Rui Filipe Cantarino Valente de Almeida

Master of Science

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

Thesis plan submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy in
Biomedical Engineering

Adviser: Pedro Vieira, Auxiliar Professor,
NOVA University of Lisbon

Contents

List of Figures

List of Tables

Acronyms

AP   Air Pollution

CEM   Continuous Emission Monitoring
COPD   Chronic Obstructive Pulmonary Disease
CT   Computed Tomography

DIAL   Differential Absorption LIDAR
DOAS   Differential Optical Absorption Spectroscopy

EEA   European Environmental Agency
EPA   Environmental Protection Agency (United States)

FBP   Filtered BackProjection
FFF   Forest Fire Finder
FST   Fourier Slice Theorem
FT   Fourier Transform

ICE   Internal Combustion Engine
IFT   Inverse Fourier Transform

ML   Machine Learning

PM   Particulate Matter

ROI   Region Of Interest
RQ   Research Question

x

SLR	Systematic Literature Review	
SMS	Systematic Mapping Study	
TDL	Tunable Diode Laser	*
VOC	Volatile Organic Compound	
WHO	World Health Organization	

Background and Motivation

1.1 Context

The idea behind this thesis was born in 2015, at NGNS-IS (a Portuguese tech startup). At the time, the company's flagship product was the Forest Fire Finder (Forest Fire Finder (FFF)). The FFF was a forest fire detection system, capable of mostly autonomous and automatic operation. The system was the first application of Differential Optical Absorption Spectroscopy Differential Optical Absorption Spectroscopy (DOAS) for fire detection, and for that it was patented in 2007 (see [56, 57]). The FFF is a remote sensing device that scans the horizon for the presence of a smoke column, sequentially performing a chemical analysis of each azimuth, using the Sun as a light source for its spectroscopic operations [54].

The FFF was deployed in several "habitats", both nationally (Parque Nacional da Peneda-Gerês and Ourém) and internationally (Spain and Brazil). One of the company's clients at the time was interested in a pollution monitoring solution, and asked if the spectroscopic system would be capable of performing such a task. The challenge resonated through the company's structure and the idea that created this thesis was born. The team then started reading about the concept of Air Pollution AP and how both populations and entities were concerned about it. It became clear that, while there were already several methods to measure AP, there was a clear market drive for the development of a system that could leverage the large area capabilities of a DOAS device while being able to provide a more spatially resolved "picture" of the atmospheric status. With this in mind, the company managed to have the investigation financed through a PT2020 funding opportunity. This achievement was a clear validation of the project's goals and of the need there was for a system with the proposed capabilities. It was, however, not enough. FFF was a very good starting point, but there was still a lot of continuous research work needed before any of the goals that had been set were achieved. This led to the publication of this PhD project, in a tripartite consortium between FCT-NOVA, NGNS-IS and the Portuguese Foundation for Science and Technology. Its main goal was to develop an atmospheric monitoring

system prototype that would be able to spectroscopically map pollutant concentrations in a two-dimensional way.

1.2 The Problem

The first step in tackling the development of the proposed system was to understand the problem it should be dealing with. Air Pollution (AP) is one of the most present concerns of people around the industrialized modern world. In Europe, it is perceived to be the second most important threat to the environment. The first is climate change, which is great part caused by AP. Scientists in many countries have established it as a major cause for premature death, disease onset and hospital visits for some decades now. Regulatory bodies of many countries have been gathered to put some legislative pressure on industries and on society itself, in order to produce a decrease in the amount of Air Pollution to which people are exposed. These policies and measures have had a dramatic influence in air quality, which is very significantly improved throughout the years. In spite of this, official reports continue to highlight the importance of keeping a vigilant eye towards AP proliferation and its possible undiscovered health effects. A better introduction to the subject of AP is produced in Section 2.1.

Our spectacular progress as a species in the last few centuries has had some unforeseen adverse consequences. Mitigating them is not only a responsibility, but also a necessity. In some regards, as is the AP case, this mitigation is only achievable with intelligent and effective action. This of course demands that we understand, trace and measure it in as many ways as possible. This project aimed to create just that: another way in which to measure and map the behavior of certain pollutants.

1.3 Objectives

From the beginning of the project, the main objective has always been to design and develop a miniaturized spectroscopic environmental analysis device. The system would have to be small and portable enough to adapt to a drone, but should be able to function if adapted to any other surface, such as a car or even a fixed station. With time, however, the goal changed somewhat. The miniaturization and drone-adaptation were kept, but there was now the need to be able to map the pollutant concentration along a geographic region through the use of tomographic reconstruction algorithms. This meant that, while the device itself would be perfectly capable of operating without a drone, the tomographic capability would be lost, unless a prohibitively large number of spectroscopic systems were deployed in that region.

The concept evolved in such a way that the overarching goal of this thesis is now to theorize and design a two-dimensional mapping tool for trace atmospheric pollutants such as NO_x and SO_x, using passive DOAS as the measurement technique. In addition to this, it should be mentioned that the system must be small and portable; use

tomographic reconstruction (and acquisition) algorithms to map the region; and that it should use only one collection point (to reduce costs and instrumental complexity). From these objectives, the research questions, detailed in Section 2.2 were derived.

Research Question

2.1 Problem Introduction

AP is a very important topic of discussion in the current days, with scientists and researchers around the globe being very well aware of the potential effects it can have on the health of individuals and populations across all ecosystems [34]. After climate change (one of the largest capital threats to life on Earth, perhaps just behind nuclear apocalypse), AP is the biggest environmental concern for Europeans, and Europe's single largest environmental health hazard [13]. It has also been established by many authors as a major cause of premature death, cardiopulmonary disease onset and hospital visits [12, 13, 16, 59].

The growing concerns about Air Pollution and its effects on human health and the world in general is an indication for the importance of measuring it correctly and with great detail. The diversity of its effects and the sheer number of variables that need to be considered establish the problem of AP as one that can only be effectively tackled if approached intelligently, highlighting the need for smart devices for the measuring and monitoring atmospheric pollutants.

At the moment, there are several solutions that can be implemented for measuring atmospheric pollution. However, these solutions are limited in their application to either large area coverage and small details (DOAS ground stations, satellite data) or very detailed information in a very localized and fixed way (in-situ electrochemical sensors). To the best of our knowledge, there are no available systems that can monitor and map pollutant concentration without requiring infrastructure installation and capable of performing at various altitudes.

The answer to this lack could be a highly mobile tomographic DOAS system, that could bridge the gap between the local monitoring capabilities of in-situ electrochemical sensors and large-area spectroscopic ground stations, while maintaining portability and flexibility.

2.2 Research Question

In Chapter 1, I have introduced the reasons which led NGNS-IS to pursue the development of an atmospheric monitoring system, and that what set it apart from other systems was the ability to spectroscopically map pollutants concentrations using tomographic methods, thus defining a primary objective for this thesis.

Two secondary objectives were born from the necessary initial research, which had a very heavy influence over the adopted methods:

- To use a tomographic approach for the mapping procedure;

- To ensure the designed system would be small and highly mobile;

- To use a single light collection point, minimizing material costs.

Taking all the above into account, we arrive at the main Research Question (Research Question (RQ)), presented in Table 2.1.

Table 2.1: Main research question.

RQ1	*How to design a miniaturized tomographic atmosphere monitoring system based on DOAS?*

This is the main research question. It gave rise to four other more detailed research questions. These secondary questions allow a better delimitation of the work at hand and are important complements to RQ1. This questions are presented in Table 2.2

Table 2.2: Secondary research questions.

RQ1.1	*What would be the best strategy for the system to cover a small geographic region?*
RQ1.2	*What would be the necessary components for such a system?*
RQ1.3	*How will the system acquire the data?*
RQ1.4	*What should the tomographic reconstruction look like and how to perform it?*

2.3 Hypothesis and Approach

This work is based on the hypothesis that a system such as the one described in Chapter 1, which responds to the RQ in Table 2.1 and Table 2.2 can be achieved by careful selection of mathematical tomographic algorithms and instrumentation that is able to implement them correctly.

The first step in answering the entirety of the research questions should be to answer RQ1.1. In fact, it is not possible to make any other decision before this matter is settled. As with any technical problem, there are several ways to create a tomographic atmospheric monitoring tool. However, each and every one of them implies some kind of compromise, which determines the system's capabilities and requirements. Will the system use retro-reflection? Shall it move during the measurement? These are the kind of questions that determine the whole project.

When the measurement strategy is determined, one could start picking parts and components. However, a better first approach would be designing a software simulation. This simulator must include all major system features, so that it correctly mimics reality and is therefore able to mathematically validate the acquisition and reconstruction approach. The results obtained from the simulation will then dictate mechanical and control requirements.

One other aspect that needs addressing is the optical section. As mentioned before, the system will be inspired in FFF's basic optical capabilities. However, the smoke detector was not conceived with spatial restrictions in mind. This important set of components will thus need redesigning, so that it is in line with the size objectives of the new system.

Literature Review

The work in this thesis involved dealing with many subjects. Air Pollution, tomographic algorithms and DOAS (in particular DOAS Tomography) were the most important. In this section, I provide a brief literary review covering the three.

3.1 Air pollution and pollutants

Daniel Vallero, in his book "Fundamentals of Air Pollution" [55] makes a very important observation: Air Pollution has no universal definition. Its meaning is intertwined with the context with which it is measured and observed, with the ecosystem in which it is perceived and even with the pollutant concentration (not every toxic compound is toxic at every concentration). The United States Environmental Protection Agency (Environmental Protection Agency (United States) (EPA)) defines Air Pollution as the following:

> *Air Pollution is the presence of contaminants or pollutant substances*
> *in the air that interfere with human health or welfare, or produce other*
> *harmful environmental effects.*

He then analyzes this definition through two possible lenses, the one that comes with the interference produced by air contaminants; and the one that comes from the harm they may cause. He notes that both points of view come with a heavy burden of ambiguity, incompatible with a scientific definition. We can thus observe that preferable to address the issue through its measurable effects and consequences. These are well-established and well known, and scientists all around the world have been publishing extensively about them for some decades now. The correlation between Air Pollution and an increased mortality in heavily industrialized areas was first established in Europe, in the 19th century, but the first time it was taken seriously was during the 1952 killer-smog incidents, in London [44]. At the time, a combination of very cold weather, an anticyclone and fireplace emissions caused a thick smog to fall over London, directly causing thousands of deaths and indirectly many more [1, 39].

The disastrous consequences of this incident had a huge impact in the civil society, resulting in a series of policies and laws, among which the Clean Air Acts of 1956 and 1968.

More than 60 years have passed since the London-smog incident, but the influence it had on the whole Western world has prevailed in time. Almost every developed country in the world has laws and protocols in place to limit and decrease AP emissions over time. In the European Union, the European Environmental Agency (European Environmental Agency (EEA)) is responsible for providing sound and independent information to the ones involved developing, adopting, implementing and evaluating environmental policy [14]. The agency is also responsible for publishing reports on the state of air quality in Europe, which are made available to the general public. In their most recent report [13], one could see that, in general, rules and regulations put in place to control atmospheric emissions have caused a clear decreasing trend to appear on AP data (see Figure 3.1), although European productivity has been rising or stable across all sectors.

Reports like this one are very important in the current day Western world. In fact, Europeans have deemed Air Pollution to be their second largest environmental concern, after climate change [53], with which it is inexorably intertwined. The population's sensitivity to this subject is undoubtedly related to the growing knowledge that exists regarding human health implications of AP. It has been known for some time that there is a direct correlation between the emission level and premature mortality, especially due to cardiovascular complications [7, 16, 30, 32]. In Europe, poor air quality is an important cause of premature death, with a estimated toll of 400000 per year [13].

A comprehensive description of the effects of Air Pollution on the human and animal bodies would be a colossal task, much beyond the scope of this document or my whole thesis. It is, however, important to mention some of these effects, not only for demonstration purposes, but also as an introductory approach to the physiological importance of AP, which justify their social and societal significance as a public health menace. I will therefore address the phenomenon impact on the respiratory system, its importance as a precursor for cardiovascular diseases and its potential impact on the most vulnerable period of human life, gestation. The following discussion will once again be based around Vallero's work [55], complemented whenever necessary by other authors.

The respiratory system's main functions are the delivery of oxygen into the blood stream and the removal of carbon dioxide from the body. Air enters the body from the upper airways and flows to the alveolar region, where oxygen diffuses across the lung wall into the blood stream, from which it is transported to the tissues where it diffuses yet again and is made available to the mitochondria in the cells, that use it for cellular respiration [37]. The whole system is in permanent interaction with the atmosphere, and is therefore exposed to all kinds of air pollutants and trace gases, increasing the
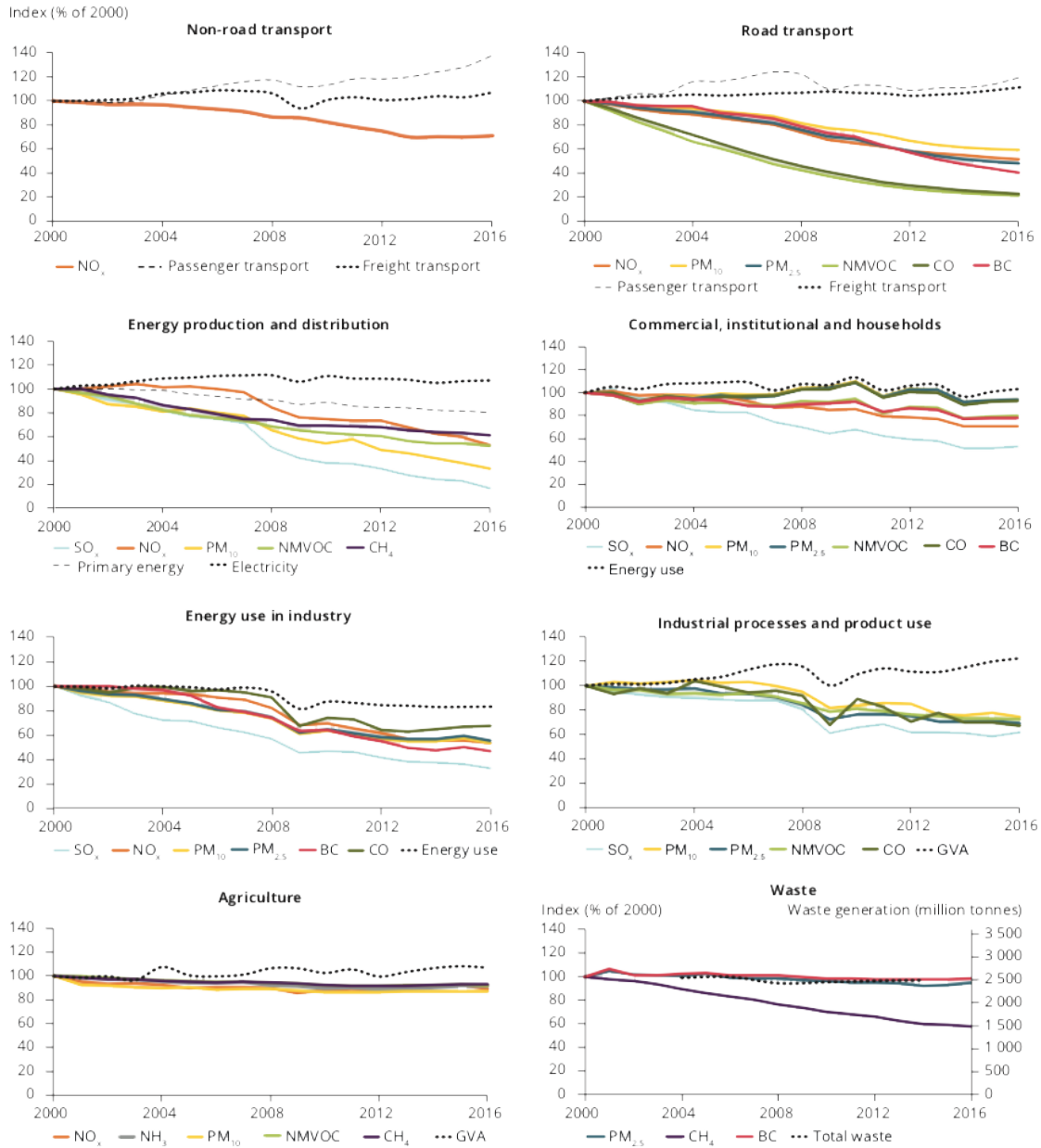
Figure 3.1: EU emissions evolution through time, since the year 2000, separated by source of emissions [13]. Notice the downward trend for emissions across all sources.

probability that some of them are delivered to this system.

Acute symptoms of AP exposure are very varied, and range from mild irritation to complete respiratory failure, depending mostly on level of exposure and individual sensitivity to the chemical compound. On a chronic level, AP has been established as cause for Chronic Obstructive Pulmonary Disease (Chronic Obstructive Pulmonary Disease (COPD)), asthma, and lung and other cancers. In addition, by entering the body through the respiratory system, air pollutants and toxins are conveyed onto the other tissues, extending the range of their damage. The next system to be affected by AP is the cardiovascular system, which is the mechanism through which the oxygen, absorbed by the respiratory system, gets transported into all the other tissues. Oxygen is not the only molecule that is able to traverse the lung's wall and therefore it is possible for several air pollutants to contaminate the body through the same paths. The link between Air Pollution and cardiovascular effects started being made during the twentieth century, given a series of incidents (like London's 1952 killer-smog) that happened in the urban areas of industrialized countries. Nowadays, this link is perfectly established and we have alredy described several mechanisms by which AP is able to interfere with cardiovascular health. For instance, scientists and doctors have managed to establish at least three pathways between AP and cardiac arrhythmia, as depicted in Figure 3.2. One of the most important cardiovascular complications that arise from Air Pollution is CO poisoning. Carbon Monoxide is a tasteless, odorless and colorless gas that is a side product from an incomplete combustion of fuels containing carbon atoms. Its properties make it impossible to detect without some kind of instrument, since our senses cannot distinguish its presence. CO can easily enter the blood stream through the lungs. When in the body, it binds with hemoglobin. Its affinity to this transport molecule is about 200 times higher than oxygen's, and this means that it starts to starve the body organs from that important molecule. Continuous exposure to Carbon Monoxide is never beneficial. Low exposure levels can cause neurobehavioral or developmental effects, but at higher concentrations, the most prominent symptoms are unconsciousness and death [41].

Mammals are in their life's most vulnerable stage while they are still developing inside their mother's womb. This is the time when there is a greater rate of tissue expansion and creation, which gives rise to the possibility of the appearance of some kind of morphological abnormality. This rate of development creates an enormous need for nutrients, provided by the mother through her blood. If the mother is exposed to a pollutant, and it gains access to the mother's blood stream, then the developing being is also exposed. The effects that come from *in utero* exposure are similar to those we see in other stages of life, but the dose at which they occur is significantly lower. Moreover, timing of exposure is also important, since there are stages to the fetus' development. Figure 3.3, taken from [55], illustrates what kind of developmental effects may come from Air Pollution exposure, and the influence of their timing.
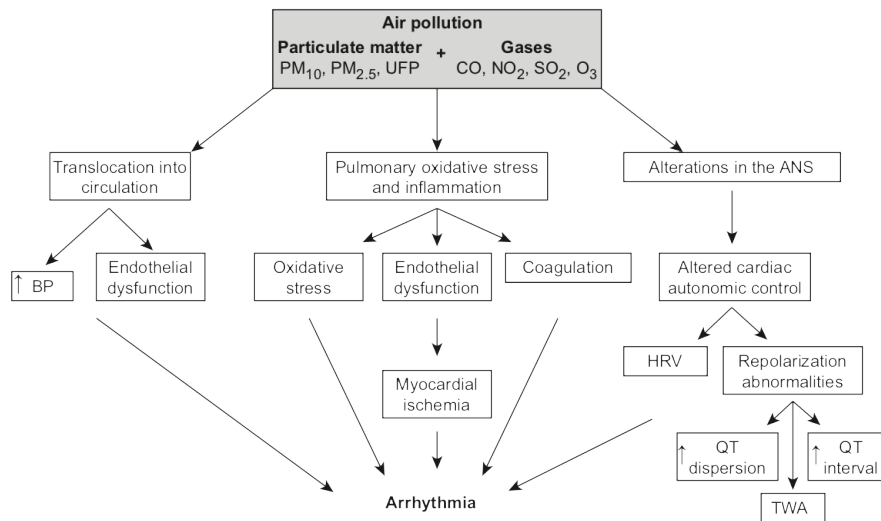
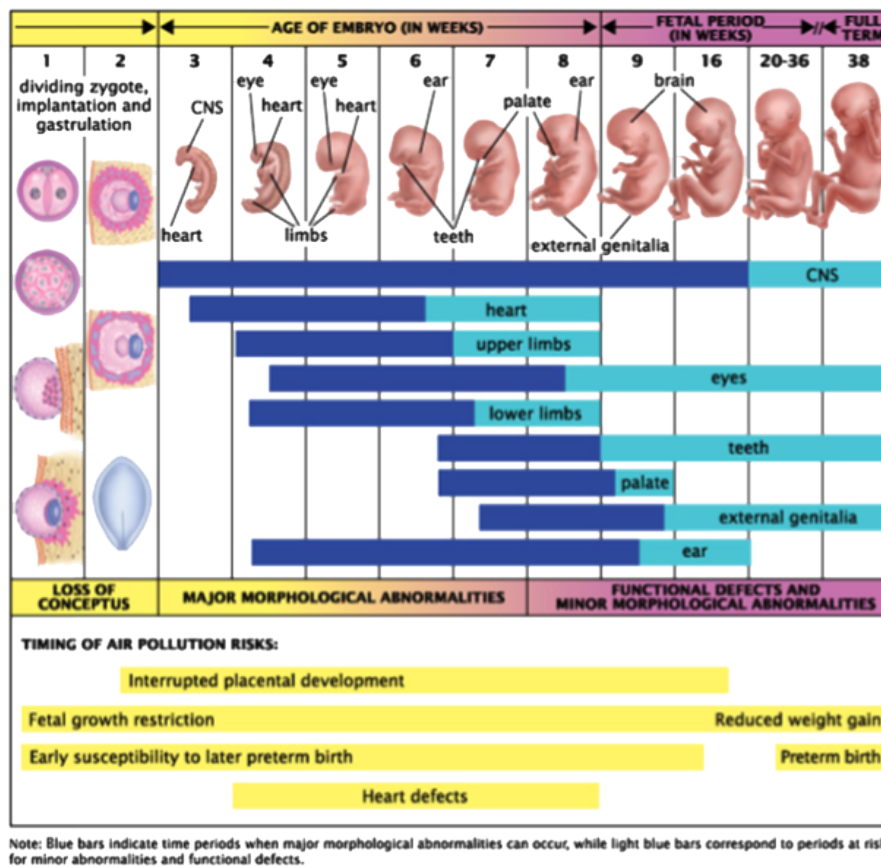Figure 3.2: Physiological pathways to arrhythmia that stem from Air Pollution exposure [55].

Figure 3.3: Possible abnormalities caused by AP exposure *in utero*. Notice that time of exposure is of critical importance [55].

The seriousness of the effects these chemicals produce create the need for quantification methods, so we can understand how we can deal with our surrounding atmosphere. The variability of this our environment, however, prevents the existence of any one-size-fits-all solution. As a result, there are many techniques and methods with which we are nowadays able to measure pollution and airborne gaseous components. In the following discussion, I will address only open path monitoring systems, since these are the most relevant for the kind of project my thesis encompasses. A more curious reader is redirected to [9, 50, 55] for a more in depth review and presentation of emission monitoring technologies.

Open path atmospheric monitoring systems exploit interaction phenomena between light and the gaseous pollutants to detect and measure pollutant concentration. One of these technologies is Tunable Diode Laser (Tunable Diode Laser (TDL)), a technique in which light is emitted at a given wavelength (particularly in the infrared region of the spectrum) and its atmospheric absorption measured on the other end of the assembly. TDL has been used extensively for methane and NO measurements, and provides robust results in real-time, being an adequate solution for Continuous Emission Monitoring (Continuous Emission Monitoring (CEM)). One of the greater disadvantages of this technique is that, since only one wavelength is measured at a time, one has to have a light source for each target compound [2, 55]. Another one of these techniques is called Differential Absorption LIDAR (Differential Absorption LIDAR (DIAL)). In this method, light is pulsed in two different wavelengths, which must be close to one another. The first of these wavelengths corresponds to an absorption band of the substance in study, while the other is just off the absorption peak. The ratio between the two received signals (one from each emitted wavelengths) is proportional to the target pollutant concentration [50]. DIAL technology allows three-dimensional mapping of pollutant concentrations in a very flexible and versatile way. It is many times used for mapping pollutant concentration in urban settings. Its main drawbacks are the fact that it is limited to a number of components and that DIAL signals are hard to analyze and can be contaminated by parasite chemical compounds, which decrease the system's sensitivity to the intended pollutants. Finally, we reach the most important open path monitoring technology in this thesis, DOAS, which will be introduced in the next section, with much more detail.

3.2 DOAS

Differential Optical Absorption Spectroscopy is a well established absorption technique that is widely used in the field of atmospheric studies [44]. In this section, I present a short introduction to the field, extracted from [54], an article we have published in 2017, marking the conclusion of the initial studies for this PhD thesis.

There are two main families of DOAS assemblies, with different goals and capabilities:
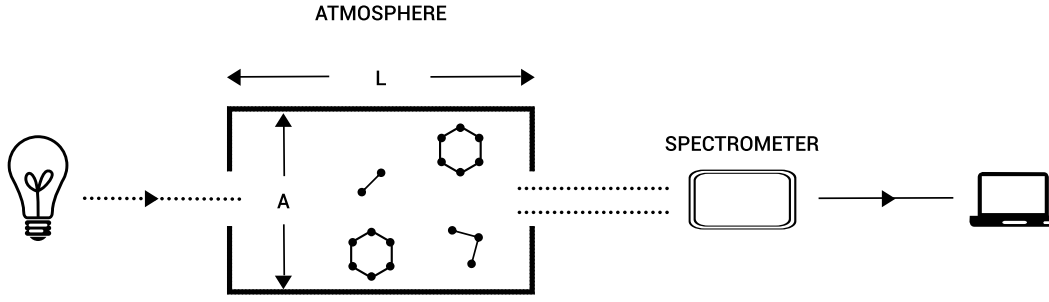
Figure 3.4: Active DOAS schematic.
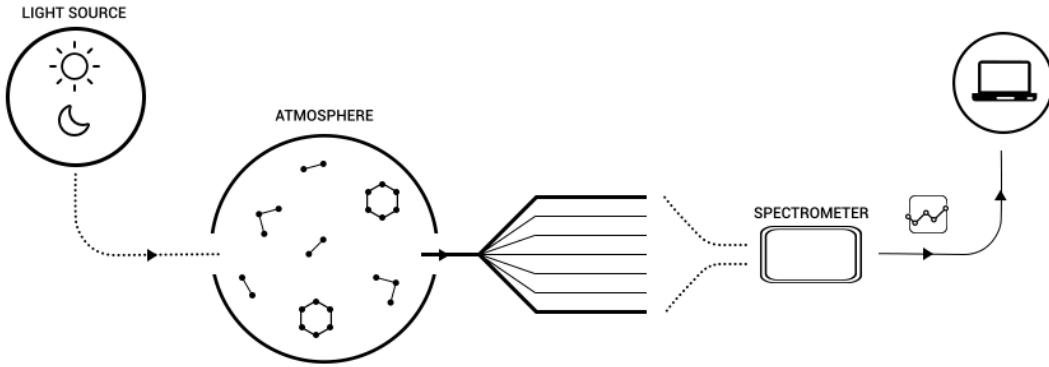
Figure 3.5: Passive DOAS schematic.

- Active systems, of which a simple illustration is presented in Fig. 3.4, are characterized 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 [44];

- Passive DOAS techniques, illustrated in Fig. 3.5, 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 [35, 44].

DOAS itself is based on Lambert–Beer's law, which can be written as [44]

$$I(\lambda) = I_0(\lambda) \cdot \exp(-\sigma(\lambda) \cdot c \cdot L) \,, \tag{3.1}$$

Where λ 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 (τ) [44]:

$$\tau(\lambda) = \ln\left(\frac{I_0(\lambda)}{I(\lambda)}\right) = \sigma(\lambda) \cdot c \cdot L. \tag{3.2}$$

In a laboratory setting, Eq. (3.1) or (3.2) 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 [44].

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

- 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. (3.3).

$$
\begin{aligned}
I(\lambda) = {}& I_0(\lambda) \cdot A(\lambda, \ldots) \cdot S(\lambda) \\
& \cdot \exp\left[-\int \left[\left(\sum_i \sigma_i(\lambda, s) \cdot c_i(s) \right) + \epsilon_{\mathrm{M}}(\lambda, s) \right.\right. \\
& \left.\left. + \epsilon_{\mathrm{R}}(\lambda, s) \right] \mathrm{d}s \right],
\end{aligned}
\tag{3.3}
$$

Where $A(\lambda, \ldots)$ 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, $\epsilon_{\mathrm{R}}(\lambda)$ represents Rayleigh's extinction coefficient and $\epsilon_{\mathrm{M}}(\lambda)$ 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 [35].

$$SC_i = \int c_i(s)\mathrm{d}s \tag{3.4}$$

This quantity, as Eq. (3.4) 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. (3.2). Instead of absolute densities, this will yield relative changes in the atmosphere. We thus arrive at Eq. (3.5).

$$\begin{aligned} \ln\left(\frac{I_{\mathrm{ref}}}{I}(\lambda)\right) = \ln\left(\frac{A_{\mathrm{ref}}}{A}(\lambda, \dots)\right) &+ \ln\left(\frac{S_{\mathrm{ref}}}{S}(\lambda)\right) \\ &+ \sum_i (\sigma_i(\lambda) \cdot \Delta SC_i(\lambda)) + \Delta\tau_{\mathrm{M}}(\lambda) \\ &+ \Delta\tau_{\mathrm{R}}(\lambda), \end{aligned} \tag{3.5}$$

Where ΔSC_i is the relative slant column of absorber i; $\Delta\tau_{\mathrm{M}}$ is the relative Mie scattering term, integrated to its optical thickness; and $\Delta\tau_{\mathrm{R}}$ 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. (3.6), we can separate these contributions [10].

$$\sigma(\lambda) = \sigma'(\lambda) + \sigma_0(\lambda) \tag{3.6}$$

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. (3.7).

$$\ln\left(\frac{I_{\mathrm{ref}}}{I}(\lambda)\right) = \sum_{i=1}^{n} \sigma_i'(\lambda) \cdot \Delta SC_i + \sum_{j=0}^{m} a_j \cdot \lambda^j, \tag{3.7}$$

Where $\sum_{i=1}^{n} \sigma_i'(\lambda) \cdot \Delta SC_i$ is the differential part (narrowband, 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. (3.7) 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 [4, 35]. 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 [19].

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. (3.7) in its matrix form:

$$\tau = \mathbf{A} \cdot X. \tag{3.8}$$

$\mathbf{A}$ 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 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 $\chi^2 = [\tau - A \cdot X] \cdot [\tau - A \cdot X]^T$.

While the Ring effect is treated as a pseudo-absorber, a synthetically produced [8] 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. (3.9) [35, 46]:

$$\ln \left(\frac{I_{\text{ref}}(\lambda)}{I(\lambda + \text{shift}) + \text{offset}} \right) = \sum_{i=1}^{n} \sigma_i{}'(\lambda) \cdot \Delta\text{SC}_i$$
$$+ \sum_{j=0}^{m} a_j \cdot \lambda^j, \tag{3.9}$$

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

3.3 Tomographic algorithms and reconstruction techniques

3.3.1 Introduction

Tomography is the cross-sectional imaging of an object through the use of transmitted or reflected waves, captured by the object exposure to the waves from a set of known angles. It has many different applications in science, industry, and most prominently, medicine. Since the invention of the Computed Tomography (Computed Tomography

(CT)) machine in 1972, by Hounsfield [20], tomographic imaging techniques have had a revolutionary impact, allowing doctors to see inside their patients, without having to subject them to more invasive procedures [28].

Mathematical basis for tomography were set by Johannes Radon in 1917. At the time, he postulated that it is possible to represent a function written in $\mathbb{R}$ in the space of straight lines, $\mathbb{L}$ through the function's line integrals. A line integral is an integral in which the function that is being integrated is evaluated along a curved path, a line. In the tomographic case, these line integrals represent a measurement on a ray that traverses the Region Of Interest (Region Of Interest (ROI)). Each set of line integrals, characterized by an incidence angle, is called a projection (see Figure 3.6). To perform a tomographic reconstruction, the machine must take many projections around the object. To the set of projections arranged in matrix form by detector and projection angle, we call sinogram. All reconstruction methods, analytical and iterative, revolve around going from reality to sinogram to image [5, 11, 24–26, 28].

Figure 3.6: A schematic representation of a projection acquisition. In this image, taken from [25], the clear line that comes down at a diagonal angle is a projection.

There are two broad algorithm families when it comes to tomographic reconstruction, regarding the physics of the problem. The problem can involve either non-diffracting sources (light travels in straight lines), such as the X-Rays in a conventional CT exam; or diffracting sources, such as micro-waves or ultrasound in more research-oriented applications [28]. In this document, I will not address the latter family, since I will not be applying them in my work. In the next few paragraphs, I will discuss the first family of algorithms, and describe how an image can be reconstructed from an object's projections when the radiation source is non-diffracting.

Let's consider the case in which we deal with a single ray of solar light entering the atmosphere at a given point. Since the atmosphere contains numerous absorbents

and comparable atmospheric effects, the ray changes from the point where it enters the atmosphere to the point at which it is measured by a detector. Total absorption will depend on the pollutant species, their cross-section and their concentration, since it obeys Lambert-Beer's law. Looking from another angle, this absorption is also the line integral that we will use to reconstruct our image. With DOAS, it is possible to measure several pollutants at the same time, but for simplicity (and since it is one of the most studied compounds in the field), let's consider that the single pollutant in our atmospheric mixture is NO_2.

3.3.2 Initial Considerations

The problem of tomographic reconstruction can be approached in a number of ways, depending mostly on the authors. In my literary search, I have found that Kak and Slaney [28] have certainly explained this problem in one of the clearer ways available. Therefore, I shall base the rest of my presentation in their writings, and complement with other authors' notes wherever necessary.

Considering the coordinate system displayed in Figure 3.7. In this schematic, the object is represented by the function $f(x, y)$. The (θ, t) parameters can be used to define any line in this schematic. Line AB in particular can be written:

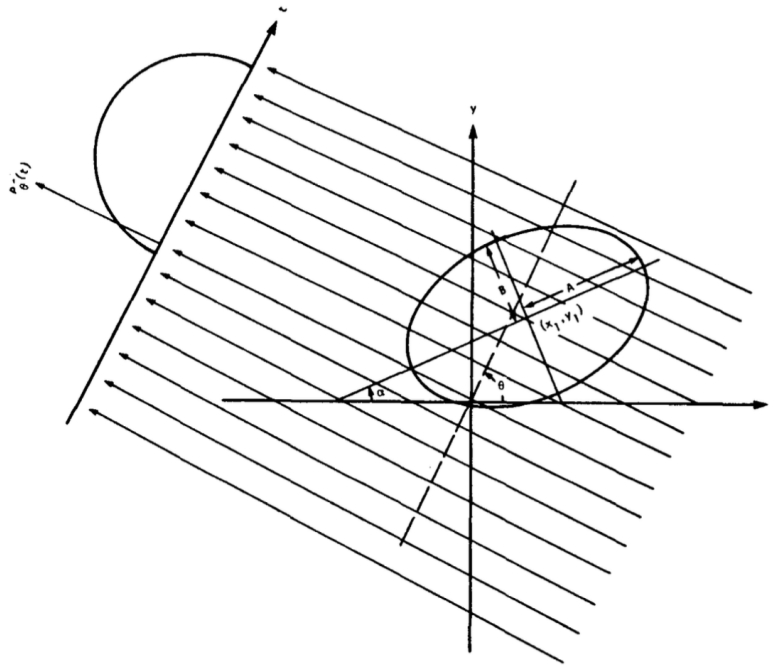$$x \cdot \cos(\theta) + y \cdot \sin(\theta) = t \tag{3.10}$$

Figure 3.7: Schematic representation for coordinate setting.

And if we were to write a line integral along this line, it would look like Equation 3.11, the Radon transform of function $f(x, y)$:

$$P_\theta(t) = \int_{-\infty}^{\infty} f(x, y) \cdot \delta(x \cdot \cos(\theta) + y \cdot \sin(\theta) - t) dx dy \tag{3.11}$$

Where δ, the delta function, is defined in Equation 3.12.

$$\delta(\phi) = \begin{cases} 1, & \phi = 0 \\ 0, & otherwise \end{cases} \tag{3.12}$$

As I have mentioned previously, a projection is a set of line integrals such as $P_\theta(t)$. Geometry plays a very important role in how the integrals are written and solved for reconstruction. The simplest case is the one where the set is acquired in a row, describing what is called a parallel geometry. Another more complex case is when a single point source is used as origin for all rays, forming a fan. This is called a fanbeam array. There are other possible geometries, but they fall out of the scope of this work and will therefore not be addressed any further.

3.3.3 The Fourier Slice Theorem

The Fourier Slice Theorem (FST) is the most important component of the most important algorithm in tomographic inversion, the Filtered BackProjection algorithm (Filtered BackProjection (FBP)). FST is based on the equality relation between the two-dimensional Fourier Transform (Fourier Transform (FT)) of the object function and the one-dimensional FT of the object's projection at an angle θ. Let's start by writing the 2D FT for the object function, Equation 3.13, and the 1D FT of projection P_θ, in Equation 3.14.

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot \exp\left[-j2\pi(ux + vy)\right] dx dy \tag{3.13}$$

$$S_\theta(\omega) = \int_{-\infty}^{\infty} P_\theta \cdot \exp\left[-j2\pi\omega t\right] \tag{3.14}$$

For simplicity, let's consider the 2D FT at the line defined by $v = 0$ in the frequency domain. We rewrite the 2D FT integral as:

$$F(u, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot \exp\left[-j2\pi\omega ux\right] dx dy \tag{3.15}$$

Notice that y is not present in the phase factor of the FT expression anymore, and this means we can rearrange the integral as:

$$F(u, 0) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \mathbf{f(x, y)dy}\right] \cdot \exp\left[-j2\pi\omega ux\right] dx \tag{3.16}$$

Now, the **bold** part of Equation 3.16 is similar to Equation 3.11. It is precisely that equation, considering $\theta = 0$ and a constant value of x, as in Equation 3.17.

$$P_{\theta=0}(x) = \int_{-\infty}^{\infty} f(x, y) dy \tag{3.17}$$

This in turn can be substituted in Equation 3.16, finally arriving at:

$$F(u, 0) = \int_{-\infty}^{\infty} P_{\theta=0}(x) \cdot \exp\left[-j2\pi ux\right] dx \tag{3.18}$$

And this is the one-dimensional FT for the projection at angle $\theta = 0$. Finally, the enunciation of the Fourier Slice Theorem:

> *The Fourier Transform of a parallel projection of an image $f(x, y)$ taken at angle θ gives a slice of the two-dimensional Fourier Transform, $F(u, v)$, subtending an angle θ with the u-axis (see Figure 3.8)*
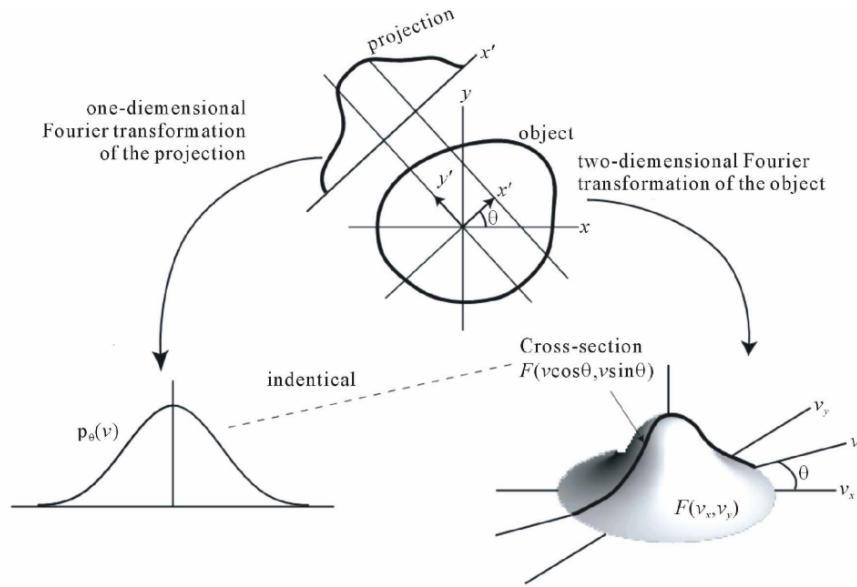
Figure 3.8: The FST, a schematic representation.

3.3.4 The Filtered BackProjection Algorithm

3.3.4.1 The rationale for FBP

If one takes the FST into account, the idea behind the FBP seems to appear almost naturally. Say one has a single projection and its Fourier transform. From the FST, this projection is the same as the object's two-dimensional FT in a single line. A crude reconstruction of the original object would result if someone were to place this projection in its right place in the Fourier domain and then perform a two-dimensional Inverse Fourier Transform (IFT), while assuming every other projection to be 0. The

result, in the image space, would be as if someone had smeared the object in the projections direction.

What is really needed for a correct reconstruction is to do this many times, with many projections. This brings a problem with the method: smearing the object in all directions will clearly produce a wrong *accumulation* in the center of the image, since every projection passes through the middle (remember we are still talking about parallel geometry projections) and are summed on top of each other, but on the outer edges, this does not occur. If one does not address this, the image intensity levels in the reconstructed image will be severely overestimated in the center and underestimated in the edges (due to normalization). The solution is conceptually easy: we multiply the Fourier transform by a weighting filter proportional to its frequency (ω) and that encompasses its relevance in the global scheme of projections. If there are K projections, then it is adequate for this value to be $\frac{2\pi|\omega|}{K}$. As an algorithm, FBP can be written as in Algorithm 1.

Algorithm 1 The Filtered BackProjection Algorithm

---

for all $\theta, \theta \in \left\{0..180, \frac{180}{K}\right\}$ **do**
  Measure projection $P_\theta(t)$;
  FT($P_\theta(t)$), rendering $S_\theta(\omega)$
  Multiply by $\frac{2\pi|\omega|}{K}$;
  Sum the IFT of the result in the image space.
end for

3.3.4.2 Fan Projections Reconstruction

Parallel projections, in which the object is scanned linearly from multiple directions, have the advantage of having a relatively simple reconstruction scheme. However, they usually result in acquisition times which are in the order of minutes. A faster way of collecting the data is one where all radiation emanates from a single point-source, which rotates around the target object (as well as the detectors). There are two types of fan beam projections: equiangular and equally spaced. In this project, I have only worked with equiangular processes, so I will not include an explanation for equally spaced fan beam projections. The reader may find this well described (much better than I would be able to) in [28] and [26].

Consider Figure 3.9. If our projection data were acquired through a parallel ray geometry, we would be able to say that ray SA belonged to a projection $P_\theta(t)$, in which θ and t would be written:

$$\theta = \beta + \gamma \quad \textbf{and} \quad t = D \cdot \sin\gamma \tag{3.19}$$

In Equation 3.19, D is the distance between the source S and the origin O; γ is the angle of a ray within a fan and β is the angle that the source S makes with a reference axis.

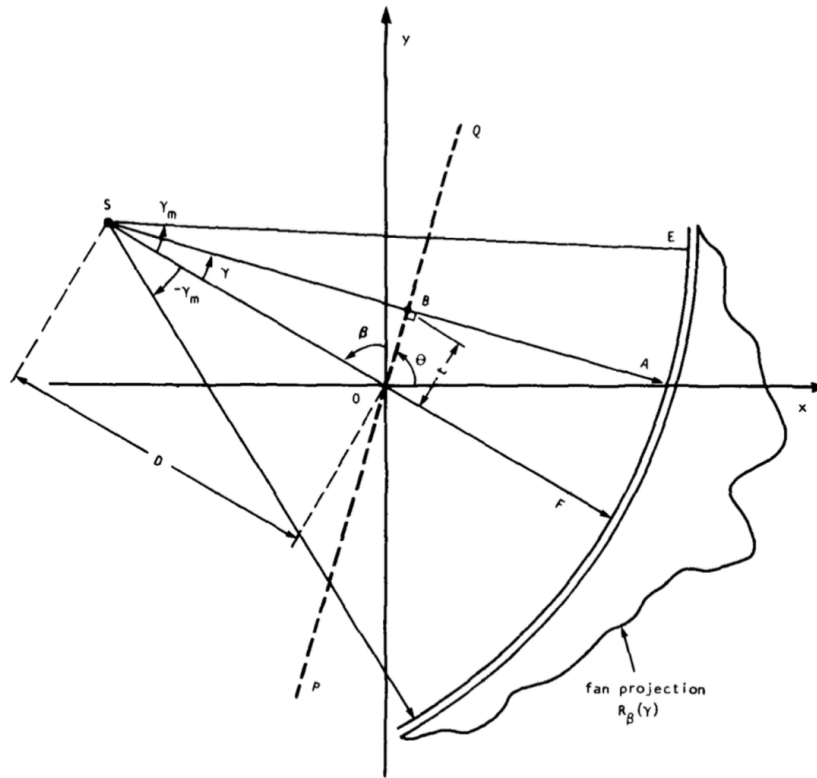Figure 3.9: Schematic representation of an equiangular fan beam projection, taken from [28].

3.4 DOAS tomography

During the course of this project, I wrote a Systematic Mapping Study (Systematic Mapping Study (SMS)) on DOAS tomography. The purpose of this study was to find the current state of affairs regarding DOAS Tomography in the literature. To avoid repetition, this article was included in full in Appendix A.

Research Methodology

4.1 Aimed contribution

This work aims to answer its main RQ by following and pursuing the proposed hypothesis (see Section 2.3). In doing so, I will contribute with a commercially viable (from a technical perspective) atmospheric monitoring system based on DOAS, capable of not only measuring pollutant contamination but also mapping their concentration in a given geographic region. This will generate several intermediate steps, which are themselves smaller contributions:

The first intermediate step will be the development of a tomographic strategy. This will of course include the selection of a sampling geometry (consider the ones presented in Section 3.3), but also the positioning of "sources and sensors" within the selected geometry. Since the system is designed to be a passive DOAS device (see Section 3.2), careful consideration must be taken in this respect. Spectral acquisition must be done in a way that allows tomographic reconstruction, but also respecting limitations and particularities that come from working with solar light. This step corresponds to answering the first secondary Research Question (see Table 2.2).

The next step will be to write a simulation tool for the tomographic procedure. The idea is for this tool to encompass the acquisition strategy that was defined in the previous step, in order to replicate the whole measurement process and then perform the reconstruction using one or more algorithms. The simulation tool has several functions. For one, it allows the fine-tuning of the approaches without having to expend any resources in purchasing material. Besides this, it also gives gives researchers the chance to experiment with different algorithms. The routine will be divided into two modules. The first module computes the projections for a given phantom, and the second performs the reconstruction with the projection data. The point is that one can write a tomographic reconstruction routine and plug it to the projection computation module, without any interference from one to the other.

Finally we reach the instrumentation definition phase. In this stage, I will have to select the optical components necessary to build this system, including telescope,

spectrometer and all the connection components. In order to maintain this system as small as possible, I will try to avoid using optical fibers, thus minimizing energy loss between telescope and spectrometer. In this stage, I will also have to design a mechanical support that allows a drone to be equipped with the optical system, respecting all positioning and pointing requirements that such a device must entail.

4.2 Detailed work plan and scheduling

In this section, I present the working plan for the proposed thesis. The complete work plan is depicted in Figure 4.1. It includes the 6 tasks, which take place in a period of 4 years of research activities and scientific writing. The order in which the tasks are performed is somewhat flexible (one could easily program the trajectory (task 5) before conducting the experiments (task 3)), but the presented sequence is the one that allows the best possible compromise between scientific results and project accomplishment, taking into account the hypothesis presented in Section 2.3.
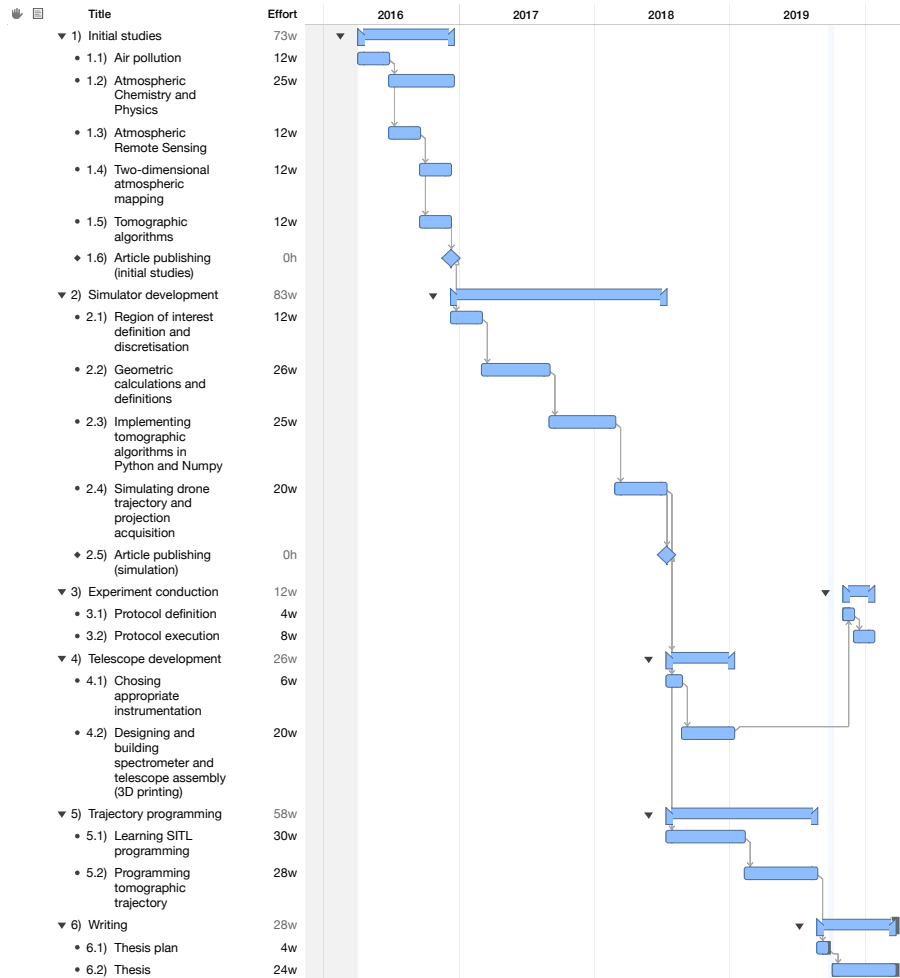

Figure 4.1: Gantt chart for the proposed thesis. Note the main milestones in 2017 and 2019, which correspond to publication moments.

Task 1 is an exploratory task, in which I started reading the literature in order to gain a good level of sensitivity to the subjects implied. The task culminated in the writing of [54] in 2017.

Task 2 encompassed the development of a simulation module for the tomographic method with which I intended to retrieve a two-dimensional concentration mapping for certain atmospheric trace gases. The module includes a precise geometric description and error estimation (using Monte Carlo methods) and intends to simulate the formation of a fan-beam measurement using a specific circular trajectory. This task should result in one publication.

The third task concerns experiment conduction. These experiments are designed to verify the acquisition method hypothesis which was simulated in Task 2. Conducting this task shall require developing the experiment protocol as well as its final execution.

The spectral acquisition strategy for the proposed project involves the development of an optical assembly without any fiber-optics. In task 4, I shall choose the optical components for this assembly and design it using 3D CAD methods. Finally, the assembly is to be manufactured: first in a 3D printer and then using traditional metal milling methods.

4.3   Validation methodology

There are several ways to validate a piece of work as the one I present here. It is important to do this in more than one, to ensure not only the validity of the project, but also its relevancy.

Before going any further, I should mention that although my project was conducted as part of the pursuit of a doctoral degree, it was not a traditional PhD, in the sense that it was not done solely in a University setting. The grant that I was awarded by the Portuguese Foundation for Science and Technology was part of a PhD Program for degrees conducted in partnership between FCT NOVA and a group of companies, named NOVA Instrumentation for Health (NOVA I4H). This is relevant as it indicates that there is at least a commercial interest in developing the technology I propose with this thesis (a market validation of sorts). Otherwise, the project would not be viable for the company.

One important way in which I will validate this project is by publishing in significant peer reviewed publications. As illustrated in Figure 4.1, during the course of my PhD, I intend to publish two relevant articles in Web-of-Science-indexed publications. The first of these was published in 2017, at the end of the preliminary studies stage of the project. The second is currently in the finishing stages of writing and should be submitted before the end of September. Conference-wise, I plan to introduce my work to the scientific community in 2020, by participating in at least one acknowledged conference in the remote sensing or atmospheric monitoring community. Table 4.1 summarizes the project's dissemination plan.

Table 4.1: Dissemination plan for the PhD Project.

4.4 Integration with other research activities

As stated in Chapter 1, this project started as a direct evolution from the Forest Fire Finder system, an autonomous forest fire detection device. In 2017, NGNS-IS (the birthplace of FFF) was absorber by the Compta group, and FFF became Bee2Fire. As an extension to the Bee2Fire, this project is integrated in Compta's research efforts, namely in the remote sensing and artificial intelligence department.

In addition to this, the present work is supported by the Portuguese Foundation for Science and Technology, through the NOVA I4H PhD Program (grant reference: PDE/BDE/114549/2016).

Bibliography

[1] M. L. Bell, D. L. Davis, and T. Fletcher. "A retrospective assessment of mortality from the london smog episode of 1952: The role of influenza and pollution." In: *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature* 6.1 (2008), pp. 263–268. DOI: 10.1007/978-0-387-73412-5_15.

[2] G. P. Bishop, J. L. E. Flack, and M. Campbell. *Sensor Systems for Environmental Monitoring, Volume 2*. 1996. ISBN: 9789401065337.

[3] N. Bobrowski, G. Hönninger, F. Lohberger, and U. Platt. "IDOAS: A new monitoring technique to study the 2D distribution of volcanic gas emissions." In: *Journal of Volcanology and Geothermal Research* 150.4 (2006), pp. 329–338. ISSN: 03770273. DOI: 10.1016/j.jvolgeores.2005.05.004.

[4] R. T. Brinkmann. "Rotational Raman scattering in planetary atmospheres." In: *Astrophys J* 154 (1968), pp. 1087–1093. ISSN: 0004-637X. DOI: 10.1086/149827.

[5] P. P. Bruyant. "Analytic and iterative reconstruction algorithms in SPECT." In: *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* 43.10 (2002), pp. 1343–58. ISSN: 0161-5505. URL: http://jnm.snmjournals.org/cgi/content/abstract/43/10/1343%7B%5C%%7D5Cnhttp://jnm.snmjournals.org/cgi/content/full/43/10/1343%7B%5C%%7D5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/12368373.

[6] R. L. Byer and L. A. Shepp. "Two-dimensional remote air-pollution monitoring via tomography." In: *Optics Letters* 4.3 (1979), pp. 75–77. ISSN: 0146-9592. DOI: 10.1364/OL.4.000075. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed%7B%5C%&%7Did=19687805%7B%5C%&%7Dretmode=ref%7B%5C%&%7Dcmd=prlinks.

[7] M. Carugno, D. Consonni, G. Randi, D. Catelan, L. Grisotto, P. A. Bertazzi, A. Biggeri, and M. Baccini. "Air pollution exposure, cause-specific deaths and hospitalizations in a highly polluted Italian region." In: *Environmental Research* 147

(May 2016), pp. 415–424. ISSN: 00139351. DOI: 10.1016/j.envres.2016.03.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0013935116300834.

[8] K. V. Chance and R. J. D. Spurr. "Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum." In: *Applied Optics* 36.21 (July 1997), p. 5224. ISSN: 0003-6935. DOI: 10.1364/AO.36.005224. URL: https://www.osapublishing.org/abstract.cfm?URI=ao-36-21-5224.

[9] A. G. Clark. *Industrial Air Pollution Monitoring*. 1997. ISBN: 9789401071437. DOI: 10.1007/978-94-009-1435-3.

[10] T. Danckaert, C. Fayt, M. Van Roozendael, I. de Smedt, V. Letocart, A. Merlaud, and G. Pinardi. *QDOAS*. 2015. URL: http://uv-vis.aeronomie.be/software/QDOAS/.

[11] M. Defrise, P. E. Kinahan, and C. J. Michel. "4 Image Reconstruction Algorithms in PET * 2D Data Organization." In: (2003), pp. 91–114.

[12] EEA. *Air pollution in Europe 1990–2004*. Tech. rep. 2. European Environment Agency, 2007.

[13] EEA. *Air quality in Europe*. Tech. rep. 28. 2016, p. 83. DOI: doi:10.2800/413142.

[14] EEA. *European Environmental Agency*. 2019. URL: https://www.eea.europa.eu (visited on 09/23/2019).

[15] T. G. Feeman. *The Mathematics of Medical Imaging*. Springer Undergraduate Texts in Mathematics and Technology. New York, NY: Springer New York, 2010. 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.

[16] A. Ghorani-Azam, B. Riahi-Zanjani, and M. Balali-Mood. "Effects of air pollution on human health and practical measures for prevention in Iran." In: *Journal of Research in Medical Sciences* 21.1 (2016), p. 65. ISSN: 1735-1995. DOI: 10.4103/1735-1995.189646. URL: http://www.jmsjournal.net/text.asp?2016/21/1/65/189646.

[17] J. V. Glowniak and I. Zak. *CT Axial Anatomy*. URL: http://www2.med.wayne.edu/diagRadiology/Limbic%20System%20RSNA%202004/brain/BrainCT.html (visited on 01/11/2019).

[18] R. Gordon. "A tutorial on art (algebraic reconstruction techniques)." In: *IEEE Transactions on Nuclear Science* 21.3 (1974), pp. 78–93. ISSN: 0018-9499. DOI: 10.1109/TNS.1974.6499238. URL: http://ieeexplore.ieee.org/document/6499238/.

[19]   J. F. GRAINGER and J. RING. "Anomalous Fraunhofer Line Profiles." In: *Nature* 193.4817 (Feb. 1962), pp. 762–762. ISSN: 0028-0836. DOI: 10.1038/193762a0. URL: http://www.nature.com/doifinder/10.1038/193762a0.

[20]   R. Gunderman. *Essential Radiology: Clinical Presentation, Pathophysiology, Imaging, 2nd ed.* 2nd ed. Thieme, 2006. ISBN: 9781588900821.

[21]   C. Haisch. "Optical Tomography." In: *Annual Review of Analytical Chemistry* 5.1 (2012), pp. 57–77. ISSN: 1936-1327. DOI: 10.1146/annurev-anchem-062011-143138. URL: http://www.annualreviews.org/doi/10.1146/annurev-anchem-062011-143138.

[22]   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." In: *Atmos. Chem. Phys* 6.3 (Mar. 2006), pp. 847–861. ISSN: 1680-7324. DOI: 10.5194/acpd-5-11781-2005. URL: http://www.atmos-chem-phys.net/6/847/2006/%20www.atmos-chem-phys.net/6/847/2006/.

[23]   A. Harzing. *Publish or Perish.* URL: http://www.harzing.com/pop.htm.

[24]   G. T. Herman. "Image Reconstruction From Projections." In: *Real-Time Imaging* 1.1 (1995), pp. 3–18. ISSN: 10772014. DOI: 10.1006/rtim.1995.1002.

[25]   G. T. Herman. *Fundamentals of Computerized Tomography.* Advances in Pattern Recognition. London: Springer London, 2009. 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.

[26]   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." In: *Journal of Theoretical Biology* 42.1 (1973). ISSN: 10958541. DOI: 10.1016/0022-5193(73)90145-8.

[27]   G. Hönninger, C. von Friedeburg, and U. Platt. "Multi Axis Differential Optical Absorption Spectroscopy (MAX-DOAS)." In: *Atmospheric Chemistry and Physics Discussions* 4 (2004), pp. 231–254. URL: www.atmos-chem-phys.org/acp/4/231/.

[28]   A. C. Kak. "Algebraic Reconstruction Algorithms." In: *Computerized Tomographic Imaging.* 2001, pp. 275–296. ISBN: 978-0-89871-494-4. DOI: 10.1137/1.9780898719277.ch7.

[29]   A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging.* 1. Society for Industrial and Applied Mathematics, Jan. 2001, pp. 49–112. ISBN: 978-0-89871-494-4. DOI: 10.1137/1.9780898719277. URL: http://epubs.siam.org/doi/book/10.1137/1.9780898719277.

[30] M. Kampa and E. Castanas. "Human health effects of air pollution." In: *Environmental Pollution* 151.2 (Jan. 2008), pp. 362–367. ISSN: 02697491. DOI: 10.1016/j.envpol.2007.06.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0269749107002849.

[31] B. Kitchenham and S. Charters. "Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3." In: *Engineering* 45.4ve (2007), p. 1051. ISSN: 00010782. DOI: 10.1145/1134285.1134500. arXiv: 1304.1186. URL: http://scholar.google.com/scholar?hl=en%7B%5C&%7DbtnG=Search%7B%5C&%7Dq=intitle:Guidelines+for+performing+Systematic+Literature+Reviews+in+Software+Engineering%7B%5C#%7D0%7B%5C%%7D5Cnhttp://www.dur.ac.uk/ebse/resources/Systematic-reviews-5-8.pdf.

[32] V. Kollanus, P. Tiittanen, J. V. Niemi, and T. Lanki. "Effects of long-range transported air pollution from vegetation fires on daily mortality and hospital admissions in the Helsinki metropolitan area, Finland." In: *Environmental Research* 151 (Nov. 2016), pp. 351–358. ISSN: 00139351. DOI: 10.1016/j.envres.2016.08.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S001393511630353X.

[33] T. Laepple, V. Knab, K.-U. Mettendorf, and I. Pundt. "Longpath DOAS tomography on a motorway exhaust gas plume: numerical studies and application to data from the BAB II campaign." In: *Atmospheric Chemistry and Physics Discussions* 4.3 (2004), pp. 2435–2484. ISSN: 1680-7375. DOI: 10.5194/acpd-4-2435-2004. URL: http://www.atmos-chem-phys-discuss.net/4/2435/2004/.

[34] G. M. Lovett, T. H. Tear, D. C. Evers, S. E. Findlay, B. J. Cosby, J. K. Dunscomb, C. T. Driscoll, and K. C. Weathers. "Effects of Air Pollution on Ecosystems and Biological Diversity in the Eastern United States." In: *Annals of the New York Academy of Sciences* 1162.1 (Apr. 2009), pp. 99–135. ISSN: 00778923. DOI: 10.1111/j.1749-6632.2009.04153.x. URL: http://doi.wiley.com/10.1111/j.1749-6632.2009.04153.x.

[35] A. Merlaud. *Development and use of compact instruments for tropospheric investigations based on optical spectroscopy from mobile platforms*. Louvain: Presses Universitaires de Louvain, 2013. ISBN: 978-2-87558-128-0. URL: http://books.google.com/books?hl=en%7B%5C&%7Dlr=%7B%5C&%7Did=inXVSyR82zwC%7B%5C&%7Doi=fnd%7B%5C&%7Dpg=PR3%7B%5C&%7Ddq=Development+and+use+of+compact+instruments+for+tropospheric+investigations+based+on+optical+spectroscopy+from+mobile+platforms+sciences%7B%5C&%7Dots=VdebeDBQMc%7B%5C&%7Dsig=4G0eVEvthJXuqd8WI3IWjjVXuXc.

[36] K. U. Mettendorf, A. Hartl, and I. Pundt. "An indoor test campaign of the tomography long path differential optical absorption spectroscopy technique."

In: *J. Environ. Monit.* 8.2 (2006), pp. 279–287. ISSN: 1464-0325. DOI: 10.1039/ B511337G. URL: http://xlink.rsc.org/?DOI=B511337G.

[37]   G. E. Nilsson. *Respiratory Physiology of Vertebrates Life with and without oxygen*. Ed. by G. E. Nilsson. Oslo, Norway: Cambridge University Press, 2010. ISBN: 9780521878548.

[38]   S. O'Driscoll, J. G. Murphy, and N. J. Smith. "Computed tomography of air pollutants in street canyons." In: *Spie* 4876 (2003), pp. 958–967. DOI: 10.1117/ 12.463985.

[39]   M. Office. *The Great Smog of 1952*. 2019. URL: https://www.metoffice. gov.uk/weather/learn-about/weather/case-studies/great-smog (visited on 09/23/2019).

[40]   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." In: *Atmospheric Environment* 150 (Feb. 2017), pp. 220–228. ISSN: 13522310. DOI: 10.1016/j.atmosenv.2016.11.058. URL: http: //linkinghub.elsevier.com/retrieve/pii/S1352231016309487.

[41]   D. G. Penney. *Carbon Monoxide Poisoning*. CRC Press.

[42]   D. Perner, D. H. Ehhalt, H. W. Pätz, U. Platt, E. P. Röth, and A. Volz. "OH - Radicals in the lower troposphere." In: *Geophysical Research Letters* 3.8 (Aug. 1976), pp. 466–468. ISSN: 00948276. DOI: 10.1029/GL003i008p00466. URL: http://doi.wiley.com/10.1029/GL003i008p00466.

[43]   K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. "Systematic Mapping Studies in Software engineering." In: *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE '08)* (2008), pp. 1–10. DOI: citeulike-article-id:3955889.

[44]   U. Platt and J. Stutz. *Differential Optical Absorption Spectroscopy*. Heidelberg, Germany: Springer, 2007. ISBN: 9783540211938.

[45]   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." In: ().

[46]   W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. 3rd Editio. 2007. ISBN: 9788578110796. arXiv: arXiv:1011.1669v3.

[47]   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." In: *Atmospheric Environment* 39.5 (2005), pp. 967–975. ISSN: 13522310. DOI: 10.1016/j.atmosenv.2004.07.035.

[48]   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 OClO from April to October 1991." In: *Journal of Geophysical Research: Atmospheres* 98.D4 (Apr. 1993), pp. 7219–7228. ISSN: 01480227. DOI: 10.1029/93JD00042. URL: http://doi.wiley.com/10.1029/93JD00042.

[49]   E. Souza, A. Moreira, and M. Goulão. "Deriving Architectural Models from Requirements Specifications: a Systematic Mapping Study." In: *Information and Software Technology* (Jan. 2019). ISSN: 09505849. DOI: 10.1016/j.infsof.2019.01.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950584919300035.

[50]   G. Stewart and M. Campbell. *Sensor Systems for Environmental Monitoring, Volume 1*. 1996. ISBN: 9789401072021. DOI: 10.1007/978-94-009-1571-8.

[51]   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." In: *Atmospheric Environment* 147 (Dec. 2016), pp. 121–132. ISSN: 13522310. DOI: 10.1016/j.atmosenv.2016.09.054. URL: http://linkinghub.elsevier.com/retrieve/pii/S1352231016307713.

[52]   E. Tacconelli. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Vol. 10. 4. 2010, p. 226. ISBN: 1900640473. DOI: 10.1016/S1473-3099(10)70065-7. arXiv: arXiv:1011.1669v3. URL: http://linkinghub.elsevier.com/retrieve/pii/S1473309910700657.

[53]   The European Union. *Eurobarometer 468*. Tech. rep. Brussels: EU, 2017. URL: https://data.europa.eu/euodp/en/data/dataset/S2156%7B%5C_%7D88%7B%5C_%7D1%7B%5C_%7D468%7B%5C_%7DENG.

[54]   R. Valente de Almeida and P. Vieira. "Forest Fire Finder – DOAS application to long-range forest fire detection." In: *Atmospheric Measurement Techniques* 10.6 (June 2017), pp. 2299–2311. ISSN: 1867-8548. DOI: 10.5194/amt-10-2299-2017. URL: https://www.atmos-meas-tech.net/10/2299/2017/.

[55]   D. Vallero. *Fundamentals of air pollution*. 5th ed. Academic Press, 2014. ISBN: 9780124017337.

[56]   P. Vieira and J. Matos. *SYSTEM FOR AUTOMATIC DETECTION OF FOREST FIRES THROUGHOPTC SPECTROSCOPY*. 2008.

[57] P. Vieira, J. Matos, and M. Mendes. *Sistema para a detecção automática de incêndios florestais por espectroscopia óptica*. 2007.

[58] "Industrial Tomography - Systems and Applications." In: *Industrial Tomography*. Ed. by M. Wang. Vol. 4. Elsevier, 2015. ISBN: 9781782421184. DOI: 10.1016/B978-1-78242-118-4.12001-5. URL: http://linkinghub.elsevier.com/retrieve/pii/B9781782421184120015.

[59] World Health Organisation Europe. "Health Aspects of Air Pollution Results from the WHO Project "Systematic Review of Health Aspects of Air Pollution in Europe"." In: *World Health* (2004), p. 30.

Systematic Review of DOAS Tomography

A.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 [42], 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 A.2 presents the context

within which this study was written; Section A.3 describes how we have planned to perform the study and the methods we have used in doing it; Section A.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 A.5 shows our conclusions and what we think might be retained from reading this paper.

A.2 Background

A.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 [44].

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 [44]:

$$I(\lambda) = I_0(\lambda) \cdot \exp[-L \cdot \sigma(\lambda) \cdot c] \tag{A.1}$$

In Equation A.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, σ is the optical cross section of the sampled element or molecule and c is the sample's concentration. λ 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 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 [35, 44].

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 [44] and [35].

In [44], 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 [42]. 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 [44].

A.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) [27], either in sequence or simultaneously, according to the schematic representation in Figure A.1.
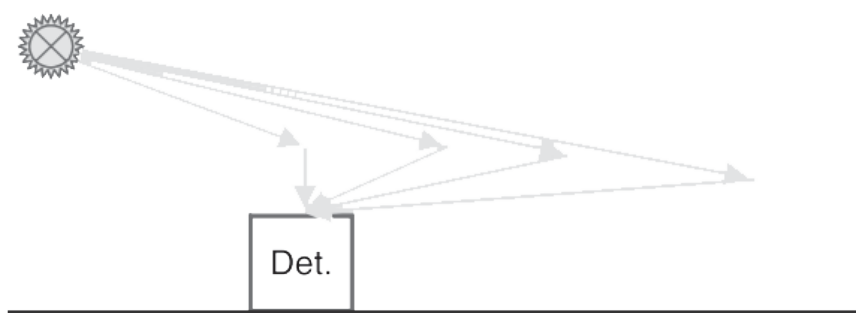
Figure A.1: MAX-DOAS schematic representation [44].

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. [48] used it to assess OClO 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 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.

A.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, as Figure A.2 illustrates.
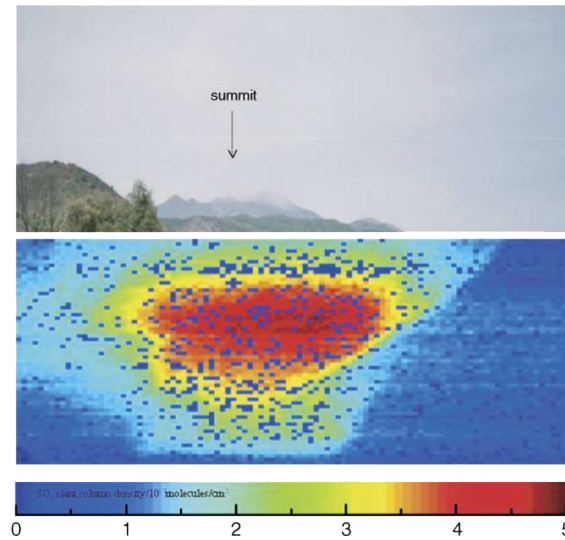
Figure A.2: IDOAS results over mount Etna, in Italy. The colour coded map under the digital photograph represents the SO_2 concentration [3].

The method, developed by Bobrowsky et al. [3], 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. A schematic representation is presented in the same article, and is here reproduced in Figure A.3.

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.

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
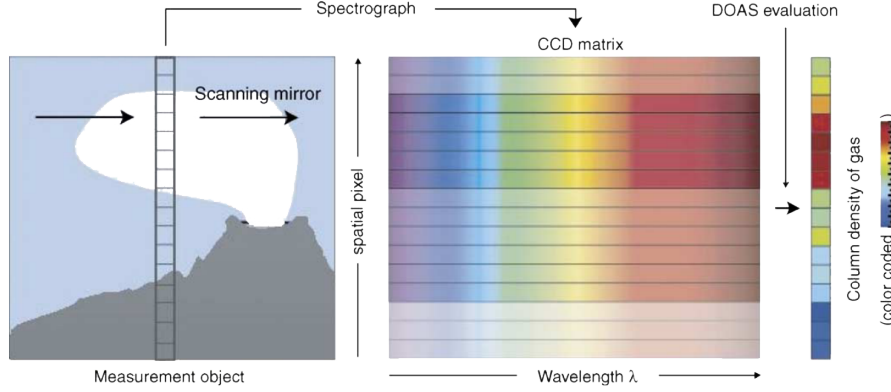
Figure A.3: IDOAS capture functioning schematic [3].

from projections, but instead comes directly from the acquisition method. Hence, we did not include articles on this method in this study.

A.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 [25, 29].

In 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 θ directions. In practice, this formulation allows the reconstruction of an image by its projections, which are nothing more than line integrals [15].

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 Feeman [15]. 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:

Algebric Reconstruction Techniques (ART) Proposed in 1970 by Gordon and Herman [26], these techniques are based on successive approximations between the actual projection data and the sum of the reconstruction elements which represent it [18]. 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 [28].

During the second half of the twentieth century, tomographic processes have had a revolutionary influence in many fields of study, but especially in medicine (see Figure A.4). 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 anihilation [15, 25, 29].
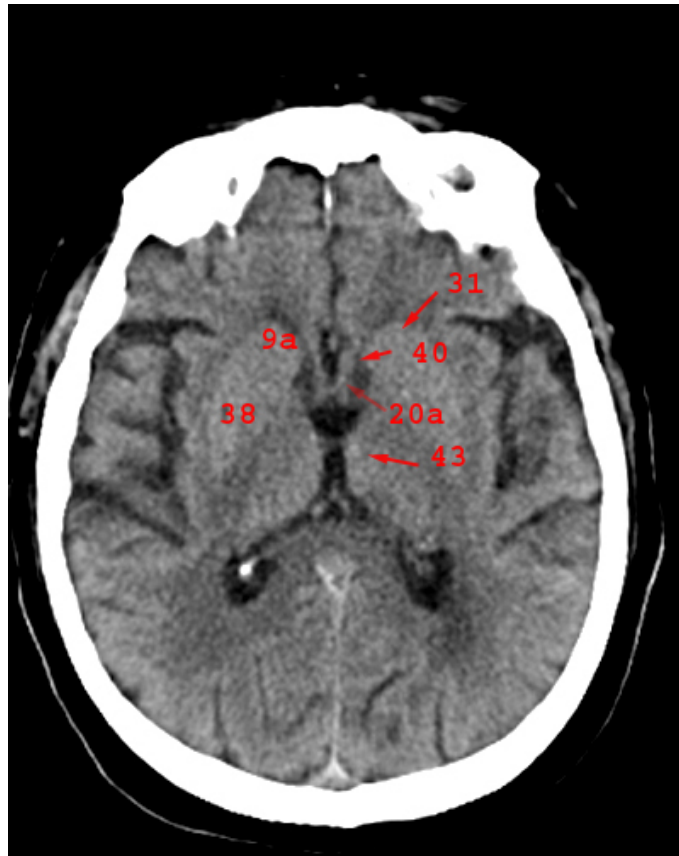
Figure A.4: Tomographic image, axial view of the human brain [17]. Note how a trained clinician can identify several structures (7 red notes) with just one slice of a computerised tomography image. Before tomographic techniques were invented, it was impossible to retrieve this this kind of detail without dissecting the patient.

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 [6, 21, 58]. 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.

A.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 [31, 43].

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 [31]. In this document, the authors establish the 3 stages which all MS and SLR generally have:

Planning This stage includes all preliminary considerations regading 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.

A.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.

A.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.

A.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 A.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 A.1: PICOC analysis.

Population	DOAS research in general.
Intervention	The papers must address tomographic DOAS.
Outcome	Status **assessment** for **DOAS tomography**.
Context	Research papers.

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 A.2.

Table A.2: Research question slicing

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

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.

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

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

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

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

After setting Table A.3 libraries, it was time to define our article selection criteria, which are summarised in Table A.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 exluding 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 A.4: Selection filters in use for this study's search.

	Criterium	Definition
Exc. Criteria	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
Inc. Criteria	IC1	Results must be articles
	IC2	Results must be about Tomographic DOAS

---

[1]The asterisk acts as a wildcard.

Table A.5: Quality assessment criteria presentation.

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

A.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).

A.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 approach [49] and adopt a similar evaluation method. Table A.5 contains the used criteria.

In our evaluation model, we took into account both general ans 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 papers 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 A.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 A.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 A.2.

$$TotalScore = \sum_i w_i \cdot S_i \tag{A.2}$$

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 descrition 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.

A.4   Conduction

A.4.1   The Search

SLR guideline literature [31, 52] recommend that the first stage of any Systematic Literature Review or Mapping Study 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 A.3.2, were run in all libraries found in Subsection A.3.3. The Google Scholar search had the particularity of being run through a specialised software called *Publish or Perish*[23], 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 A.5. 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 A.3.3.
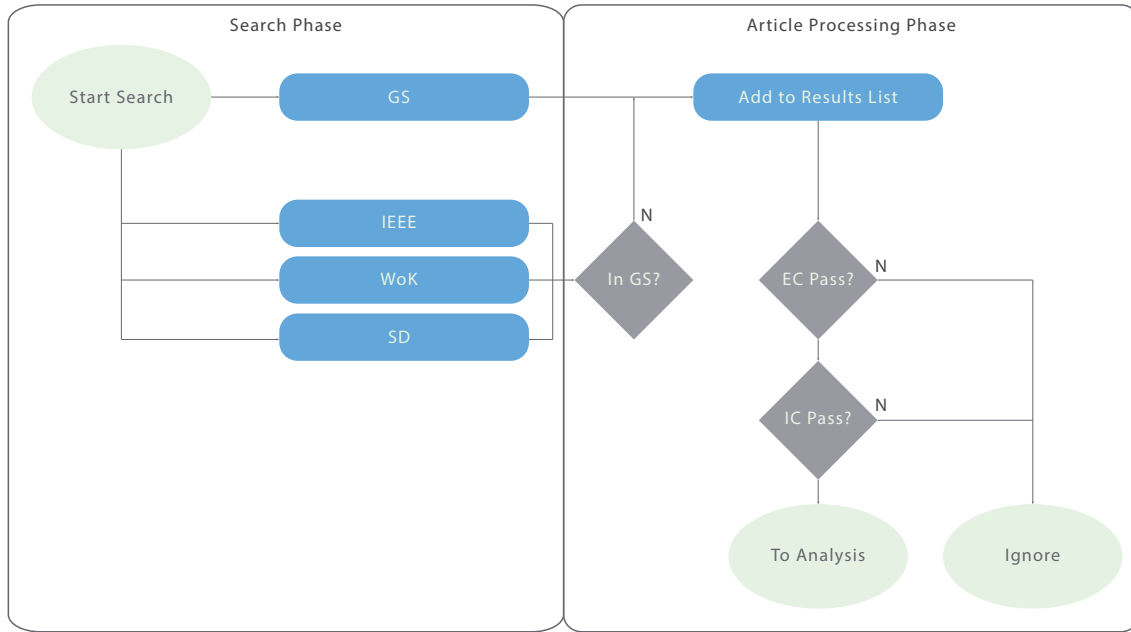
Figure A.5: Conduction stage flowchart. Notice Google Scholar's prevalence.

A.4.2 Results and Discussion

A.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$80%, see Figure A.6). Selection criteria (Inclusion and Exclusion) application resulted in the exclusion of 701 results ($\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 A.6.

Table A.6: Search results. For a paper to reach the righmost 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
								TOTAL	9

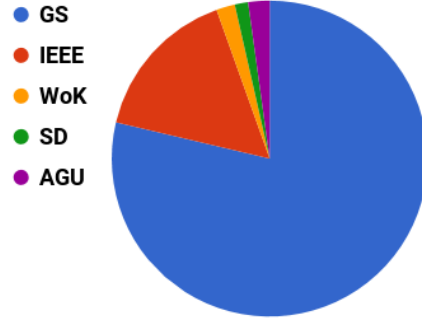| Library	# Results	Percentage
GS	576	78.7%
IEEE	116	15.8%
WoK	14	1.9%
SD	10	1.4%
AGU	16	2.2%

Figure A.6: Results distribution by library.

Table A.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 A.3.5, and presents the 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.

A.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.

Table A.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 [22]	T	1	0,4	0	0,5	0,73
Laepple2004 [33]	T	1	0	0	1	0,70
Mettendorf2006 [36]	E	0,4	1	1	0,25	0,57
ODriscoll2003 [38]	T	1	0	0	0,25	0,63
Olaguer2017 [40]	E	0	0,2	0	0,25	0,07
Poehler (unpub) [45]	E	0	0,4	0	0,25	0,11
Pundt2005 [47]	E	0,4	1	1	1	0,64
Stutz2016 [51]	E	0	1	1	0,25	0,33

A.4.3.1 Instrument

Instrumentation description is present in 7 of the 8 ([22, 33, 36, 40, 45, 47, 51]) selected articles. Stutz2016 [51], Pundt2005 [47] and Mettendorf2006 [36] present the highest level of detail.

In Stutz2016 [51], 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 ≈0,3 nm spectral resolution, which was stabilised to -35ºC.

Pundt2005 [47] was conducted during the BAB II motorway campaign. The team was working with the goal of performing a tomographic measurement of vehicle polution 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 pixe Photo Diode Array (PDA), thermally stabilised at -15ºC. The telescopes were pointed towards two towers which beared the reflectors, set at heights of 10, 20, 30 and 40 m from the ground. The system's

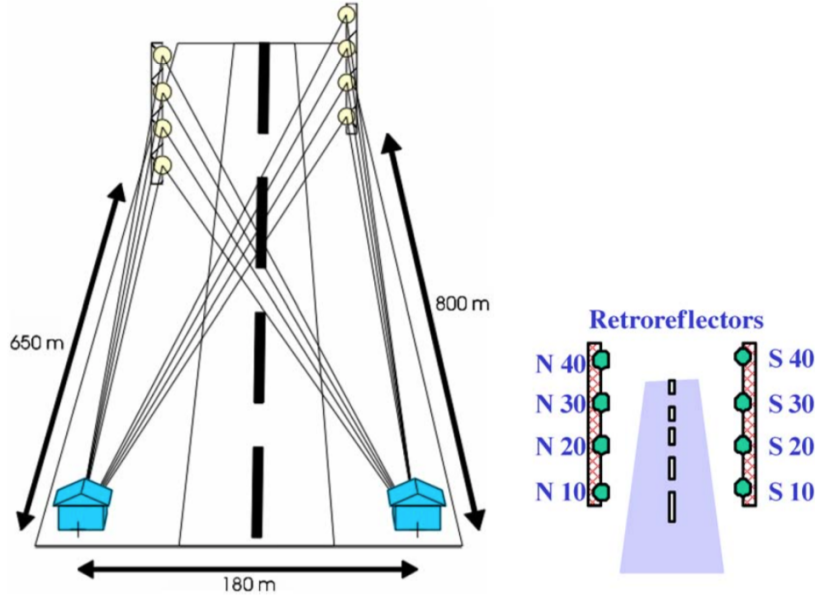geometry and distances are shown in Figure A.7.

Figure A.7: Schematic representation of the BABII campaign experiment [47].

In Mettendorf2006 [36], 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). Telescope, mirror and reflector positions are illustrated in Figure A.8.

As for the other 4 less detailed instrument description, three (Hartl2006 [22], Poehler [45] and Laepple2004 [33]) are from the same group as Pundt2005 [47] and Mettendorf2006 [36], and therefore use the same or similar hardware. Olaguer2017 [40], on the other hand, is the companion paper of Stutz2016 [51], and therefore gives a description of the same instrumentation, though in a less detailed manner.

A.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 A.3.5). Algorithm descriptions are present in 6 of the 8 selected articles: [22, 33, 36, 38, 40, 47]. The most complete descriptions are featured in Hartl2006 [22], Laepple2004 [33] and ODriscoll2003 [38]. Mettendorf2006 [36], Olaguer2017 [40] and Pundt2005 [47] approach the reconstruction algorithms with less emphasis or in a less detailed way.
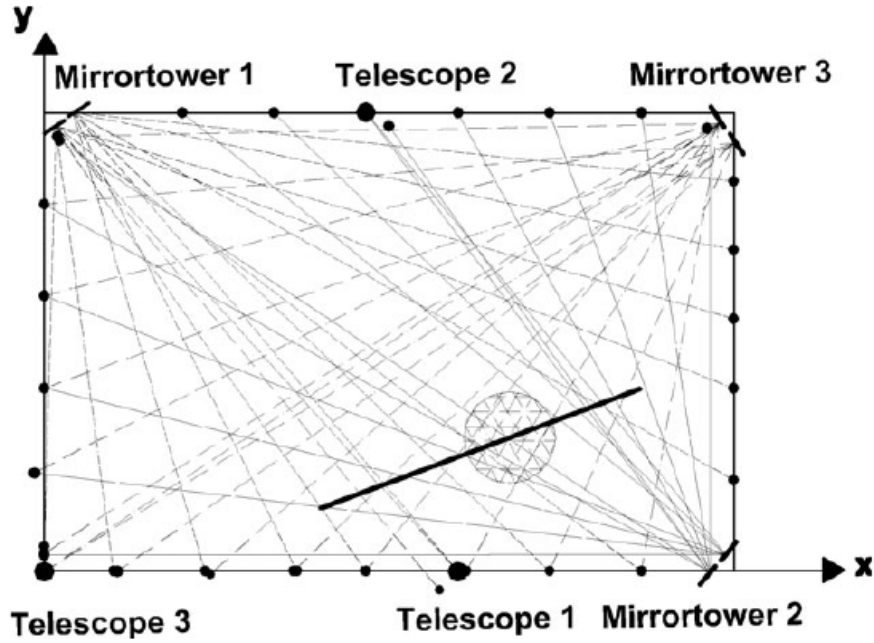
Figure A.8: Mettendorf2006 [36] experiment geometry, detailing position of telescopes (large filled dots), mirrors (small lines in upper corners), reflectors (small dots along edges), test samples (hatched circle), light paths (thin lines) and the movement path for the sample (thick diagonal line).

In Hartl2006 [22], 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 [22] 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 [33]. 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 . 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.

ODriscoll2003 [38] also covers the algorithm extensively. While this paper is considerably shorter that the previous two, it provides a detailed (on an iteration basis) description for ART and SIRT. In addition, and perhaps of greater interest, the paper's authors suggest a different approach to solving the reconstruction matrix, different from the algebric 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 [36] and Pundt2005 [47]. Both these studies were conducted under the same project as Laepple2004 [33] and Hartl2006 [22] and therefore their algorithm descriptions and methods draw heavily on these two studies.

A.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 [36], the team used TOMOLAB for the calculation of the modelled column densities of their experiment. In Pundt2005 [47], spectral analysis was performed using the *MFC Software*. Finally, Stutz2016 [51] used the DOASIS software for control and automation purposes, and does not explicitly mention its use for spectral analysis purposes, although this is likely.

A.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 wwe had to analyse.

The first important mention is the BABII campaign. This study, which ran in 2001 and aimed to quantify polution 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 algebric methods for solving the discretisation and reconstruction problem, namely ART, SART and SIRT.

A.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 purposefuly 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 studies which we have found.

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 A.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 remaining focused only to the techonologic 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.

A.5 Conclusions

Our initial goal was to evaluate the existing literature for tomographic DOAS, assessing the technologic status of this analytical technique. We have started by devising a search protocol and applying it to a set of 5 electronic libraries which were previously determined to be relevant.

This approach has rendered 732 articles, of which 724 were eliminated through the application of our selection criteria (inclusion and exclusion). We have thus ended with 8 articles, then evaluated with regard to the 3 aspects that we have set out to assess: equipment, algorithm and software used.

Although we have to recognise that we have a very low number of selected articles, we believe that they are representative of the current literature, since tomographic DOAS is a relatively new technique, which has not yet been thoroughly researched.

Our results suggest that there is a great prevalence of active DOAS for tomographic reconstruction. We have not found a single example of passive DOAS in our search. In addition, we also detected that despite being similar in many ways, there is still no uniformly designed equipment, software or geometry for DOAS tomography.