract:** TomoSim comes as part of project ATMOS, a miniaturised Differential Optical Absorption Spectroscopy (DOAS) tomographic atmospheric evaluation device, designed to fit a small drone. During the development of the project, it became necessary to write a simulation tool for system validation. TomoSim is the answer to this problem. The software has two main goals: to mathematically validate the tomographic acquisition method; and to allow some adjustments to the system before reaching final product stages. This measurement strategy was based on a drone performing a sequential trajectory and gathering projections arranged in fan beams, before using some classical tomographic methods to reconstruct a spectral image. The team tested three different reconstruction algorithms, all of which were able to produce an image, validating the team's initial assumptions regarding the trajectory and acquisition strategy. All algorithms were assessed on their computational performance and their ability for reconstructing spectral "images", using two phantoms, one of which custom made for this purpose. In the end, the team was also able to uncover certain limitations of the TomoSim approach that should be addressed before the final stages of the system.

**Keywords:** atmospheric monitoring; DOAS; tomography; UAV; drone



**Citation:** Valente de Almeida, R.; Matela, N.; Vieira, P. TomoSim: A Tomographic Simulator for Differential Optical Absorption Spectroscopy. *Drones* **2021**, *5*, 3. <https://doi.org/10.3390/drones5010003>

Received: 13 November 2020

Accepted: 21 December 2020

Published: 29 December 2020

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background and Motivation

This article details the construction of TomoSim, which is the software simulation package used in the ATMOS project, a Portuguese European Union (EU) funded initiative that aimed at creating a miniaturised spectroscopy platform for atmospheric monitoring and trace gas mapping. The project was a joint effort from Compta, S.A., one of the oldest IT groups in Portugal; and FCT NOVA, from the NOVA University of Lisbon, also one of the largest and most important science schools in the country.

The idea of using a tomographically capable drone with spectroscopic equipment for mapping trace gas concentrations in defined geographical regions made the need for a simulation tool evident from a very early stage. Any monitoring device that depends on drones for the actual measurements must take into account a number of restrictions, of both physical and logistical order (weight, battery life, etc.). These considerations are even more important if the measurements themselves depend on the ability of the drone to move and position itself a certain way and the precision with which it can do so. For the ATMOS project, they were paramount. Simulation is a simple and inexpensive way of determining the technical feasibility of the project and the mathematical validity of the reconstruction strategy.

This strategy is based on two premises:

- A custom-equipped drone should be able to measure trace gas column density in a carefully programmed sequence;

- One can then organise these measurements into an array.

These column density values could then be used as projections in a tomographic reconstruction. The resulting image would correspond to the geographic distribution of the target trace gases.

The simulator, built in Python [1] and using NumPy [2], uses the system's premise and applies three algorithms for reconstruction: fan beam Filtered BackProjection (FBP), Simultaneous algebraic Reconstruction Technique (SART) and Maximum Likelihood Expectation Maximisation (MLEM). Algorithm verification and analysis was performed using a custom-designed phantom for DOAS spectral measurements. All three mathematical algorithms managed to reconstruct the image. FBP was shown to be significantly more computationally cost-effective than the other two. The custom coded MLEM routine was not on-par with the other two solutions, which were plugged directly from the SciPy library.

Besides runtime and other significant software engineering considerations, the development of TomoSim concluded that the assumptions of the team regarding the acquisition strategy for spectral geographic images were correct, but lifted the veil on some limitations that should be addressed before building the final system.

The paper is structured as follows: the section after this introduction, Section 1, targets the theoretical background with which the paper was built; Section 2 describes the design, the rationale behind it and the technical choices that have been taken in the making of TomoSim; in Section 3 there is a description of the results that were obtained through the simulator, including reconstructions, running times and Section 5 is dedicated to the conclusions that were taken using this piece of software, as well as some foreseeable future developments.

## 1.2. Differential Optical Absorption Spectroscopy (DOAS)

DOAS is a widely used atmospheric analysis technique, developed in the late 1960s. Fundamentally, it is an absorption spectroscopy technique, therefore based on Lambert-Beer's law. This law was actually first formulated by Pierre Bouguer in 1729. At the time, he wrote that "in a medium of uniform transparency, the light remaining in a collimated beam is an exponential function of the length of the path in the medium" [3]. This theory can thus be written as in Equation (1).

$$I(\lambda) = I_0(\lambda) \cdot \exp(-\sigma(\lambda) \cdot c \cdot L). \quad (1)$$

$I_0(\lambda)$  is the source intensity of a light beam,  $I(\lambda)$  the intensity of the light that reaches the detector,  $\sigma$  is the absorption cross-section of the chemical compound being measured,  $c$  its concentration and  $L$  the optical path of the light (the length of the path that is traversed by the ray of light). Finally,  $\lambda$  is the wavelength of the radiation.

The rearrangement of Equation (1) gives us the concentration, and more importantly, allows us the creation of a new quantity,  $\tau$ , which is called *optical depth*. This is explicit in Equation (2) [4,5].

$$c = \frac{\ln\left(\frac{I_0(\lambda)}{I(\lambda)}\right)}{\sigma(\lambda) \cdot L} = \frac{\tau}{\sigma(\lambda) \cdot L}. \quad (2)$$

In the laboratory, this equation can be (and is) used directly and with few obstacles, since there are very few uncontrolled variables, and that which exists is controlled for. In the open atmosphere, this is not the case at all.

There are many factors that influence the degree to which we can apply Lambert-Beer's rule in atmospheric measurements. The most important one is the fact that in order to know the source intensity ( $I_0(\lambda)$ ), one would have to remove any and every absorber from optical path, and this is clearly infeasible in the great majority of circumstances. To surpass this problem, DOAS measures differences between absorption structures at many different wavelengths [5].

There are many different DOAS systems, but the single greatest distinction among them is whether they are active or passive. Active systems use artificial light sources such

as a Xenon lamp to measure chemicals in the air. Passive systems use natural sources, such as the Sun or the Moon (or other celestial bodies). This paper will focus on passive systems, since the ATMOS project is a scattered sunlight DOAS system, and thus passive.

Scattered sunlight DOAS systems have their own specific particularities and special properties. For instance, the optical path is unknown as the last scattering event cannot be determined. Scattering also implies that there is a fraction of the source's light that does not reach the detector. Moreover, one must account for all the other passive DOAS common effects, such as the fact that there are many absorbers that have spectral structures overlapping the target species; that there are Fraunhofer lines, Mie and Rayleigh scattering, turbulence and instrumental effects, and so forth [5,6].

All these influential effects mean that, in the open atmosphere, Equation (3) is a more plausible model than that presented in Equation (1).

$$I(\lambda) = I_0(\lambda) \cdot A(\lambda, \dots) \cdot \exp \left[ - \int_s \left( \sum_i \sigma_i(\lambda, s) \cdot c_i + \epsilon_M(\lambda, s) + \epsilon_R(\lambda, s) \right) ds \right]. \quad (3)$$

In Equation (3), there is more than one absorber, which is denoted by index  $i$  in the sum.  $A(\lambda, \dots)$  denotes the fraction of light that gets scattered into the detector,  $\epsilon_M$  and  $\epsilon_R$  are Mie and Rayleigh scattering coefficients, and the integral is performed on the whole optical path,  $s$ .

Typically, we measure a trace gas's atmospheric contribution by its total column. This quantity is essentially the integral of the compound's number density, in molecules/cm<sup>3</sup>, over a column that goes from the ground to the Top Of the Atmosphere (TOA). In Equation (4), the number density for molecule  $x$  is denoted  $c_x$ .

$$C_x = \int_0^{TOA} c_x(z) dz. \quad (4)$$

One can look again at Equation (3) with the interest of quantifying only the trace gas contribution. This contribution should be somehow related to the total column, since the only real difference between what the DOAS instruments see and the total column is the dependence on the optical path of the former. In fact, we can separate the integral to get the trace absorbers optical density expression, displayed in Equation (5). Since we can consider the cross-section to be fairly constant with the optical path, it is possible to separate this integral, and we arrive at the very important value of the slant column,  $SC_i$ , which is the integral of the gas's number density along the optical path.

$$SC_i = \int_s c_i(s) ds. \quad (5)$$

If we integrate the absorption coefficients inside the exponential term of Equation (3) to their optical densities, we arrive at the expression in Equation (6). This equation cannot be solved without some very seldomly available data, such as the original light source  $I_0$ . DOAS allows us to overcome this problem by making relative measurements (hence the differential in the name of the technique), using another scattered light spectrum as reference and by observing that most interesting trace gases have very narrow spectral structures, while effects like Mie and Rayleigh scattering have broad spectral features.

$$\ln \left( \frac{I_{ref}}{I} (\lambda) \right) = \ln \left( \frac{A_{ref}}{A} \right) + \sum_i \sigma_i(\lambda) \cdot \Delta SC_i + \tau_R(\lambda) + \tau_M(\lambda). \quad (6)$$

$I_{ref}$  and  $A_{ref}$  are, respectively, the reference light intensity and reference scattered light ratio and  $\tau_R$  and  $\tau_M$  the integrated optical depth of the Rayleigh and Mie scattering phenomena. It is thus possible to separate the original optical depth in two parts, a differential part, which is comprised of the narrowband contributions coming mainly from trace gases; and the “continuous” part, which contains the slow-variation, broadband contributions

in Equation (3). The latter can be approximated by fitting a low order polynomial to the original optical depth signal. If one subtracts this polynomial from the optical depth, the result is the narrowband signal that constitutes the differential part. Target absorbers column densities can then be found by fitting their cross-sections to the differential signal, using a numerical algorithm such as Levenberg-Marquardt's [7].

This small introduction to the topic of DOAS is more than enough to explain and understand the rest of this article regarding the spectroscopic procedures. References [5,6] contain a more thorough coverage of the topic, both with reference to the technique's physics and instrumental considerations.

### 1.3. The Tomography Problem

#### 1.3.1. Introduction

Tomography is the cross-sectional imaging of an object through the use of transmitted or reflected waves, collected by the object exposure to the waves from a set of known angles. Tomography has many different applications in science, industry, and in particular, medicine. Since the invention of the Computed Tomography (CT) machine, in 1972 by Hounsfield [8], tomographic imaging techniques have had a revolutionary impact, allowing doctors to see inside their patients without having to subject them to surgical interventions. Examples of tomographic applications for scientific and industrial purposes include radio-astronomy and certain kinds of tomographic non-destructive test techniques [9].

Tomography has had its mathematical basis set by Johannes Radon, a German mathematician that proved that it is possible to represent a function in  $\mathbf{R}$  (the domain of the real numbers) in the space of straight lines  $L$  through its line integrals. In the tomographic case, these integrals represent a measurement on a ray that is traversing the field of analysis. Each set of line integrals (rays), characterised by a given projection angle, is called a projection. The set of all projections, arranged in matrix form by projection angle and detector is called a sinogram. All reconstruction methods revolve around this matrix [9–11].

#### 1.3.2. Discretisation

Siddon's algorithm [12], published in 1985, is one of the most common and most studied ways to go from the naturally analogical presentation of the real world into computer-operable discrete geometric fields of vectors, in the field of tomography. The algorithm considers each pixel to be the intersection area of orthogonal sets of equally spaced parallel planes. With this in mind, it is possible to parametrise a ray of light, with the ray being written (in two dimensions):

$$\begin{aligned} X(\alpha) &= X_1 + \alpha(X_2 - X_1) \\ Y(\alpha) &= Y_1 + \alpha(Y_2 - Y_1). \end{aligned} \quad (7)$$

In Equation (7),  $X_1$  and  $Y_1$  are the coordinates for the entry point (of the ray in the analysis field) and  $X_2$  and  $Y_2$  are the exit point coordinates.  $\alpha$  is the parametric value. If the ray is totally contained within the field of analysis, this value varies between 0 and 1; otherwise, it has its minimum at the entry point and maximum at the exit.

The parametrical representation of the line integral, allows one to recursively calculate all intersections between the ray and the grid defined by the orthogonal lines described above. The differences between intersection points render the lengths of each ray contained within each pixel. The sum of all the lengths for one ray yields the total value of the line integral, which corresponds to a projection. The algorithm steps are presented in Algorithm 1 [12,13].

#### 1.3.3. Geometry

The application of analytical algorithms such as FBP depends on the type of geometric arrangement of the tomographic problem. TomoSim uses two tomographic geometries in its reconstructions: parallel and fan beam.

---

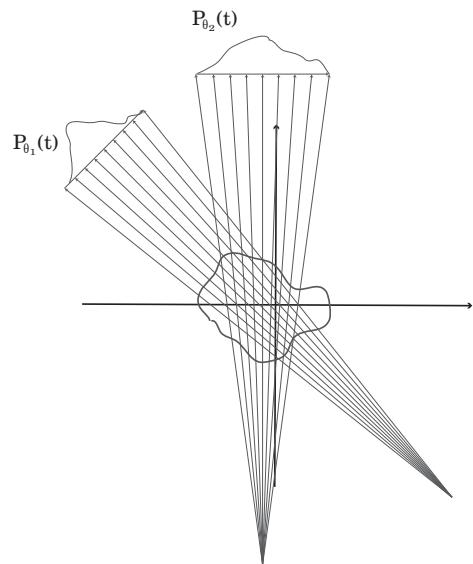
**Algorithm 1:** Siddon's algorithm's procedural steps. After running this algorithm, one is able to represent any continuous ray through the analysis field as a sum of discrete lengths

---

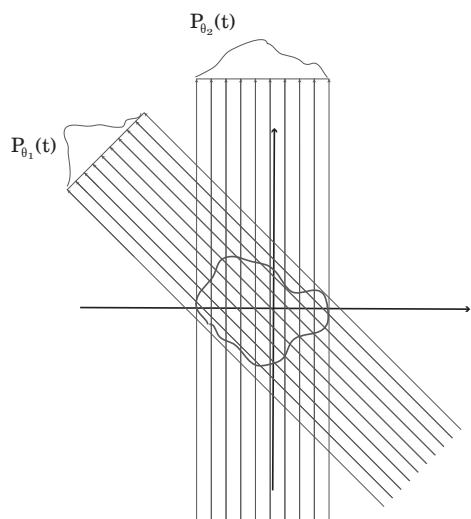
**Result:** Discretised Region of Interest(ROI).

- calculate range of parametric values;
  - calculate range of pixel indices;
  - calculate parametric sets;
  - merge sets;
  - calculate pixel(or voxel) lengths;
  - calculate pixel indices;
- 

Parallel Projection geometry (see Figure 1) is the most basic assembly. In this arrangement, each projection is a set of parallel line integrals, as can be seen in Figure 1. The radiation sources can be assembled in a linear array, facing detectors in the same number on the opposite side of the target object, or, in alternative, a single source can move in a linear trajectory, directing its rays onto an array of detectors, linearly arranged on the opposite side of the target object. Projections of this sort are characterised by the projection angle, which is the angle each ray makes with the vertical axis. The other relevant geometric assembly is the fan beam arrangement (see Figure 1). In this projection mode, all radiation in the projection comes from a single source, with rays being directed outwards onto a set of detectors, which may be arranged on a circumference arc (equiangular rays) or on a straight line (equally spaced rays) [9].



**Figure 1.** *Cont.*



**Figure 1.** Fanbeam and Parallel geometric assemblies for a tomographic experiment (respectively, left to right). Note that the figures are only meant to exemplify how these assemblies are widely different, and therefore do not have scales or any angular data.

#### 1.3.4. Reconstruction

The overarching division between tomographic reconstruction algorithms is on the level of their nature, which can be analytical or algebraic (iterative). Other subdivisions come from the geometry and the type of technology used for the particular application on which reconstruction is being run. In medical imaging, the most common analytical method is the Filtered BackProjection algorithm (FBP). FBP is based on Fourier's Slice Theorem, which states that the one dimensional Fourier Transform (FT) of a projection at a given angle is the two dimensional FT of the reconstructed image through that same angle [9,14]. If a sufficient number of projections is gathered, one can create a good estimate of the image by performing a 1D FT of the projections, and inverting them in 2D, before summing them in the image space. It so happens that this direct inversion process produces heavily distorted images due to the fact that calculation errors are typically larger the higher the frequency of the image component [9]. This is overcome in FBP by the application of a weighing filter before the inversion process.

This sequence of steps is sufficient for parallel projection reconstruction, but for fan beam projections, the FBP can only be applied after a series of somewhat cumbersome geometric transformations. If this is not acceptable, for some reason, there is an alternate solution: the fan beam sinogram can be reorganised, based on the observation that a ray in a fan beam is equal to a ray in a parallel projection in another given angle. Of course this resorting operation will not render a perfect sinogram for this new parallel projection (since not every fan ray has a direct parallel ray correspondent), but imperfections can be normally corrected through interpolation. After this procedure is run, one can proceed as if the geometry were parallel [9,11,15].

Iterative algorithms are based on simpler premises, but require a different mindset. The set of projections can be thought of as a matrix, called sinogram, as has been introduced in this same section. In this matrix, the lines refer to the projection number, and the columns deal with the detectors (for instance, the first line of this matrix corresponds to all detections in the first projection). The image can also be thought of as a matrix, in which each pixel has a given value, which gives it its intensity (and/or colour). Finally, there is the system matrix, which is the matrix that contains the lengths of every ray in each projection contained in each one of the image's pixels, lengths which are obtained, in this case, through Siddon's algorithm, already presented. Iterative algorithms, in general, attempt to solve Equation (8). In it,  $\mathbf{g} \in \mathbb{R}^{m,1}$  is the column vector sinogram,  $\mathbf{a} \in \mathbb{R}^{m,n}$  is the system matrix and  $\mathbf{f} \in \mathbb{R}^{n,1}$  is the column vector image.  $m$  is the number of measurements (projections times detectors)

and  $n$  is the number of pixels in the image. As their designation implies, iterative algorithms produce an estimation for  $\mathbf{f}$  which is updated in the direction of error minimisation in every iteration [9,10].

$$\mathbf{g} = \mathbf{a} \cdot \mathbf{f}. \quad (8)$$

The popularity of algebraic reconstruction methods has not remained constant throughout the years. For a long time, they have been considered too computationally intensive to use in a clinical setting (paradoxically, Hounsfield's machine used this kind of algorithm). This was in direct opposition to the fact that researchers know that these methods are better able to model reconstruction since Shepp and Vardi published the maximum likelihood tracer estimation in 1982. Nowadays, and since the mid nineties, these algorithms are the first choice whenever the reconstruction dataset is not too large to process using the available computational capabilities [15].

The general goal of iterative reconstruction algorithms is to solve Equation (8) [10]. In principle, any method that solves it can be used for image reconstruction in tomography. In reality, however, only a few are currently in use by the community. Of these, TomoSim uses two of the most prominent: Simultaneous Algebraic Reconstruction Technique (SART) and Maximum Likelihood Expectation Maximisation (MLEM).

SART was presented in 1984 by Andersen and Kak [16] and the global idea is that the estimated image is corrected for all projections at the same time (in opposition to the original algebraic Reconstruction Technique, in which corrections were applied for each single projection). Iterations in SART change the estimated image according to Equation (9), iterating on  $k$ .

$$\mathbf{g}_i^{(k+1)} = \mathbf{g}_i^{(k)} + \frac{\sum_j \left[ \mathbf{a}_{ij} \cdot \frac{p_j - \mathbf{a}_j^T \cdot \mathbf{g}^{(k)}}{\sum_{i=1}^n \mathbf{a}_{ij}} \right]}{\sum_j \mathbf{a}_{ij}}. \quad (9)$$

MLEM algorithms were first published in the medical imaging community in 1982, by Shepp and Vardi [17]. With this algorithm, image corrections are ruled by Equation (10), which also iterates over  $k$ .

$$\mathbf{f}_j^{k+1} = \frac{\mathbf{f}_j^k}{\sum_{i=1}^n \mathbf{a}_{ij}} \sum_{i=1}^n \frac{\mathbf{g}_i}{\sum_{j'=1}^m \mathbf{a}_{ij'} \mathbf{f}_{j'}^k}. \quad (10)$$

This equation is very easy to implement computationally, if one observes that the sums of the second multiplication term expand neatly onto matrix products. In the end, this equation is the equivalent of writing Equation (11), as explained in Reference [10], in which  $\text{IMG}^{(k)}$  is the estimated image in the  $k$ th iteration, NBP is the Normalised Backprojection operation, RSNG the real sinogram (as in coming from the detector hardware) and SSNG the simulated sinogram, calculated through the previous iteration.

$$\text{IMG}^{(k+1)} = \text{IMG}^{(k)} \times \text{NBP} \left( \frac{\text{RSNG}}{\text{SSNG}^{(k)}} \right). \quad (11)$$

#### 1.4. DOAS Tomography

DOAS tomography is a relatively new field of study within DOAS. It involves, as the name implies, the application of tomographic techniques to the atmospheric studies that are normally conducted through DOAS. The concentration values retrieved through the spectroscopic technique are essentially line integrals in themselves. Therefore, they can be almost immediately considered projections. If one gathers enough of these integrals from a sufficient number of angles, any tomographic algorithm is able to reconstruct an image, which corresponds to a map of concentrations of the target trace gas in the study.

One of the first suggestions of a technique which could be adapted to the DOAS procedures was made in 1979 [18]. However, the first study that applied tomography to DOAS in a significant manner was the BAB-II campaign [19,20]. This was a research

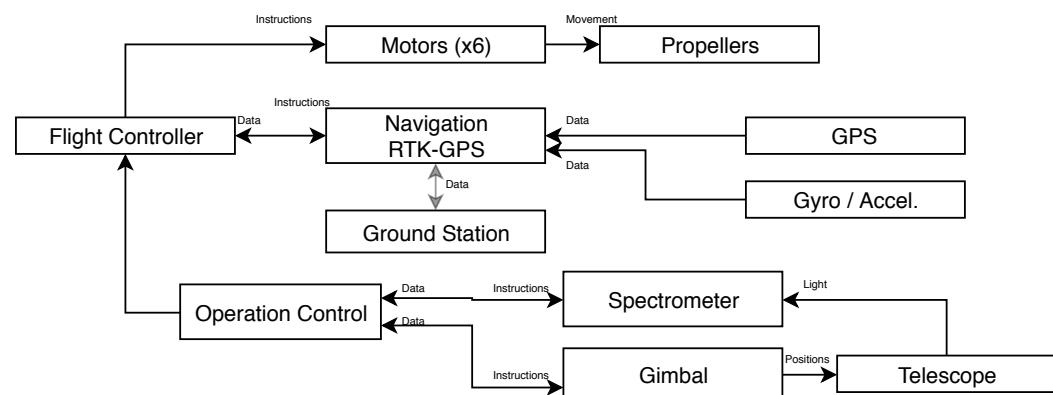
initiative that involved people from the Heidelberg DOAS group and intended to study the temporal evolution of the NO<sub>2</sub> concentration in a bi-dimensional way, along the motorway that connects Heidelberg to Mannheim, in Germany. This campaign led to several publications and is to this day the main “contributor” to tomographic studies with DOAS. More recently, in 2016, Stutz and his team have built and used a similar set-up to study the atmospheric profiles of aromatic hydrocarbons near an refinery plant, in Texas. Their system was composed of a dual-light emitting diode light source, a telescope which acted as emitter and receiver of light and retro-reflector arrays, positioned strategically in the geographic region that was being studied. Although this study was not as extensive as the previously mentioned Heidelberg study, it is also very important, as it proves the practical applicability of the technique to real world problems [21]. Finally, it is worth to mention the paper by Erna Frins, who in 2006 used sun-illuminated targets to perform a tomographic analysis of the region in which her system was positioned, which coincidentally is also Heidelberg. This study is important because it is one of the few that uses scattered sunlight with this technique. Moreover, it also features a very good description of the physics and mathematical approximations that are inherent to the experiments at hand [22].

These studies and more are addressed in another paper, which should be submitted shortly, and in which the authors have conducted a deeper and more systematic literature review on the subject.

## 2. Materials and Methods

### 2.1. Device Description

TomoSim is a simulation platform for a drone-mounted atmospheric monitoring system based on DOAS. Although the physical device has not yet been assembled, the team has already compiled a final (or very close) design, which is schematically represented in Figure 2. The reasoning behind the custom design was to increase the maximum payload and allow longer flight times. The team chose to use a DJI S900 frame (hexacopter), manufactured by DJI in China, with custom-made 368 mm carbon fibre arms, longer and lighter than the original. The increased empty space allows the replacement of the default propellers by 17" carbon fibre units, coupled to 6 E1200 motors. This propeller-motor configuration is not only significantly more powerful than the default assembly, but also more efficient. According to the manufacturer [23], this configuration is able to lift and work with payloads exceeding 8 kg, which is much more than we need for data acquisition platform, comprised of the gimbal, a Celera SSIN-06 [24] unit with a maximum pointing error of 2 arcseconds; the telescope, an Omegon MightyMak [25]; and an Avantes Mini spectrometer with 2048 spectral channels [26].



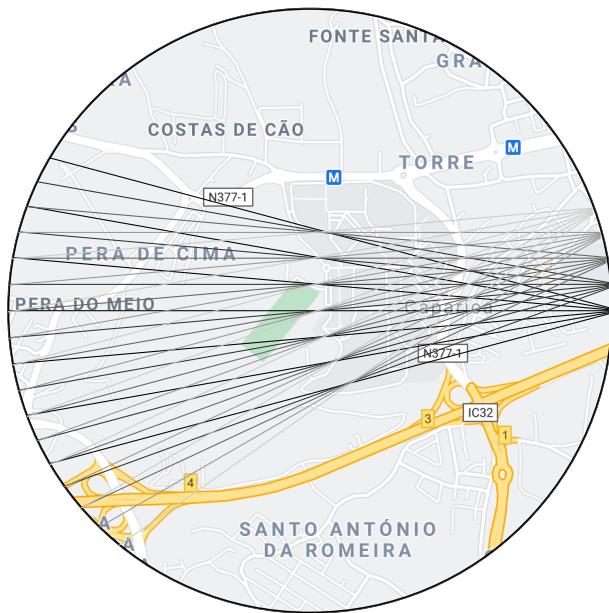
**Figure 2.** Drone system schematic representation, with component relations.

A Pixhawk flight controller is used to handle and manage aerial dynamics, and also to gather every sensor output in the device. The controller comes with integrated gyroscopes, magnetometers and barometers. The only external sensor that needs to be adapted and purchased separately is the navigation (GPS) unit. The Pixhawk supports Real Time Kinematic GPS (RTK-GPS), a combination of inertial sensors and satellite navigational data that can grant the UAV a positioning precision under 20 cm [27,28]. The flight controller is in permanent communication with the operation controller, which is a Raspberry Pi 0 (or similar) single board computer. This computer stores the flight program and directs the flight controller to each necessary position, and also controls data acquisition through a USB connection to the spectrometer. The device's trajectory will be planned using Arducopter's Python libraries and their Software In The Loop (SITL) simulation platform.

## 2.2. Data Acquisition

The simulated drone is to describe an unobstructed horizontal circular trajectory with a maximum diameter of 1 km at the intended measurement height, the interior of which is the Region Of Interest (ROI). There are two moments to the data gathering process. Figure 3 attempts to illustrate both.

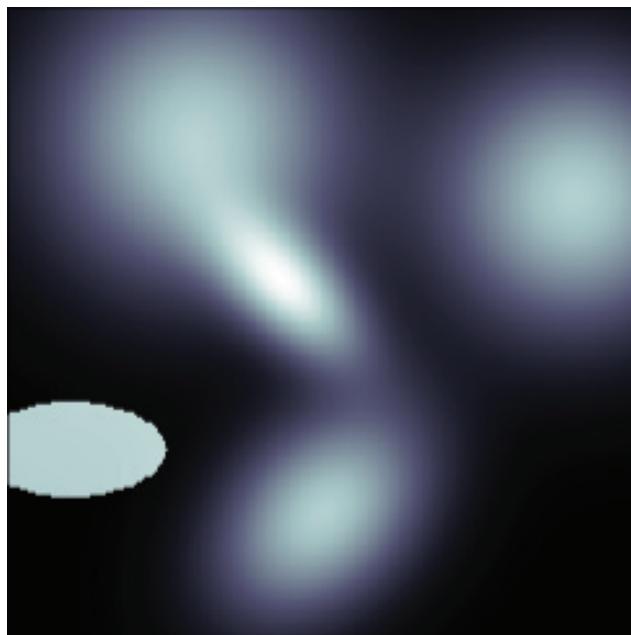
- **First moment** While flying in this circle, the device stops in a series of positions at a given fixed angular interval ( $\Delta$ ) from each other. The number of stops is defined by  $\Delta$  and by fan beam information requirements (see Reference [9]) At each one of these stops, the gimbal turns towards the trajectory's interior and points in a series of angular directions that describe an arc. For procedural simplicity, the angle between these directions is also  $\Delta$ . At each one of these angles, the device's operational controller instructs the spectrometer to acquire a given number of spectra, which depends on configuration and conditions. Besides spectral data, the system algebraically calculates and stores the point in which the light will exit the ROI (see Appendix A).
- **Second moment** The device positions itself in each of the points in which light has exited the ROI in the first moment and the gimbal is pointed towards the entrance point, effectively aiming in the opposite direction to which a spectral measurement took place in the 1st moment. Light that comes from the sun is scattered somewhere in the atmosphere and enters the ROI (at a given angle) in point A. It then traverses the distance AB and leaves the ROI in point B. At these distances and with this kind of geometry, light scattering can be considered negligible [22,29,30] and therefore light extinction will primarily be due to absorption by components between A and B [5]. It should then be possible to apply Lambert-Beer's law to extract trace gases concentrations in the ROI, by considering light at point A as the source intensity ( $I_0$  in Equation (1)) and light at B the final intensity ( $I$  in Equation (1)). When the 2nd moment is complete, the system has a set of fan beam distributed spectra, which can be equated to projections in a tomography problem.



**Figure 3.** Simplified schematic representation of the first five fans acquired by the drone over the circular trajectory. The drone stops at the vertex of each fan, represented on the right. These points correspond to point A in the explanation of the two measurement moments in Section 2.2. Each line in the drawing represents a ray within a fan, i.e., a direction, in which a projection is taken. The “exit points” of each ray from the ROI correspond to point B, and are represented on the left. In the second measurement moment, the drone moves to each of these points and takes another spectrum in the same direction as the ray itself. Here, both fans and rays are separated by an angular interval of 5 degrees, and there are only 3 rays within each fan, for graphical simplicity. Both values are customisable at runtime. The map, included for example purposes, was retrieved from Google Maps in 2019 ©Google.

### 2.3. Phantoms

In medicine, a phantom is a model that emulates certain properties of human or animal tissue. Researchers use these models to evaluate therapeutic or diagnostic methods. In the imaging field, phantoms are known matrices with a given size that were designed to mimic the types of bodies that are to be reconstructed with the technique or algorithm being tested. Most phantoms described in the literature were constructed specifically for medical imaging, since this is clearly the most prominent application field for tomographic methods. Computed Tomography phantoms, for instance, intend to mimic the X-ray absorption of the human body, or of part of the human body. Since the distribution of gases in the atmosphere is entirely different from biological tissue's, these phantoms are not adequate for TomoSim. This implied the design of a new phantom, based on the idea that a two-dimensional (in this case) Gaussian peak is more appropriate to describe the smoother nature of the distribution of a gas than a series of sharply defined ellipses [31]. The new phantom, designed with TomoPhantom [32], is comprised of 5 bivariate Gaussian profiles, depicting a static gas mixture, and an ellipse near one corner of the image, which serves as a reference point. This new spectral phantom can be seen in Figure 4, and a descriptive summary is provided in Table 1.



**Figure 4.** A graphical representation of the new spectral phantom, custom built for the TomoSim application.

**Table 1.** Table summarising the new phantom's construction details, as a sum of 5 Gaussian profiles and an ellipse designed using TomoPhantom. In this table, C0 is the object's amplitude, X0 and Y0 are its center coordinates, and a and b are the objects half-widths. The table is constructed using TomoPhantom's particular syntax and more information can be obtained in Reference [32].

| Type     | C0 | X0   | Y0   | a    | b    | Angle |
|----------|----|------|------|------|------|-------|
| Gaussian | 1  | -0.1 | -0.1 | 0.25 | 0.5  | -45   |
| Gaussian | 1  | 0.6  | 0    | 0.65 | 0.45 | -45   |
| Gaussian | 1  | -0.6 | -0.4 | 0.8  | 0.8  | 0     |
| Gaussian | 1  | -0.4 | 0.8  | 0.7  | 0.7  | 0     |
| Ellipse  | 1  | 0.4  | -0.8 | 0.3  | 0.15 | 0     |

During simulation, a phantom is totally contained within the ROI. A gaussian filter (kernel size 5, auto standard deviation) is applied to the phantom image to simulate noise. The phantom shares the same grid as the discretised ROI and each pixel has a value comprehended between 0 and 255. This value is linearly correlated with the number of molecules of the target trace gases in the ROI. Software configuration allows the definition of the maximum number of molecules per pixel. Default value for NO<sub>2</sub>, the test case presented in Section 3 is  $1 \times 10^{15}$  molecules.

#### 2.4. Reconstruction

Any tomographic reconstruction requires the previous and detailed knowledge of the ray geometry of the problem. This implies that the space being reconstructed is discretised, so that it can be addressed through computational routines. In this case, the discretisation consists in overlaying a  $100 \times 100$  pixel grid (10 m square pixels, considering a 1 km diameter circular drone trajectory). By applying Siddon's algorithm to this geometry, the lengths that each ray traverse in each pixel of the grid are retrieved, assembling the system matrix. The system matrix is a complete description of the problem's geometric properties, and is therefore characteristic of each experiment, depending on the angular intervals between projections ( $\Delta$  in this case) and on the size and number of the pixels in the discretisation grid.

TomoSim then performs a resorting operation on the sinogram, in order to transform the fan beam projections into parallel projections, greatly simplifying image reconstruction.

Since the angular interval between the fans and the rays within the fans is the same ( $\Delta$ ), resorting is greatly simplified [9]. After discretisation and the necessary resorting steps are taken, the software reconstructs the images with three different algorithms—MLEM, FBP and SART.

Finally, after the images are calculated, the simulator must convert the pixel values back into molecule numbers. For this, the software runs Siddon algorithm on the reconstructed images for a selected number of angles. Resulting projection values are then compared with the projection values of the original images in order to find a converting parameter that allows the presentation of accurate concentration values.

### 2.5. Error Estimation

There are three major error sources in TomoSim—geometric errors, spectroscopic errors and reconstruction errors. Geometric errors come from the fact that the device's positioning has an associated error: the drone is not where it thinks it is, nor does it point to where it thinks it points. TomoSim addresses this kind of error in a Monte Carlo like fashion. Positioning and pointing errors are assumed to be normal. Each time a point is calculated, the software generates a normally distributed 0 mean random number, with a standard deviation equal to the rated error of the positioning system and sums it to the intended point (error calculations illustrated in Appendix A). Given the ratios between the linear distances involved in the trajectory and the positioning errors, geometric errors have a very small contribution towards the end results.

On the spectroscopic level, errors come from the instruments used for capturing the data. TomoSim takes this noise into account by adding Gaussian noise spectra to each measurement, for which the magnitude is configurable via its standard deviation, a method previously followed in Reference [33]. This approach is only valid insofar as the captured spectra are perfectly calibrated regarding spectral shift and squeeze, which is an acceptable assumption for a simulation.

Finally, the software has to deal with the reconstruction errors. In image reconstruction from projections, it is common to use techniques such as Mean Squared Error (MSE) as a metric with which to assess the algorithm's performance. This simulator was also evaluated in this light, in two different ways. First, the MSE for the whole reconstructed image was calculated. This enables the possibility to look at the reconstruction as a whole and visually tell where it is lacking and where it is better performing. Secondly, a score was calculated according to Equation (12). In this equation, and with reference to this simulator,  $f$  is the original image and  $g$  is the reconstructed image.

$$E = \sqrt{\frac{\sum |g(x, y) - f(x, y)|^2}{\sum |f(x, y)|^2}}. \quad (12)$$

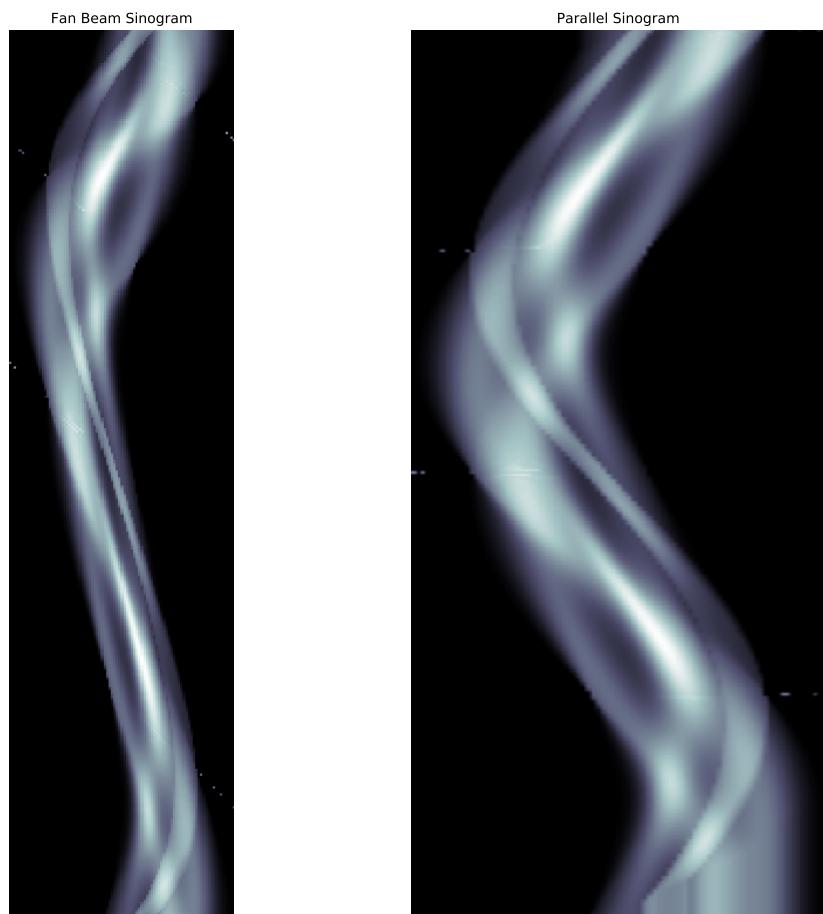
Finally, there is an additional source of error that was not explored in this simulation. It is the temporal error that comes from the difference in time between measurement moments (see Section 2.2), which can introduce a larger error than those considered above. In the real world, this can be easily mitigated by the introduction of a second vehicle, which would only conduct 2nd moment measurements. In the simulation, it was considered that there were no changes in the field of measurement with time.

## 3. Results

This section presents, analyses and discusses results obtained by the application of the techniques and methods described in the previous two sections, that is, the tomographic reconstruction of the phantoms which were also presented in Section 2.

### 3.1. Projection Calculations

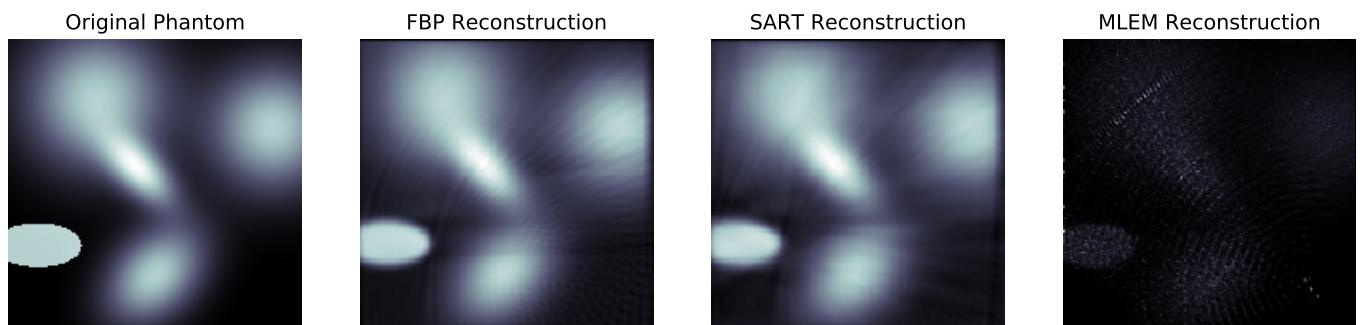
In TomoSim, a projection is the sum of the pixel lengths (the lengths of the rays that traverse each pixel) for each ray and for the grid mentioned in Section 2. Unlike a real life situation, the contents of the ROI are completely known and correspond to the phantoms also described in Section 2 multiplied by a given maximum number of molecules. Siddon's algorithm is used in this process, and the final results of its application are the sinogram and the system matrix. Figure 5 contains some examples of these matrices, before and after the resorting operation described in Section 2.4.



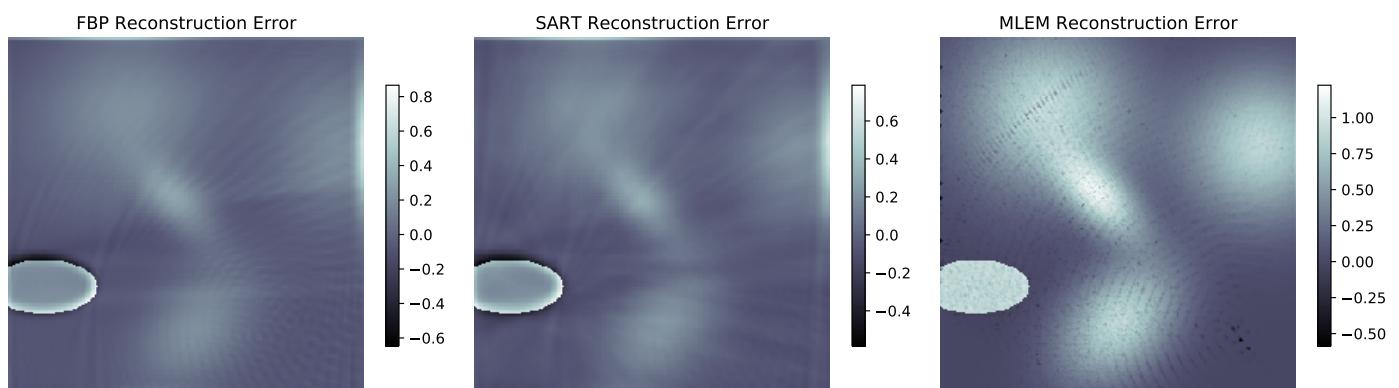
**Figure 5.** Sinogram examples: the new spectral phantom projection data at a projection interval of 1 degree. On the **left**, the projection data before resorting; on the **right**, the parallel projection data obtained after resorting the fan-beam line integrals.

### 3.2. Reconstruction Results

Images corresponding to the trace gas distribution within the ROI were reconstructed using iterative and analytical methods. In Figure 6, one can see the reconstruction results for the three tested methods when applied to the new spectral phantom; Figure 7 shows the graphical representation of the reconstruction errors for the spectral phantom and is accompanied by Table 2; and in Figure 8, a comparison between reconstructions with different  $\Delta$  values is presented, also for the new spectral phantom.



**Figure 6.** Tomographic reconstruction results, projection interval of 1 degree. From **left** to **right**: original phantom, Filtered BackProjection algorithm (FBP) reconstruction, Simultaneous Algebraic Reconstruction Technique (SART) reconstruction, Maximum Likelihood Expectation Maximisation (MLEM) reconstruction



**Figure 7.** Tomographic reconstruction errors. Each one of these images was constructed by subtracting the respective reconstruction matrix, displayed in Figure 6, from the original phantom matrix. Error value is normalised to the pixel values, i.e., 32 bit floating point numbers with values between 0 and 1.



**Figure 8.** Reconstruction degradation: the projection interval was crucial for reconstruction. Note the image degradation going from a projection interval of 1 degree to 5 degrees (**left** to **right**). Images reconstructed using the FBP algorithm.

**Table 2.** Reconstruction error table for the new spectral phantom at several projection intervals. The MLEM routine used pure fan-beam data while the other two used resorted parallel information. Errors presented were calculated using the Root Mean Square Errors, normalised to the range of the reconstructed image.

| Algorithm   | Projection Intervals |        |        |        |        |
|-------------|----------------------|--------|--------|--------|--------|
|             | 1                    | 2      | 3      | 4      | 5      |
| <b>FBP</b>  | 0.2365               | 0.2408 | 0.2609 | 0.2948 | 0.3465 |
| <b>SART</b> | 0.2225               | 0.2278 | 0.2771 | 0.3537 | 0.3302 |
| <b>MLEM</b> | 0.8705               | 0.9723 | 0.9986 | 0.9744 | 0.9890 |

#### 4. Discussion

The results presented in the previous subsection raise a series of pertinent observations that should be addressed in discussion. The first remark goes to the fact that the number of projections used was adequate to perform the tomographic reconstruction in adequate fashion, as expected. There are already several examples of studies in which a considerably lower number of projections was used, still producing satisfactory results [20]. Moreover, and since the simulation software automatically includes errors in the calculations, this figure also proves that geometric error plays a very limited role in changing the result of the reconstruction. Given the difference in size between the drone's trajectory and the geometric error, this was also a predicted result.

Even with the relatively low number of projections produced by the drone's trajectory and measurement strategy, all reconstruction algorithms were able to produce a reconstructed image that resembled the phantom that generated it.  $\Delta$ , the angular interval between projections, revealed itself to be crucial. This was expected as the number of captured projections is obtained by dividing 360 by  $\Delta$ . This is, also as expected, confirmed by the upward trend of the error when increasing the value of the angular projection interval, as can be seen numerically in Table 2 and qualitatively in Figure 8. With respect to the algorithms used, the custom-made MLEM routine produces clearly outlier results, which are not on par with the other two reconstruction methods used. This is plain to see both in Figure 6 and in Table 2, in which this algorithm's NRMSE is almost four times the second best result (FBP) for the smallest projection interval. This difference could to some extent be expected. SART and FBP algorithms were implemented using some of the most relevant and consistently used Python libraries (SciPy, for instance [2]). Given the amount of attention these libraries get from the scientific programming community, levels of optimisation are extremely high. Although it is nowhere near the other two approaches, the MLEM routine is still useful, as it is the only truly geometry-independent algorithm in this study (SART is also geometry independent, but this particular implementation expects a parallel projection sinogram as an input).

As stated in Section 2.5, three different kinds of error influence the reconstruction results: geometric, spectroscopic and reconstruction errors. The first kind of error is directly included in projection calculations, through the application of a Monte Carlo-like method to the geometry described in Appendix A. The second kind of uncertainty comes from the spectrum acquisition process itself, which is not perfect. If one considers there are no systematic errors present in the results, which is an acceptable premise in a simulation, then these errors can be simulated by the inclusion of Gaussian noise in the spectral measurements. This approach is based on the one used in Reference [33], in which a Gaussian noise spectrum is added to the spectrum of interest in order to simulate how the error behaves with a degraded signal. Finally, reconstruction errors come from the finite precision of the calculations that render the images. These errors were presented in Section 3.

The three methods were also evaluated as to how they perform computationally, by measuring the time it took to produce the images in Figure 6 using a Paperspace P4000 cloud computing instance. In this regard, the fastest method was FBP, which took around 3 s to reconstruct. The second was MLEM, with around 50 s for 1000 iterations, and finally came SART, with 1 min and 50 s for 1 iteration. One relevant observation comes from the fact that MLEM was significantly faster than SART, even taking into account the difference in optimisation, which was not an expected result and may indicate some reconstruction enhancing technique on SART's side, as the literature seems to indicate that this technique is faster than MLEM [15].

All things considered, the FBP algorithm produces a very good reconstruction, equivalent to SART's, while being more than 10 times faster, indicating that for this kind of application and with this kind of projection information, it is the best reconstruction algorithm.

## 5. Conclusions

The initial goal of the TomoSim software project was to develop a simulation platform to create the tomographic reconstruction of the column density distribution for a number of target atmospheric trace gases.

The software program was written using the Python language and some numeric calculation libraries, such as NumPy and SciPy. Using these two libraries had two main effects: on the one hand, it enabled the programmers to easily create and manipulate matrices and vectors (images, for instance), and on the other, they greatly improved the running speed of the code, since their core is written in lower level languages (namely C).

The simulations that the software performs prove that, if the final device is programmed to comply to trajectory and acquisition requirements, reconstruction is perfectly achievable, even with relatively low projection numbers (comparing with medical imaging procedures). This brings another significant conclusion which is that the devised acquisition definitions, which produce a set of fan beam arrays, provide sufficient projection information to run the reconstruction and achieve plausible results.

TomoSim runs three algorithms on the projection data in order to produce the spectral mapping of the target pollutants—FBP (analytical), SART and MLEM (both algebraic). SART offered the best results, at the expense of time. The analytical algorithm produced very nearly the same results, but took a fraction of the time when comparing with either SART or MLEM. The MLEM algorithm cannot be directly compared to the SART algorithm, due to differences in the optimisation levels of both routines, but had nonetheless a reasonable time performance altogether, although producing the poorest reconstruction results.

Regarding future developments, there are three main avenues that should be explored:

- **Other phantoms:** Presently, TomoSim only includes tomographic reconstruction for two different phantoms. While this is sufficient for simulation, it would be desirable to have some more phantoms, which could mimic other concentration distributions of interest.
- **Paradigm shift:** This simulation software was developed under the passive DOAS analysis model. Active measurements are much more versatile and accurate, and it would be interesting to develop this same technique using an artificial light source. Of course this would require many adaptations, namely regarding equipment and trajectory (probably even algorithms and interpolations).
- **Three-dimensional reconstruction:** TomoSim was developed to produce the reconstruction of a two dimensional image corresponding to the spatial distribution of an array of target trace gases. It would be much more interesting to have a three dimensional equivalent. As far as simulation goes, this is one of the most immediate developments for this project. On a more tangible level, the additional dimension would make the problem much more complex, mainly because of trajectory and battery logistics.

**Author Contributions:** Conceptualization, R.V.d.A., N.M. and P.V.; Investigation, R.V.d.A.; Methodology, R.V.d.A. and P.V.; Project administration, P.V.; Software, R.V.d.A.; Supervision, P.V.; Validation, N.M.; Writing—original draft, R.V.d.A.; Writing—review & editing, N.M. and P.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by Fundação para a Ciência e Tecnologia, grant number PDE/BDE/114549/2016 and UIDB/00645/2020.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Geometric Calculations

### Appendix A.1. Light ROI Exit Point ( $P_2$ ) Determination

Figure A1 is a schematic snapshot of a point in which the drone is taking a spectrum in one of its stops. Here, the drone's position,  $P_1$ , is given by the distance  $D$  and the angle  $\beta$ , while the gimbal is pointing at a direction at an angular distance of  $\gamma$  from line  $0P_1$ .

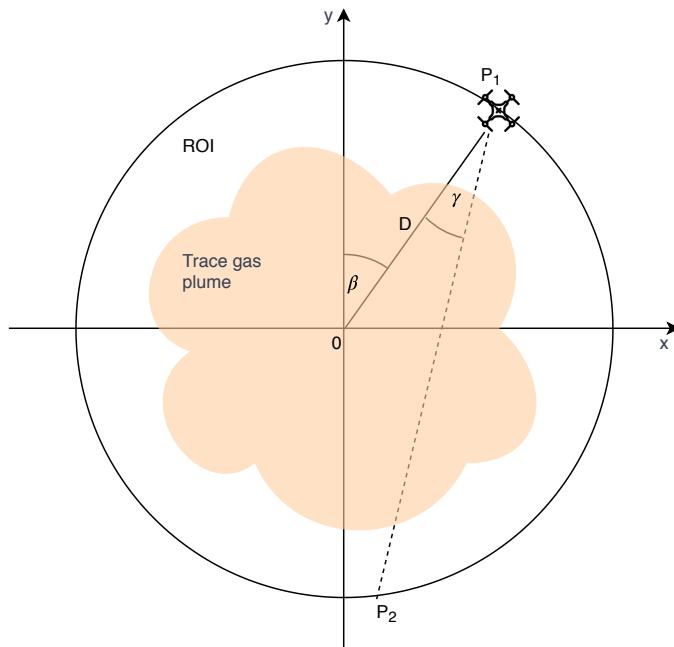
Point  $P_2$ , which is not known, is at the intersection between the trajectory's circumference and line  $P_1P_2$ . Parametrically, any point in this line can be expressed as  $P_1 + t \cdot (P_2 - P_1)$ , with  $t$  being a scalar. Moreover, to say a point  $X$  is on this circumference is the same as writing  $X - Q = D$ . Here,  $Q$  is the centre of the trajectory and can therefore be eliminated from the equation. If one is to expand these equations, the situation can be accurately described by Equation (A1).

$$\begin{aligned} P_2 &= P_1 + t \cdot (P_2 - P_1) \\ |P_2| &= D^2 \end{aligned} \quad (\text{A1})$$

Unravelling the expressions in Equation (A1), and making use of the algebraic property that says  $|A|^2 = A \cdot A$ , the expression becomes a two degree equation, as stated in Equation (A2), if one writes  $P_2 - P_1$  as  $V$ .

$$t^2 V^2 + 2 \cdot V \cdot P_1 \cdot t + P_1^2 - D^2 = 0 \quad (\text{A2})$$

If line  $P_1P_2$  non-tangentially intersects the circumference, solving Equation (A2) renders two values for  $t$  (which correspond to  $P_1$  and  $P_2$ ). Selection is made by determining the returned value of  $t$  that maximises the euclidean distance between the produced point and  $P_1$ .

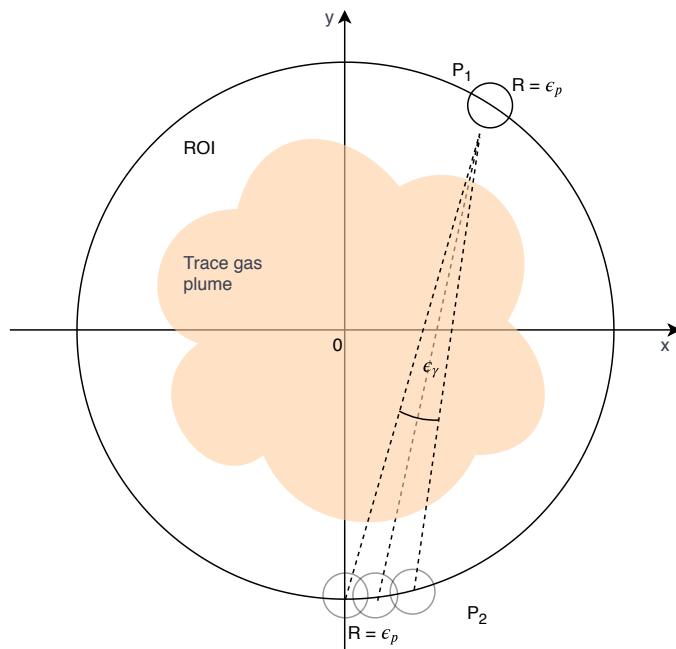


**Figure A1.**  $P_2$  Calculation.

#### Appendix A.2. Geometric Error Determination

Figure A2 is a graphical representation for the reasoning behind the geometric error estimation. There are two types of error in this image: the ones that come from the RTK-GPS system (positioning error, denoted  $\epsilon_p$ ) and the ones that come from the gimbal (pointing error, denoted  $\epsilon_\gamma$ ). TomoSim considers these errors to be normally distributed, and the two  $\epsilon$  values correspond to their standard deviation. To introduce the error into the simulation, the software calculates the theoretical  $P_1$  from the  $\beta$  and  $D$  values (see Figure A1) and then adds a normally distributed random number that respects  $\epsilon_p$ , retrieving the true  $P_1$ . This new point is used to draw the theoretical line  $P_1P_2$  and the pointing error is added using the same process as in  $P_1$ . Given the very low gimbal error, the small angles approximation ( $\sin \theta = \theta$ ) is used to determine the theoretical value of  $P_2$ , on the drone's circular trajectory. Finally, the software adds again the positioning error, in the same

manner as it had on  $P_1$ . As a finishing remark, it is important to note that in Figure A2, all errors are extremely exaggerated as they would not be visible otherwise, due to the huge size difference between them and the trajectory.



**Figure A2.** Error estimation graphical representation. Note errors are extremely exaggerated for visualisation purposes.

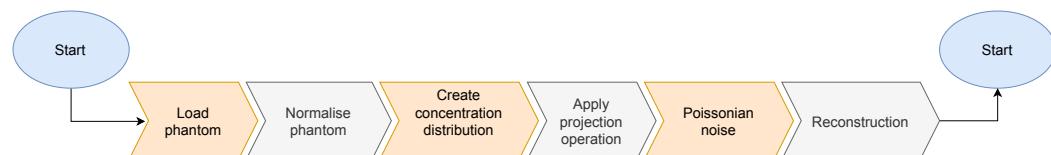
## Appendix B. Simulation Data Characterization

As stated in Section 2, the simulation uses a phantom matrix to perform its reconstructions. However, this is not purely the input to the simulator routine. Some transforms, introducing random variation, are run beforehand.

As described in Figure A3, the simulator starts by loading the phantom data. This is a 300 by 300 pixels image, which is created by running the TomoPhantom [32] software. This matrix, which is parametrically stored, is always the same. The data it contains are 64 bit floating point numbers, ranging from 0 to 1 (0 is black; 1 is white). This image is then normalised to range between 0 and 255, and its data type is changed to unsigned 8 bit integers.

The extremes of concentration values are randomly taken from a uniform distribution between  $1 \times 10^{15}$  and  $1 \times 10^{17}$ , rendering `min_val` and `max_val`. These two values will be used as concentration limits for the phantom matrix, with `min_val` representing the baseline (0 in the phantom matrix) and `max_val` representing the maximum value in the ROI (255 in the phantom matrix).

The projection operator which geometrically describes the projection system is applied to the phantom matrix, rendering the sinogram. Before reconstruction, the program adds poissonian noise to this matrix.



**Figure A3.** The data flowchart of the simulation routine. The phantom is entered to the routine as a fixed matrix, but is transformed so that each time the program runs, the reconstruction is randomly different and more in line with what happens in nature.

## References

- Perez, F.; Granger, B.E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21–29. [[CrossRef](#)]
- Oliphant, T.E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20. [[CrossRef](#)]
- APS. August 15, 1758: Death of Pierre Bouguer. *Am. Phys. Soc. News* **2011**, *20*, 2.
- Struve, W.S. *Fundamentals of Molecular Spectroscopy*; Wiley-Interscience: Hoboken, NJ, USA, 1989.
- Platt, U.; Stutz, J. *Differential Optical Absorption Spectroscopy—Principles and Applications*; Springer: Berlin/Heidelberg, Germany, 2007.
- Merlaud, A. Development and Use of Compact Instruments for Tropospheric Investigations Based on Optical Spectroscopy from Mobile Platforms. Ph.D. Thesis, Faculté des Sciences—Université Catholique de Louvain, Louvain, Belgium, 2013.
- Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes: The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2007.
- Gunderman, R. *Essential Radiology: Clinical Presentation, Pathophysiology, Imaging*, 2nd ed.; Thieme: New York, NY, USA, 2006.
- Kak, A.; Slaney, M. *Principles of Computerized Tomographic Imaging*; Society of Industrial and Applied Mathematics: Philadelphia, PA, USA, 2001.
- Bruyant, P.P. Analytic and iterative reconstruction algorithms in SPECT. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* **2002**, *43*, 1343–1358.
- Herman, G.T. Image Reconstruction From Projections. *Real-Time Imaging* **1995**, *1*, 3–18. [[CrossRef](#)]
- Siddon, R.L. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med. Phys.* **1985**, *12*, 252–255. [[CrossRef](#)] [[PubMed](#)]
- Gao, H. Fast parallel algorithms for the x-ray transform and its adjoint. *Med. Phys.* **2012**, *39*, 7110–7120. [[CrossRef](#)]
- Herman, G.T. *Fundamentals of Computerized Tomography: Advances in Pattern Recognition*; Springer: London, UK, 2009. [[CrossRef](#)]
- Defrise, M.; Kinahan, P.E.; Michel, C.J. *Positron Emission Tomography*; Springer: London, UK, 2005; pp. 63–91.
- Andersen, A. Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrason. Imaging* **1984**, *6*, 81–94. [[CrossRef](#)]
- Shepp, L.A.; Vardi, Y. Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Trans. Med Imaging* **1982**, *1*, 113–122. [[CrossRef](#)] [[PubMed](#)]
- Byer, R.L.; Shepp, L.A. Two-dimensional remote air-pollution monitoring via tomography. *Opt. Lett.* **1979**, *4*, 75. [[CrossRef](#)] [[PubMed](#)]
- Laepple, T.; Knab, V.; Mettendorf, K.U.; Pundt, I. Longpath DOAS tomography on a motorway exhaust gas plume: Numerical studies and application to data from the BAB II campaign. *Atmos. Chem. Phys. Discuss.* **2004**, *4*, 2435–2484. [[CrossRef](#)]
- Pundt, I.; Mettendorf, K.U.; Laepple, T.; Knab, V.; Xie, P.; Lösch, J.; Friedeburg, C.V.; Platt, U.; Wagner, T. Measurements of trace gas distributions using Long-path DOAS-Tomography during the motorway campaign BAB II: Experimental setup and results for NO<sub>2</sub>. *Atmos. Environ.* **2005**, *39*, 967–975. [[CrossRef](#)]
- Stutz, J.; Hurlock, S.C.; Colosimo, S.F.; Tsai, C.; Cheung, R.; Festa, J.; Pikelnaya, O.; Alvarez, S.; Flynn, J.H.; Erickson, M.H.; et al. A novel dual-LED based long-path DOAS instrument for the measurement of aromatic hydrocarbons. *Atmos. Environ.* **2016**, *147*, 121–132. [[CrossRef](#)]
- Frins, E.; Bobrowski, N.; Platt, U.; Wagner, T. Tomographic multiaxis-differential optical absorption spectroscopy observations of Sun-illuminated targets: A technique providing well-defined absorption paths in the boundary layer. *Appl. Opt.* **2006**, *45*, 6227–6240. [[CrossRef](#)] [[PubMed](#)]
- DJI. DJI E1200 Pro Tuned Propulsion System. 2015. Available online: <https://www.dji.com/pt/e1200> (accessed on 28 December 2020).
- Celera. UAV Gimbal. Available online: <https://www.celeramotion.com/applications/satcom-uav/uav-gimbal/> (accessed on 28 December 2020).
- Omegon. MightyMak 60 Specifications. Available online: [https://www.omegon.eu/telescopes/omegon-maksutov-telescope-mightymak-60/p,46442#tab\\_bar\\_1\\_select](https://www.omegon.eu/telescopes/omegon-maksutov-telescope-mightymak-60/p,46442#tab_bar_1_select) (accessed on 28 December 2020).
- Avantes. AvaSpec-Mini: Small and Powerful OEM Spectrometer. Available online: <https://www.avantes.com/products/spectrometers/compactline/avantes-spectrometer-mini-2048cl/> (accessed on 28 December 2020).
- Triantafyllou, D.; Aspragathos, N.A. *Intelligent Robotics and Applications—Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 7101, pp. 509–519. [[CrossRef](#)]
- Zimmermann, F.; Eling, C.; Klingbeil, L.; Kuhlmann, H. Precise Positioning of Uavs—Dealing with Challenging Rtk-Gps Measurement Conditions during Automated Uav Flights. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *4*, 95–102. [[CrossRef](#)]
- Casaballe, N.; Osorio, M.; Di Martino, M.; Frins, E. Comparison Between Regularized Optimization Algorithms for Tomographic Reconstruction of Plume Cross Sections in the Atmosphere. *Earth Space Sci.* **2017**, *4*, 723–736. [[CrossRef](#)]
- Kern, C.; Deutschmann, T.; Vogel, L.; Wöhrbach, M.; Wagner, T.; Platt, U. Radiative transfer corrections for accurate spectroscopic measurements of volcanic gas emissions. *Bull. Volcanol.* **2010**, *72*, 233–247. [[CrossRef](#)]
- Stachniss, C.; Plagemann, C.; Lilienthal, A.J. Learning gas distribution models using sparse Gaussian process mixtures. *Auton. Robot.* **2009**, *26*, 187–202. [[CrossRef](#)]

32. Kazantsev, D.; Pickalov, V.; Nagella, S.; Pasca, E.; Withers, P.J. TomoPhantom, a software package to generate 2D–4D analytical phantoms for CT image reconstruction algorithm benchmarks. *SoftwareX* **2018**, *7*, 150–155. [[CrossRef](#)]
33. Stutz, J.; Platt, U. Numerical analysis and estimation of the statistical error of differential optical absorption spectroscopy measurements with least-squares methods. *Appl. Opt.* **1996**, *35*, 6041. [[CrossRef](#)] [[PubMed](#)]

C

## SYSTEMATIC REVIEW OF DOAS TOMOGRAPHY

The following is a review article written as a requirement for completion of a course on Advanced Software Development, as part of the curricular segment of my PhD. The article was submitted to Elsevier Remote Sensing Applications: Society and Environment, having been rejected on the grounds of featuring an insufficient number of articles and excluded important papers, namely some written in Japanese and Russian, according to the reviewers. I have discussed with the other authors the possibility of resubmission to a different journal, but have not yet been able to do this.

# Tomographic DOAS - a Systematic Mapping Study on its technological status

Rui Valente de Almeida<sup>a,c,\*</sup>, Ana Moreira<sup>b</sup>, João Araújo<sup>b</sup>, Vasco Amaral<sup>b</sup>,  
Pedro Vieira<sup>a</sup>

<sup>a</sup>*Physics Department, FCT NOVA, Campus da Caparica, 2829-416 Caparica*

<sup>b</sup>*Computer Science Department, FCT NOVA, Campus da Caparica, 2829-416 Caparica*

<sup>c</sup>*Compta, S.A.; Alameda Fernão Lopes, 12, 10th floor, 1495-190 Algés*

---

## Abstract

In atmospheric chemistry and physics, DOAS is one of the most commonly used analytical techniques. The method allows the application of tomographic reconstruction procedures for 2D or 3D mapping of target trace gases' concentrations.

Tomographic DOAS applications have been documented for at least 20 years. The number of articles detailing its implementation, however, is low, and literature is sparse. This paper aims to catalogue publications on this subject, painting a quick picture of the field's technological landscape.

Our search has found that there is a great prevalence of active DOAS systems in Tomographic DOAS research activities, since we were unable to find a passive application. We can also infer that there is no current commercial application using this technology and while it is true that there are similarities between different groups' apparatus, there is still no uniform application.

Our initial search has rendered more than 700 articles. However, the application of our inclusion and exclusion criteria resulted in a final set of 8 papers detailing the tomographic equipment, reconstruction algorithm and software. From them, we were able to identify some common practices and possible research gaps.

*Keywords:* DOAS, Tomography, Systematic Mapping Study

---

\*Corresponding Author

Email address: rf.almeida@campus.fct.unl.pt (Rui Valente de Almeida)

## 1. Introduction

Differential Optical Absorption Spectroscopy (DOAS) is one of the most prominent methods for analysing and quantifying atmospheric chemistry, namely in what concerns trace gas concentrations. The technique, developed during the 70s by Perner and Platt [20], was popularised in the following decades by its use in detecting Ozone, Nitrogen Oxydes and studies of cloud radiative transport. DOAS is a type of absorption spectroscopy, which uses a clever mathematical and physical observation to overcome the difficulties of spectral measurement in the open atmosphere.

Through the setting of very careful geometric considerations, it is possible to combine DOAS with tomographic reconstruction methods in order to assemble a map of the gaseous concentrations in a given geographic region. Tomography is the process of reconstructing an image through projections obtained by subjecting a given target (in our case, the atmosphere) to being traversed by any kind of penetrating or reflecting wave, which in our case is visible light.

With this study, we have intended to capture the current literary landscape surrounding the usage of tomographic DOAS, assessing this technique's technological status. For this purpose, we have employed a review methodology originary from Evidence Based Medicine. This method, which has migrated to engineering through Software Engineering, is called a Systematic Mapping Study (MS). It provides a framework that allows researchers to produce detailed and systematic search protocols, which are used to catalogue literature information and identify research gaps within a determined subject.

The search procedure that we have defined was carefully engineered to cover all tomographic DOAS research relevant to urban, rural or industrial environments. Through it, we were able to find several different applications, all pertaining to scientific research, which are similar regarding physical principle, but differ in equipment assembly, algorithms, software, and geometry.

The rest of this paper has the following structure: Section 2 presents the context within which this study was written; Section 3 describes how we have planned to perform the study and the methods we have used in doing it; Section 4 describes the application of the said methods in the pursuit of our goals and presents the results we have obtained, as well as our evaluation of our processes; Section 5 shows our conclusions and what we think might be retained from reading this paper.

## 2. Background

### 2.1. Differential Optical Absorption Spectroscopy

Absorption Spectroscopy is the term used to identify all techniques that use radiation absorption by matter to assess and quantify elements or molecules in a given spectroscopic sample. It had, and still has, a very important role in the study of the Earth's atmosphere [22].

It is, as many other spectroscopic techniques, based on Lambert-Beer's law, which states that 'in a medium of uniform transparency the light remaining in a collimated beam is an exponential function of the length of the path in the medium', as described originally by Pierre Bouguer in 1729, and can be written [22]:

$$I(\lambda) = I_0(\lambda) \cdot \exp[-L \cdot \sigma(\lambda) \cdot c] \quad (1)$$

In Equation 1,  $I$  is the light intensity as measured by the spectrometer,  $I_0$  the original light intensity at the source,  $L$  is the optical path in which the sample is exposed to the light,  $\sigma$  is the optical cross section of the sampled element or molecule and  $c$  is the sample's concentration.  $\lambda$  is the radiation's wavelength.

Lambert-Beer's equation, while valid in a laboratory setting, is generally not enough to determine gaseous concentrations in an open atmosphere experiment.  $I_0$  determination would require any absorbant from the medium, which is impossible. Besides, in this medium, there are many factors that influence measurements: Rayleigh's scattering, Mie's scattering, thermal variations, turbulence and instrumental transmissivities. All these play an important part in altering atmospheric light [22, 16].

Differential Optical Absorption Spectroscopy (DOAS) overcomes these difficulties by capitalising on cross section's differences between interfering phenomena (normally broad spectral features) and certain trace gases (usually narrow spectral structure). The mathematical formulations behind the technique are well beyond the scope of this article, but suffice it to say that the broad structures are removed through subtraction of a fitted low order polynomial, and a fitting algorithm (such as Levenberg-Marquardt) is used to retrieve concentrations. Detailed presentations of these procedures are presented in [22] and [16].

In [22], the authors split the DOAS method into two fundamental families: passive and active. The passive family is characterised by being designed to capture and analyse natural light, whether from the Sun, the Moon or any other celestial body. This kind of measurement has the advantage of being simple to assemble, but natural light usage implies an additional technical effort for the retrieval of atmospheric concentrations. Active DOAS

applications, on the other hand, use artificial light sources to make their measurements. This has been used extensively in the identification of several atmospheric components. Its concentration extraction procedure is simpler, at the expense of a more complex assembly.

DOAS has had a number of applications throughout the years. The technique was first applied in the 1970s. At that time, Perner used an active setup with a laser light source to identify the OH radical in the atmosphere [20]. More recently, researchers around the world have been employing broadband sources (such as Xenon lamps) to measure trace gases like Ozone, Nitrogen Dioxide or Sulphur Dioxide. Almost simultaneously, passive systems have been used to study stratospheric chemistry and radiative transport in clouds [22].

### 2.2. Multi-Axis DOAS

Multi-Axis DOAS (MAX-DOAS) is one of the more recent applications of the DOAS technique. It represents a significant progress regarding zenith scattered sunlight measurements, a well established atmospheric analysis technique. It performs a series of passive DOAS measurements in several telescope elevations (typically 4 to 10) [11], either in sequence or simultaneously.

MAX-DOAS stems from another set of techniques called *off-axis*, which in this case means that the telescope is pointed at another angle than the zenith. Off-axis DOAS was first employed in 1993 when Sanders et al. [25] used it to assess OCIO in Antarctica. During this experiment, the team concluded that the off-axis geometry greatly improves sensitivity for tropospheric species, but does not change the system's ability to quantify stratospheric absorbers.

By evaluating several directions, the technique allows researchers to measure not only stratospheric contributors, as zenith sky assemblies, but also to detect absorbers at ground level, as an active DOAS instrument would.

We mention MAX-DOAS in this paper because one could argue that these systems would be able to be adapted to perform tomographic measurements, if more than one system would analyse the same region from more than the same number of observation angles. MAX-DOAS tomography is a special case, and would probably deserve to be investigated fully. However, since this was not the object of our study, we chose not to specifically target this method in our search.

### 2.3. Imaging DOAS

Imaging DOAS combines spectral and spatial information by combining an imaging spectrometer with a scanning system. The resulting data clearly resembles that of a hyperspectrum ??.

The method, developed by Bobrowsky et al. [2], employs a 2D CCD detector. One dimension measures spectral information, while the other

contains spatial information for one direction. The other spatial direction is obtained by scanning the field of view with the pushbroom method.

DOAS is used to yield slant column density values for the absorbers for each pixel. The values are colour coded and produce an image describing the gas distribution.

Mention to this technique is included in this paper because it exists in order to produce a two-dimensional image from spectral information. This image, however, does not come from a tomographic reconstruction procedure, nor is spatial information recovered from projections, but instead comes directly from the acquisition method. Hence, we did not include articles on this method in this study.

#### 2.4. Tomography

Tomography refers to the set of techniques that aim to produce a cross sectioning image from data collected by exposing a given target body to some kind of penetrating or reflecting wave from many different directions [9, 13].

The initial theories that gave rise to tomography were laid out by Johannes Radon in 1917, with a mathematical operation that would later be known as the Radon transform. This process maps a function  $f$ , defined in the plane, to the function  $Rf$ , comprised of the values of the line integrals of  $f$ , taken in  $\theta$  directions. In practice, this formulation allows the reconstruction of an image by its projections, which are nothing more than line integrals [4].

Tomographic image reconstruction can be achieved by running one of several algorithms through a computer program. The presentation of these algorithms is completely beyond the scope of this article, but a good starting point for learning about these operations is *The Mathematics of Medical Imaging*, by Timothy Freeman [4]. It is in the scope of this article, however, to make a small introduction to a particular set of reconstruction methods. The reason for this being the prevalence of these methods in the field of DOAS tomography, which is the main subject of this study. These techniques are thus:

**Algebraic Reconstruction Techniques (ART)** Proposed in 1970 by Gordon and Herman [10], these techniques are based on successive approximations between the actual projection data and the sum of the reconstruction elements which represent it [5]. The process is conducted line by line, until a satisfactory convergence condition is met.

**Simultaneous ART** Simultaneous ART is very similar to the ART algorithm. The difference being that the iterative changes occur for all lines at the same time, instead of in only one.

**Simultaneous Iterative Reconstruction Techniques (SIRT)** The main difference between SIRT and SART is that in the former, cell changes are not reflected immediately after one calculation. Updates occur at the end of each iteration. At this point, the change for each cell is the average correction calculated for it taking all equations into account [12].

During the second half of the twentieth century, tomographic processes have had a revolutionary influence in many fields of study, but especially in medicine. Computational tomography scanners allow doctors to see their patients interior in a highly detailed and extremely safe fashion. At first, tomographic imaging was performed only with X-Rays. Their attenuation throughout the patient's body being used as a projection. Nowadays, there are much more methods of image retrieval, such as radioisotopes, ultrasound or particle annihilation [13, 4, 9].

Although it was the field of medicine was more influenced by tomographic procedures than any other, the applications of these methods are not restricted to it. One can find numerous industrial and research applications [1, 6, 3]. One of which is the application to atmospheric research, namely in conjunction with DOAS. In recent years, scientists have been working on tomographic methods for measuring atmospheric trace gas concentration values. The field is interesting because it allows for 2D or even 3D mapping of a given region, with respect to those trace gases. This article aims to make an assessment of the status of this tomographic application, by analysing current literature on the subject.

### *2.5. Mapping Study*

A Systematic Mapping Study (MS) is a type of secondary study designed to determine the general features of the research landscape in the subject they are addressing [14, 21].

An MS is driven by broad (and often multiple) research questions and applies an also broad data extraction protocol. This is in line with the fact that this kind of study aims to summarise its findings, answering the research questions, and in-depth analysis is not required. It is common for an MS to be a precursor to a Systematic Literature Review (SLR), which is a much deeper kind of systematic study. Guidelines for performing studies of both kinds can be found in a report made by Kitchenham and Charters in 2007 [14]. In this document, the authors establish the 3 stages which all MS and SLR generally have:

**Planning** This stage includes all preliminary considerations regarding the MS or SLR in the making. All protocols, from search to evaluation, through data extraction, are devised;

**Conduction** During this staged, researchers apply what they have planned in the previous phase. Protocols are *actually* run, and data is synthesised;

**Reporting** In this phase, the team has to define their dissemination strategy, and implement it. It is in this stage that a final report is written and evaluated.

Although it is logical (and fundamentally correct) to assume that these steps are sequential, this may not be, and usually is not, accurate. Many of these stages and their intermediate steps require iteration. For instance, some inclusion or exclusion criteria may only be found necessary once the search protocol is implemented.

### 3. Methods

In the elaboration of this article, we took the three normal stages of SLR conduction: planning, conduction and reporting. The first stage involves making the decisions that guide the rest of the process; the conduction phase is comprised of the actual gathering of data, using the protocol defined in the first stage. The final section of the study is basically the writing and the publishing of the results. In this section, we present the methods used in the study and their rationale, which roughly corresponds to the planning stage.

#### 3.1. Objectives

An MS always aims to answer its research questions in a broad but definite way. It is a way of understanding a given field of research, and being able to systematise how this understanding is achieved.

As stated before, this is an MS aimed at characterising DOAS tomography general status. In doing this, we pretend to get a clearer image of what has been done and what should be attempted next, hopefully managing a sort of roadmap for future research contributions.

#### 3.2. Research Questions

We have begun by defining the goals for our study, and structuring them with a PICOC (Population, Intervention, Context, Outcome and Comparison) analysis, which is summarised in Table 1. This analysis led us to our research goal: *to assess the technological status of the DOAS tomography technique*.

We used this goal statement as a primer to our research question, which was then formulated as: **what is the current status of the technology used in tomographic DOAS?**

**Table 1:** *PICO analysis.*

|                     |   |
|---------------------|---|
| <b>Population</b>   | DOAS research in general.                             |
| <b>Intervention</b> | The papers must address tomographic DOAS.             |
| <b>Outcome</b>      | Status <b>assessment</b> for <b>DOAS tomography</b> . |
| <b>Context</b>      | Research papers.                                      |

**Table 2:** *Research question slicing*

|                 |  |
|-----------------|--|
| <b>Original</b> | What is the current status of the technology used in tomographic DOAS? |
| <b>RQ1</b>      | Is there a typical hardware setup used in tomographic DOAS studies?    |
| <b>RQ2</b>      | Is there a standard software used to perform these analysis?           |
| <b>RQ3</b>      | What are the algorithms more commonly used?                            |

Now, this question is too vague to pursue in a systematic fashion, so we had to slice it into smaller and more objective chunks. This sectioning is presented in Table 2.

The research question is one of the most important steps in planning a Systematic Literature Review, but it cannot be entered into a library's search box. Therefore, we have to define our search terms before we can make any effort of answering our questions.

### 3.3. Search Query Definition, Library Selection and Filter Definition

In the case of this study, the search terms were selected in order to purposefully maintain a broad scope, so that we could retrieve a high number of relevant studies. The selected search terms were: **DOAS atmospher\*** **tomography**<sup>1</sup>. The search query was entered into 5 academic search engines, as shown in Table 3.

**Table 3:** *Electronic libraries used in this study.*

| Library                | URL   |
|------------------------|---|
| Google Scholar (GS)    | <a href="https://scholar.google.com/">https://scholar.google.com/</a>                                 |
| Web of Knowledge (WoK) | <a href="https://webofknowledge.com">https://webofknowledge.com</a>                                   |
| Science Direct (SD)    | <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>                             |
| IEEE                   | <a href="http://ieeexplore.ieee.org/">http://ieeexplore.ieee.org/</a>                                 |
| AGU Publications (AGU) | <a href="http://agupubs.onlinelibrary.wiley.com/hub/">http://agupubs.onlinelibrary.wiley.com/hub/</a> |

<sup>1</sup>The asterisk acts as a wildcard.

After setting Table 3 libraries, it was time to define our article selection criteria, which are summarised in Table 4. We began with 2 Inclusion Criteria (IC) and 3 Exclusion Criteria. The IC determined that our selected papers would have to be journal articles (thus excluding thesis, white papers, patents and other documents) and that these articles should be on the topic of Tomographic DOAS. The EC dictated that no selected paper should include volcanology studies or satellite data analysis (these have particularities which we do not want to approach in this study) and that no other language than English will be accepted.

During the course of the search, however, we had to include another two EC. The first was included in the Google Scholar search, where we understood that papers from a certain publisher were not accessible. The second came in the subsequent searches, when it became clear that most papers had already been retrieved by the GS search.

**Table 4:** Selection filters in use for this study's search.

| Criterium Definition |     |  |
|----------------------|-----|--|
| <b>Exc. Criteria</b> | EC1 | Duplicate in Scholar                     |
|                      | EC2 | Non English articles are not accepted    |
|                      | EC3 | Volcanology papers are not accepted      |
|                      | EC4 | Satellite data papers are not accepted   |
|                      | EC5 | CNKI published articles are not accepted |
| <b>Inc. Criteria</b> | IC1 | Results must be articles                 |
|                      | IC2 | Results must be about Tomographic DOAS   |

### 3.4. Data Extraction Strategy

The data extraction process is a key part of any systematic review, whether an SLR or an MS. It determines how each article is approached with regard to its content, before any information is retrieved. In our case, our strategy took place in two separate moments: an initial screening, in which we would assess contents as expressed by the articles' abstract; and a second moment, in which we performed a full article read. Special attention was given to explicit sections covering our target topics (equipment, algorithm and software).

### 3.5. Quality Assessment

It is very difficult to assess a paper's quality, and to rank it accordingly. However, for this review in particular, we have decided to follow Souza's

**Table 5:** *Quality assessment criteria presentation.*

| Criterium Type           | Criterium<br>(Weight)          | Decision Factor                          | Score |
|--------------------------|--------------------------------|--|-------|
| <b>General Criteria</b>  | Contribution to this SLR (0,2) | cited in study:<br>more than three times | 1     |
|                          |                                | cited in study:<br>three times           | 0,75  |
|                          |                                | cited in study:<br>twice                 | 0,5   |
|                          |                                | cited in study:<br>once or less          | 0,25  |
|                          |                                | Detailed                                 | 1     |
| <b>Specific Criteria</b> | Algorithm description (0,6)    | Semi-Detailed                            | 0,4   |
|                          |                                | Mentioned                                | 0,2   |
|                          |                                | None                                     | 0     |
|                          |                                | Detailed                                 | 1     |
|                          | Instrument description (0,2)   | Semi-Detailed                            | 0,4   |
|                          |                                | Mentioned                                | 0,2   |
|                          |                                | None                                     | 0     |
|                          |                                | Mentioned                                | 1     |
|                          | Software Description (0,1)     | None                                     | 0     |

approach [26] and adopt a similar evaluation method. Table 5 contains the used criteria.

In our evaluation model, we took into account both general and specific criteria. The former addresses an article's contributions to our particular SLR; the latter targets the actual content of that article.

In order to measure the contribution of each individual paper to our study (our general criterium), we have assessed its number of citations in all the other selected papers. This is a valid measurement of a paper's impact in the study, but it might become difficult to implement if a high number of articles are selected for the final stage. Contentwise (specific criteria), we have defined our scoring model according to the Research Question separation explained in Subsection 3.2.

In our study's case, distinction between Specific and General was not sufficient to adequately separate scores according to importance. It was necessary to introduce scoring weights for that end. These weights are also shown in Table 5 and were set according to the goals of our SLR, meaning

that the tomographic element is the most important.

In the end, a paper's total score comes from the formula described by Equation 2.

$$TotalScore = \sum_i w_i \cdot S_i \quad (2)$$

Where  $S_i$  and  $w_i$  are a paper's score and weight for a given criterium, respectively.

Finally, we shall discuss the different weights given to each criterium and the different ways in which they are evaluated. The most important aspect that we are trying to assess is the algorithm, which defines the whole tomographic process and the results achievable by the studies.

A detailed algorithmic description includes the mathematical basis as well as a complete description of required adaptations, both on the mathematical level and on a conceptual method.

Instrument description is also an important criterium, since it is with it that scientists retrieve the information they will afterwards process tomographically, through the algorithm.

It is sometimes difficult to establish how good an instrument description is. A too detailed description can be just as bad as a non-sufficient one, if the equipment options are not correctly presented.

That being said, we have considered a detailed description one that includes explicit mention to the composition of the optical system and its assembly details, together with the analysing hardware (e.g. spectrometers) configuration and capabilities.

The least important of the technical features under evaluation is the software. This is because theoretically, results would be the same independently of the software in use. We have included this feature in the study as a way of identifying if there was some kind of software prevalence in the community. In this study, software is binarily assessed: either the scrutinised study mentions it or not.

Finally, we evaluate the contribution of each article to this study. Since DOAS tomography is a field with a relatively low number of players, it can be expected that there are many cross citations. We have introduced this as a method of measuring an article's relative importance within this mapping study, simply by counting the number of times a cross citation occurs.

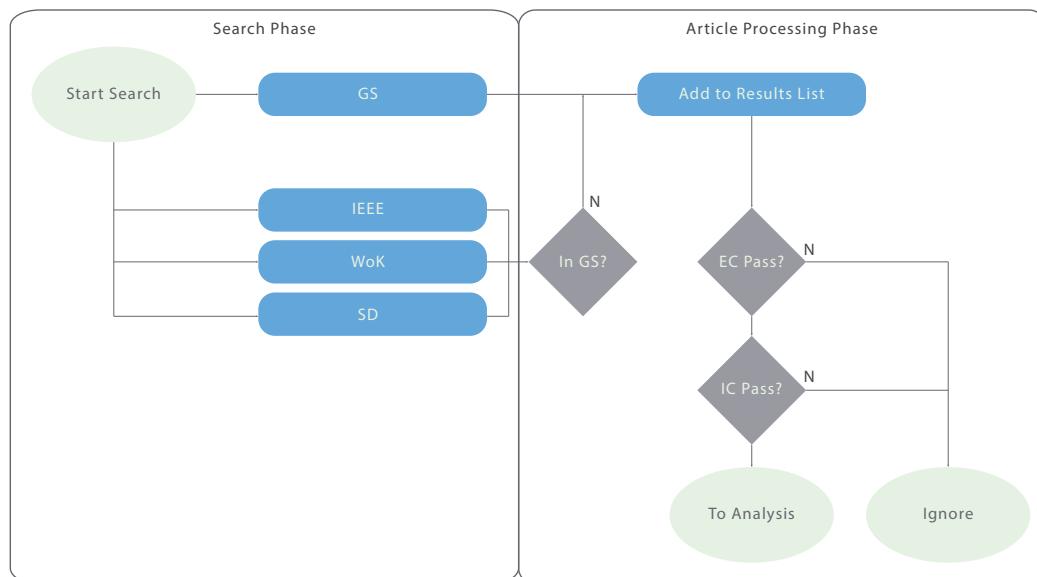
## 4. Conduction

### 4.1. The Search

SLR guideline literature [14, 28] recommend that the first stage of any Systematic Literature Review be the search for previous literature of this kind, since a recent systematic study may render the execution of a new study disencourageable. In our case, none of the several libraries used appeared to have any article of the sort.

The search terms, derived in Subsection 3.2, were run in all libraries found in Subsection 3.3. The Google Scholar search had the particularity of being run through a specialised software called *Publish or Perish*[8], which allowed the search results to be exported to a comma separated values file, which made the process a lot easier, since it was then possible to work the data directly in a spreadsheet program (Microsoft Excel, in this particular case).

The conduction phase of our study followed the flowchart illustrated by Figure 1. Notice that Google Scholar is the first library to be searched. This is motivated by the fact that the vast majority of the articles were retrieved by Google's academic search engine. In fact, GS-retrieved articles were so predominant that we had to create a special EC, as described in Subsection 3.3.



**Figure 1:** Conduction stage flowchart. Notice Google Scholar's prevalence.

## 4.2. Results and Discussion

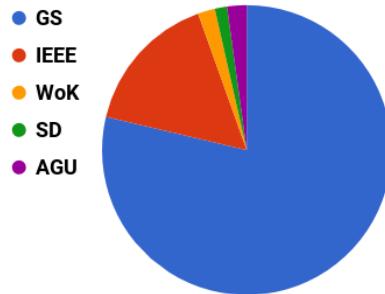
### 4.2.1. Presentation and analysis

Our search returned 732 results, of which 709 were distinct ( $\approx 97\%$ ). Of these, 601 were journal articles ( $\approx 82\%$ ). The vast majority of the results came from GS ( $\approx 80\%$ , see Figure 2). Selection criteria (Inclusion and Exclusion) application resulted in the exclusion of 701 results ( $\approx 99\%$ ), thus leaving 8 articles reaching the content analysis stage (the attempt to answer the Research Question). A summary of these findings can be seen in Table 6.

**Table 6:** Search results. For a paper to reach the rightmost column, which means it is selected, it must verify both IC1 and IC2 as well as none of the EC, ranging from EC1 to EC5.

| Articles which trigger criteria |            |     |     |     |     |     |     |     |               |
|---------------------------------|------------|-----|-----|-----|-----|-----|-----|-----|---------------|
| Source                          | # Articles | IC1 | IC2 | EC1 | EC2 | EC3 | EC4 | EC5 | Rem. Articles |
| GS                              | 576        | 455 | 142 | -   | 25  | 82  | 53  | 18  | 8             |
| IEEE                            | 116        | 116 | 1   | 0   | 0   | 1   | 1   | 0   | 0             |
| WoK                             | 14         | 14  | 6   | 13  | 0   | 1   | 1   | 0   | 0             |
| SD                              | 10         | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0             |
| AGU                             | 16         | 1   | 1   | 1   | 1   | 1   | 1   | 0   | 1             |
| <b>TOTAL</b>                    |            |     |     |     |     |     |     |     | <b>9</b>      |

| Library | # Results | Percentage |
|---------|-----------|------------|
| GS      | 576       | 78.7%      |
| IEEE    | 116       | 15.8%      |
| WoK     | 14        | 1.9%       |
| SD      | 10        | 1.4%       |
| AGU     | 16        | 2.2%       |



**Figure 2:** Results distribution by library.

Table 7 presents the 8 selected articles. It categorises them with respect to their covered topics and whether they are empirical or theoretical in nature. The same table summarises article scores according to the criteria defined in Subsection 3.5, and presents the keys with which we will refer to each article from this point forward.

The 8 papers averaged a score of 0,48 and a median score of 0,6. There is a strong difference between the average and the median suggesting that there are some outliers or some kind of clustering. Although this is actually the case, it is statistically irrelevant, given the small size of the sample.

**Table 7:** Scoring results for the selected articles. In the second column, a T means the article is theoretical and an E means the article is empirical.

| Key                  | Type | Alg. | Inst. | Soft. | Cit. | Score |
|----------------------|------|------|-------|-------|------|-------|
| Hartl2006 [7]        | T    | 1    | 0,4   | 0     | 0,5  | 0,73  |
| Laeapple2004 [15]    | T    | 1    | 0     | 0     | 1    | 0,70  |
| Mettendorf2006 [17]  | E    | 0,4  | 1     | 1     | 0,25 | 0,57  |
| ODriscoll2003 [18]   | T    | 1    | 0     | 0     | 0,25 | 0,63  |
| Olaguer2017 [19]     | E    | 0    | 0,2   | 0     | 0,25 | 0,07  |
| Poehler (unpub) [23] | E    | 0    | 0,4   | 0     | 0,25 | 0,11  |
| Pundt2005 [24]       | E    | 0,4  | 1     | 1     | 1    | 0,64  |
| Stutz2016 [27]       | E    | 0    | 1     | 1     | 0,25 | 0,33  |

### 4.3. Discussion

In this subsection, we will use the 8 articles that were selected in order to try and answer the Research Questions. For clarity, we will approach with respect to the instrument, algorithm and software in a separate manner.

Different articles present data in different ways, and with different levels of detail. This has been taken into account in designing our evaluation method, and it should also be observed when discussing results. Therefore, our general approach in this subsection will be to address the more detailed articles first and then complement that information with what we can gather from the less detailed papers.

#### 4.3.1. Instrument

Instrumentation description is present in 7 of the 8 ([7, 15, 17, 19, 23, 24, 27]) selected articles. Stutz2016 [27], Pundt2005 [24] and Mettendorf2006 [17] present the highest level of detail.

In Stutz2016 [27], the authors used a newly developed Long Path DOAS instrument for the study of atmospheric concentration of Benzene, Toluenes and Xylenes. This instrument's main innovation is its light source, which consists in a double LED (255nm and 265nm) assembly. This system's telescope is a homebuilt telescope with a focal length of 120 cm and a 12 inch diameter aluminum coated main mirror, mounted on a high accuracy motorised pan and tilt unit from Newark Systems. The telescope is used both

as emitter and receiver, therefore the system needs a reflector. Stutz used a quartz corner cube reflector array, with an individual reflector diameter of 57 mm and the number of reflector ranging from 10 to 25 (depending on the path length). For detection, the system relied on a UV-enhanced PIXIS 256 CCD detector from Princeton Instruments on an Acton spectrometer with 300 grating and  $\approx 0.3$  nm spectral resolution, which was stabilised to -35°C.

Pundt2005 [24] was conducted during the BAB II motorway campaign. The team was working with the goal of performing a tomographic measurement of vehicle pollution along a certain motorway between Heidelberg and Mannheim. For that, they used an assembly of two telescopes and eight reflectors, rendering a total of 16 light paths, then used to perform a tomographic reconstruction of the trace gas detection in that region. The telescopes used had a focal length of 150 and 80 cm, with respective diameters of 300 and 200 mm. Both assemblies used Acton spectrometers. One used the Acton 500, with 0.5 nm spectral resolution in the range between 295 and 375 nm; the other used an Acton 300, with 0.4 nm spectral resolution between 295 and 355 nm. In both cases, the sensor used was a 1024 pixel Photo Diode Array (PDA), thermally stabilised at -15°C. The telescopes were pointed towards two towers which bore the reflectors, set at heights of 10, 20, 30 and 40 m from the ground.

In Mettendorf2006 [17], the authors validated two-dimensional LP-DOAS tomography through an indoor experiment. To this end, they have used three multibeam instruments, which consisted in a telescope with a focal length of 1.5 m and 300 mm in diameter, which was also used as emitter and receiver. The system used a broad spectrum Xenon lamp as light source, though no details are given. The experiment assembly included the careful positioning of plane mirrors and 6 cm diameter corner cube reflectors, used to create a total of 39 light paths (13 for each multibeam instrument).

As for the other 4 less detailed instrument description, three (Hartl2006 [7], Poehler [23] and Laepple2004 [15]) are from the same group as Pundt2005 [24] and Mettendorf2006 [17], and therefore use the same or similar hardware. Olaguer2017 [19], on the other hand, is the companion paper of Stutz2016 [27], and therefore gives a description of the same instrumentation, though in a less detailed manner.

#### *4.3.2. Algorithm*

The reconstruction algorithm is the most important part of our study, as we already demonstrated by the weight it is given in our quality evaluation model (see Subsection 3.5). Algorithm descriptions are present in 6 of the 8 selected articles: [7, 15, 17, 18, 19, 24]. The most complete descriptions are featured in Hartl2006 [7], Laepple2004 [15] and ODriscoll2003 [18]. Mettendorf2006 [17],

Olaguer2017 [19] and Pundt2005 [24] approach the reconstruction algorithms with less emphasis or in a less detailed way.

In Hartl2006 [7], the research team describe their discretisation process, reconstruction methods, grid translation methods and error estimation and quality assessment, with the greatest level of detail being given to the latter.

The paper also focuses in the comparison SIRT and ART results for the test samples, which consisted in up to four Gaussian concentration profiles, which were randomly arranged in a 100x100 (a.u.) test field, in six different geometries and with up to 36 known light paths.

Furthermore, Hartl2006 [7] discusses how the choice of the reconstruction grid affects both the reconstruction error and reconstruction area integrals, the possibility of the existence of background concentration influencing equation constraints and reconstruction results, and how the whole system would behave were its geometry any different, namely regarding light paths and number of telescopes.

The next algorithm-oriented paper is Laepple2004 [15]. In this article, the group discussed several discretisation approaches, their drawbacks and advantages. Still on discretisation, they approach the problem of resolution, and the necessary balance between physical accuracy and the need for *a priori* information which arises from increasing it. Afterwards, the group presents some strategies for solving the linear system that results from discretising the concentration field and how to take error into account.

For their reconstructions, the group chose to adapt ART, SIRT, and SART (see Subsection 2.4). These adaptations were described and detailed in the article's third section, before the error estimation procedures adopted in their case. Finally, the team presents how they chose to optimise reconstruction in several aspects, including the generation of test plumes and optimisation for the BABII campaign, which was the parent project of this article.

O'Driscoll2003 [18] also covers the algorithm extensively. While this paper is considerably shorter than the previous two, it provides a detailed (on an iteration basis) description for ART and SIRT (see Subsection 2.4). In addition, and perhaps of greater interest, the paper's authors suggest a different approach to solving the reconstruction matrix, different from the algebraic methods already presented: an evolutionary algorithm.

An evolutionary algorithm is a mathematical method of solving complex problems, which mimics or is in any form based on the process of natural selection. These algorithms have, according to the paper's authors and their references, been shown to be extraordinarily powerful.

The research team have applied a Differential Evolution algorithm to the reconstruction process and provide a detailed description of how they have done this.

The other two articles which mention the algorithm are Mettendorf2006 [17] and Pundt2005 [24]. Both these studies were conducted under the same project as Laepple2004 [15] and Hartl2006 [7] and therefore their algorithm descriptions and methods draw heavily on these two studies.

#### 4.3.3. Software

Of the 8 selected articles, only 3 mention the software used. Even these, do not go into any detail of the reasons that led to that specific usage.

In Mettendorf2006 [17], the team used TOMOLAB for the calculation of the modelled column densities of their experiment. In Pundt2005 [24], spectral analysis was performed using the *MFC Software*. Finally, Stutz2016 [27] used the DOASIS software for control and automation purposes, and does not explicitly mention its use for spectral analysis purposes, although this is likely.

#### 4.3.4. General Observations

While this is not a part of the discussion *per se*, we believe it makes sense to make some general observations about the data which we had to analyse.

The first important mention is the BABII campaign. This study, which ran in 2001 and aimed to quantify pollution from the A656 motorway between Heidelberg and Mannheim produced a significant part of the literature which we analysed.

Another point which should be addressed is that all DOAS tomography efforts detected in this search were based on active DOAS technology. This only means that the DOAS systems all employed an artificial light as a light source.

A final remark is due to the prevalence of algebraic methods for solving the discretisation and reconstruction problem, namely ART, SART and SIRT.

#### 4.4. Validity Threats

When writing an MS or an SLR, authors always have to analyse their findings and methods in order to mitigate potential sources of error or lack of validity. This is called a validity threat analysis.

There are two main families of validity threats. They can be internal, i.e., they come from the methods employed used in conducting the study; or external, which means that the threat comes from the applicability (or lack thereof) of the effects observed in the study, outside of its scope.

On the level of internal validity of our study, two main observations come to mind:

**Relevant papers left out** The very low number of found studies could be an indication that our inclusion and exclusion filters were set in a too restrictive manner.

It could also happen that some relevant papers were not found due to being written in such a way that the libraries' search engines did not find them with our search phrase. This same problem would also occur if for some reason, an important library was left out of the study, and therefore not searched.

We mitigate all these risks by selecting a purposefully broad search phrase, by using powerful general search engines (eg. Google Scholar) and by running several undocumented test-runs with other search phrases.

A common strategy used for tackling this kind of threat is to extend the study through snowballing. In our case, we have opted to not perform this operation because of the very high cross-reference pattern between the found studies.

**Quality of selected papers** While it is true that we do not have any control over the quality of the articles rendered by the search engines, and there is no standard regarding it, we must address the issue that it entails. We have tried to mitigate this risk, as far as we can, by using strict and strong selection criteria in systematic fashion (see Section 3).

On the external threat plane, we contend with the applicability of our findings outside our study. We have tackled this issue by trying to remain focused only to the technologic aspect of the Tomographic DOAS technique, both in respect to its instrumentation and to the mathematical methods involved.

With this in mind, and even if the internal validity threats were all verifiably concerning, this study's finding are of great use to anyone wanting to understand how the field is working or wishing to design and build an analysis system.

## 5. Conclusions

The essential goal of this MS was to assess the technological status of the tomographic DOAS technique as described in the relevant academic literature.

By performing a systematic search using the phrases and libraries enumerated in Section 3, we have retrieved more than 700 articles. The elimination process took place through the application of inclusion and exclusion criteria, as described in full in Subsection 3.3.

In the end, 8 articles were identified as relevant. There are several possible justifications for such a low number of retrieved articles. Some of those reasons are addressed in Subsection 4.4. Mostly, we believe that DOAS tomography

is a relatively new field of study, which has not yet been sufficiently explored, and that results in a small *corpus* of literature.

Our analysis was performed on three distinct levels: **algorithms**, **hardware** and **software**. We have found that, while there are commonalities amongst almost all papers (such as the fact that they are all active DOAS applications), on the hardware and software planes, there is no *standard model* for the used devices. On the algorithm level, however, and due to data quantity restrictions related to the low number of used projections (several tens), iterative reconstruction methods (namely ART) are almost universal.

## 6. Acknowledgments

This study was funded by the FCT grant PDE/BDE/114549/2016 and by the ATMOS Project (PORLisboa FEDER 3491).

## References

- [1] Industrial Tomography - Systems and Applications. In M. Wang, editor, *Industrial Tomography*, volume 4. Elsevier. ISBN 9781782421184. doi: 10.1016/B978-1-78242-118-4.12001-5. URL <http://linkinghub.elsevier.com/retrieve/pii/B9781782421184120015>.
- [2] N. Bobrowski, G. Hönninger, F. Lohberger, and U. Platt. IDOAS: A new monitoring technique to study the 2D distribution of volcanic gas emissions. 150(4):329–338. ISSN 03770273. doi: 10.1016/j.jvolgeores.2005.05.004.
- [3] R. L. Byer and L. A. Shepp. Two-dimensional remote air-pollution monitoring via tomography. 4(3):75–77. ISSN 0146-9592. doi: 10.1364/OL.4.000075. URL <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19687805&retmode=ref&cmd=prlinks>.
- [4] T. G. Feeman. *The Mathematics of Medical Imaging*. Springer Undergraduate Texts in Mathematics and Technology. Springer New York. ISBN 978-0-387-92711-4. doi: 10.1007/978-0-387-92712-1. URL <http://link.springer.com/10.1007/978-0-387-92712-1>.
- [5] R. Gordon. A tutorial on art (algebraic reconstruction techniques). 21(3):78–93. ISSN 0018-9499. doi: 10.1109/TNS.1974.6499238. URL <http://ieeexplore.ieee.org/document/6499238/>.

- [6] C. Haisch. Optical Tomography. 5(1):57–77. ISSN 1936-1327. doi: 10.1146/annurev-anchem-062011-143138. URL <http://www.annualreviews.org/doi/10.1146/annurev-anchem-062011-143138>.
- [7] A. Hartl, B. C. Song, and I. Pundt. Atmospheric Chemistry and Physics 2-D reconstruction of atmospheric concentration peaks from horizontal long path DOAS tomographic measurements: parametrisation and geometry within a discrete approach. 6(3):847–861. ISSN 1680-7324. doi: 10.5194/acpd-5-11781-2005. URL [http://www.atmos-chem-phys.net/6/847/2006/](http://www.atmos-chem-phys.net/6/847/2006/www.atmos-chem-phys.net/6/847/2006/).
- [8] A. Harzing. Publish or Perish. URL <http://www.harzing.com/pop.htm>.
- [9] G. T. Herman. *Fundamentals of Computerized Tomography*. Advances in Pattern Recognition. Springer London. ISBN 978-1-85233-617-2. doi: 10.1007/978-1-84628-723-7. URL <http://link.springer.com/10.1007/978-1-84628-723-7>.
- [10] G. T. Herman, A. Lent, and S. W. Rowland. ART: Mathematics and applications. A report on the mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques. 42(1). ISSN 10958541. doi: 10.1016/0022-5193(73)90145-8.
- [11] G. Hönniger, C. von Friedeburg, and U. Platt. Multi Axis Differential Optical Absorption Spectroscopy (MAX-DOAS). (4):231–254. URL [www.atmos-chem-phys.org/acp/4/231/](http://www.atmos-chem-phys.org/acp/4/231/).
- [12] A. C. Kak. Algebraic Reconstruction Algorithms. In *Computerized Tomographic Imaging*, pages 275–296. ISBN 978-0-89871-494-4. doi: 10.1137/1.9780898719277.ch7.
- [13] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Number 1. Society for Industrial and Applied Mathematics. ISBN 978-0-89871-494-4. doi: 10.1137/1.9780898719277. URL <http://epubs.siam.org/doi/book/10.1137/1.9780898719277>.
- [14] B. Kitchenham and S. Charters. Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3. 45(4ve):1051. ISSN 00010782. doi: 10.1145/1134285.1134500. URL <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Guidelines+for+performing+Systematic+Literature+Reviews+in+Software+Engineering%20%Cnhttp://www.dur.ac.uk/ebse/resources/Systematic-reviews-5-8.pdf>.

- [15] T. Laepple, V. Knab, K.-U. Mettendorf, and I. Pundt. Long-path DOAS tomography on a motorway exhaust gas plume: numerical studies and application to data from the BAB II campaign. 4(3):2435–2484. ISSN 1680-7375. doi: 10.5194/acpd-4-2435-2004. URL [www.atmos-chem-phys.org/acp/4/1323/](http://www.atmos-chem-phys.org/acp/4/1323/) <http://www.atmos-chem-phys-discuss.net/4/2435/2004/>.
- [16] A. Merlaud. Development and use of compact instruments for tropospheric investigations based on optical spectroscopy from mobile platforms. URL <http://books.google.com/books?hl=en&lr=&id=inXVSyR82zwC&oi=fnd&pg=PR3&dq=Development+and+use+of+compact+instruments+for+tropospheric+investigations+based+on+optical+spectroscopy+from+mobile+platforms+sciences&ots=VdebeDBQMc&sig=4G0eVEvthJXuqd8WI3IWjjVXuXc>.
- [17] K. U. Mettendorf, A. Hartl, and I. Pundt. An indoor test campaign of the tomography long path differential optical absorption spectroscopy technique. 8(2):279–287. ISSN 1464-0325. doi: 10.1039/B511337G. URL <http://xlink.rsc.org/?DOI=B511337G>.
- [18] S. O'Driscoll, J. G. Murphy, and N. J. Smith. Computed tomography of air pollutants in street canyons. 4876:958–967. doi: 10.1111/12.463985.
- [19] E. P. Olaguer, J. Stutz, M. H. Erickson, S. C. Hurlock, R. Cheung, C. Tsai, S. F. Colosimo, J. Festa, A. Wijesinghe, and B. S. Neish. Real time measurement of transient event emissions of air toxics by tomographic remote sensing in tandem with mobile monitoring. 150:220–228. ISSN 13522310. doi: 10.1016/j.atmosenv.2016.11.058. URL <http://linkinghub.elsevier.com/retrieve/pii/S1352231016309487>.
- [20] D. Perner, D. H. Ehhalt, H. W. Pätz, U. Platt, E. P. Röth, and A. Volz. OH - Radicals in the lower troposphere. 3(8):466–468. ISSN 00948276. doi: 10.1029/GL003i008p00466. URL <http://doi.wiley.com/10.1029/GL003i008p00466>.
- [21] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. Systematic Mapping Studies in Software engineering. pages 1–10. doi: citeulike-article-id: 3955889.
- [22] U. Platt and J. Stutz. *Differential Optical Absorption Spectroscopy*. Springer. ISBN 9783540211938.

- [23] D. Poehler, B. Rippel, A. Stelzer, K. U. Mettendorf, A. Hartl, U. Platt, and I. Pundt. Instrumental setup and measurement configuration for 2D tomographic DOAS measurements of trace gas distributions over an area of a few square km.
- [24] I. Pundt, K. U. Mettendorf, T. Laepple, V. Knab, P. Xie, J. Lösch, C. V. Friedeburg, U. Platt, and T. Wagner. Measurements of trace gas distributions using Long-path DOAS-Tomography during the motorway campaign BAB II: Experimental setup and results for NO 2. 39(5): 967–975. ISSN 13522310. doi: 10.1016/j.atmosenv.2004.07.035.
- [25] R. W. Sanders, S. Solomon, J. P. Smith, L. Perliski, H. L. Miller, G. H. Mount, J. G. Keys, and A. L. Schmeltekopf. Visible and near-ultraviolet spectroscopy at McMurdo Station, Antarctica: 9. Observations of OCIO from April to October 1991. 98(D4):7219–7228. ISSN 01480227. doi: 10.1029/93JD00042. URL <http://doi.wiley.com/10.1029/93JD00042>.
- [26] E. Souza, A. Moreira, and M. Goulão. Deriving Architectural Models from Requirements Specifications: a Systematic Mapping Study. ISSN 09505849. doi: 10.1016/j.infsof.2019.01.004. URL <https://linkinghub.elsevier.com/retrieve/pii/S0950584919300035>.
- [27] J. Stutz, S. C. Hurlock, S. F. Colosimo, C. Tsai, R. Cheung, J. Festa, O. Pikelnaya, S. Alvarez, J. H. Flynn, M. H. Erickson, and E. P. Olaguer. A novel dual-LED based long-path DOAS instrument for the measurement of aromatic hydrocarbons. 147:121–132. ISSN 13522310. doi: 10.1016/j.atmosenv.2016.09.054. URL <http://linkinghub.elsevier.com/retrieve/pii/S1352231016307713>.
- [28] E. Tacconelli. *Systematic reviews: CRD's guidance for undertaking reviews in health care*, volume 10. ISBN 1900640473. doi: 10.1016/S1473-3099(10)70065-7. URL <http://linkinghub.elsevier.com/retrieve/pii/S1473309910700657>.

D

TESTING LAMBERTIAN ABSORPTION -  
HANDOUT PROTOCOL FOR THE  
EXPERIMENT OF SECTION 1.2

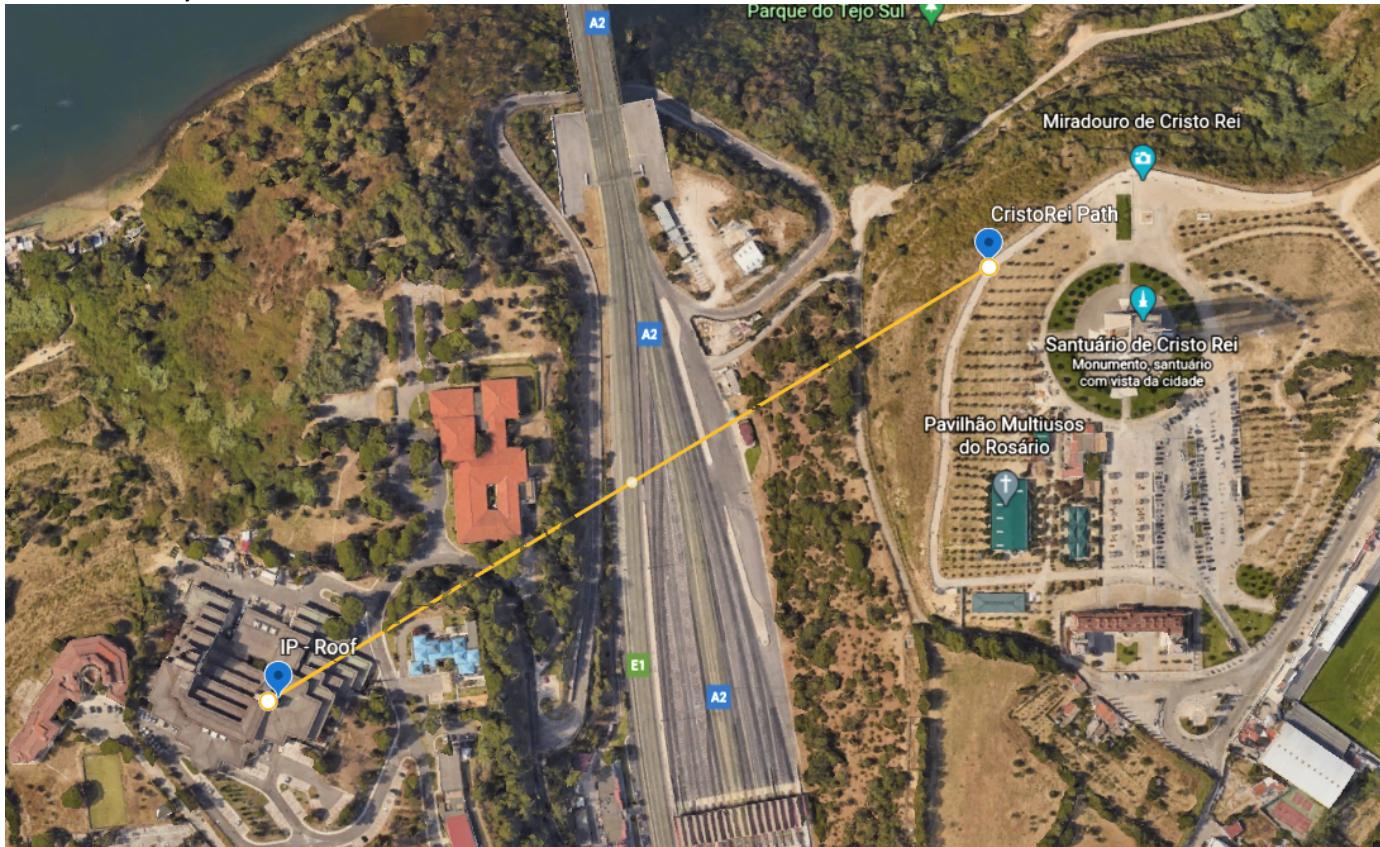
# Testing Lambertian Absorption

|                             |  |
|-----------------------------|--|
| Experiment plan and results | This experiment is designed to prove that we can use the lambertian notion that light absorption depends on the optical path and that we can use this to determine a projection in a tomographic imaging system. |
| Experiment owner            | @ Rui Almeida  |
| Jira ticket(s)              | <a href="https://rfa123c.atlassian.net/browse/PHD-3">https://rfa123c.atlassian.net/browse/PHD-3</a>  |
| Status                      | DONE   |
| On this page                | <ul style="list-style-type: none"><li>• <a href="#">Experiment planning</a></li><li>• <a href="#">Results</a></li><li>• <a href="#">Conclusions</a></li></ul>  |

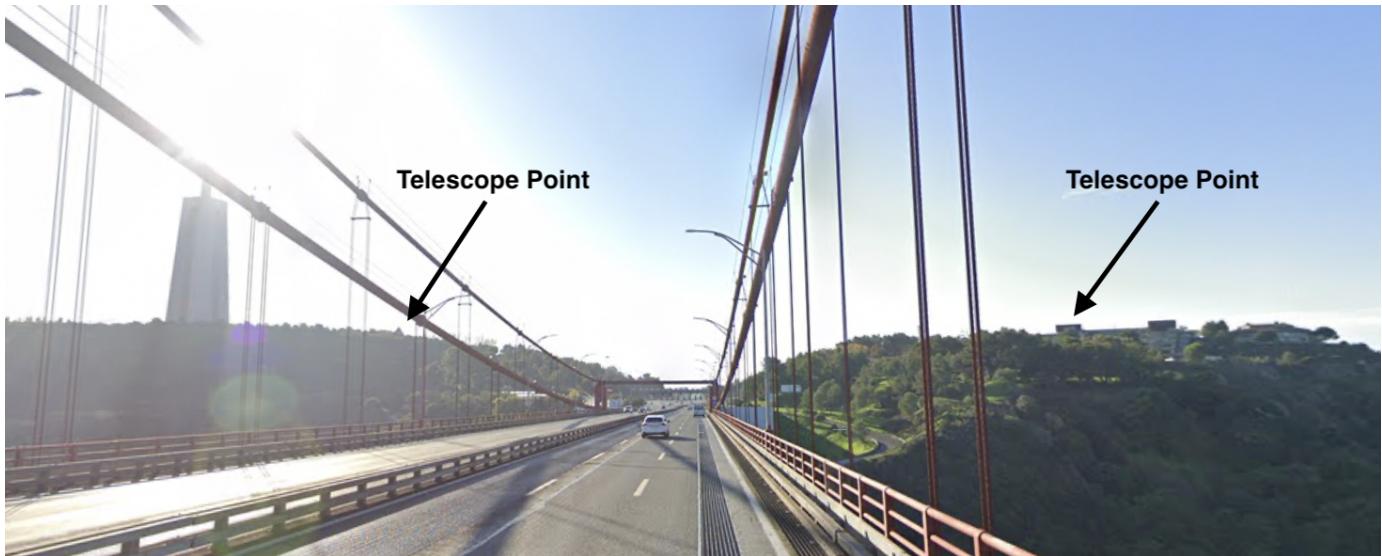
## Experiment planning

### Overview

The experiment was designed to validate that we are able to measure trace gas concentrations with our current instruments and using passive DOAS for an optical path of around 1km. We will assemble a telescope on the roof of the IEP building and another one in a specific point of the Cristo Rei sanctuary.

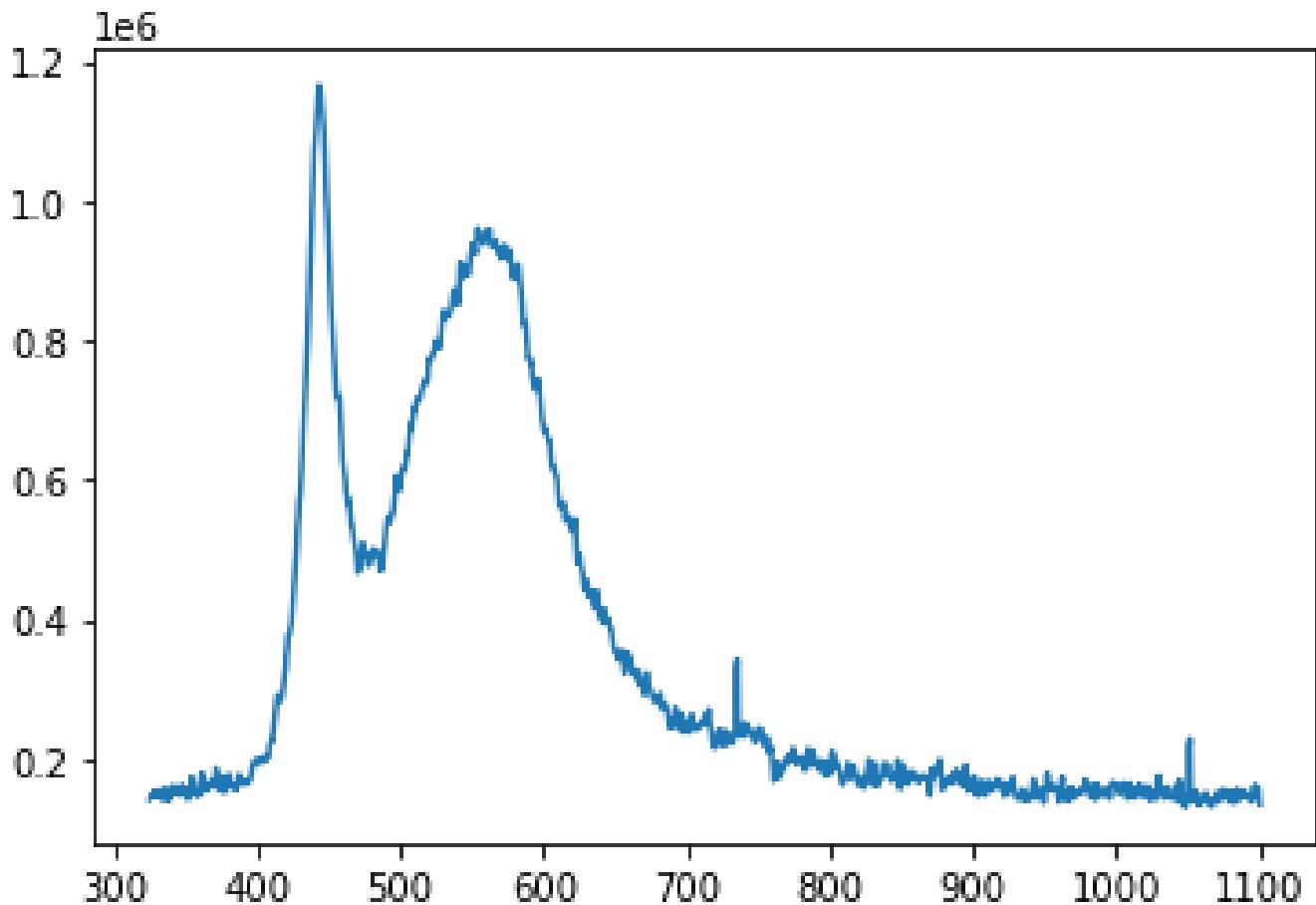


Both points are located in Pragal - Almada, Portugal. They are, however, on opposite banks of the valley created by the A2 motorway, which is one of the country's busiest roads, especially during rush hour.



In addition to the telescopes, we will use a collimated light source pointing from one telescope to the other. An artificial light source will give us a baseline with which to compare our target data, the passive DOAS trace gas concentrations.

The selected light source is a bright 20W XP50.2 LED lamp collimated through the telescope located at the sanctuary. The relative spectrum of this white lamp is presented in the figure below.



### Hypothesis

We hypothesize that light absorption between points A and B will be equal to the difference in absorptions in A and B and we can measure this difference with our equipment.

## Actions

| Action ID | Action  | Description   |
|-----------|---|---|
| A         | Active trace gas concentration determination  | With the two telescopes facing each other, we collect spectra for two minutes with the light source turned off and another 2 min with the light source turned on. |
| B         | Passive trace gas concentration determination | With the two telescopes aligned and approximately facing East, we collect spectra for 2 minutes.  |
| C         | Passive reference collection                  | With the two telescopes pointed upwards, we collect spectra for 2 minutes.  |

## Experiment day

When we get to the sites of the experiment, we must start by preparing the equipment for operation. This will entail:

1. Mount the tripods and the telescope on it;
2. Point the telescope in the general direction of the other experiment site;
3. Connect all fiber optics to the spectrometer and the optical adapter on the telescope;
4. Connect the spectrometer to the USB port and check AVASOFT detects it;
5. Finish pointing the device and focus telescope and take note of its geographical position (LAT LON).

## Additional Material

The experiment starts in the dead of the night and goes on for 6 hours. Bring:

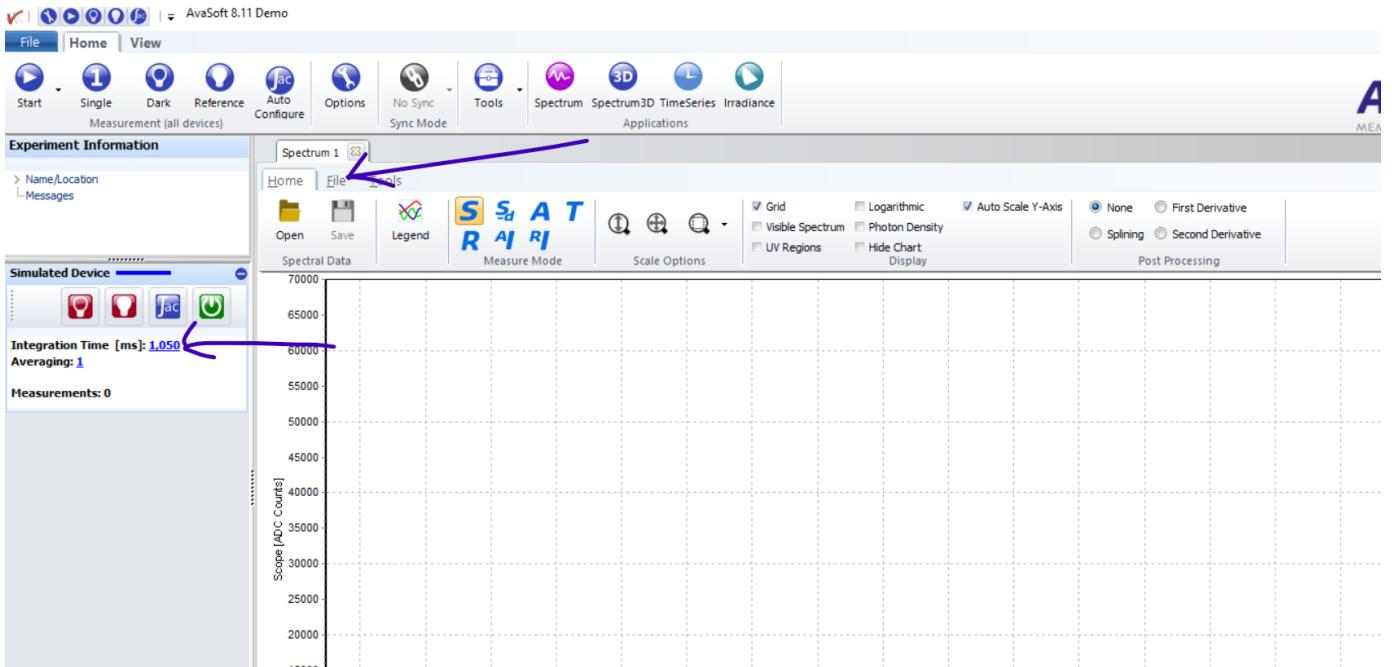
- warm clothes;
- food;
- water;
- something to pass the time (most of the time there will be nothing to do for the experiment);

The following table presents the list of prescribed actions and their periodicity. Note that only the IEP telescope moves (natural light exclusive).

| Time Frame      | Period     | Action  |
|-----------------|------------|---------|
| 05:00 - Sunrise | 15 minutes | A       |
| Sunrise         | Once       | C       |
| Sunrise - 11:00 | 15 min     | A and B |

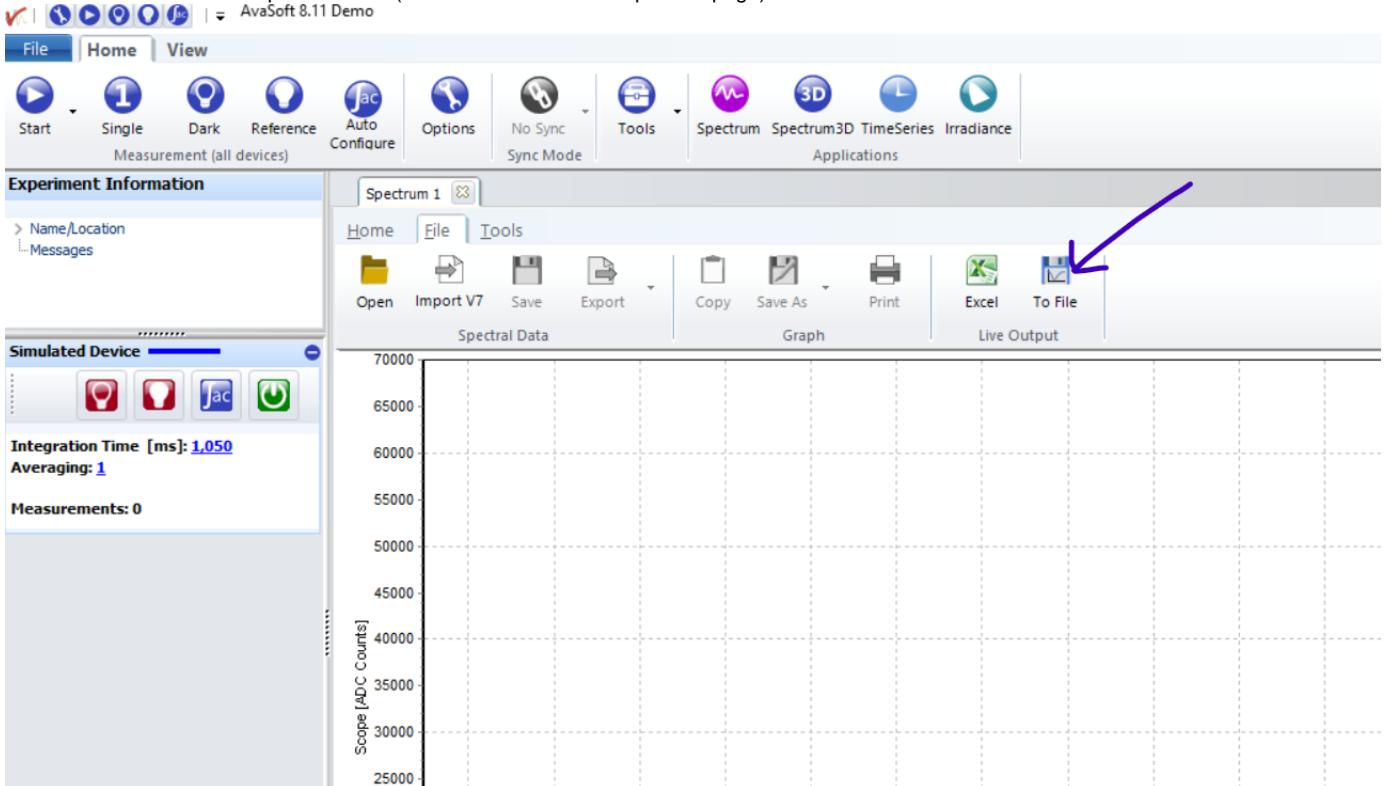
## Spectrometer Software Settings

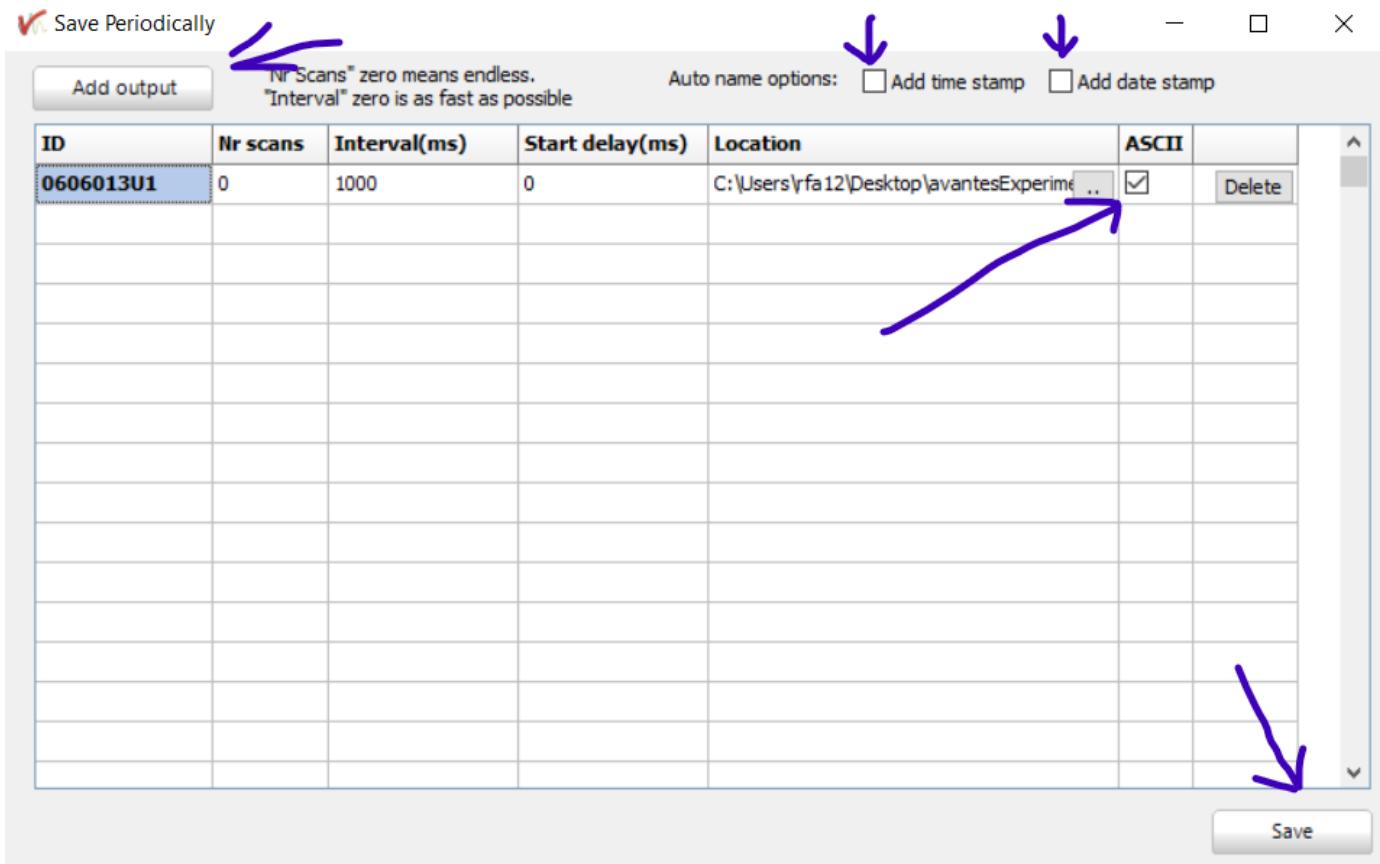
|                          |   |
|--------------------------|---|
| Calibration Coefficients | From spectrometer's memory (no action required) |
| Integration Time         | 500 ms  |
| Number of measurements   | -1 (non stop)                                   |
| Reference Spectrum       | None  |



## File saving settings

At the start of the experiment, setup file saving in the AvaSoft software to save a file to disk each second from the spectrometer output. Check that both time and datestamps are active (two checkboxes in the top of the page).





## Results

### 1st experiment run

|                  |              |
|------------------|--------------|
| Experiment start | 07 Jun 2021  |
| Conclusion       | INCONCLUSIVE |

The first attempt at the experiment was run on Jun 7 2021. All weather conditions were optimal for the kind of data acquisition we were conducting. The experiment started with a small delay, caused by last minute conditions due to not having a laptop battery charger (on the sanctuary side). This lack of a charger came to not influence the results, since it was the other laptop (on IEP's roof) that failed first, closing the experiment for this day.

The acquisition table follows:

| # | Passive meas. | Passive ref. | Active mean. | Active ref. |
|---|---------------|--------------|--------------|-------------|
| 1 | 06:20         | 06:23        | 06:15        | --          |
| 2 | 06:46         | 06:50        | 06:41        | --          |
| 3 | 07:04         | 07:07        | 07:00        | --          |
| 4 | 07:25         | 07:27        | 07:21        | --          |
| 5 | 07:48         | 07:50        | 07:45        | --          |
| 6 | 08:10         | 08:12        | 08:06        | 08:08       |

On the one hand, it is clear from this table that we do not have data on the active reference spectra (likely due to some operation error on my part); on the other hand, we have too few data points to make any claim of validation or refutation of the principle we were to measure.

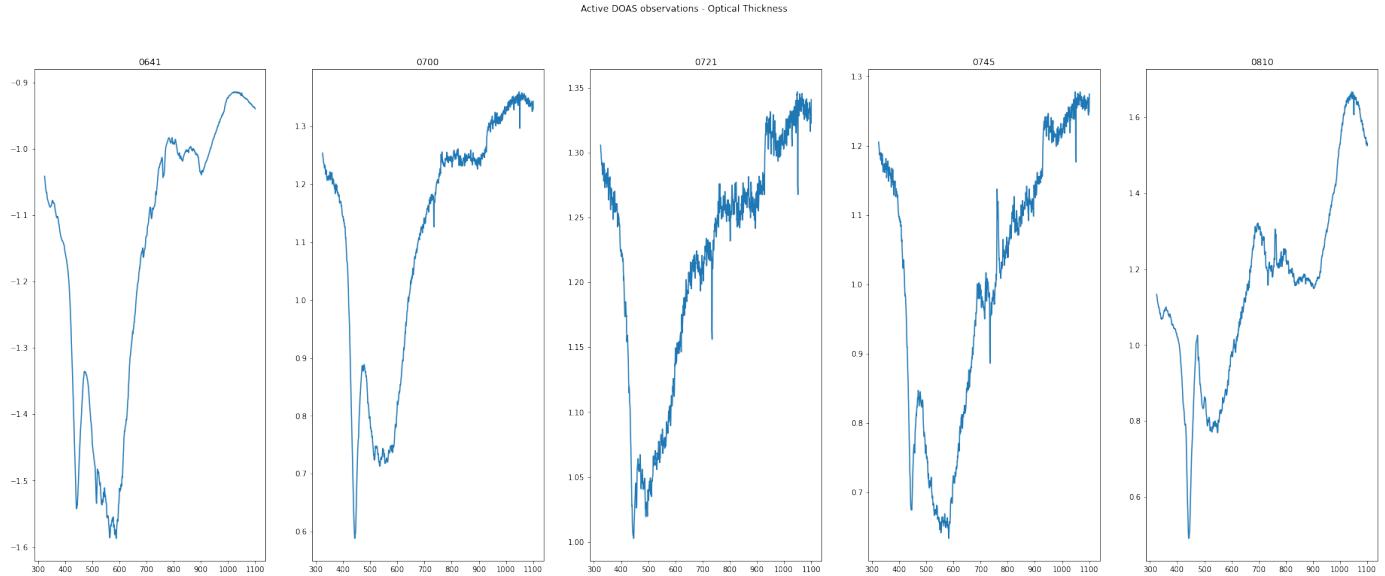
The acquired data makes it clear that we need to take another go at the experiment, from the top. Logistically, this is already being prepared (UPS and authorizations).

## Preliminary results - data analysis

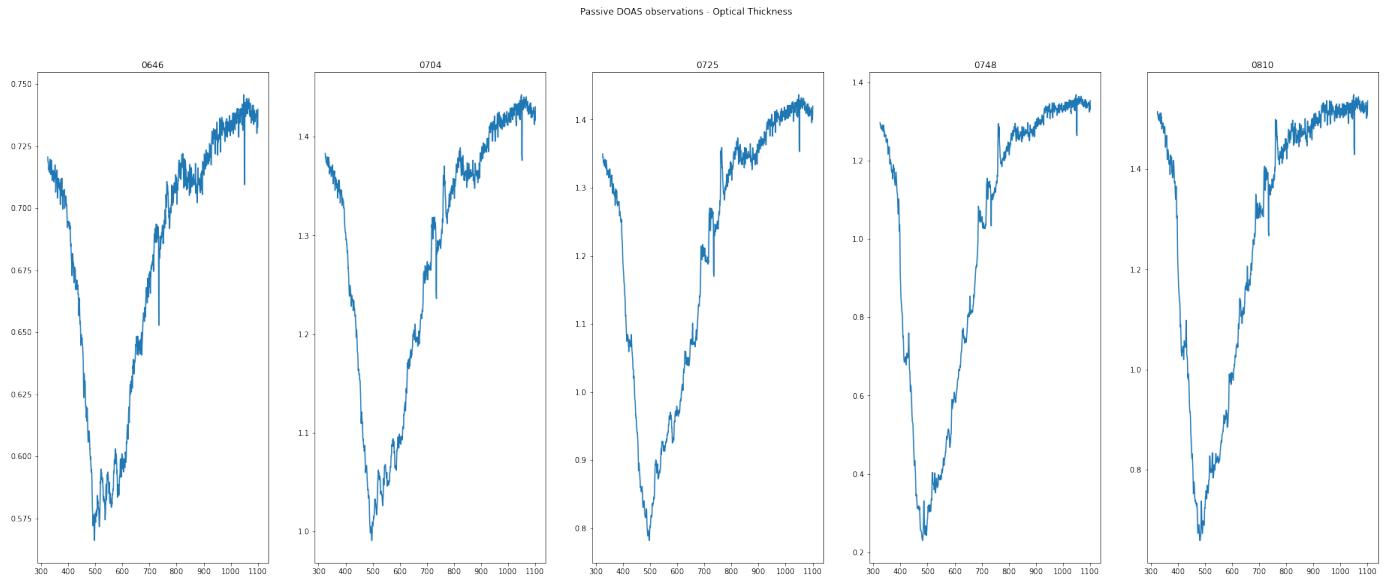
Although there is insufficient data to produce a convincing conclusion at this point, the data that was collected was of good quality and allowed some interesting analysis.

The collected spectra were correctly displayed, having the expected shape. In terms of optical density, here are the charts for the produced signals:

## Active DOAS



## Passive DOAS

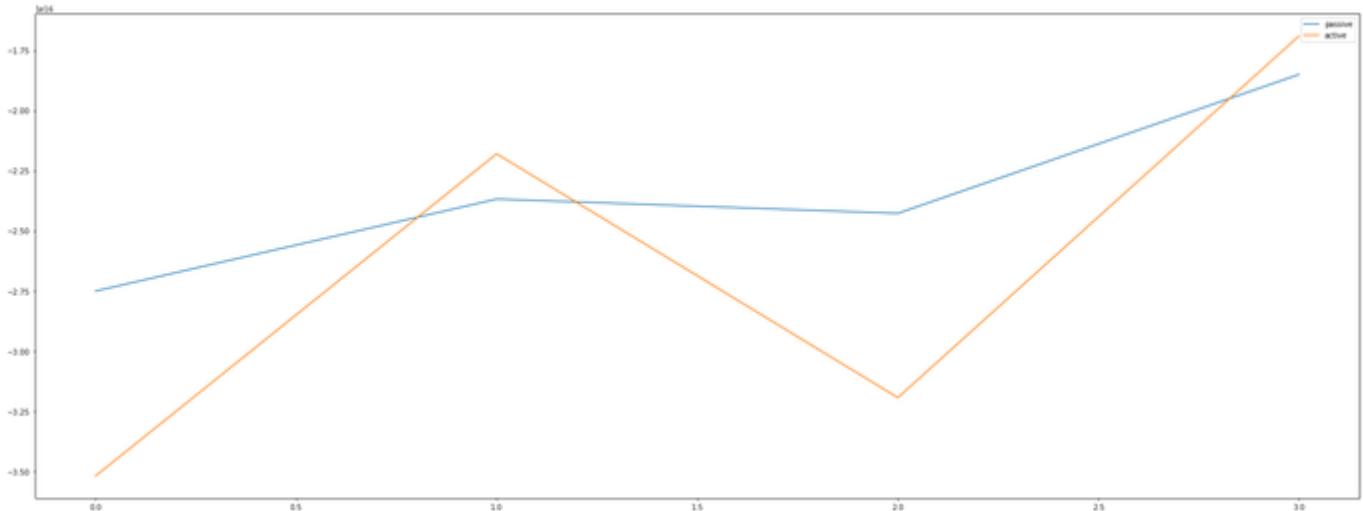


## NO<sub>2</sub> Concentration

NO<sub>2</sub> concentration results were a bit surprising, but this is probably due to a lack of reference or some contamination in the spectra, so it's not a definitive conclusion.

The first datapoint, at around 0620, is decisively the highest. This is quite surprising, given that the number of cars that were on the road at that time was minute, compared with what came next. To add to the confusion, the second datapoint is the lowest of the dataset, and after this, measurements behave as they should (rising).

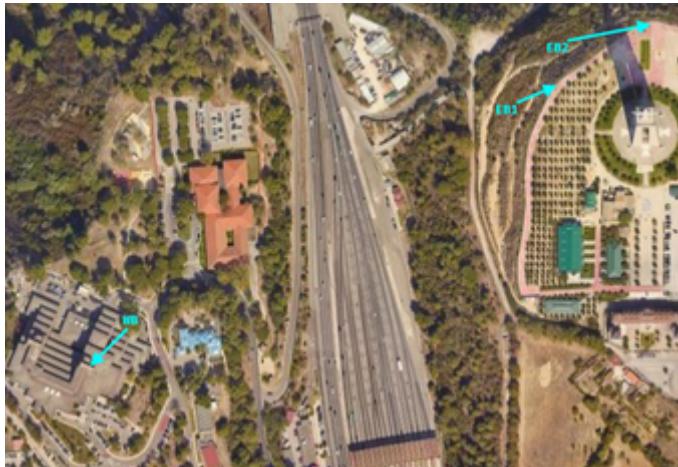
A good result was that in any case, both passive and active measurements seem to have a good agreement, indicating that our premise is valid, as it should. The following chart demonstrates this point.



## 2nd experiment run

|                         |                  |
|-------------------------|------------------|
| <b>Experiment start</b> | 22 Jun 2021      |
| <b>Conclusion</b>       | <b>PLAUSIBLE</b> |

The second run of the experiment took place on the 23rd of June. The weather was not as good as during the first run, namely because of the wind. The East Bank assembly was moved behind a few meters. This improved the measurements but made them more difficult. The figure shows how the East Bank was relocated.



We were able to make many more measurements, that are presented in the following table.

| #  | Active | Active Reference | Passive |
|----|--------|------------------|---------|
| 1  | 06:07  | 05:57            | 06:14   |
| 2  | 06:35  | 06:30            | 06:40   |
| 3  | 06:50  |                  | 06:50   |
| 4  | 07:10  | 07:15            | 07:00   |
| 5  | 07:50  | 07:40            | 07:10   |
| 6  | 08:08  | 08:03            | 07:25   |
| 7  | 08:18  | 08:21            | 07:53   |
| 8  | 08:35  | 08:30            | 08:35   |
| 9  | 09:00  | 08:53            | 09:02   |
| 10 | 09:45  | 09:50            | 09:24   |

11

10:00

10:05

09:53

DOAS analysis of the collected spectra produced the following charts:



passive\_densitie...converted-to.pdf



active\_densities...converted-to.pdf

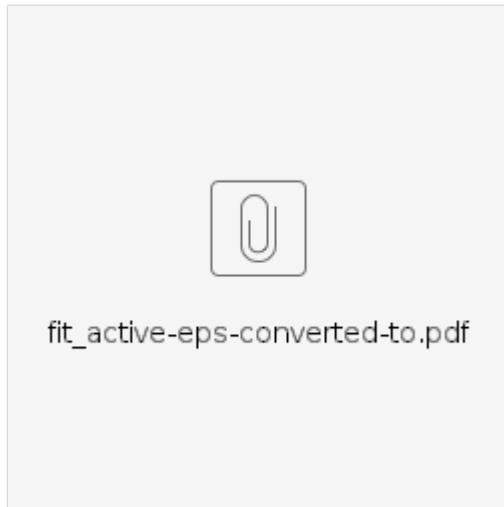
These density values were obtained from the following fits:



fit\_passive\_fron...converted-to.pdf



fit\_passive\_back...converted-to.pdf



fit\_active-eps-converted-to.pdf

## ⭐ Conclusions

- Good quality measurements produced coherent results (passive vs active) that disagree with the official numbers.

- In the field, we did see an increase in the number of cars that were passing the bridge, but we don't know if that would be enough to produce these changes.
- The fits are good, particularly the ones obtained in the West Bank, indicating that we were in fact measuring what we thought we were measuring, thus confirming the densities in the previous plots.
- The fact that this is a manual procedure makes it very difficult to ensure proper alignments, so it is possible that this error played some part in the disagreement between the two sources.
- The wind also played a role in destroying alignments.

Results do disagree with the official concentrations, but are quite coherent within themselves. The hypothesis in test can be considered plausible, but without improving the acquisition process (at least making it automatic) it is difficult to assert that it is verified.

E

## TELESCOPE - SPECTROMETER COUPLING

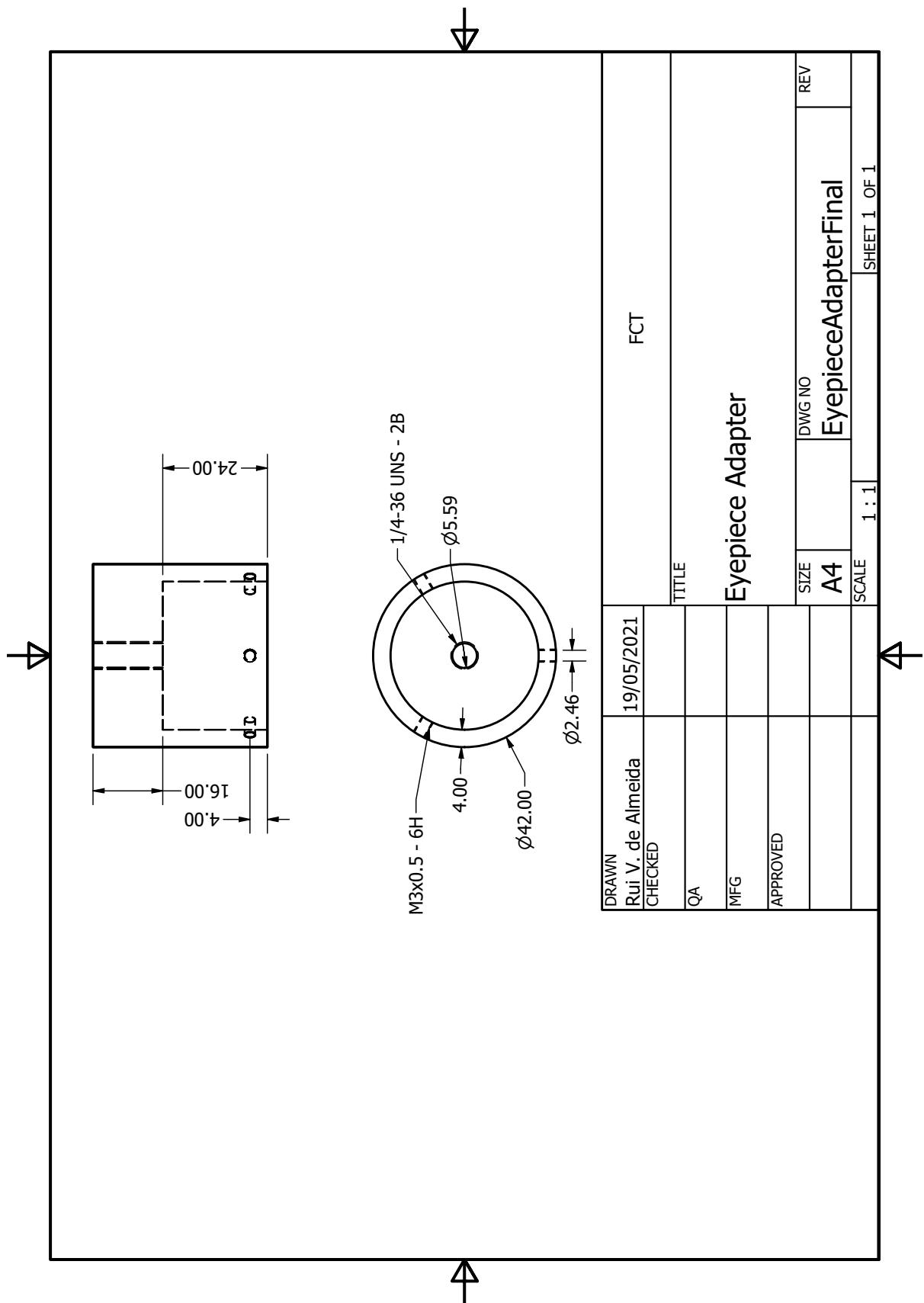


Figure E.1: Mechanical drawing for the telescope-spectrometer coupling device.

F

## REFRACTOR TELESCOPE DESIGN

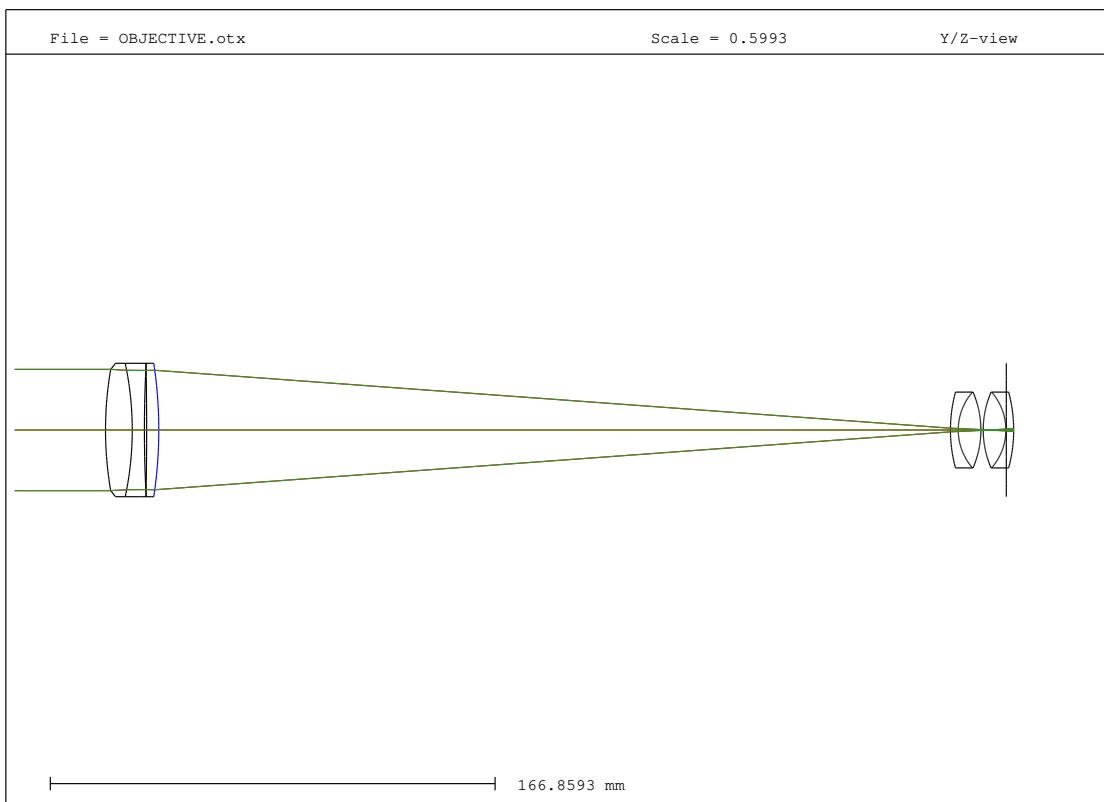


Figure F.1: Refractor telescope design: basic schematic

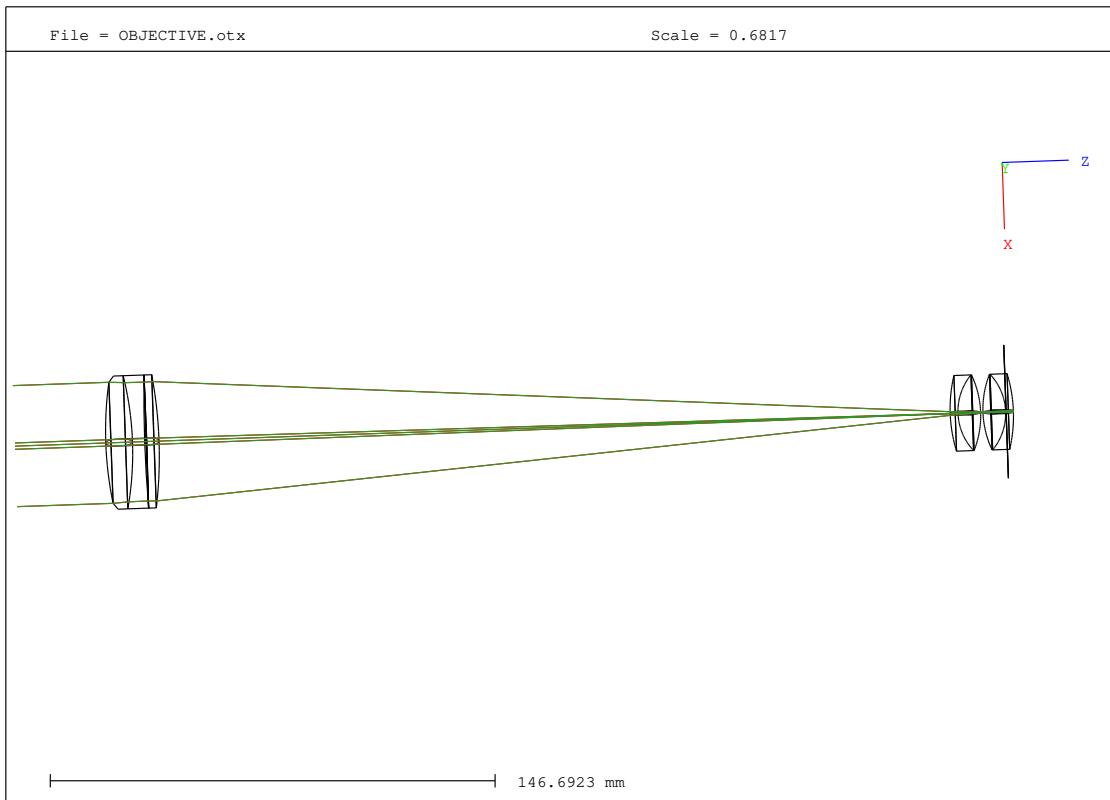


Figure F.2: 3D schematic for the refractor telescope design.

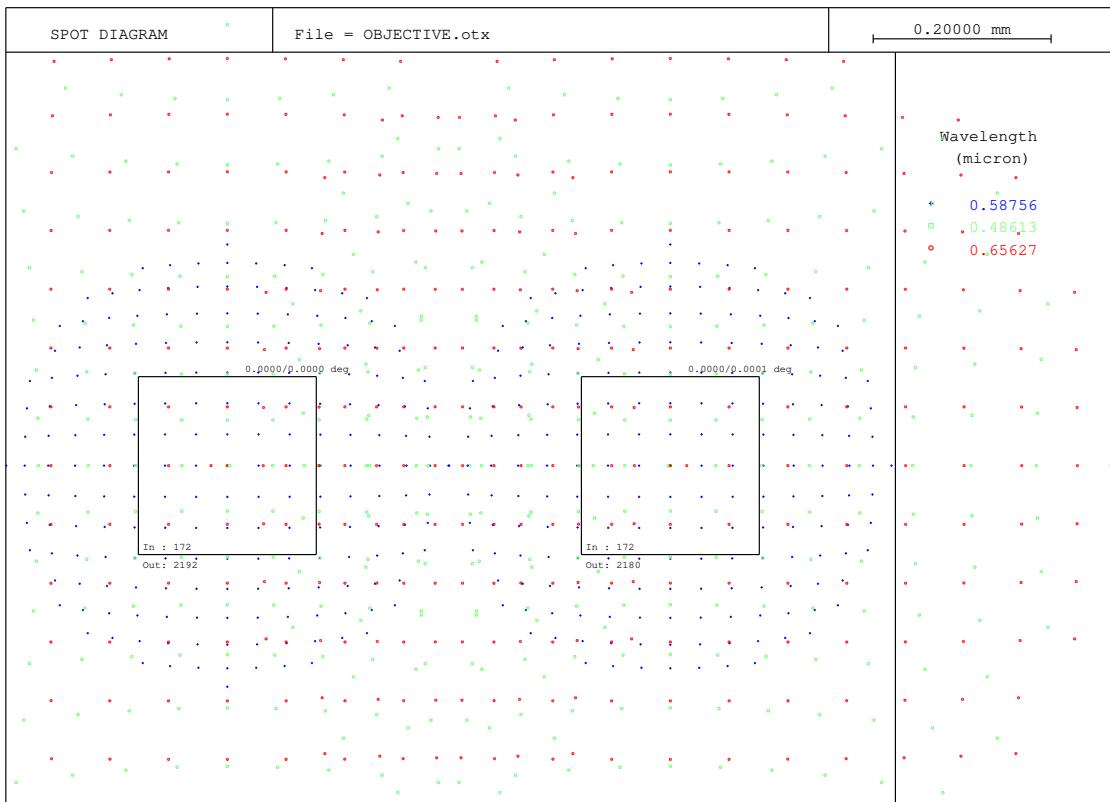


Figure F.3: Spot diagram for the refractor telescope.

## APPENDIX F. REFRACTOR TELESCOPE DESIGN

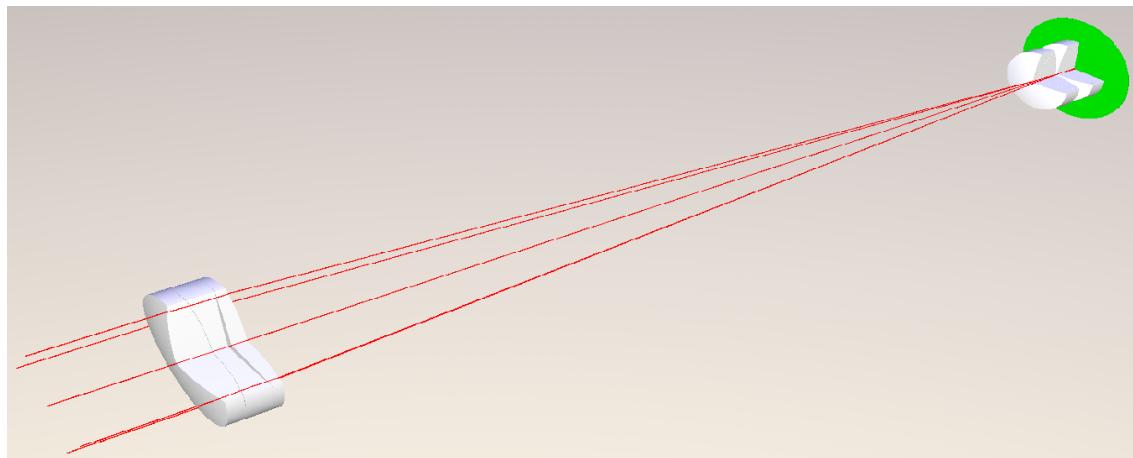


Figure F.4: 3D rendering of the optical system

# Refractor Telescope Prescription Data

## objective\_prescription.txt

FILE = OBJECTIVE.otsx 8.Dec.2021 11:24

Wavelength : 0.58756 0.48613 0.65627

Weight : 1 1 1

REF = 1

XAN 0.00000 0.00000

YAN 0.00000 0.00006

FWGT 100 100

## FACT 1

PTM = no

SYM = yes

EPD = 45 4991

| #     | TYPE | RADIUS     | DISTANCE    | GLASS  | INDEX    | APE-Y | AP | CP | DP | TP | MP | GLB |
|-------|------|------------|-------------|--------|----------|-------|----|----|----|----|----|-----|
| OBJ   | S    | 10.0000    | 0.27535E+21 |        | 1.000000 | 20.00 | C  | 0  | 0  | 0  | 0  | 0   |
| STO>S |      | 134.1587   | 10.02911    | PRI    | 1.528554 | 22.75 | C  | 0  | 0  | 0  | 0  | 0   |
| 2     | S    | -120.1177  | 4.55869     | M-FDS1 | 1.922860 | 25.02 | C  | 0  | 0  | 0  | 0  | 0   |
| 3     | S    | 481.5796   | 0.84488     |        | 1.000000 | 25.02 | C  | 0  | 0  | 0  | 0  | 0   |
| 4     | S    | -1271.7005 | 4.55869     | SF57   | 1.846663 | 25.02 | C  | 0  | 0  | 0  | 0  | 0   |
| 5     | S    | -166.6807  | 296.98319   |        | 1.000000 | 25.02 | C  | 0  | 0  | 0  | 0  | 0   |
| 6     | S    | 54.6123    | 2.92821     | SF59   | 1.952497 | 14.22 | C  | 0  | 0  | 0  | 0  | 0   |
| 7     | S    | 21.5860    | 8.49192     | SK1    | 1.610250 | 14.22 | C  | 0  | 0  | 0  | 0  | 0   |
| 8     | S    | -35.8273   | 0.87842     |        | 1.000000 | 14.22 | C  | 0  | 0  | 0  | 0  | 0   |
| 9     | S    | 35.8273    | 8.49191     | SK1    | 1.610250 | 14.22 | C  | 0  | 0  | 0  | 0  | 0   |
| 10    | S    | -21.5814   | 2.92821     | SF59   | 1.952497 | 14.22 | C  | 0  | 0  | 0  | 0  | 0   |
| 11    | S    | -54.6123   | -2.76731    |        | 1.000000 | 14.22 | C  | 0  | 0  | 0  | 0  | 0   |
| IMG   | S    | Infinity   |             |        | 1.000000 | 25.02 | C  | 0  | 0  | 0  | 0  | 0   |

FIRST ORDER DATA AT INFINITE CONCENTRATES:

EFL 300.00000 SH1 Princ.Plane 1 3292.81155  
 FNO 6.59353 SH2 Princ.Plane 2 -306.47497

---

FIRST ORDER DATA AT USED CONJUGATE:

|                    |             |                      |           |
|--------------------|-------------|----------------------|-----------|
| MAG Magnification  | 0.00000     | SEP Entr.Pup.Loc.    | 0.00000   |
| NAO Num.ape.object | 0.00000     | EPD Entr.Pup.Dia.    | 45.49912  |
| NA Num.ape.image   | 0.07583     | APD Exit Pup.Dia.    | 4.56084   |
| BFL                | -6.47497    | SAP Exit Pup.Loc.    | 23.59709  |
| DEF Defocus        | 3.70766     | PRD pupil relay dist | 364.29030 |
| IMD Image distance | -2.76731    | OAL S1->Image        | 337.92591 |
| OID Object->Image  | 0.27535E+21 | SYL System Length    | 340.69322 |

---

