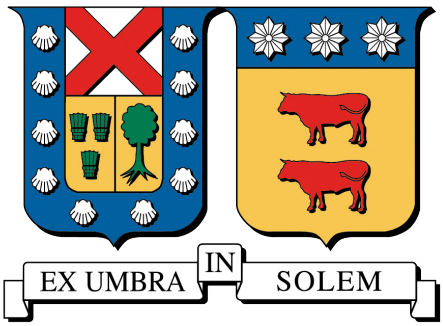
2022

Characterization of soot properties in coflow laminar axisymmetric diffusion flames using image processing and machine learning

Rodríguez Barreda, Alonso Antonio

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



CHARACTERIZATION OF SOOT PROPERTIES IN COFLOW LAMINAR AXISYMMETRIC DIFFUSION FLAMES USING IMAGE PROCESSING AND MACHINE LEARNING

THESIS SUBMITTED BY

ALONSO ANTONIO RODRÍGUEZ BARREDA

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE PROFESSIONAL TITLE OF ELECTRONIC ENGINEER AND THE DEGREE OF MASTER OF SCIENCE IN ELECTRONIC ENGINEERING

THESIS SUPERVISOR : GONZALO CARVAJAL B.
THESIS CO-SUPERVISOR : ANDRES FUENTES C.

VALPARAÍSO, 2022

A mi familia ...

Abstract

Traditional non-invasive optical techniques for estimating soot volume fraction and soot temperature in coflow laminar axisymmetric diffusion flames require solving ill-posed inverse problems from convoluted signals. Applying these techniques using inversion and regularization makes them highly susceptible to experimental noise and the choice of adjustable regularization parameters, which can prevent obtaining consistent and accurate estimations.

This thesis proposes replacing these techniques with a machine learning model that implements the inversion step in the estimation of these soot properties. We develop a framework to generate physically-grounded simulated signals of reference soot fields and their corresponding convoluted projections in the camera plane, thus enabling a supervised learning approach to train our machine learning model. We focus on line-of-sight attenuation measurements for the characterization of the soot volume fraction field and Broadband Emission measurements for the characterization of the soot temperature field, as these techniques represent a promising approach to low-cost characterization of their respective soot property in flames.

Using our framework, we generate datasets for each of the two soot properties to train a model based on U-Net, a fully convolutional neural network previously used in similar inversion problems. We then compare the performance of our models over the synthetic dataset versus traditional techniques and perform a final validation over data obtained during experimental campaigns. These results show that our machine learning models outperform traditional inversion techniques when processing noisy measurements, especially in areas of interest for soot formation, such as the flame center and its path of maximum soot concentration along the flame. The resilience to noise shown by machine learning models makes them attractive for implementing low-cost techniques to characterize soot properties in flames using experimental equipment of different quality, representing a promising research avenue.

Acknowledgements

I would first like to thank my supervisor, Prof. Gonzalo Carvajal, and co-supervisor, Prof. Andres Fuentes, for their guidance, advice, and support during this thesis work. I would also like to acknowledge the contributions of Jorge Portilla, with whom I collaborated in most of the research presented in this report.

Additionally, I would like to acknowledge the vast contributions of members of the Energy Conversion and Combustion Group from the Universidad Técnica Federico Santa María in the results presented in this thesis. Particularly Felipe Escudero, who developed the framework used to generate the physically-grounded synthetic dataset to train our models, and Juan Jose Cruz, who led the experimental campaigns to capture data to validate our models and implemented the code for the traditional methodologies used to compare our results.

This thesis was partially funded by ANID Chile through the projects ACT210052, Basal FB0008, PCI/NSFC 190009.

Contents

List of Figures

List of Tables

1 Introduction

This document presents the design, implementation, and validation of a novel framework based on machine learning (ML) to characterize soot volume fraction and soot temperature in canonical Coflow Laminar Axisymmetric Diffusion (CLAD) flames. Specifically, we focus on retrieving a 2D representation of the soot volume fraction field from line-of-sight attenuation (LOSA) measurements and the soot temperature field from light emission measurements obtained with an RGB camera.

This chapter covers the motivation for our work, describes the general principles for the characterization of the soot properties of interest, and introduces our proposed methodology for the characterization of the soot volume fraction and soot temperature fields using ML. This chapter also describes the scope and contributions of this thesis and describes the organization of this report.

1.1 Context and motivation

Combustion processes are an essential part of multiple industrial activities of significant importance for the global economy. However, combustion processes also carry the negative effect of releasing contaminant materials into the atmosphere, such as soot. Soot is the name given to the agglomerates of nanometric carbonaceous particles formed as a byproduct of the incomplete burning of the fuel within a flame [1]. Soot formation is particularly strong in environments with a high presence of fuel and high-temperature combustion, such as carbon-based electrical plants, chemical factories, motorized vehicles, among others [2].

Soot formation is a natural process in combustion that is not inherently negative or undesired. Soot is one of the main contributors to energy transmission by radiation in the flame, as soot particles that are burned inside the flame carry the beneficial effect of increasing the energy output of the combustion process. However, if the amount of soot particles generated is higher than the rate that can be oxidized inside the flame, then the excess particles are released into the environment. Soot released to the environment from inefficient combustion processes plays an important role in climate change, being the second contributor to the greenhouse effect, only behind Carbon Dioxide (CO_2) [3]. Additionally, regular exposition to soot particles over time may produce multiple adverse effects on human health, causing respiratory and circulatory diseases [4]. As a result, there has been a recent push for regulators around the world to impose rules on the amounts of soot release allowed in industrial processes.

The global effort to move energy generation and consumption towards clean and renewable energy solutions like solar, wind, geothermal, etc., can be perceived as a solution to reduce the emission of contaminants to the environment from industrial processes. However, these new energy paradigms require the development and installation of new infrastructure that will take time to be implemented and adopted on a global scale. Moreover, fossil fuels still have a higher energy density than current renewable energy technologies, and thus some applications of combustion processes such as, energy generation for critical infrastructure (hospitals, banking, defence, etc.) and aerial and maritime transportation, will still rely on fossil fuels for the foreseeable future. This is

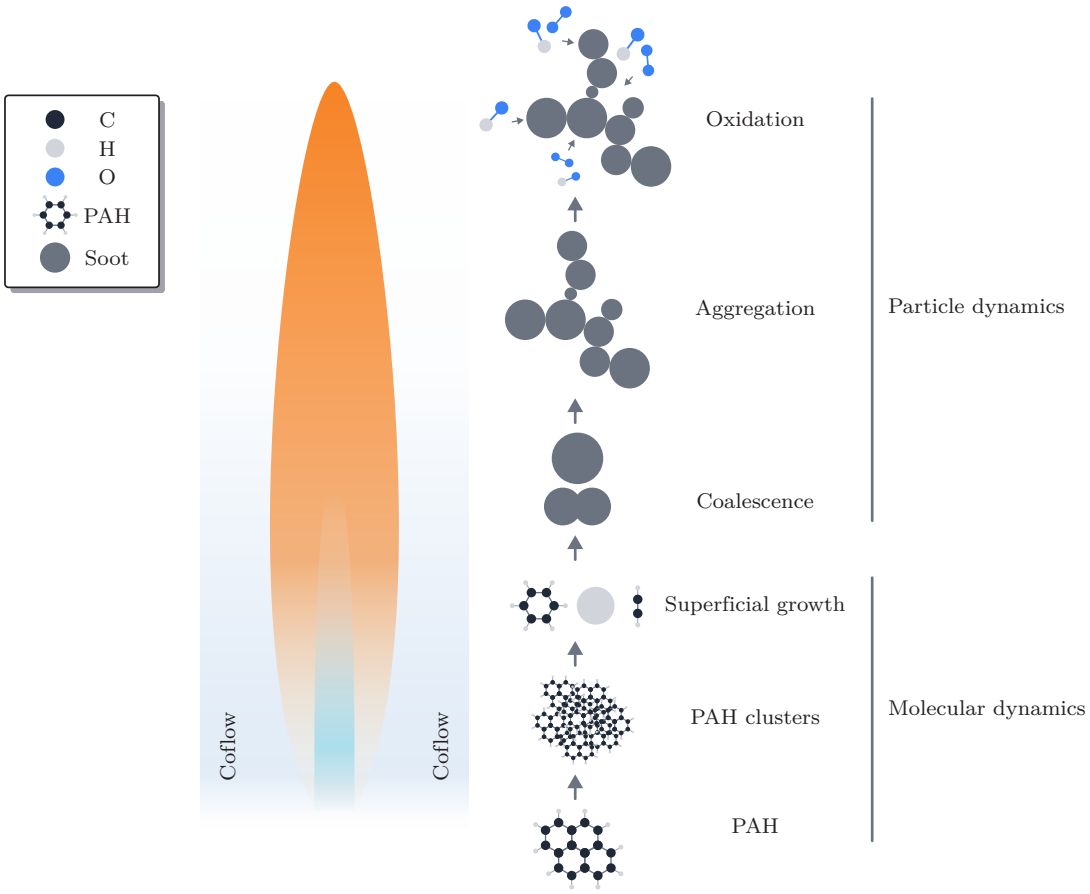


Figure 1.1: Diagram of the stages of soot production in a flame.

particularly important in developing economies, where population and economic growth can create an energy demand that outpaces the development of infrastructure for renewable sources of energy. In consequence, the generation and utilization of energy based on contaminant fuels will not be eliminated in the short term, and thus its use must be conducted in a way that reduces the release of polluting agents to the environment as much as possible [5].

Soot formation is a highly complex physico-chemical phenomenon with multiple steps that are not yet fully understood and take place simultaneously along the different regions of a flame, which might in itself be turbulent and hard to characterize [6, 7, 1]. As a first step towards understanding the dynamics of flames generated in industrial combustion processes, it is necessary to study, characterize and model the behavior of flames generated in controlled laboratory environments that allow researchers to perform repeatable and reproducible observations. The academic community dedicated to studying the phenomena associated with combustion processes, including soot characterization, has defined a set of reference flames generated using standardized burners and operating conditions to promote the repeatability of the experiments in different laboratories [8, 9]. In this thesis, we refer to these flames as canonical Coflow Laminar Axisymmetric Diffusion (CLAD) flames. In diffusion flames, the fuel and oxidant combine at the reaction zone, generating more soot than in premixed flames due to the limited mixing of fuel and oxidant, making them more suited to study the behaviour of soot formation. Additionally, CLAD flames are stabilized by a continuous surrounding airflow (coflow), which gives them enough stability to assume axisymmetry [9, 10, 11].

Figure 1.1 shows a depiction of a stable CLAD flame and some of the stages of soot production

that simultaneously occur in different areas of the flame [1]. Currently, it is accepted that the first soot particles emerge from Poly-cyclic Aromatic Hydrocarbons (PAHs) collisions. Afterwards, as soot particles move through the flame, they grow in size and number due to surface growth and aggregation processes, while oxidative agents on the gaseous phase contribute to a reduction in size and number of these aggregates [7, 1]. CLAD flames represent a simplified 3D flame geometry that facilitates the study of the different physico-chemical processes that lead to soot production distributed in radial and axial directions. Results obtained for CLAD flames can provide the basis to understand the process of soot formation in more complex regime [12, 13, 14].

Even though CLAD flames simplify the modeling and analysis of soot properties like the soot volume fraction and soot temperature fields, there are additional challenges related to the experimental characterization of these properties in a laboratory setting. Traditional non-intrusive optical methodologies to characterize soot properties in CLAD flames obtain projected measurements of the relevant variables along the line-of-sight of an optical sensor (ratio of transmitted light intensity from LOSA for the soot volume fraction and light intensity at different wavelengths for the soot temperature). Then, to estimate local information to the geometry of the flame, researchers apply a deconvolution process with regularization to the projected image, which constitutes an ill-posed inverse problem that presents low tolerance to measurement noise and high dependence on the tuning of regularization parameters [15, 16, 17]. In general, researchers adjust the regularization parameters based on trial and error and the application of the L-curve method [18] to obtain smooth curves for the desired variables. The L-curve method is used to aid the researcher in obtaining a regularization parameter that sets a compromise between smoothness and accuracy of the solution for a specific profile; however, the line-by-line nature of the deconvolution procedure and the abundance of measurement noise in experimental conditions may result in different regularization parameters for different points along the height of the flame, leading to discontinuities in the predicted 2D field. Additionally, the choice of the regularization parameter can be a subjective process that affects the repeatability of the estimations and can result in predictions that lack physical sense. This problem is magnified considering that these methods aim to estimate properties of highly complex processes, and in an experimental setting the actual value of the estimated variables is unknown, which precludes quantifying the quality of the estimations.

This thesis presents a novel approach for the characterization of the soot volume fraction and soot temperature fields in CLAD flames. Based on the generation of a physically-grounded synthetic dataset, we implement ML techniques under a supervised learning framework as a replacement to traditional iterative deconvolution techniques. We show that the proposed framework captures the relationship between the measured signals and the soot properties we want to obtain, computing the soot volume fraction field directly from LOSA measurements and the soot temperature field from intensity measurements as captured by an RGB camera on its three spectral bands. To achieve this, we adapted and tested neural network architectures available in the literature that were developed for spatially structured data, such as the 2D measurements obtained by the optical sensors. We train our ML models with a physically-grounded synthetic dataset, generated by considering large variability on flame conditions and soot distribution, and then validate our models with the synthetic dataset as well as using measurements captured during an experimental campaign.

1.2 Problem definition

This thesis focuses on the development of a framework for the characterization of soot properties in CLAD flames based on ML. In particular, we focus on the following soot properties:

- The soot volume fraction field f_s, which characterizes the fraction of volume of soot versus the aggregated volume of the soot and mix of gases in the flame.
- The soot temperature field T_s, which characterizes the temperature of the soot particles distributed along the geometry of the flame.

These properties can be computed using traditional methodologies available in the literature.

These techniques can be separated into those that are performed by the extraction of a sample for its future analysis (ex-situ) and those that are based on direct measurements (in-situ). The method proposed in this thesis lies in the category of direct measurements using optical sensors, and thus we will compare it with existing methods in this category.

The soot volume fraction is usually computed from LOSA measurements. By projecting a narrowband light source through the flame and measuring it at the other end using an optical sensor, we can compute a coefficient for the light that enters the flame versus the light that exits the flame. Then, from these measurements, we can use a known procedure, described in detail in Section 2.1.1, to obtain the concentration of soot particles contained in an area of the flame.

The characterization of the soot temperature field using traditional non-intrusive optical methodologies (described in detail in Section 2.1.2) can be separated into two techniques commonly employed in the literature:

- The Modulated Absorption and Emission (MAE) technique considers the soot radiation captured by an optical sensor at multiple wavelengths, while also including in the calculations the self-absorption of the soot particles inside of the flame.
- The Emission technique (EMI) is a simplification of the MAE technique that dismiss the effects of the soot self-absorption, and thus it does not require LOSA measurements.

In order to apply the EMI technique, we require measurements at two or more wavelengths known to contain radiation emitted by soot, while the MAE technique also requires information of the soot absorption coefficient, which can be obtained from LOSA measurements in a similar way that the soot volume fraction. Additionally, the scientific community has developed versions of these techniques that use RGB cameras as the optical sensors, which can capture multiple wavelengths simultaneously and thus reduce the amount of laboratory equipment necessary for their application. We refer to these techniques as broadband MAE (BMAE) and broadband EMI (BEMI).

To compute the soot volume fraction and soot temperature fields, we first have to perform a deconvolution step over the measurements captured with an optical sensor to project the information from the sensor plane to the flame symmetry plane, then we can use a known procedure described in detail in Section 2.1.2 to obtain the desired variables. This deconvolution process is the crucial step that we aim to replace by using a ML approach. We propose to treat this as a supervised learning problem; for this effect, we first develop a framework to systematically generate physically-grounded synthetic data based on the expansion of a numerical solution obtained with software tools widely used by the combustion community. Using this framework, we generate a separate dataset for each of the two properties that we aim to predict. We then use these datasets to train and evaluate a ML model based on a convolutional neural network architecture previously used in similar inverse problems. By using the model trained with each specific dataset, we are able to directly compute the soot volume fraction field from LOSA measurements and the soot temperature field from light intensity measurements captured with an RGB camera.

Considering the above discussion, we can formally state the hypothesis of this thesis as follows: *Under a machine learning framework, we can model the reconstruction of the soot volume fraction and soot temperature fields from noisy and convoluted measurements as a multi-dimensional regression problem. We can then generate a physically-grounded synthetic dataset from the explicit resolution of known mathematical equations to train an artificial neural network (ANN) to replace the traditional deconvolution methodologies in the estimation of these fields in CLAD flames.*

1.3 Scope and contributions

As stated previously in Section 1.2, we limit the scope of our research to the following constraints:

- We limit our analysis to a number of canonical Coflow Laminar Axisymmetric Diffusion flames.
- We focus exclusively on non-intrusive optical measurement techniques. We also limit the comparison of our technique to classical non-intrusive optical techniques that use onion-peeling deconvolution with Tikhonov regularization as a deconvolution step, as these are the most common non-intrusive techniques used in the existing literature.
- When performing soot pyrometry, we limit the results presented in this thesis to techniques based on measurements from an RGB camera. This approach favors more modest laboratory equipment than other techniques, such as EMI, which requires capturing two or more wavelengths simultaneously, or BEMI, which requires a laser to perform LOSA calculations.

From our results, we recognize the following contributions:

- Evaluation of the utilization of onion-peeling deconvolution with Tikhonov regularization for the characterization of soot volume fraction and soot temperature in CLAD flames and its shortfalls.
- Development of a framework to generate a physically-grounded synthetic dataset used in the training of supervised learning models for the characterization of soot volume fraction and temperature in CLAD flames.
- Development and evaluation of an ANN architecture for for the characterization of soot volume fraction and temperature in CLAD flames.
- Comparison between our proposed ANN-based framework and the chosen traditional methodology for the characterization of soot volume fraction and soot temperature in CLAD flames.

As a result of our research, we published the following peer-reviewed scientific papers:

- Rodríguez, A., Portilla, J., Cruz, J. J., Escudero, F., Demarco, R., Fuentes, A., & Carvajal, G. (2021). Improving Broadband Emission-Based Soot Pyrometry Using Convolutional Neural Networks. 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). `https://doi.org/10.1109/I2MTC50364.2021.9460106`.
- Rodríguez, A., Escudero, F., Cruz, J. J., Carvajal, G., & Fuentes, A. (2021). Retrieving soot volume fraction fields for laminar axisymmetric diffusion flames using convolutional neural networks. Fuel, 285. `https://doi.org/10.1016/j.fuel.2020.119011`.
- Rodriguez, A., Diomedi, A., Portilla, J., Garces, H., & Carvajal, G. (2019). Automatic Classification of Soot Propensity in Flames Using Image Processing and Machine Learning. IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies, CHILECON 2019. `https://doi.org/10.1109/CHILECON47746.2019.8988106`.

Additionally, the outcomes of this research were presented at the *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* in May of 2021, and at the *Jornada de Mecanica Computacional 2021 (JMC2021)* in October of 2021.

1.4 Organization of the report

The rest of this report is organized as follows:

- Chapter 2 presents an overview of traditional methodologies for the characterization of soot properties in CLAD flames using non-intrusive optical measurements, as well as a literature overview of previous ML-based approaches similar the one proposed.
- Chapter 3 serves as an introduction to machine learning and artificial neuronal networks, as well as introducing our machine learning approach for obtaining the soot volume fraction and

soot temperature fields for CLAD flames. We present an overview of the architecture of the proposed ML model, as well as a systematic methodology to generate physically-grounded synthetic data to train the model.

- Chapter 4 presents the results obtained when comparing our proposed ANN-based methodology with a traditional approach. We compare our results for synthetically generated data as well as data obtained from an experimental campaign.
- Chapter 5 summarizes the main conclusions of this report and provides directions for future work.

2 Background and Related Work

Characterizing the properties of the soot released in CLAD flames can be performed using intrusive and non-intrusive methodologies. In practice, non-intrusive techniques are preferred over intrusive methods, as they provide high temporal and spatial resolution without disturbing the system. Traditional non-intrusive methodologies usually rely on optical sensors (i.e., arrays of filters and cameras). Because of the nature of optical sensors, the signals captured contain the aggregated information along the optical path between the flame and the camera, with an additional noise component inherent to the nature of the measurement system. We can use the assumption of axisymmetry to extend the 2D optical measurements to the 3D geometry of the flame. In order to recover the information generated at the flame plane, we have to perform a deconvolution step along the optical path to the data captured at the camera plane. This deconvolution step usually includes regularization to improve the predictions given the levels of noise contained in the measured signals. Once we have the deconvoluted signals, we can use known equations from combustion theory to obtain the desired variables from the measurements.

This chapter describes the equations used in classical methodologies for the characterization of the soot volume fraction field f_s and the soot temperature field T_s, as well as the laboratory setups required to obtain measurements that allow the computations of these signals. We will also mention previous attempts to implement methodologies based on ML to obtain these quantities.

2.1 Traditional non-intrusive optical methodologies for soot characterization

2.1.1 Soot volume fraction

The soot volume fraction field f_s describes the ratio between the volume of soot present in the flame and the aggregated volume of the soot and the mix of gases in the flame. By using a controlled light source pointed at one side of the flame and an optical sensor at the other side, we can relate f_s to the amount of light that to passes through the flame using line-of-sight Attenuation (LOSA) measurements. LOSA works on the premise that a laser beam of a given wavelength λ will be extinguished by the absorption and scattering of distributed soot particles when traversing the flame [19, 20].

Figure 2.1 illustrates the schematic of a typical experimental setup for LOSA measurements. The light emitted from a laser passes through the flame described in cylindrical coordinates (r, z) and is captured with a camera at the other end of the optical path. Using the signals captured by the camera, we construct a 2D map of the fraction of transmitted light τ_λ described in Cartesian coordinates (y, z) representing the quotient between the intensity of the light that exits the flame I and the nominal intensity emitted by the source I_0 as shown in Figure 2.2 using the equation

$$\tau_\lambda = \frac{I}{I_0}. \tag{2.1}$$

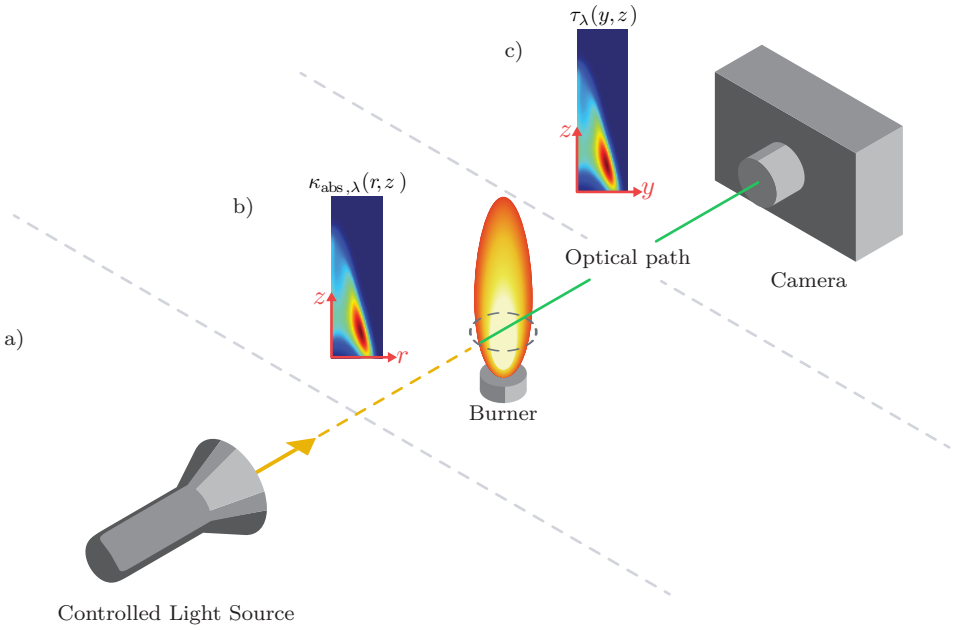


Figure 2.1: Schematic of the setup for LOSA measurements to determine the soot extinction coefficient of a flame. Section a) shows the light source projected through the flame. Section b) shows the 3D CLAD flame represented by its 2D symmetry plane and the local absorption coefficient. Section c) shows the optical sensor and the integrated light attenuation field projected to the camera plane.

To determine the soot volume fraction, we need to measure the absorption coefficient of the soot particles κ_{abs} within the target flame. Assuming that soot particles are small enough to behave in the Rayleigh regime, we can relate f_s to the absorption coefficient as [21, 22]

$$f_s = \kappa_{\text{abs},\lambda} \frac{\lambda}{C_\lambda}, \tag{2.2}$$

where λ is the wavelength of the reference source of light and C_λ is the soot absorption function assumed to be a known value that can be computed using different methods [23, 24]. By assuming that soot particles are small in relation to λ, we can neglect the scattering of light and assume the absorption coefficient to be equal to the extinction coefficient ($\kappa_{\text{abs}} = \kappa_{\text{ext}}$). We can then relate the coefficient τ_λ measured using LOSA techniques to κ_{ext} through the expression [1]

$$-\ln\left[\tau_\lambda(y,z)\right] = \int_{l_0(y)}^{l_1(y)} \kappa_{\text{ext},\lambda}(l,z)\mathrm{d}l, \tag{2.3}$$

where $l_0(y)$ and $l_1(y)$ form the section where the flame is present along the optical path between the source of light and the camera. Therefore, in order to compute f_s from LOSA measurements, we need to invert the Abel type integral from Equation (2.3). The inversion of the integral constitutes an ill-posed problem which is usually solved by using onion-peeling deconvolution [15] in conjunction with Tikhonov regularization to maintain numerical precision in the presence of measurement noise. As shown in Figure 2.3, the attenuation measurements at different points along the y-axis contain the aggregated information of all the soot particles contained in a line defined by the optical path along the x-axis as represented by the integral in Equation (2.3). By considering an axisymmetric flame, we can represent the measurements in polar coordinates and directly apply onion-peeling deconvolution to solve the inversion problem [1].

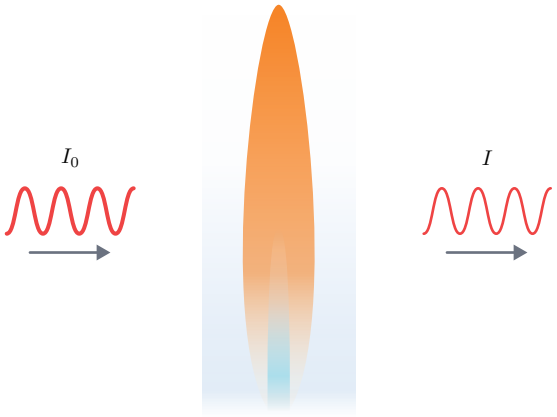


Figure 2.2: Light passing through the flame in LOSA measurements to compute the fraction of transmitted light.

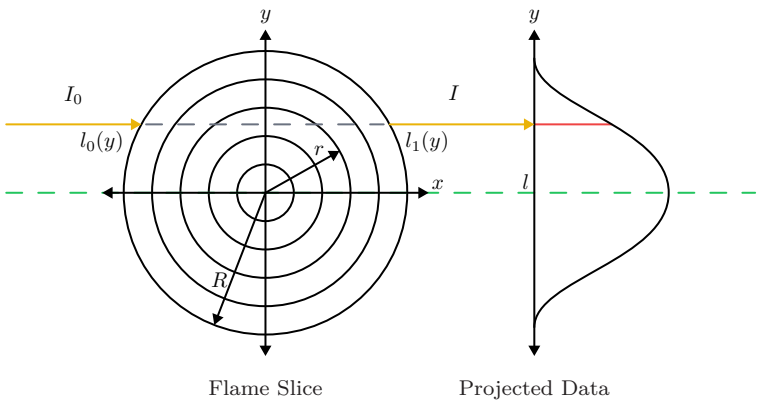


Figure 2.3: 2D slice of a flame showing LOSA measurements for a specific flame height. The light ray in yellow is attenuated as it passes through the flame section defined by $l_0(y)$ and $l_1(y)$. The optical sensor captures the projected attenuation of all the soot particles in that section of the flame. This projection is represented in red in the projected data.

2.1.2 Soot temperature

The measurement of the soot temperature field (also known as soot pyrometry) can be computed from models based on the radiation emitted by soot particles inside of the flame at multiple wavelengths [1], as shown in Figure 2.4. The Modulated Absorption and Emission (MAE) technique computes the temperature by comparing measurements from two or more wavelengths captured by an optical sensor through equations based on the black-body radiation and the decay of light intensity along the optical path of the sensor, while also considering the self-absorption and scattering of soot radiation with adjacent soot particles inside the flame. To compensate for the self-absorption and scattering of soot particles, the MAE technique requires additional LOSA measurements to determine the amount of soot inside the flame, which can be obtained by determining the soot extinction coefficient as described in Section 2.1.1. The Emission (EMI) technique simplifies the

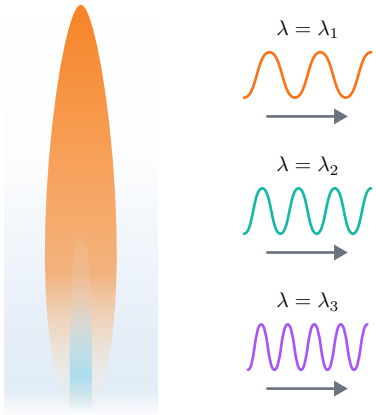


Figure 2.4: Radiation emitted at multiple wavelengths by the soot particles inside of the flame and captured to perform soot pyrometry.

model used in MAE by neglecting the effects of soot self-absorption and scattering, thus simplifying the related calculations. The inclusion of interference between soot particles in MAE leads to better estimations in the absence of measurement noise, but requires a controlled light source that passes through the flame and is thus better suited for highly controlled laboratory environments. For this reason, the MAE technique is only used as a validation step during the development of our methodology, and is not presented as a direct comparison to our ML models. To simplify even further the requirements of the optical setup, we consider the Broadband Emission (BEMI) technique, which replaces the optical measurements at two wavelengths from EMI with a single RGB camera. From these measurements we can perform corrections to treat each red, green and blue channel as separated measurements and apply similar equations to EMI. Figure 2.5 illustrates the generalized setup and variables used for estimating the soot temperature field T_s in a CLAD flame using classical deconvolution techniques.

For wavelength λ known to contain soot radiation, the local emissions from soot J_λ at the flame plane are given by

$$J_\lambda = \kappa_{\text{abs},\lambda} I_\lambda^{bb}, \tag{2.4}$$

where $\kappa_{\text{abs},\lambda}$ is the soot absorption coefficient and I_λ^{bb} is the black-body radiation intensity established by Planck's law as

$$I_\lambda^{bb} = \frac{C_1}{\lambda^5 \left(e^{(C_2/\lambda T)} - 1\right)}, \tag{2.5}$$

with C_1 and C_2 being known constants. The camera receives a 2D projection I_λ of local soot emissions integrated along the line-of-sight l. By neglecting extinction from scattering effects, the line-of-sight radiation intensity detected by the camera becomes [21]

$$I_\lambda(y,z) = \int_l J_\lambda(l,z) \exp\left[-\int_{l'} \kappa_{\text{abs},\lambda}(l',z)\mathrm{d}l'\right]\mathrm{d}l, \tag{2.6}$$

where l' is the section of the optical path from position l to the camera sensor. Equation (2.6) represents the formulation of the MAE technique, in which $\kappa_{\text{abs},\lambda}$ can be computed using LOSA measurements and Equation (2.3). When using EMI, the sooting region of the flame is assumed to behave only as an emitting medium, neglecting the self-absorption and scattering effects on neighboring soot particles that would affect the amount of radiation exiting the flame [1]. Therefore,

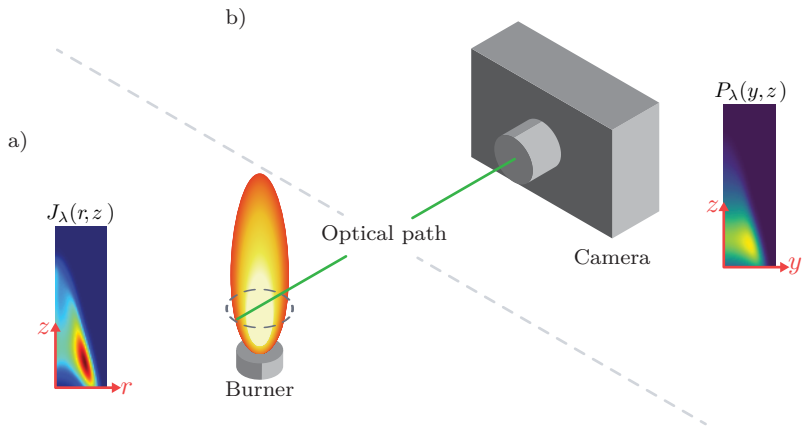


Figure 2.5: Schematic of the setup for emission measurements to perform soot pyrometry using an optical sensor to capture soot radiation emissions at a wavelength λ. Section a) Shows the 3D CLAD flame represented by its 2D symmetry plane and the local local soot emissions field. Section b) shows the optical sensor and the integrated soot emissions projected to the camera plane.

the exponential term in Equation (2.6) is omitted, simplifying the radiation intensity to

$$I_\lambda(y,z) = \int_l J_\lambda(l,z)\mathrm{d}l. \tag{2.7}$$

If we consider an RGB camera for the BEMI technique, when I_λ reaches the camera it is weighted by the spectral sensitivity of the sensor η_λ, and integrated across the spectral band of each color channel. The expression for the fields captured by the camera is given by [25]

$$P_i(y,z) = \int_{\lambda_i} \eta_\lambda I_\lambda(y,z)\mathrm{d}\lambda = \int_l H_i(l,z)\mathrm{d}l, \tag{2.8}$$

with $i \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$ for the red, blue and green channels of the RGB camera. $H_i(l,z)$ represents the spectral integration of soot emissions in the wavelengths captured around the corresponding color channel, which can be recovered from Equation (2.8) through a deconvolution operation.

For this thesis we will focus on BEMI, as it the most promising methodology for low-cost implementations of emission techniques. In a BEMI framework, we can use different functions to relate the $H_{\{R,G,B\}}$ fields measured at different wavelengths to soot temperature. In this thesis, we use the following expression that combines the three RGB channels [26]

$$\frac{H_G(l)H_B(l)}{\left(H_R(l)\right)^2} = \frac{\left(\int_{\lambda_G} F_\lambda I_\lambda^{bb}(l,T_s)\,\mathrm{d}\lambda\right)\left(\int_{\lambda_B} F_\lambda I_\lambda^{bb}(l,T_s)\,\mathrm{d}\lambda\right)}{\left(\int_{\lambda_R} F_\lambda I_\lambda^{bb}(l,T_s)\,\mathrm{d}\lambda\right)^2}, \tag{2.9}$$

where $F_\lambda = \eta_\lambda \frac{C_\lambda}{\lambda}$, with C_λ being the soot absorption function which can be derived from known literature and is thus assumed to be known [1, 23, 24], η_λ the sensitivity of each RGB channel (calculated for our specific experimental setup as described in Section 4.2.2) and I_λ^{bb} a function of T_s. Dependence on the flame height z was removed for simplicity as this method is usually applied at the profile level (line-by-line). The left-hand-side (LHS) of Equation (2.9) represents the three-color ratio from deconvoluted experimental measurements, whereas the right-hand-side (RHS) describes a mathematical model depending on T_s for the same relationship, which considers the spectral transmissivity of the RGB sensor through F_λ. To determine T_s from the experimental signals, we

evaluate the black-body radiation intensities in the RHS for different temperatures until it matches the LHS value within a given tolerance.

As mentioned before, a major difficulty in the methodology described above is that the deconvolution operation that retrieves the $H_{\{R,G,B\}}$ fields from the $P_{\{R,G,B\}}$ measurements captured by the camera is highly sensitive to the choice of regularization parameters to reduce the effect of experimental noise.

2.2 Related work

During the first stages of this thesis work, T. Ren et al. [27] published an article that aimed to characterize soot properties using ML, which we used as a baseline comparison in the development of our models. In this article, the authors aim to use a fully-connected feed forward ANN to predict soot volume fraction and soot temperature profiles (i.e., single horizontal lines) at a particular flame height above the burner. For this, they assume that the soot volume fraction and soot temperature profiles can be described by three different polynomial functions. Then, using these polynomials, they generate synthetic datasets to train the ANN and validate their models using both synthetic and experimental data.

The main differences with our work is that we perform the prediction of the soot volume fraction and soot temperature fields as a two-dimensional field instead of single-line profiles, thus increasing the structural similarity of the predictions. Additionally, we employ a model based on a convolutional neural network that has been widely used by the community in inversion problems, and we train our model based on the comprehensive framework described in Section 3.2 instead of predicting polynomial functions based on previously studied soot distributions.

During the development of this thesis, multiple research teams started using similar techniques for the characterization of soot properties in CLAD flames that were published in the same time frame that our publications described in Section 1.3. For example, T. Ren et al. in [28] trained a neural network using hyper-spectral measurements to predict soot temperature and concentration of different species in a laminar flame, in [29] they used a similar approach based on hyper-spectral measurements to train a ML model, this time to predict three-dimensional fields in laminar flames, and in [30] they use near-infrared emission signals captured with a monochromatic camera to train a neural network and validated their model using two kinds of flames. Q. Wang et al. in [31] used a neural network tuned using Bayesian optimization to predict the soot volume fraction, temperature, and primary particle diameter fields, and validated their research using both synthetic and experimental data. These publications use a fully-connected feed forward ANN to learn a mapping function for single-line radial profiles between optical measurements and different soot characteristics for each case. By retrieving single-line profiles at different heights, it is possible to assemble the entire 2D fields of the soot characteristic being predicted. However, since each line is processed independently without considering the neighboring profiles, the retrieved fields tend to show horizontal discontinuities, especially in zones around the base and tip of the flame, where more pixels correspond to the background instead of the flame.

Additionally, multiple research papers published in the last years have utilized ML techniques to compute different variables in combustion processes, which shows the interest of the combustion community in this field. A comprehensive report of the state of the art of ML used in combustion has been performed by L. Zhou et al. in [32].

3 Soot Characterization Using Machine Learning

Machine Learning (ML) is the scientific field that studies algorithms that allow computers to improve their performance on a specific task through an iterative learning process rather than by explicitly programmed rules. We can provide a generic description of a ML algorithm as follows: Given a hypothesis space $\mathcal{H}$ defined by the ML model and its parameters, a set of examples $\mathbf{x}^{(i)} \in T$ for $i = 1, \ldots, S$ to feed to the model for its training, and a cost function $\mathcal{J}(\mathbf{h}(\mathbf{x}^{(i)}))$ that quantifies the quality of the choice of a hypothesis $\mathbf{h} \in \mathcal{H}$ for an example $\mathbf{x}^{(i)}$, the general objective of a ML algorithm is to find an optimal hypothesis $\mathbf{h}^*$ such that

$$\mathbf{h}^* = \arg\min_{\mathbf{h} \in \mathcal{H}} \frac{1}{S} \sum_{i=1}^{S} \mathcal{J}(\mathbf{h}(\mathbf{x}^{(i)})). \tag{3.1}$$

In a supervised learning framework, the ML model adapts a set of internal parameters using labeled input-output pairs to obtain a mapping function or hypothesis $\mathbf{h}$ that links the inputs and outputs according to the cost function $\mathcal{J}$. Given a training set $T = \{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)})), \ldots, (\mathbf{x}^{(S)}, \mathbf{y}^{(S)})\}$ of S input vectors $\mathbf{x}^{(i)}$ with useful features for the learning task and their corresponding outputs $\mathbf{y}^{(i)}$, indexed by pairs as $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})$ with $i = 1, \ldots, S$, and considering a set of parameters $\boldsymbol{\theta}$ derived from the architecture of the model to characterize the space of mapping functions $\mathbf{h}(\mathbf{x}; \boldsymbol{\theta}) \in \mathcal{H}$, the general objective of a supervised learning algorithm is to find a set of parameters $\boldsymbol{\theta}^*$ that generate a good predictor such that

$$\boldsymbol{\theta}^* \approx \arg\min_{\mathbf{h}(\mathbf{x};\boldsymbol{\theta}) \in \mathcal{H}} \frac{1}{S} \sum_{i=1}^{S} \mathcal{J}(\mathbf{h}(\mathbf{x}^{(i)}; \boldsymbol{\theta}), \mathbf{y}^{(i)}), \tag{3.2}$$

in which the cost function is usually defined in a way that the predictions $\mathbf{h}(\mathbf{x}^{(i)}; \boldsymbol{\theta})$ approximate the actual value $\mathbf{y}^{(i)}$ such that $\mathbf{h}^*(\mathbf{x}^{(i)}; \boldsymbol{\theta}^*) \approx \mathbf{y}^{(i)}, \forall i = 1, \ldots, S$. The process of training consists in the systematic modification of the set of parameters $\boldsymbol{\theta}$ to explore the hypothesis space defined by the ML model in order to minimize the cost function.

The task of reconstructing the soot volume fraction and soot temperature fields from noisy images can be modeled as a 2D regression problem, for which we can leverage the extensive knowledge from the image processing community in dealing with spatially structured data. To solve this regression problem, we propose using Artificial Neural Networks (ANNs) under a supervised learning framework based on physically-grounded synthetic data. ANNs are a family of widely used ML algorithms that define a graph of neural units connected by weights and then trained to obtain an input-output mapping function [33].

The following sections provide a brief overview of the underlying concepts of traditional and convolutional neural networks, followed by a description of the specific architecture that we use in this thesis. We then describe the training process of the ML models and their utilization to reconstruct the soot volume fraction and soot temperature fields, ending this chapter by describing

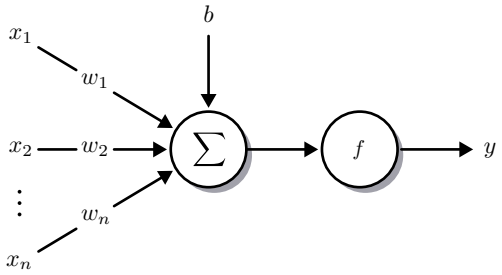


Figure 3.1: Schematic representation of a single artificial neuron.

the framework used to generate the physically-grounded synthetic dataset that enables the training of these models.

3.1 Artificial Neural Networks

Figure 3.1 shows the representation of single a artificial neuron that computes its output y as the result of applying a non-linear function f to the dot product between the input feature vector $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ and the neuron weights $\mathbf{w} = [w_1, \ldots, w_n]^{\mathrm{T}}$ with an added bias term b, such that

$$y = f(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b). \tag{3.3}$$

By selecting a proper non-linearity f, the objective of the ML algorithm is to train the neuron to obtain a set of parameters $\boldsymbol{\theta} = \{\mathbf{w}, b\}$, that map the feature vector x to a desired output y as in described in Equation (3.2).

3.1.1 Fully-connected feed forward artificial neural networks

To be able to learn more complex mappings between inputs and outputs than when using a single artificial neuron, multiple neurons can be stacked together to form a neuronal layer, then multiple layers can be connected to form an ANN. Figure 3.2 shows the generalized structure of a fully-connected feed forward ANN, also called Multi Layer Perceptron (MLP). The MLP maps an n-dimensional input vector $\mathbf{x}$ to an m-dimensional output vector $\mathbf{y}$. Signals travel from the input layer to the output layer passing through several intermediate hidden layers, with all neurons in a layer commonly using the same non-linear activation function. The number of neurons in the input and output layers are directly set according to the dimensions of $\mathbf{x}$ and $\mathbf{y}$ respectively. The number of hidden layers and neurons, the activation functions, as well as other parameters that affect the training of the network are known as hyperparameters, which must be tuned experimentally during multiple training sessions according to the complexity of the model and the characteristics of the training dataset [34].

To train the network after defining its structure, it is necessary to define a cost function $\mathcal{J}(\boldsymbol{\theta})$ that quantifies the prediction cost of selecting a set of parameters $\boldsymbol{\theta}$, corresponding to the weights and bias for all neurons in the model. Given an initial value $\boldsymbol{\theta}_0$, for each training step k, the learning algorithm performs a forward pass with all the training examples to obtain the current value of $\mathcal{J}(\boldsymbol{\theta}_{\boldsymbol{k}})$. Then, using gradient descent techniques, each parameter is updated in the direction that minimizes the cost function according to

$$\theta_{i,k+1} = \theta_{i,k} - \eta \nabla_{\theta_{i,k}} \mathcal{J}(\boldsymbol{\theta}_k) \quad \forall \theta_i \in \boldsymbol{\theta}, \tag{3.4}$$

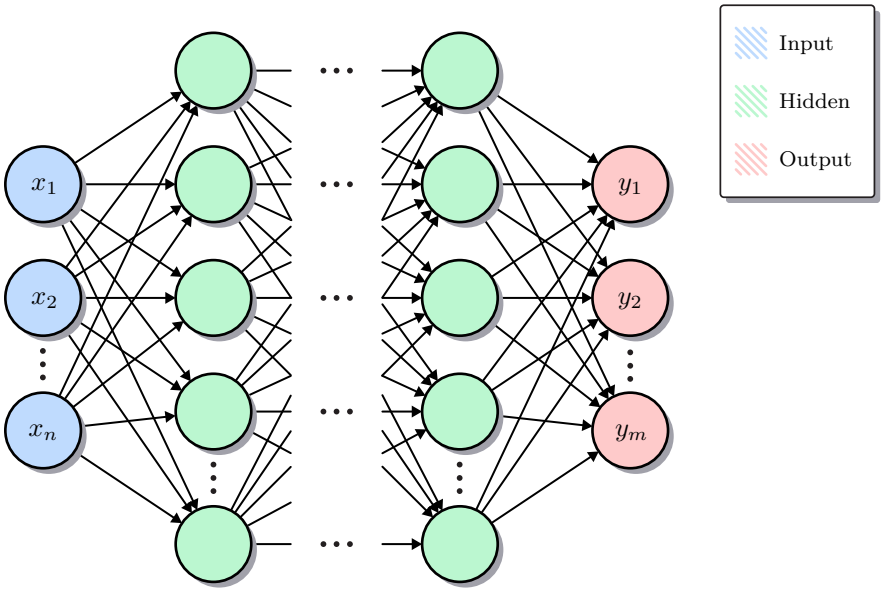


Figure 3.2: Schematic representation of a fully-connected feed forward ANN. The input layer in blue represents the input vector $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$, the output layer in red represents the output vector $\mathbf{y} = (y_1, \ldots, y_m)^{\mathrm{T}}$ and the hidden layers in green define the depth of the network.

where η is a scalar factor known as the learning rate. The process of determining each partial derivative in the computational graph is a direct application of the chain rule for derivation known as the back-propagation algorithm [35]. Learning usually stops when additional passes of Equation (3.4) over the training set do not generate significant reductions in the cost function.

As described in Section 2.2, other authors have used ML models based on an MLP architecture to compute the soot volume fraction and soot temperature fields [27, 28, 29, 30, 31]. However, as discussed before, these articles describe single-line predictions, resulting in fields with discontinuities between adjacent horizontal lines. For this reason, we employ ML models that exploit the spatial structure of the 2D fields measured by the optical sensors.

3.1.2 Convolutional neural networks

Convolutional neural networks (CNNs) are ANNs with one or more convolutional layers. Convolutional layers replace the dot product used in fully-connected layers with a two-dimensional convolution operator between a learned kernel or filter $g(u, v)$ and the input two-dimensional array $h(u, v)$ [36], where u and v are spatial coordinates. The 2D convolution operator is formally defined as

$$r(u, v) = h(u, v) * g(u, v) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} h(n_1, n_2) g(u - n_1, v - n_2). \tag{3.5}$$

Intuitively, Equation (3.5) maps a 2D input to a 2D output, computing for each input element the sum of a vicinity of neighboring elements delimited by the size of the kernel and weighted by the kernel coefficients to produce each element of the output $r(u, v)$. Figure 3.3 shows the convolution operation for a single element, while the final output is obtained by sliding the kernel to cover the entire input array. Each convolutional layer often contains more than one kernel with learned weights to extract different sets of useful features from the input, generating as many outputs (called feature maps or channels) as kernels in the convolutional layer. CNNs are commonly used for processing spatially structured data, as found in images and videos.

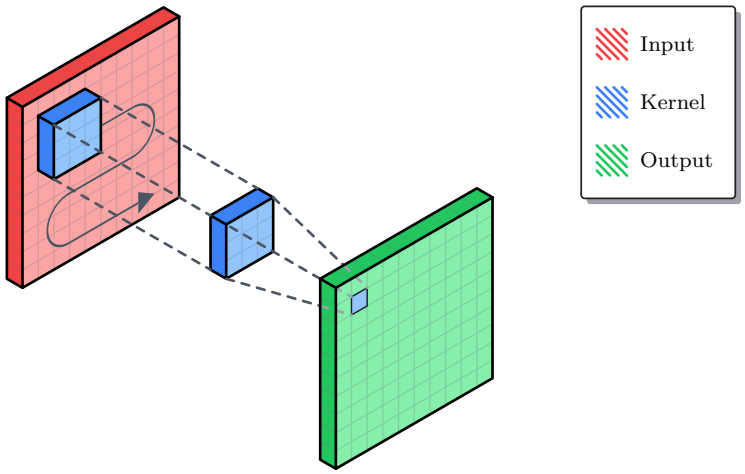


Figure 3.3: Two-dimensional filtering using convolution. The input array (red) is filtered by the kernel (blue) through the 2D convolution operator, generating the output array (green).

Convolutional layers are often used in conjunction with pooling layers, which implement a sub-sampling operation to reduce the input dimension. In particular, Max Pooling layers that select the maximum value in a vicinity of pixels have been used to implement CNNs that show resilience to scale and positional changes in the input images and good noise tolerance [37].

3.1.3 U-Net

For the results presented in this thesis, we use a CNN based on the U-Net architecture [38] with small modifications to adapt it to our application.

U-Net is a fully convolutional encoder/decoder ANN architecture with a contracting and an expanding path forming a symmetric U-shaped graph. The contracting path consists of a stack of neural layers separated in stages, each stage perform a downsampling operation, halving the geometric dimensions of the input data to identify relevant features at different scales. The expanding path consists of a stack of neural layers that perform an upsampling operation, that doubles the geometric dimensions at each stage. By combining the features extracted from the contracting path with the upsampled output from the expanding path at each stage, a properly trained network can assemble a new output image based on the relevant features of the original image. U-Net was originally designed for contextual image segmentation [38], but multiple variations of the architecture have been used for solving inverse problems in imaging in different domains [39, 40].

Figure 3.4 shows a diagram of the implemented U-Net architecture for our application. The network receives as its input a tensor formed by the data captured by the camera for each case: τ_λ as an $m \times n$ tensor to predict f_s, and $P_{\{R,G,B\}}$ as an $m \times n \times 3$ tensor to predict T_s, with m and n being the width and height of the images captured by the camera respectively. After processing the data, the network outputs an $m \times n$ tensor representing the predicted field for each case. Note that Figure 3.4 shows the architecture for predicting the soot temperature field, as the architecture for predicting the soot volume fraction field is the same, only with a single input image corresponding to τ_λ.

The contracting path of U-Net (teal boxes in Figure 3.4) consists of multiple downsampling stages that operate on smaller versions of the feature maps extracted from the input image. After the input passes through a first convolutional layer, each stage contains two convolutional layers

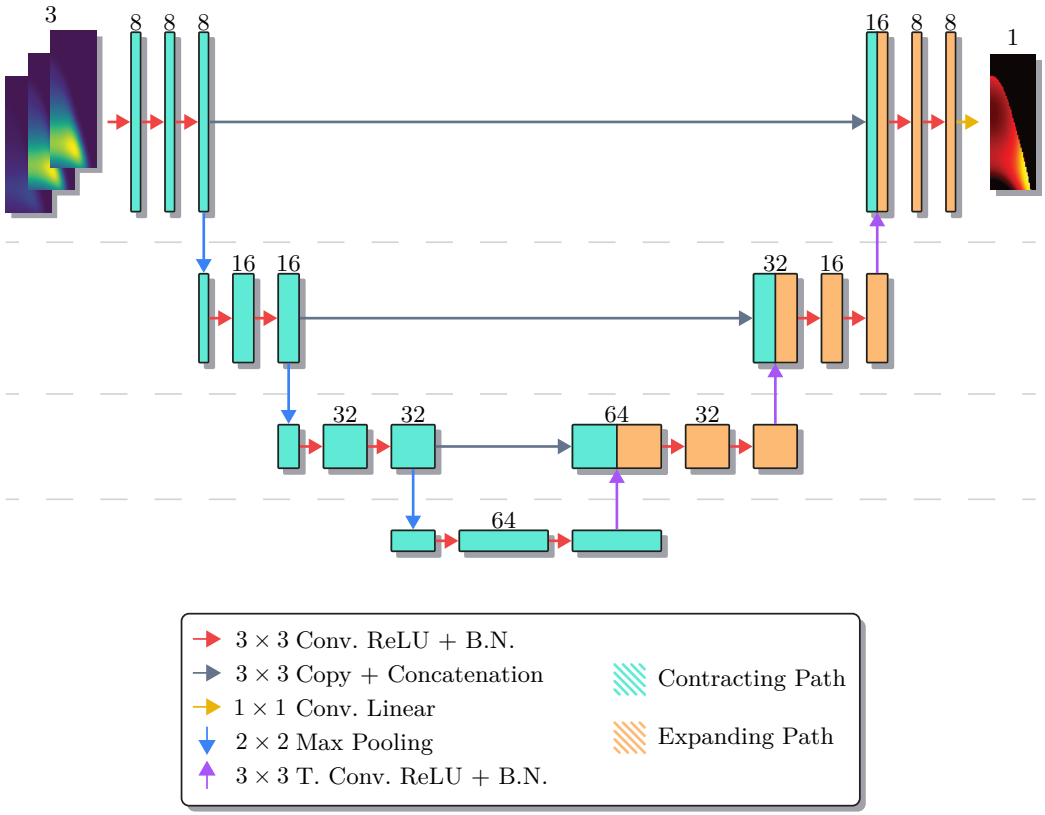


Figure 3.4: Architecture based on U-Net with four stages for the prediction of T_s from measurements of $P_{\{R,G,B\}}$. The output after each operation is represented as a box, with its number of feature maps on top. The physical dimensions of each channel are halved through each stage in the contracting path (green) and doubled through each stage in the expanding path (orange).

of kernel size 3×3 and the same number of filters (red arrows), followed by a 2×2 Max Pooling operator (blue arrows), which reduces the size of its input by half. The number of filters of the convolutional layer doubles at each stage. The output of each operator in the contracting path is represented as a teal box, with its number of feature maps on top.

Each stage of the expanding path (orange boxes in Figure 3.4) consists of a transposed convolution layer (purple arrows), a slight modification of a convolutional layer that applies a convolution with learned kernel weights over a padded version of its input, generating an up-sampled version of the input and reducing its number of filters by half. The output of each upsampling stage is concatenated with the corresponding feature map of the matching stage in the contracting path (gray arrows), and then filtered through two convolutional layers of kernel size 3×3 and the same number of filters as the corresponding stage of the contracting path (red arrows). The output of each operator in the expanding path is represented as an orange box, with its number of feature maps on top. At the end of the expanding path, we use a 1×1 convolutional layer of one filter and a linear activation function to condense the information of all the filters into the final prediction of the f_s or T_s field (yellow arrow).

All of the convolutional and transposed convolutional layers except the last one use the ReLU activation function defined as [41]

$$\mathrm{ReLU}(x) = \max\{0, x\}, \tag{3.6}$$

with an additional Batch Normalization (BN) step, that standardizes the layer inputs to stabilize the learning process and accelerate the training of the network [42].

3.2 Simulation of soot measurements, dataset generation and ANN training

This section describes our methodology for the systematic generation of physically-grounded synthetic measurements. This framework was developed by members of the Energy Conversion and Combustion Group (EC^2G) from the Universidad Técnica Federico Santa María, and was modified under their supervision during the curse of this research to fit the specific needs of our ML models. Our focus in this thesis is limited to generating a proper dataset to train our ML models; we assume that the solutions to the transfer equations generated by CoFlame are correct and that they accurately describe the flames generated in laboratory environments.

3.2.1 Dataset generation

Using the framework, we simulate LOSA experiments by generating synthetic soot volume fraction f_s fields at the flame plane and their corresponding convoluted τ_λ fields for a specific wavelength as captured by an optical sensor with a narrow-band filter. We also used this framework to simulate the soot temperature field T_s for a given flame and the corresponding projected intensity signals $P_{\{R,G,B\}}$ at the three spectral bands of an RGB camera. For simplicity, we will describe the generation of synthetic data for both soot volume fraction and soot temperature at once, as the first steps of these processes are the same for both cases. This framework is composed of three main steps as shown in Figure 3.5:

1. We first use CoFlame [43, 44, 45, 46] to obtain a converged solution of the canonical Gülder CLAD flame [47] as shown in Step 1 of Figure 3.5. By solving the radiative transfer equation [21] with a discrete ordinates method [48] coupled with a statistical narrow-band correlated-k model [49], CoFlame obtains the radiative source term of the energy equation and solves the transport equations for momentum, continuity, fraction of mass by species, soot particles and energy. CoFlame also implements a sectional method for solving the population balance equation of the number density of soot aggregates and primary particles [50, 51]. The solution of these equations by CoFlame compiles the models and understanding of soot formation developed by the combustion community; however, these techniques are computationally costly, making them unsuitable for their utilization in the generation of a large dataset as required for the training of a ML model.

   CoFlame allows the user to specify settings to represent different conditions of the process, including the dimensions of the numerical domain, boundary conditions, number of control volumes of the 2D grid (CVs), and parameters controlling the different soot processes. We select conditions to generate a flame of similar geometry to the ones captured during experimental campaigns in the laboratory. For this, we specify a numerical domain consisting of a non-uniform grid of 202 CVs in the axial coordinate (z-axis) and 94 CVs in the radial coordinate (r-axis), extending to 14.3 cm and 2.74 cm, respectively. We describe a flame in which pure ethylene is injected through the central tube of the burner (internal radius $r_{in} = 5.45$ mm) at a rate of 3.465 cm/s and an oxidizer coflow composed of 21 % O_2 and 79 % N_2 at a rate of 61.255 cm/s through the annular section between both concentric tubes of the axisymmetric burner. We impose flat inlet velocity profiles as boundary conditions for the fuel and oxidizer flows, entering the domain at 300 K. Finally, we model the flame chemistry using the DLR mechanism described in [52] and [53]. The specified conditions produce a stable, non-smoking flame of approximately 7.2 cm of height. The converged simulation provides predictions for the soot volume fraction and the soot temperature fields among other flame properties, from these solutions we extract the necessary fields required to train the ML model as 2D images residing in the symmetry plane described by the z and r coordinates.

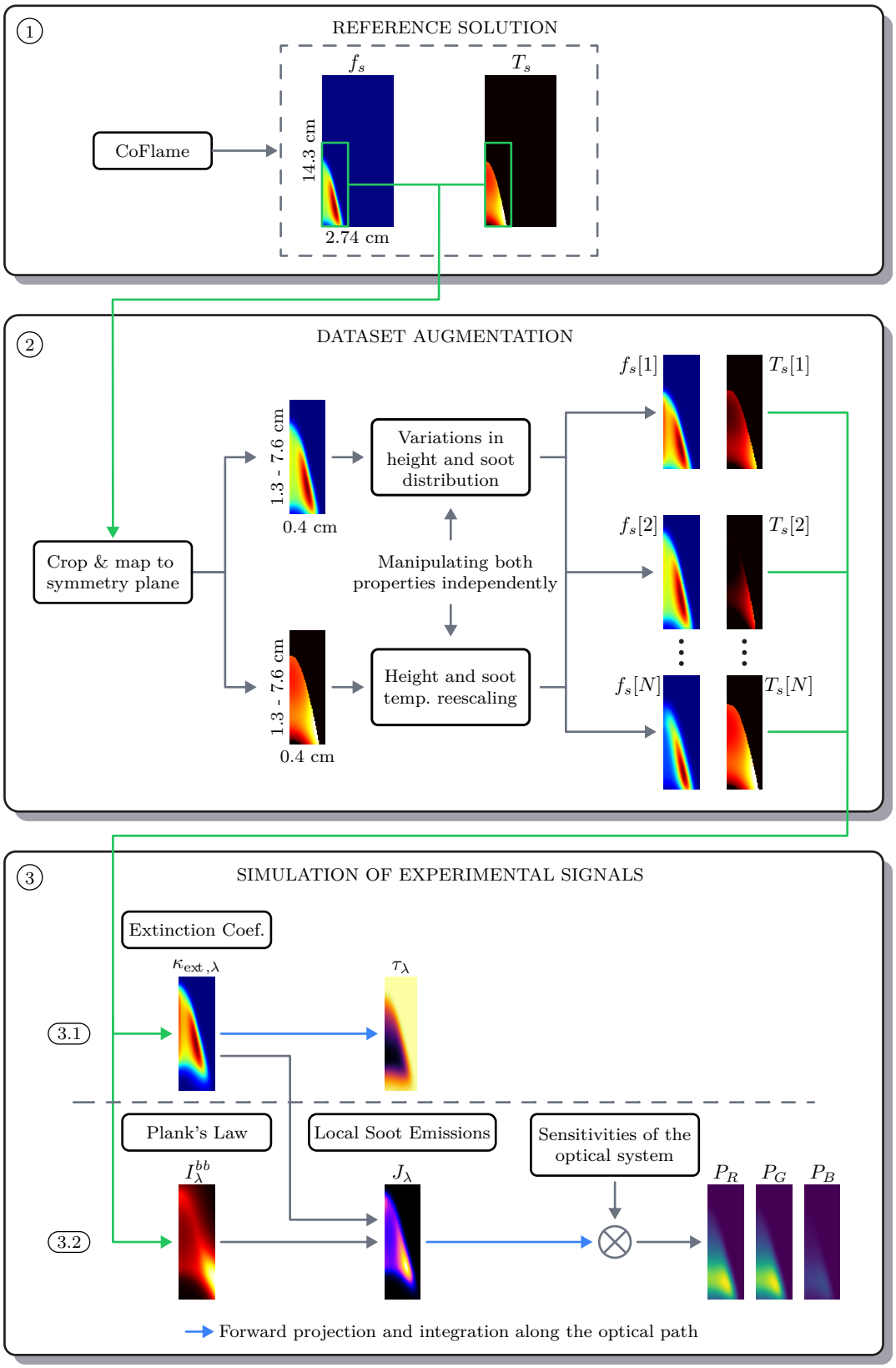


Figure 3.5: Steps for the generation of the physically-grounded synthetic datasets.

2. We then manipulate the converged numerical solution described in Step 1 to generate new instances of flames that incorporate statistical variations in height, soot loading, and soot volume fraction distribution (See Step 2 of Figure 3.5). This step aims to broaden the values of soot conditions generated in the dataset to better represent the values expected in different configurations of CLAD flames in a reasonable time frame. Although in real flames both f_s and T_s fields are intrinsically coupled through the contribution of radiant emission from soot particles [14], we manipulate both properties independently to facilitate the generation of large datasets covering broad ranges of spatially correlated values typically found in canonical flames, which is a key requirement for the proper training of our model.

   2.1. To generate new instances of synthetic f_s fields, we apply the following sequence of steps over the original reference solution:

      i. From the original converged solution, we select the region of interest that encloses the flame, cropping the 2D image from $z = 1$ cm to $z = 10$ cm and $r = 0$ cm to $r = 0.4$ cm. We then map this region into a uniform grid of 200×40 CVs using bicubic interpolation. The new grid represents the typical spatial resolution of the sensors used to capture data during our experimental campaigns.

      ii. We modify the height of the flame to a new value $h_f^* = h_f \pm 2N_r$, where N_r is a random number between 0 and 1 extracted from a uniform distribution, and h_f is the height of the original flame ($h_f = 7.2$ cm in the original solution used for these experiments).

      iii. We modify the original f_s distribution using three different approaches: (i) linear, (ii) regularized Dirac function [54] centered at $r = 0$ cm and (iii) regularized Dirac function centered at the position $r = r^*$, where r^* follows the path of maximum f_s at each height along the z-axis. These three approaches result in a new distribution obtained as $f_s^* = f_s p_i$, with $i = \{1, 2, 3\}$ for the three approaches and p_i being one of the following expressions selected randomly:

$$p_1 = A\frac{r - r_c}{r_c}, \tag{3.7}$$

$$p_2 = \frac{1}{\gamma\sqrt{\pi}} \exp\left(-\frac{r^2}{\gamma^2}\right), \tag{3.8}$$

$$p_3 = \frac{1}{\gamma\sqrt{\pi}} \exp\left(-\frac{(r - r^*)^2}{\gamma^2}\right), \tag{3.9}$$

      where A is a random number between 2 and 4 extracted from a uniform distribution, $r_c = 5$ mm is the radius at which the linear function becomes 0, and $\gamma = (0.75 \pm 0.25N_r) \times 10^{-3}$ is a regularization parameter that controls the radial gradient of the regularized Dirac function. Both p_2 and p_3 functions are normalized by its radial maximum.

      iv. We scale the f_s field so that its maximum value ranges from 1 ppm to 10 ppm to generate different soot loadings. This new maximum soot volume value is extracted from a uniform distribution.

      v. We then repeat Step i to Step iv to generate additional instances of synthetic f_s fields.

   2.2. The procedure for applying statistical variations to the reference T_s fields of the CoFlame solution is summarized as follows:

      i. We map the region that encloses the flame, covering between $z = 1.3$ cm to $z = 8.6$ cm and $r = 0$ cm to $r = 0.4$ cm, to a symmetry plane in an image of 128×40 pixels using bicubic interpolation.

ii. We modify the height of the flame to a new value $h_f^* = h_f \pm 2N_r$, where N_r is a random number between 0 and 1, and h_f is the height of the original flame ($h_f = 7.2$ cm in the original solution used for these experiments).

iii. We scale the T_s field to a new random maximum soot temperature ranging from 1650 K to 2250 K to simulate different flame conditions. This new maximum temperature value is extracted from a uniform distribution.

iv. We then repeat Step i to Step iii to generate additional instances of synthetic T_s fields.

The limiting cases for these conditions are designed to encompass the conditions of about 90% of the ISF-4 catalog of coflow flames [12, 8]. Manipulating an already converged CoFlame solution to generate new instances of soot volume fraction and soot temperature fields representing different operational conditions is orders of magnitude faster than computing a new converged solution for each condition. It is important to note that the new instances that incorporate statistical variations may lead to soot distributions that do not necessarily represent real flames. Instead, the purpose of this step is to facilitate the generation of large datasets that statistically cover the range of values found in different canonical flames. The capacity to generate this statistically representative dataset is the key enabler for the use of a supervised learning approach to soot characterization.

3. We now apply the equations described in Chapter 2 to obtain the convoluted signals that we would measure for each of the soot volume fraction and soot temperature fields generated in Step 2 (See Step 3 of Figure 3.5).

3.1. To simulate experimental LOSA measurements of τ_λ, we use the reference f_s fields obtained in Step 2, and then compute the τ_λ projections according to the following steps:

i. We map the 2D-axisymmetric reference f_s field into 3D-Cartesian coordinates, using linear interpolation for the newly mapped values of f_s.

ii. We compute $\kappa_{\text{abs},\lambda}$ from the f_s field using Equation (2.2) for a wavelength $\lambda = 800$ nm and a soot absorption function C_λ derived from [23].

iii. We perform the line-of-sight integration of Equation(2.3) using discretization to obtain the fraction of transmitted light τ_λ.

3.2. To simulate light intensity measurements as observed from an RGB camera, we take the reference T_s fields obtained in Step 2, and then compute the projections according to the following steps:

i. From each synthetic T_s and f_s field in the augmented dataset, we obtain the I_λ^{bb} and $\kappa_{\text{abs},\lambda}$ fields using Equation (2.5) and Equation (2.3) respectively. We then solve the integral in Equation (2.6) to obtain the corresponding simulated spectral flame emissions I_λ. We perform this step for multiple wavelengths λ to simulate the wide spectrum captured by an RGB camera.

ii. We perform a second integration over multiple wavelengths λ as described in Equation (2.8). These wavelengths are selected to represent the spectrum captured by the RGB camera (400 nm to 700 nm), and are weighted by the sensitivity extracted from the characteristics of the optical system shown in Figure 3.6. From these calculations we obtain the projected broadband soot emissions $P_{\{R,G,B\}}$.

It is important to note that although the BEMI technique neglects the effects of self-absorption in the flame, our simulations do account for this effect to provide a better representation of the real phenomena.

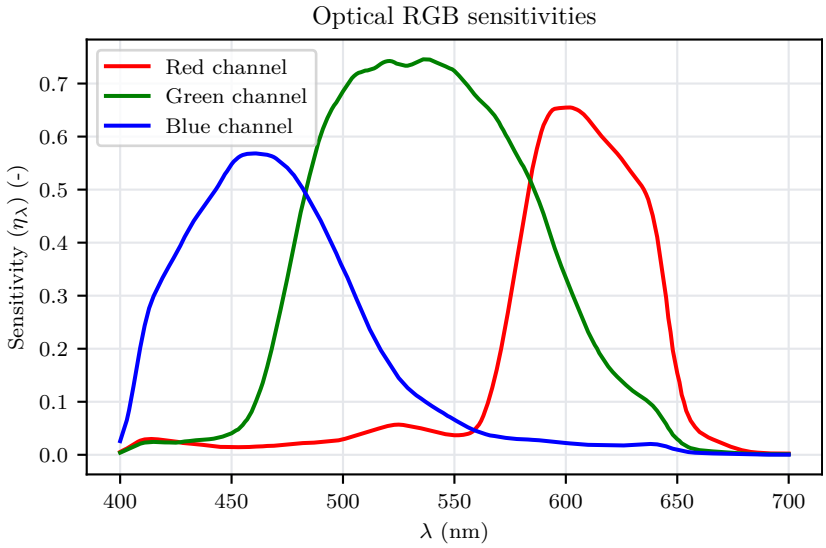


Figure 3.6: Sensitivity of the optical sensor for the red, green and blue channels.

3.2.2 ANN training

Using the previously described steps, we generate the dataset for both the characterization of the soot volume fraction and soot temperature fields. To improve the representation of different capture conditions and distortions in the experimental equipment, we added noise to both dataset proportional to the intensity of the input signals for each case:

- For the f_s dataset we added zero-mean Gaussian noise to the τ_λ fields, where the standard deviation of the noise for each image is drawn from a uniform distribution from 0 to 1% of the maximum signal value.
- For the T_s dataset we added a zero-mean Gaussian noise to the $P_{\{R,G,B\}}$ values of the generated images. To make the noise level proportional to a representative quantity for the entire dataset, we set the variance of the noise proportional to the average maximum value of the blue channel for all images in the dataset (the blue channel has the least average intensity).

  Each image in the training and validation subsets is subject to a different noise level, with variances uniformly distributed between 0% and 1% of the maximum value of P_B. We then normalized the $P_{\{R,G,B\}}$ fields on each image by their maximum value in the red channel (the red channel has the highest maximum intensity), so that the pixel intensities are bounded between 0 and 1

To train and evaluate the network, we generate for each case an input tensor $\mathbf{X}$ and an output tensor $\mathbf{Y}$, consisting of all of the input-output pairs of each dataset. To prevent instabilities in the training process, we standardized the input and output tensors by their means $\mu_\mathbf{X}, \mu_\mathbf{Y}$ and standard deviations $\sigma_\mathbf{X}$, $\sigma_\mathbf{Y}$ as

$$\bar{\mathbf{X}} = \frac{\mathbf{X} - \mu_\mathbf{X}}{\sigma_\mathbf{X}}, \quad \bar{\mathbf{Y}} = \frac{\mathbf{Y} - \mu_\mathbf{Y}}{\sigma_\mathbf{Y}}. \tag{3.10}$$

Additionally, to train the network as is common in supervised learning [34], we separate the dataset into three different subsets following commonly used partitions:

- The training set contains the examples used in the iterative training process that adapts the internal network parameters for a fixed network architecture. This set must be large enough to be statistically representative of the entire input space.

- The validation set is used to evaluate the performance of the model after a training pass. The performance over the validation set may provide information to adjust the network hyperparameters if necessary. After adjusting the hyperparameters, a new training pass must be performed.
- The test set contains samples not used during the network design and training. This set is used to assess the performance once the network architecture is fully defined and trained. We do not add noise to the test set during the dataset generation stage, instead we add different noise levels during the evaluation of the different methodologies to asses their performance under different conditions.

It is important to note that given that the different subsets are extracted from the same distribution of synthetically generated measurements, the sizes of the validation and test sets do not greatly affect the generalization capabilities of our network; instead, they are mainly used as a sanity check. The actual performance assessment of the network is done by comparing the retrieved fields with the ones obtained with classical deconvolution methods in synthetic and experimental data in Chapter 4.

For the multi-dimensional regression problem of computing the soot volume fraction and soot temperature fields from convoluted measurements, we use the Mean Squared Error (MSE) as our cost function $\mathcal{J}(\boldsymbol{\theta})$, defined as

$$\mathcal{J}(\boldsymbol{\theta}) = \text{MSE}\left(\hat{\mathbf{Y}}(\mathbf{X};\boldsymbol{\theta}), \mathbf{Y}\right) = \text{mean}\left\{\left(\hat{\mathbf{Y}}^{(i)} - \mathbf{Y}^{(i)}\right)^2\right\}, \tag{3.11}$$

which quantifies the average square difference between the correct output tensor $\mathbf{Y}$ and the predicted output tensor $\hat{\mathbf{Y}}$ for the current value of the set of parameters $\boldsymbol{\theta}$ defined by the network architecture. The operator mean$\{\cdot\}$ represents the average value over all dimensions of the tensor. Finally, we implement the network using the Keras API [55], and trained the model under the TensorFlow framework [56] using the Adam optimizer [57] as the training algorithm. The Adam optimizer is a widely used stochastic version of the gradient descent algorithm described in Equation (3.4).

4 Results

This chapter compares the results obtained when retrieving soot characteristics using traditional methodologies versus our proposed ANN-based method. This chapter is divided into two sections, in Section 4.1 we discuss the results obtained when predicting the f_s field from τ_λ measurements, and in Section 4.2 we discuss the results when predicting the T_s field from $P_{\{R,G,B\}}$ measurements. In both cases, we will first discuss the results obtained when predicting the corresponding soot characteristic from simulated signals and then from an experimental campaign.

During the development of this, we also evaluated other ANN architectures and different MLP configurations similar to the ones used by other authors when approaching similar deconvolution problems in combustion. The results obtained using an MLP architecture are similar to those obtained with U-Net. However, MLP implementations described in the literature for the characterization of soot properties only predict at the profile level (line-by-line) [27, 28, 29, 30, 31] and therefore, as discussed before, this implementations might not be well suited for spatially structured data such as images captured with a camera. Because of this, the different lines predicted with an MLP might look slightly skewed, with a lack of correlation between contiguous rows of the image and worse structural similarity than those obtained using a model based on a CNN.

We have not yet published the results of an in-depth comparison between different ANN architectures and experimental capture conditions; therefore, we will limit the results presented in this chapter to the ones already published at the time of the presentation of this thesis. We limit the results presented to the comparison between onion-peeling deconvolution with Tikhonov regularization and our ANN approach based on the U-Net architecture.

4.1 Soot volume fraction

This section compares the results obtained when retrieving the soot volume fraction field f_s in a canonical Gülder CLAD flame from LOSA measurements of the fraction of transmitted light τ_λ in both synthetic and experimental measurements. For these experiments, we consider a laser wavelength of $\lambda = 800$ nm.

4.1.1 Synthetic dataset

Using the framework described in Section 3.2 for the generation of f_s and τ_λ pairs, we generated a dataset of 6000 entries of synthetic fields, from which we took 50 pairs as the test set and split the remaining entries in 80% for the training set and 20% for the validation set. Using the training and validation sets, we trained the ANN described in Section 3.1.3, and evaluated the performance of the different deconvolution methodologies using the test set.

In this section we first show examples of the simulated τ_λ fields, and then compare the groundtruth soot volume fraction field (f_s GT) with multiple predicted f_s fields using onion-peeling deconvolution with Tikhonov regularization (f_s OP) for different values of the regularization parameter α. Finally, we compare the f_s OP field using the best-performing value of α with the results obtained from our ANN-based methodology (f_s ANN).

Table 4.1: Mean Absolute Error for the predictions of the soot volume fraction f_s field using onion-peeling deconvolution with Tikhonov regularization ($\alpha = \{0, 10^{-6}, 10^{-5}, 10^{-4}\}$) and our proposed ANN-based method.

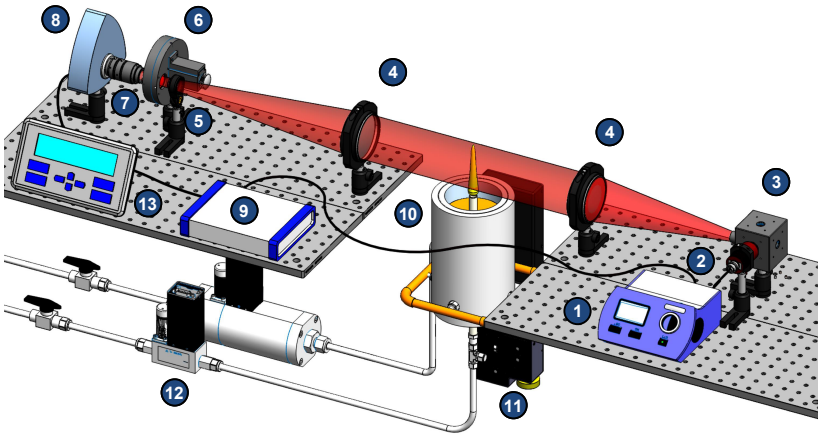
Method	Mean Absolute Error
f_s OP $\alpha = 0$	0.4846 ppm
f_s OP $\alpha = 10^{-6}$	0.1252 ppm
f_s OP $\alpha = 10^{-5}$	0.1146 ppm
f_s OP $\alpha = 10^{-4}$	0.2794 ppm
f_s ANN	0.0308 ppm

Figure 4.2 shows an example from the test set of the τ_λ field with 0%, 0.25%, 0.5% and 1% of added noise. We can observe the visual degradation of the input signals, which represents different experimental conditions in the synthetic dataset. For all of the following results we will consider a noise level of 0.5%, as we found this level of noise to be representative of the noise perceived during experimental measurements.

Figure 4.3 shows the groundtruth soot volume fraction field, and predicted fields for different values of the regularization parameter α for three examples of the test set. Figure 4.4 shows the same signals for three different examples of the test set at a height above burner (HAB) of $z = 2.8$ cm. Using the L-Curve criterion [18], we found a base value of α that was then tuned to obtain a value that provides enough smoothing without affecting the general shape of the predicted profiles, arriving at a value of $\alpha = 10^{-5}$. From the results presented in the figures, we note that the reconstructed profile without regularization ($\alpha = 0$) amplifies the experimental noise contained in the simulated τ_λ measurements, while $\alpha = 10^{-6}$ shows under-regularization as a noisy output and $\alpha = 10^{-4}$ shows over-regularization as an over-smoothed f_s field. The following results will use a value of $\alpha = 10^{-5}$, as it shows an adequate balance between smoothness and accuracy in the reconstructed f_s fields.

Figure 4.5 shows the input fields τ_λ for three examples of the test set, as well as the groundtruth soot volume fraction fields f_s GT, the prediction used traditional methodologies f_s OP with the previously selected regularization parameter ($\alpha = 10^{-5}$) and the predictions of our method f_s ANN. Figure 4.6 and Figure 4.7 show the f_s profiles at a HAB of 2.8 cm and 2.0 cm respectively. Figure 4.8 shows the f_s profiles along the flame centerline and Figure 4.9 shows the f_s profile along the path of maximum soot concentration. We can see that in general our proposed method predicts f_s fields that are closer to the groundtruth than those obtained by the traditional methodology. Additionally, since onion-peeling deconvolution is applied independently to each line of the τ_λ field along the z-axis, the retrieved fields are slightly skewed as they contain horizontal shifts in the transition between rows of the resulting images, while in our methodology we do not see these shifts as we use a CNN that considers the 2D geometry of the flame. We also note that, as shown in Figure 4.8 and Figure 4.9, our methodology provides notably better predictions for the centerline of the flame and its path of maximum soot concentration along the z-axis, these areas are of particular interest to the combustion community for validating numerical soot models and are known to be difficult to characterize [58].

To provide a global perspective of the results, Figure 4.10 shows the comparison of the mean value over all the predicted fields for the groundtruth versus the retrieved values of f_s for each entry of the test set, with the black line representing a perfect reconstruction. Additionally, to summarize the results, Table 4.1 shows the Mean Absolute Error for the predictions using f_s OP and f_s ANN. We also provide the Mean Absolute Error for other values of α to validate the value chosen for this parameter. We compute the Mean Average Error over all the values of each field in the test set, and then average these results to obtain the MAE over the entire set. From these results, we can conclude that the proposed method has a lower Mean Average Error than the predictions using onion-peeling deconvolution with Tikhonov regularization.



1. LED driver
2. LED source
3. Integrating sphere
4. Achromatic lens
5. ND filters
6. Spectral filter wheel
7. Camera lens
8. Andor CCD camera
9. Pulse generator
10. Burner
11. Translation stage
12. Flowmeters
13. Flow readout module

Figure 4.1: Schematic of the optical arrangement for the two-dimensional line-of-sight attenuation (LOSA) technique. This schematic was created by members of the EC^{2}G, and is being used with their permission.

4.1.2 Experimental measurements

We performed an experimental campaign of LOSA measurements to evaluate our model. We established a CLAD flame on a Gülder type burner [47] by delivering 0.194 slpm of pure ethylene and 284 slpm of airflow (21.1°C, 1 bar). Figure 4.1 shows the experimental setup used for the extinction measurements using LOSA. An LED coupled to an integrating sphere provides a diffused light source, which is collimated by an achromatic lens and directed to the center of the burner. The transmitted light that passes through the medium is converged by a second achromatic lens and captured by a Navitar lens ($f2.8/50$ mm) coupled to an Andor Luca R CCD camera with a resolution of 1004×1002. We used neutral density filters to avoid camera saturation and band-pass filters to reduce the noise in the signals. We took 150 images at a detection wavelength $\lambda = 800 \pm 5$ nm using an exposure time of 0.01 s, and then computed the fraction of transmitted light as described in Section 2.1.1 using the average of the captured images to reduce high-frequency noise.

Figure 4.11 shows the reconstructed f_s fields generated from the experimental signals using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our proposed methodology. Additionally, Figure 4.12 shows the profiles of the predicted fields at a HAB of 2.8 cm, HAB of 2.0 cm and along the flame centerline. Although we do not have a groundtruth value to determine the exact quality of the predictions, we can comment on the general trends and shapes of the predicted fields and profiles. We can see that the profiles reconstructed with our ANN model are similar to the ones reconstructed using OP deconvolution, with the wings (areas of maximum soot concentrations) of the flame being slightly bigger in the case of the ANN. Additionally, we can see that OP deconvolution provides noisier results along the flame centerline, this could be due to the presence of more measurement noise in these regions and the fact that OP deconvolution works in a line-by-line basis and thus does not maintain the correlation between different heights of the flame.

4.2 Soot temperature

This section compares the results obtained when retrieving the soot temperature field T_s of a canonical Gülder CLAD flame from BEMI signals using onion-peeling deconvolution with Tikhonov regularization and our proposed ANN-based method.

Table 4.2: Mean Absolute Error for the predictions of the soot temperature T_s field using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our proposed ANN-based method.

Method	Mean Absolute Error
T_s BEMI-OP $\alpha = 10^{-5}$	19.6626 K
T_s BEMI-ANN	3.5937 K

4.2.1 Synthetic dataset

Using the framework described in Section 3.2 for the generation of T_s and BEMI signals ($P_{\{R,G,B\}}$), we generate a dataset of 8000 entries of synthetic fields, from which we took 100 pairs as the test set and split the remaining entries in 80% for the training set and 20% for the validation set. Using the training and validation sets, we trained the ANN described in Section 3.1.3 and evaluated the performance of the different deconvolution methodologies using the test set.

In this section, we will first show examples of the simulated BEMI signals $P_{\{R,G,B\}}$ as captured by an RGB camera, and then compare the groundtruth soot temperature field (T_s GT) with the T_s fields predicted using the traditional methodology of onion-peeling deconvolution with Tikhonov regularization (T_s BEMI-OP) and the results obtained from our ANN-based methodology (T_s BEMI-ANN). To simulate similar noise levels to those observed during experimental campaigns, we set the noise level of the $P_{\{R,G,B\}}$ signals in the test set such that the variance of the added noise is fixed at 0.25% of $\hat{P}_{B,\max}$. Figure 4.13 shows the normalized $P_{\{R,G,B\}}$ fields for three examples of the test set after adding the noise.

Using a similar analysis to the one done for f_s in Section 4.1, we find that a value of $\alpha = 10^{-5}$ represents a good basis for the comparison of BEMI-OP with our method BEMI-ANN. Figure 4.14 shows the groundtruth soot temperature fields T_s GT, the predictions using onion-peeling deconvolution T_s BEMI-OP and the predictions using our method T_s BEMI-ANN for three examples of the test set. Figure 4.15 and Figure 4.16 show the predictions for the T_s profiles at a HAB of 2.8 cm and 1.85 cm respectively. Figure 4.17 shows the T_s profiles along the flame centerline and Figure 4.18 shows the T_s profiles along the path of maximum soot concentration (path with maximum f_s for each point in the z-axis).

We can see that, in general, BEMI-ANN provides smoother and more accurate reconstructions for the T_s fields. We also see shifts between rows in the results of BEMI-OP similar to those described in Section 4.1, these horizontal shifts are generated due to the lack of correlation between rows of the predicted T_s field when using BEMI-OP. Adittionally, the predictions using BEMI-OP have a negative offset when compared to the groundtruth as expected by not considering the soot self-absorption and scattering.

The negative offset from the groundtruth in BEMI-OP is an expected consequence derived from the use of a simplified model that neglects the effect of soot self-absorption in the optical path [59]. BEMI-ANN captures the relationship between the inputs and outputs of the training set generated using our framework, without requiring an assumption about the underlying model and thus does not present this negative offset. In this sense, the results generated using BEMI-ANN appear more similar to the results obtained when using Broadband Modulated Absorption and Emission (BMAE), a more complex technique that also requires the calculation of the soot volume fraction f_s to account for the soot self-absorption. The absence of this negative offset in the predictions using BEMI-ANN is a promising feature that could be exploited to obtain predictions with a similar quality to BMAE, while avoiding the need for LOSA measurements to determine f_s. However, these results need more scrutiny, and we have not yet published an in-depth comparison between all these techniques, so we will continue the presentation of the results by comparing exclusively BEMI-OP and BEMI-ANN.

Figure 4.19 shows the mean value over each example of the test set for the groundtruth versus the predicted T_s fields, with the black line representing a perfect reconstruction. Table 4.2

shows the Mean Absolute Error for the predictions using BEMI-OP and BEMI-ANN. From the table we can conclude that the BEMI-ANN has a lower Mean Absolute Error than BEMI-OP, and thus provides closer reconstruction of the T_s fields.

4.2.2 Experimental measurements

We conducted an experimental campaign using a CLAD flame generated on a Gülder burner [47], delivering 194 slpm of pure ethylene and 284 slpm of airflow (21 °C, 1 bar). We collected the radiation emitted by soot particles that pass through an achromatic lens of focal distance $f = 400$ mm, and then by a spectral filter Thorlabs F65600 in the visible range. We used a Basler acA3088-16gc camera with a Sony IMX178 CMOS sensor and a 35 mm Thorlabs MVL35M23 lens with an aperture of $F/11$. The images were recorded in 12 bit TIFF format using a pixel format Bayer RG 12 with a resolution of 3088×2064 pixels and using an exposure time of 4 ms. We averaged multiple measurements in a controlled environment to reduce distortions from high-frequency ambient noise.

Figure 4.20 shows the T_s fields retrieved from the experimental measurements using BEMI-OP and BEMI-ANN. When evaluating the performance of a new inversion technique for soot pyrometry over laboratory measurements, it is common to first verify that the estimations stay in the same range compared to the ones obtained with a known technique, and then evaluate the improvements in terms of smoothness and representation of expected physical behavior in regions of interest [58, 60]. Similar to what we observed in the simulated data, the T_s field retrieved using BEMI-OP presents an irregular behavior in the center region of the flame, showing discontinuities along the z-axis. On the other hand, the temperature field retrieved using BEMI-ANN is smoother and shows higher values toward the edge of the flame. Figure 4.21 shows single-line profiles retrieved from the experimental image, whose behavior is consistent with the results obtained from the simulated dataset. The first two plots show the retrieved radial profiles at a HAB of 2.8 cm and 1.85 cm respectively, where we see that both estimations follow a similar trend, but BEMI-OP provides results with more irregular and lower values for the estimated temperature across the radial axis. The third plot shows the axial profile along the centerline of the flame, in which we see that BEMI-ANN provides a smoother profile for the flame centerline, which is a region of particular interest to understand soot formation.

It is important to note that the ML model used in BEMI-ANN will estimate temperatures using the inversion model implicitly inferred during training with simulated data, without requiring explicit assumptions about the physical process. This means that the performance of our model is dictated by the quality of the data used during the training stage. Therefore, if the computational model that generates the simulated data is representative of experimental CLAD flames, we expect the fields obtained with BEMI-ANN to be closer to the actual soot temperature.

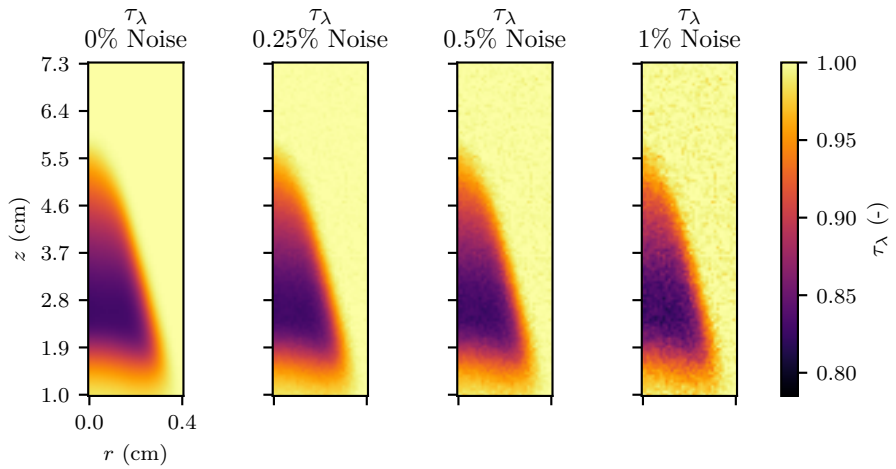


Figure 4.2: Synthetic τ_λ fields with different level of added noise. Three examples.

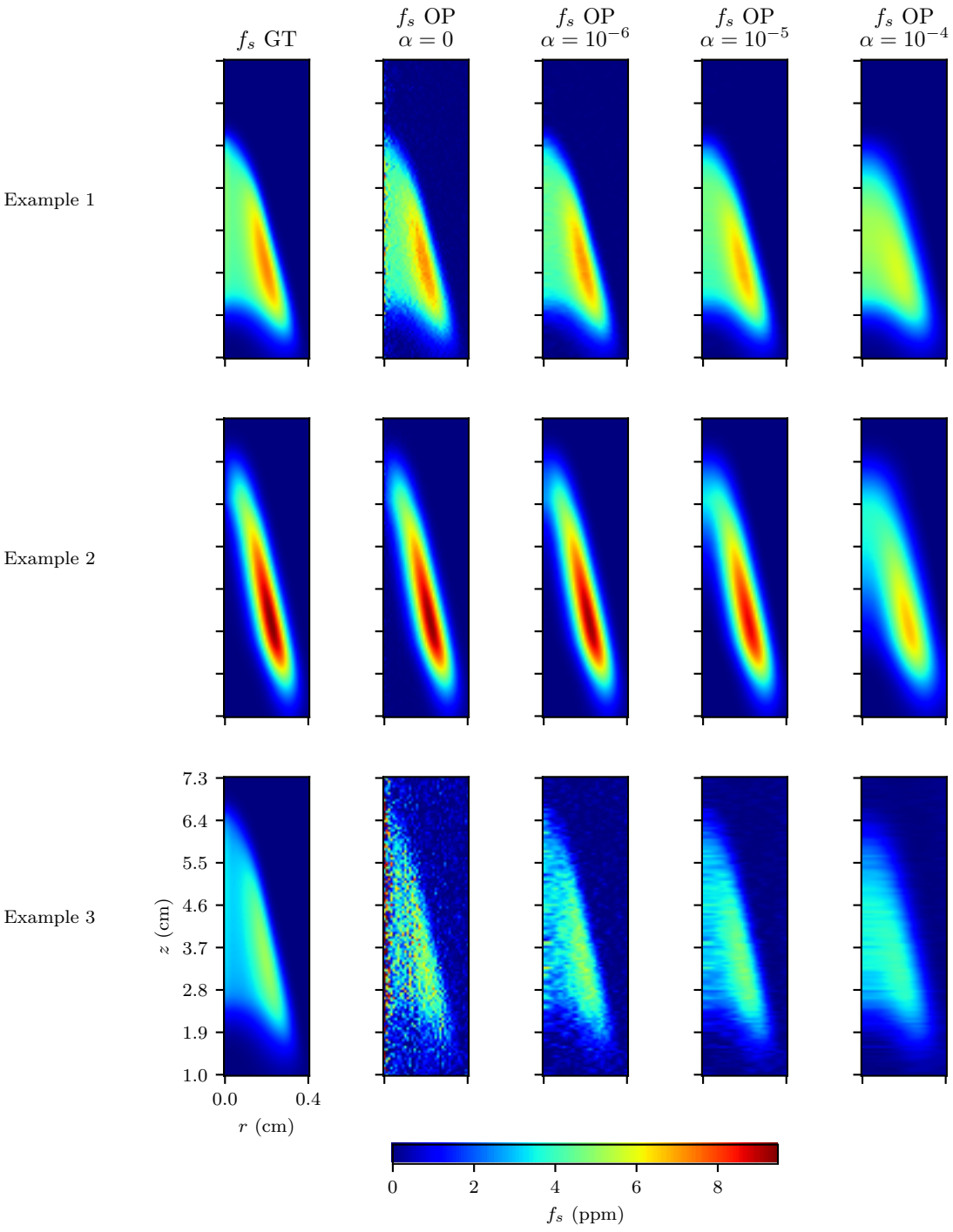


Figure 4.3: Groundtruth and predicted f_s fields using onion-peeling deconvolution with Tikhonov regularization and different values for the regularization parameter α. Three examples.

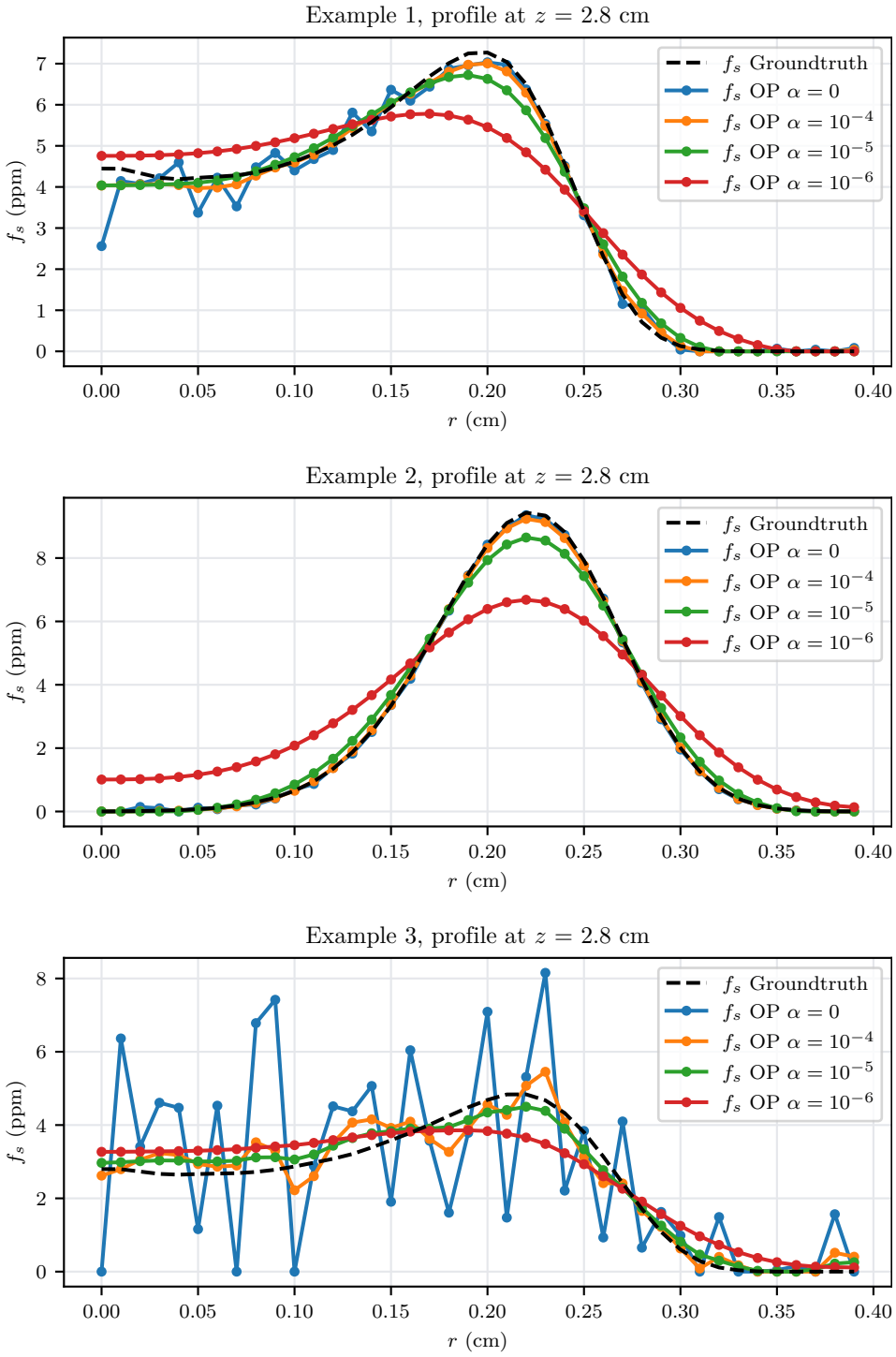


Figure 4.4: Groundtruth and predicted profile of f_s at a HAB of $z = 2.8$ cm using onion-peeling deconvolution with Tikhonov regularization and different values for the regularization parameter α. Three examples.

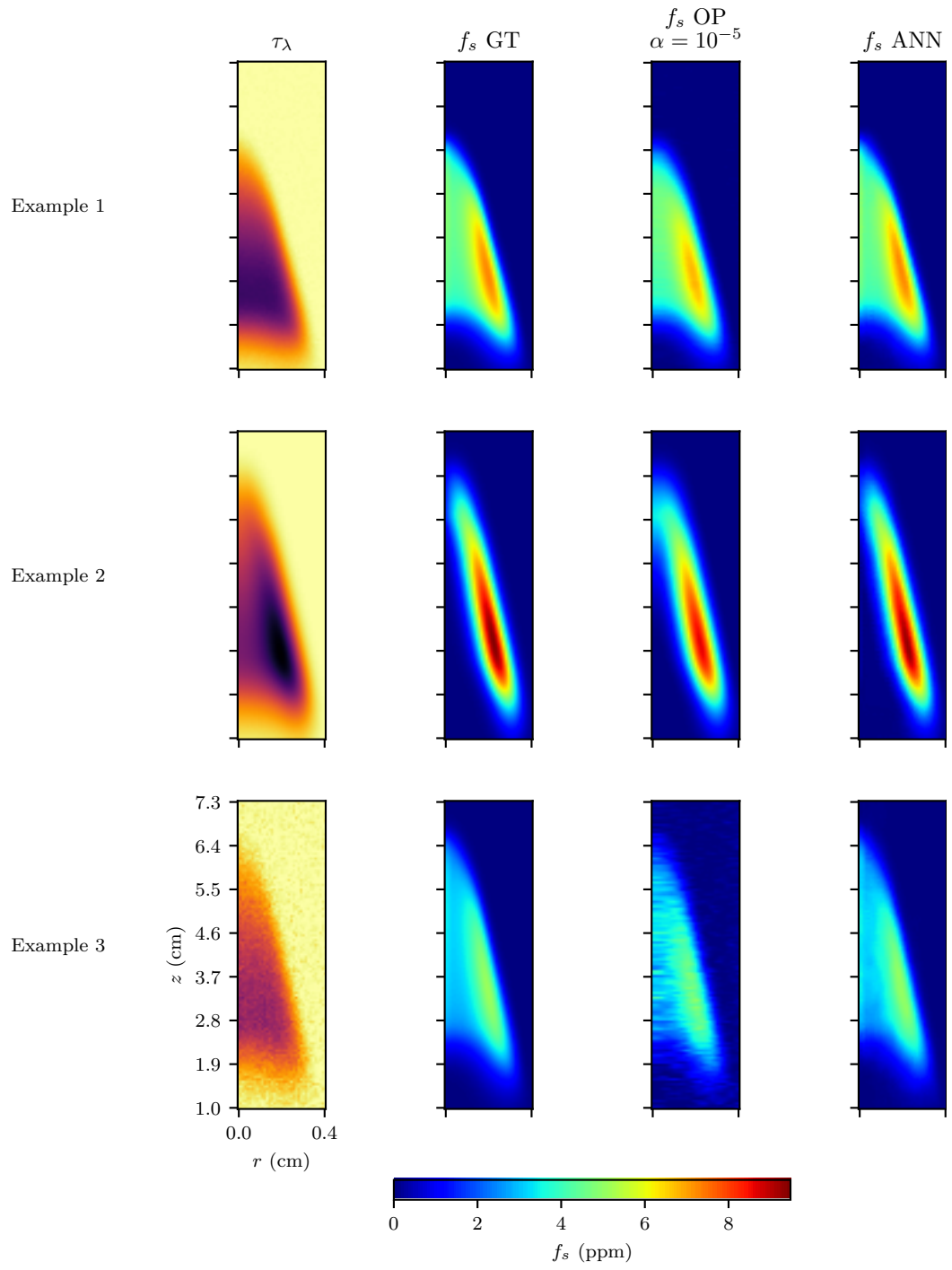


Figure 4.5: Groundtruth and predicted f_s fields using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method. Three examples.

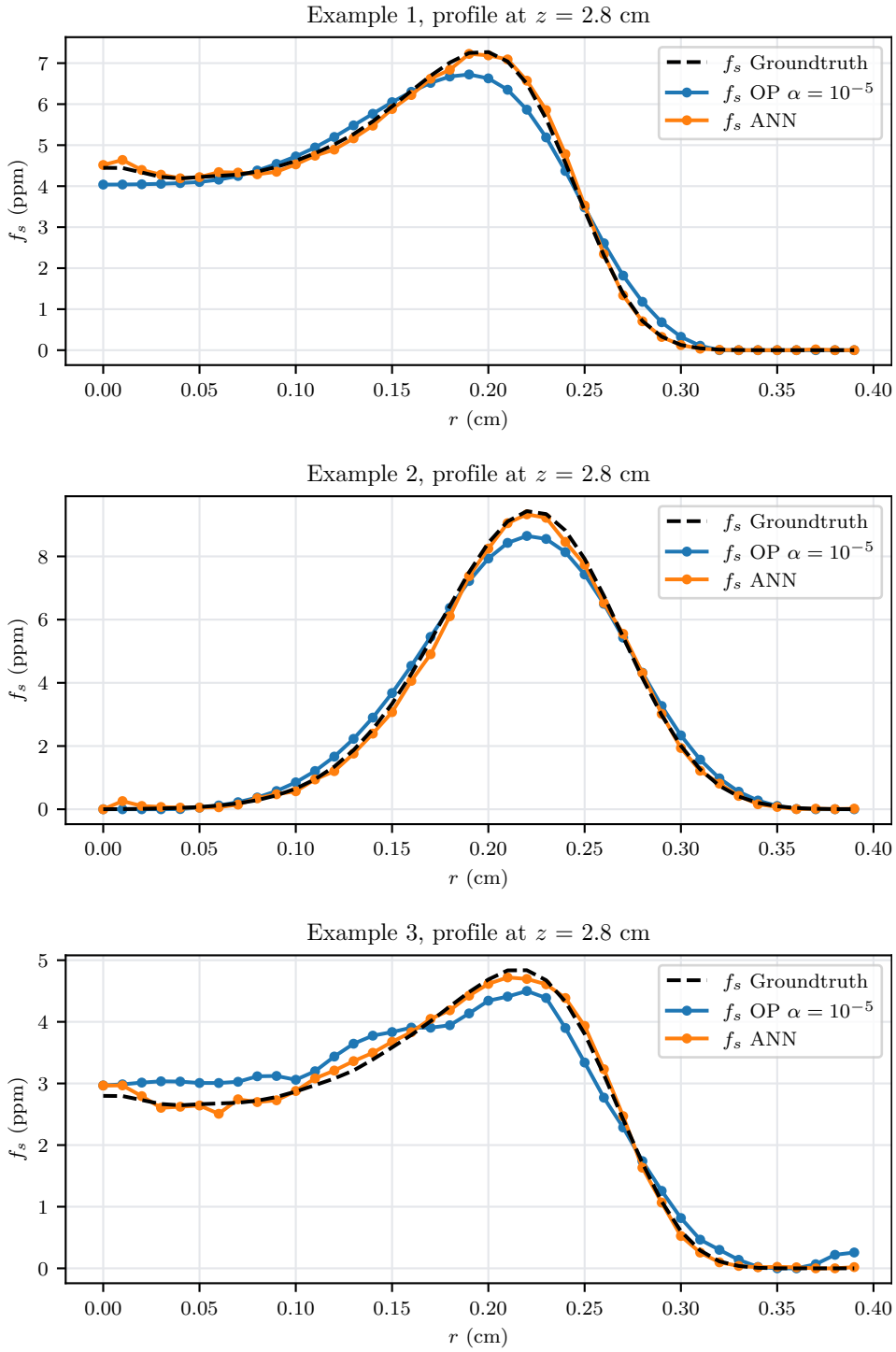


Figure 4.6: Groundtruth and predicted profile of f_s at a HAB of $z = 2.8$ cm using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method. Three examples.

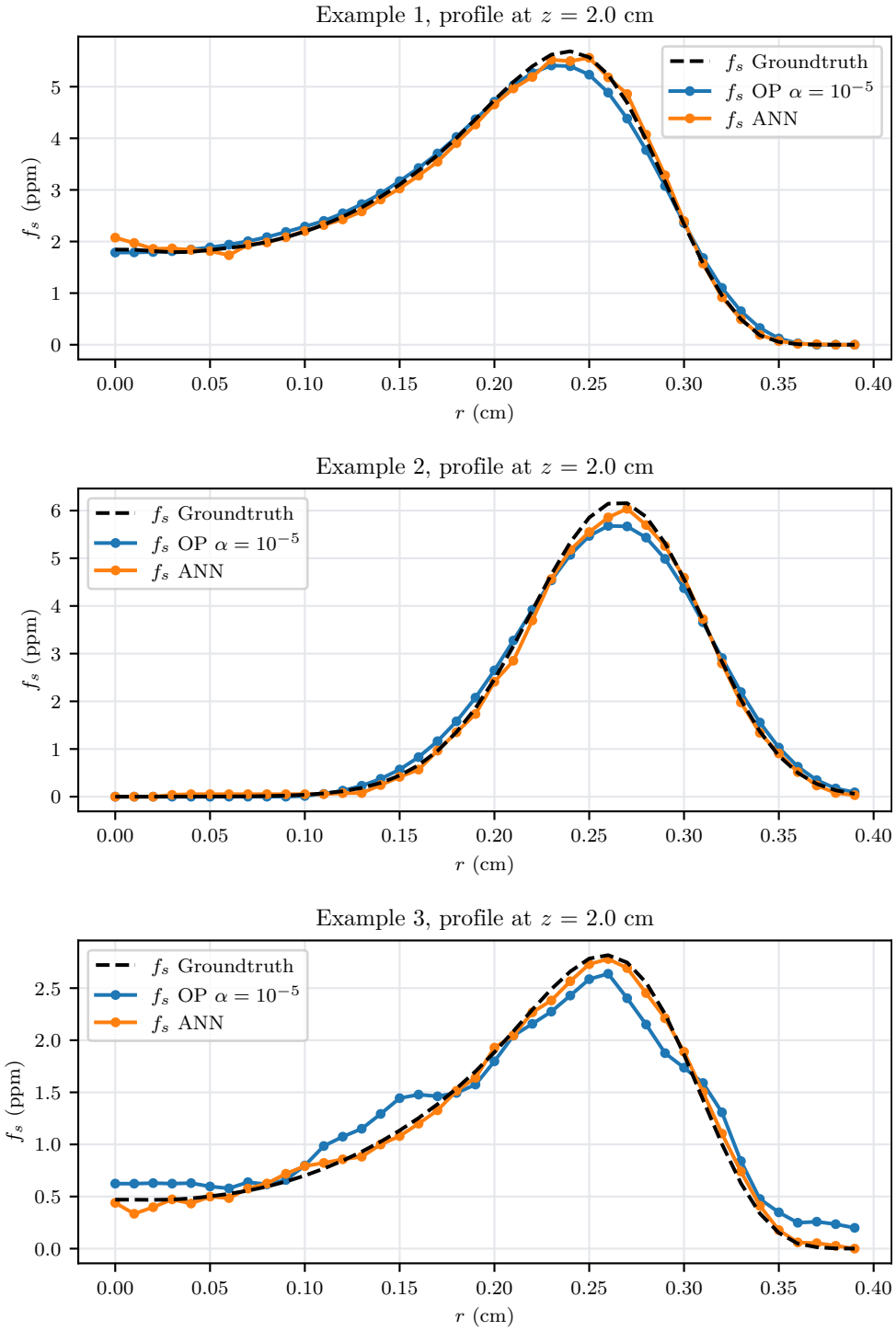


Figure 4.7: Groundtruth and predicted profile of f_s at a HAB of $z = 2.0$ cm using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method. Three examples.

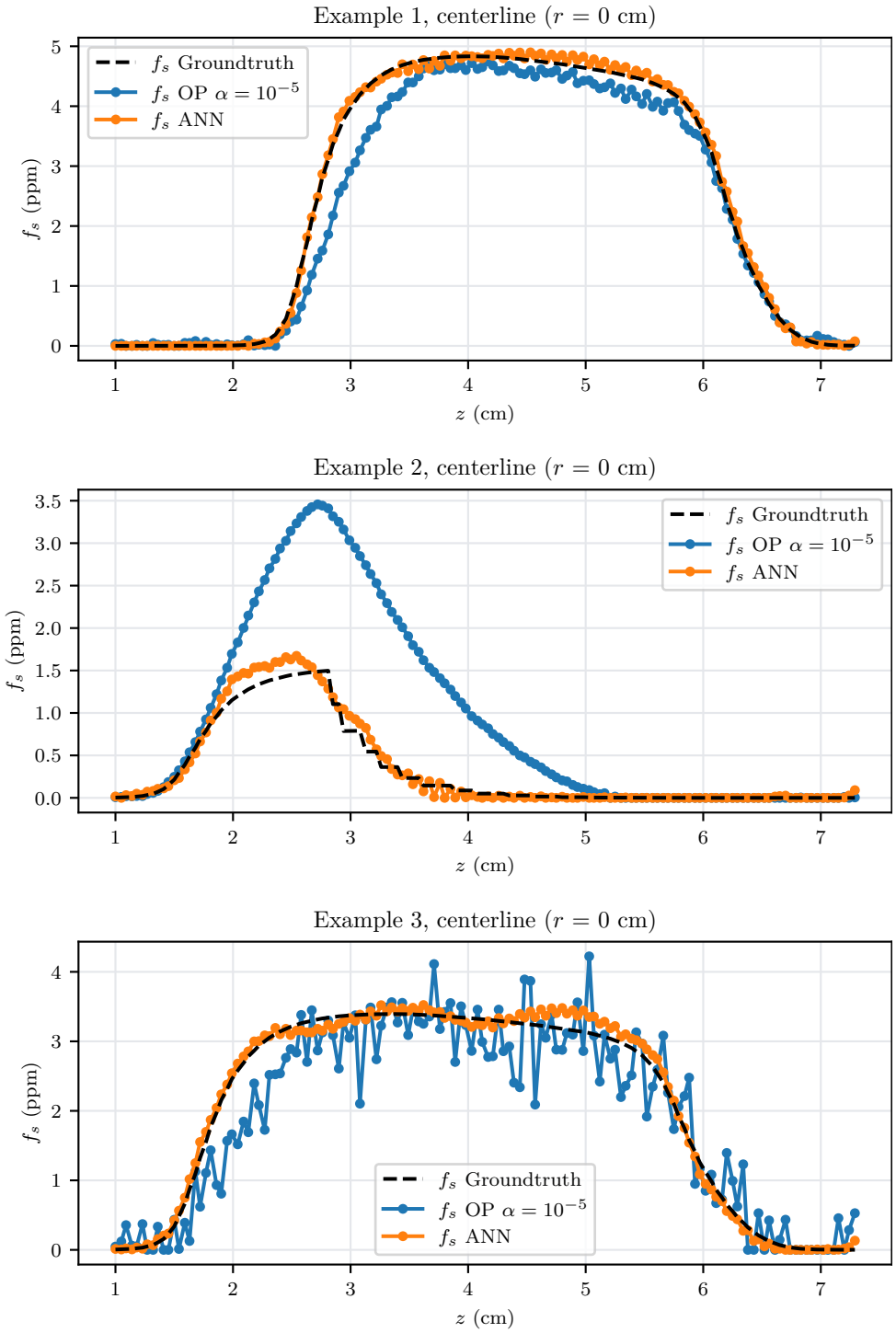


Figure 4.8: Groundtruth and predicted profile of f_s along the flame centerline ($r = 0$ cm) using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method. Three examples.

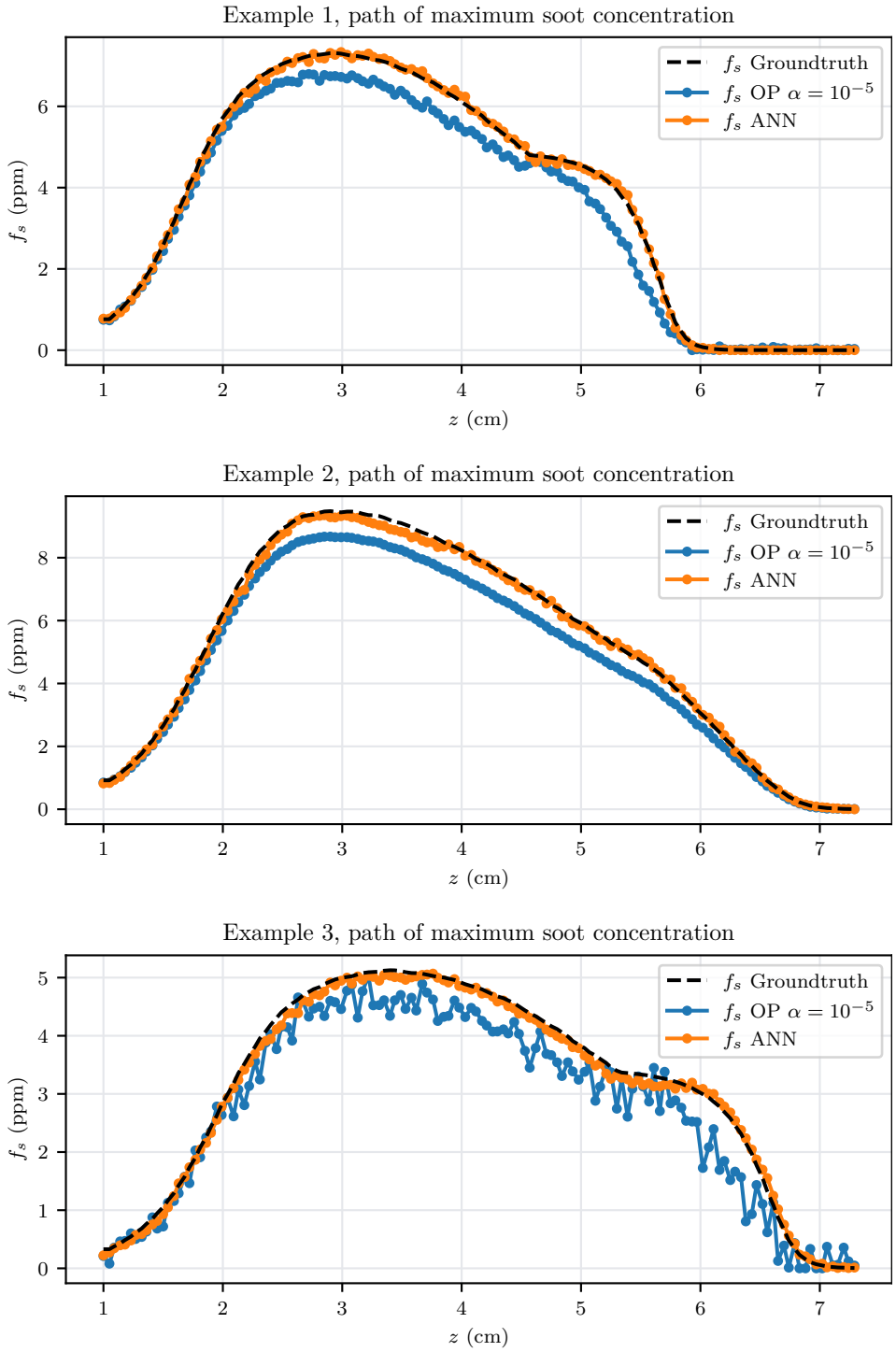


Figure 4.9: Groundtruth and predicted profile of f_s along the path of maximum soot using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method. Three examples.

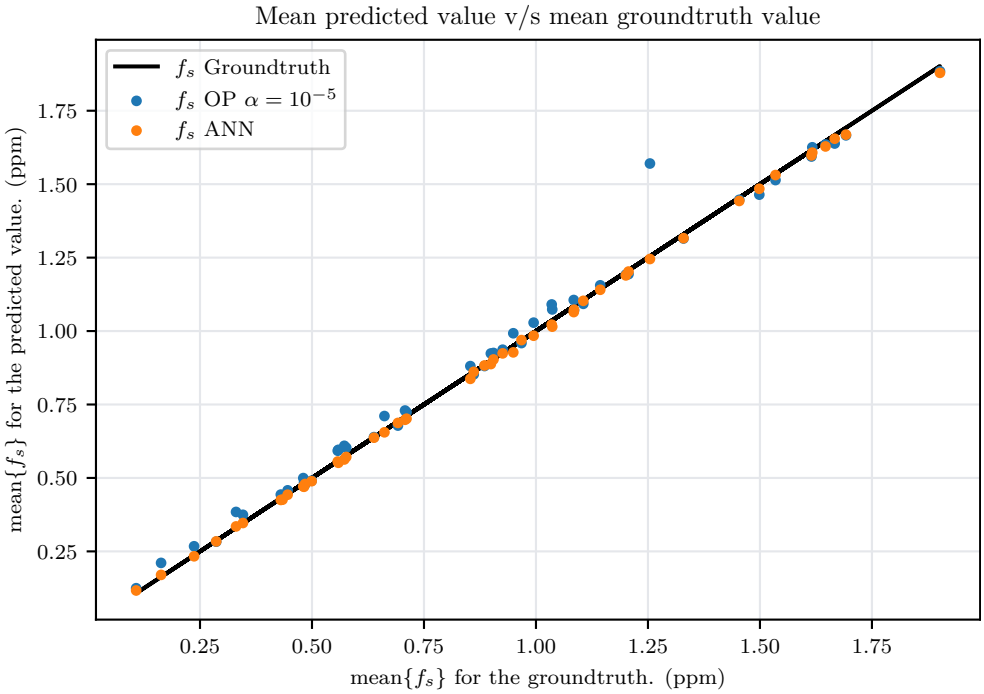


Figure 4.10: Comparison between the mean predicted value and the mean groundtruth value for the f_s fields over the entire test set using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method.

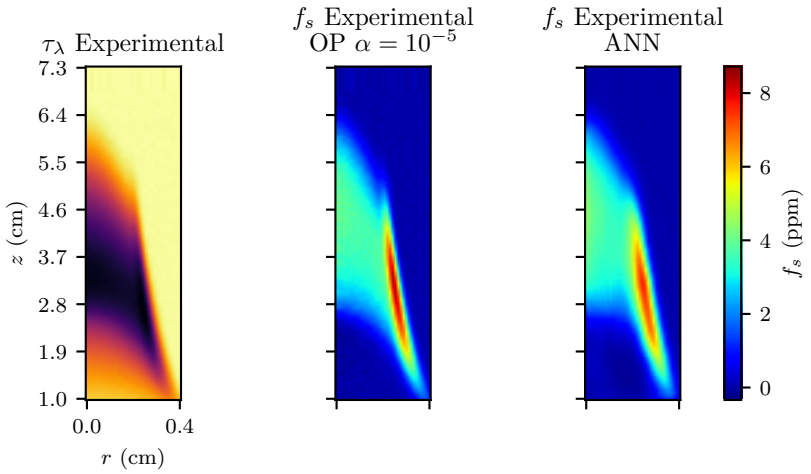


Figure 4.11: τ_λ captured during an experimental campaign and predicted f_s field using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method.

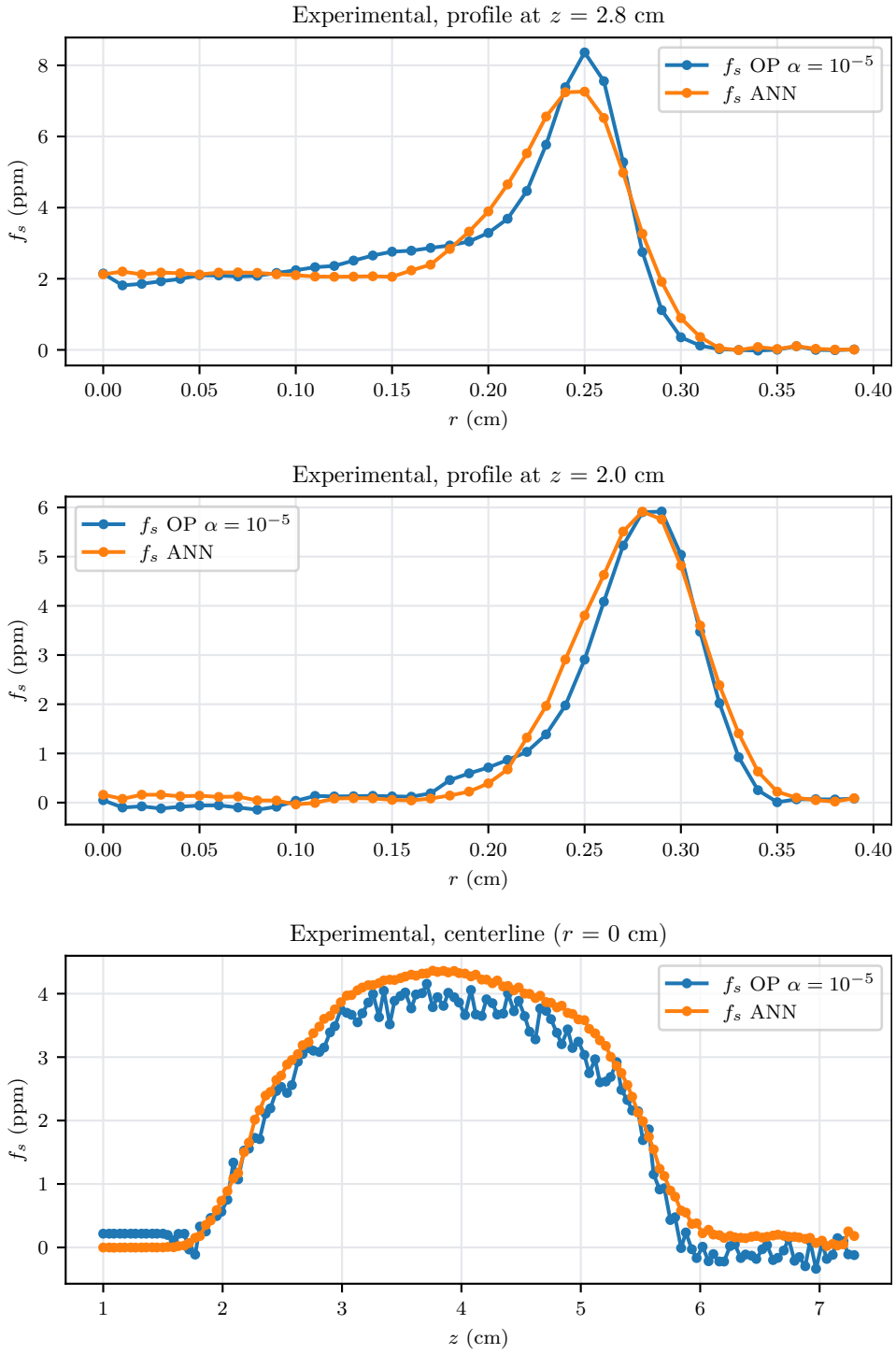


Figure 4.12: Predicted f_s profiles using onion-peeling deconvolution with Tikhonov regularization ($\alpha = 10^{-5}$) and our ANN-based method for data captured during an experimental campaign. Profiles at a HAB of $z = 2.8$ cm, HAB of $z = 2.0$ cm and along the flame centerline ($r = 0$ cm).

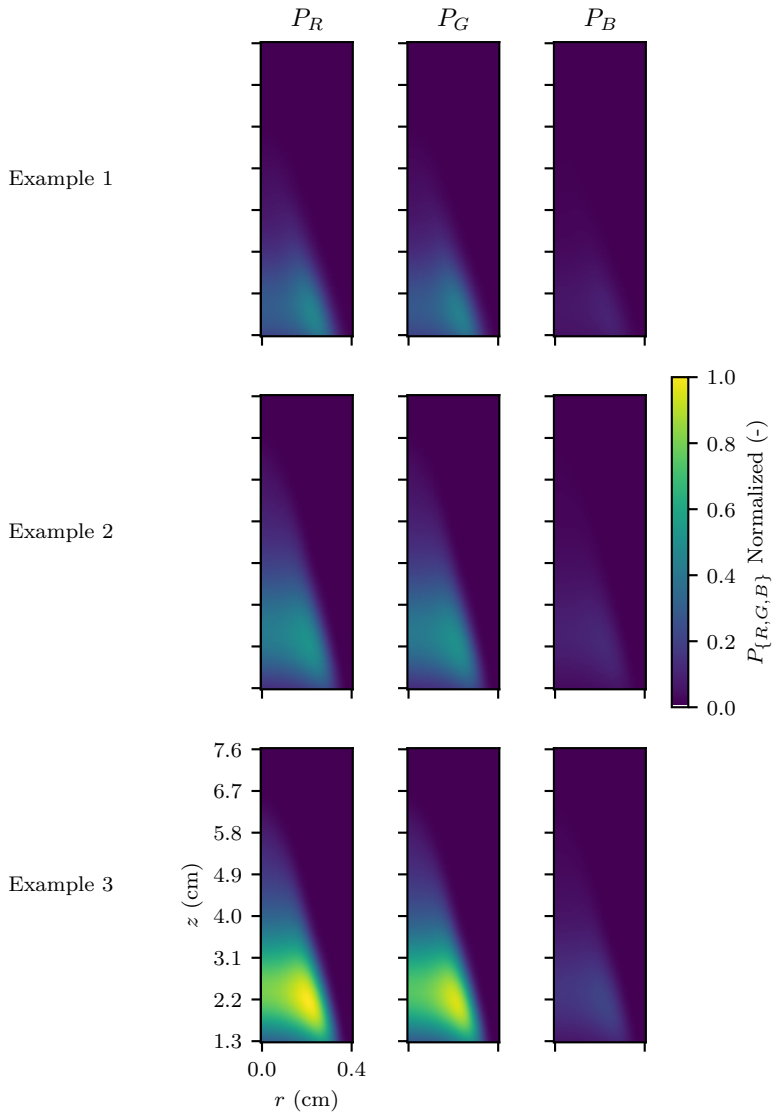


Figure 4.13: Synthetic normalized $P_{\{R,G,B\}}$ fields. Three examples.

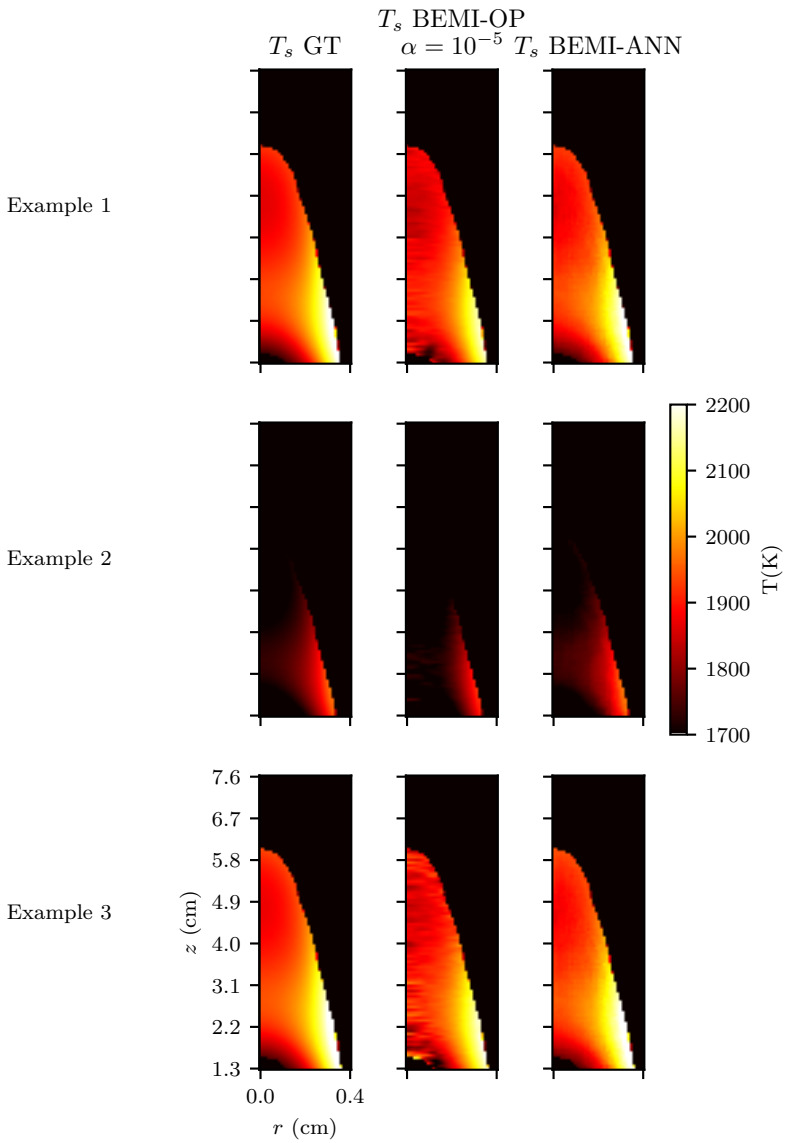


Figure 4.14: Groundtruth and predicted T_s fields using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN. Three examples.

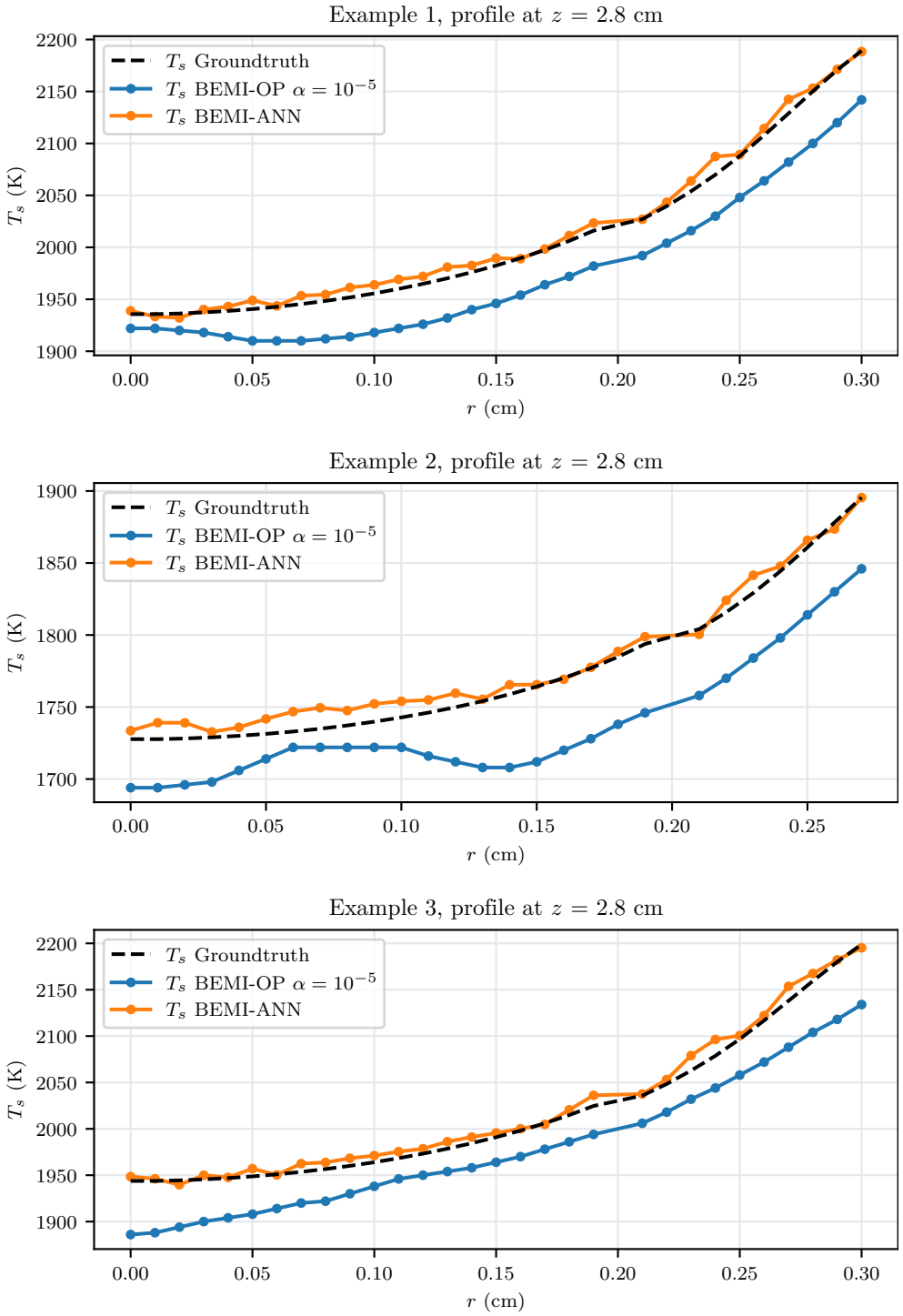


Figure 4.15: Groundtruth and predicted profile of T_s at a HAB of $z = 2.8$ cm using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN. Three examples.

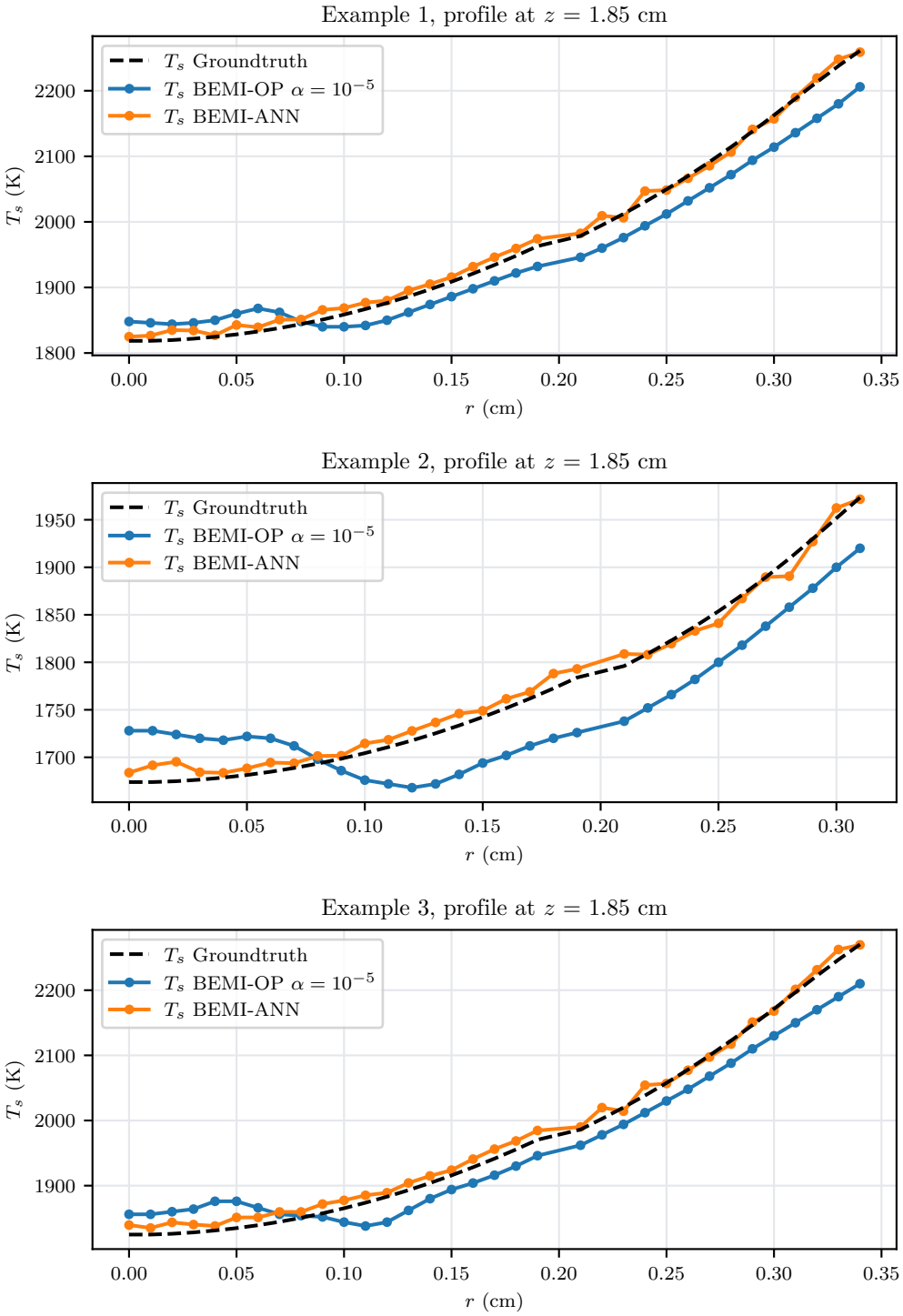


Figure 4.16: Groundtruth and predicted profile of T_s at a HAB of $z = 1.85$ cm using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN. Three examples.

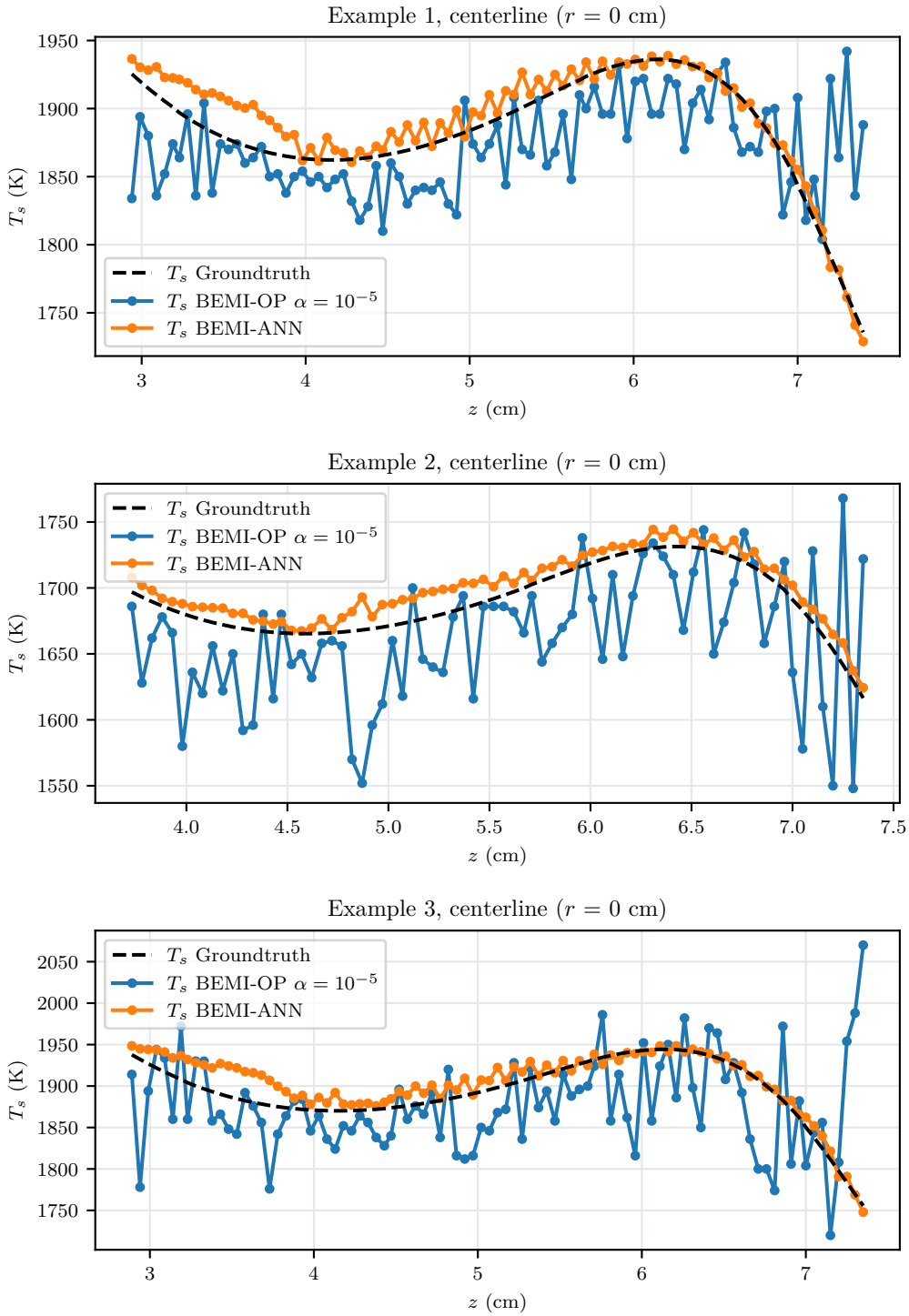


Figure 4.17: Groundtruth and predicted profile of T_s along the flame centerline ($r = 0$ cm) using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN. Three examples.

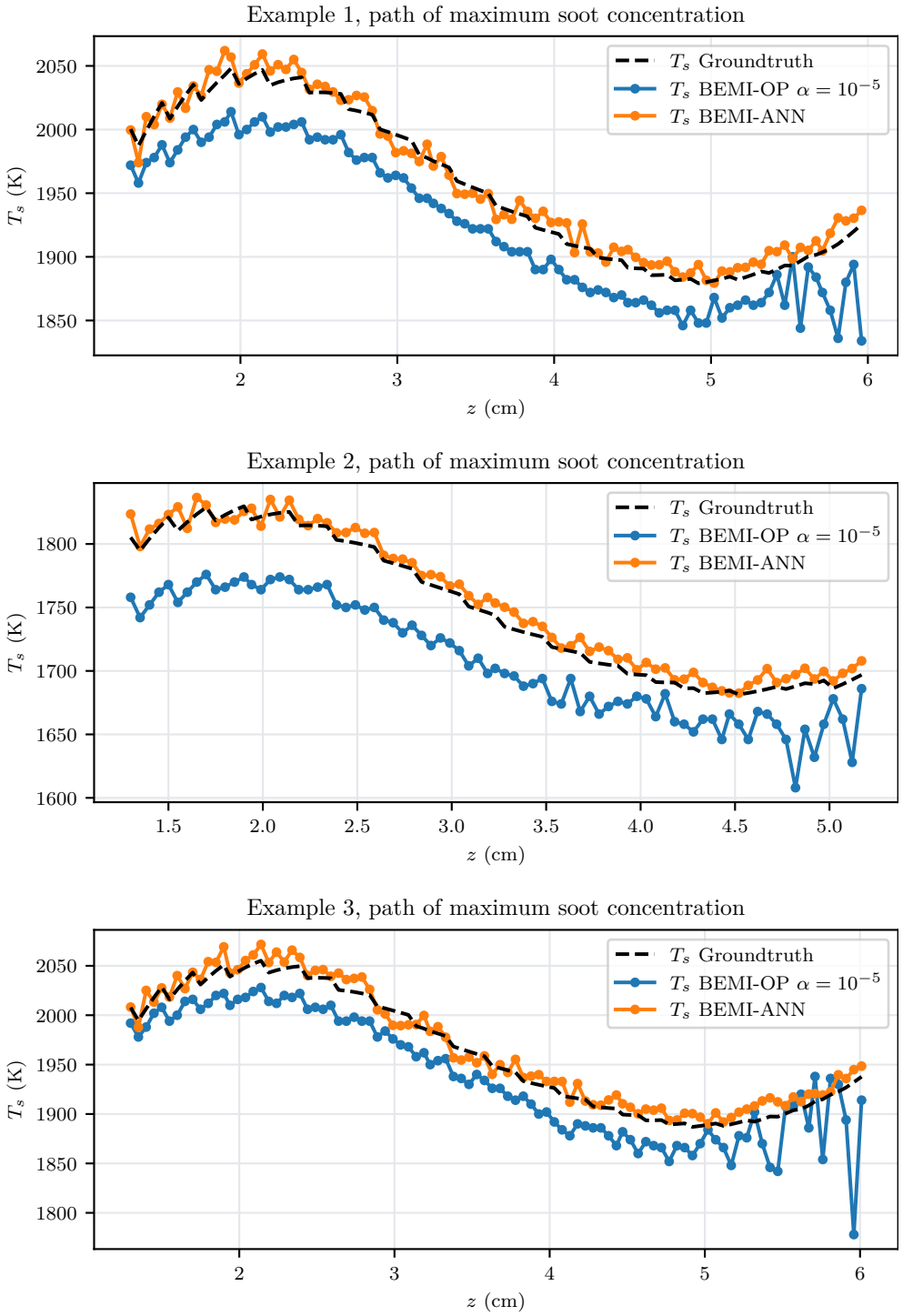


Figure 4.18: Groundtruth and predicted profile of T_s along the path of maximum soot using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN. Three examples.

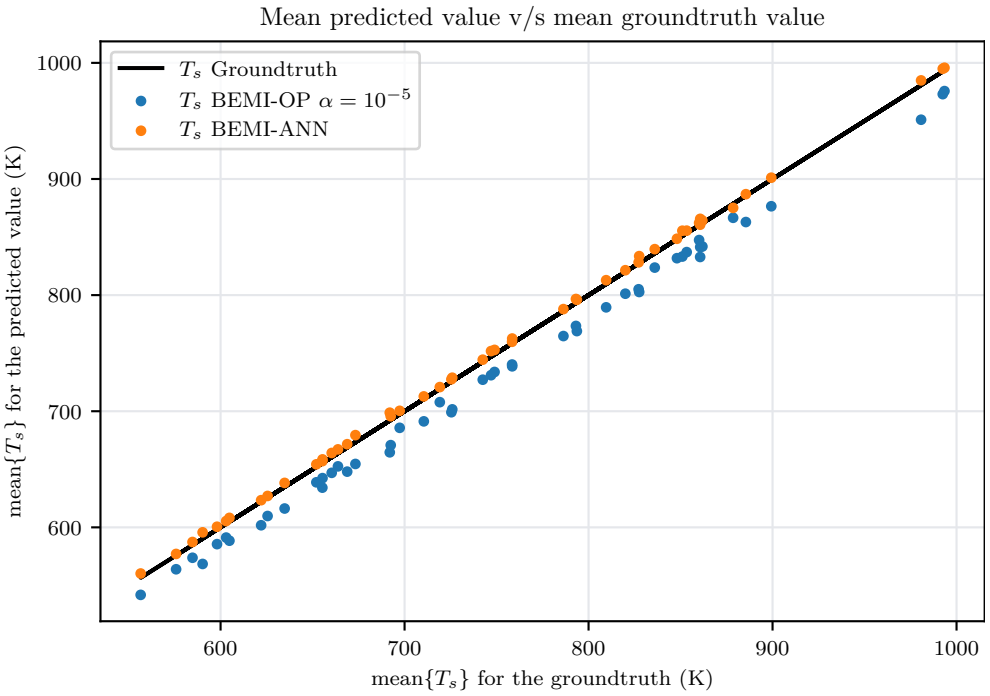


Figure 4.19: Comparison between the mean predicted value and the mean groundtruth value for the T_s fields over the entire test set using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN.

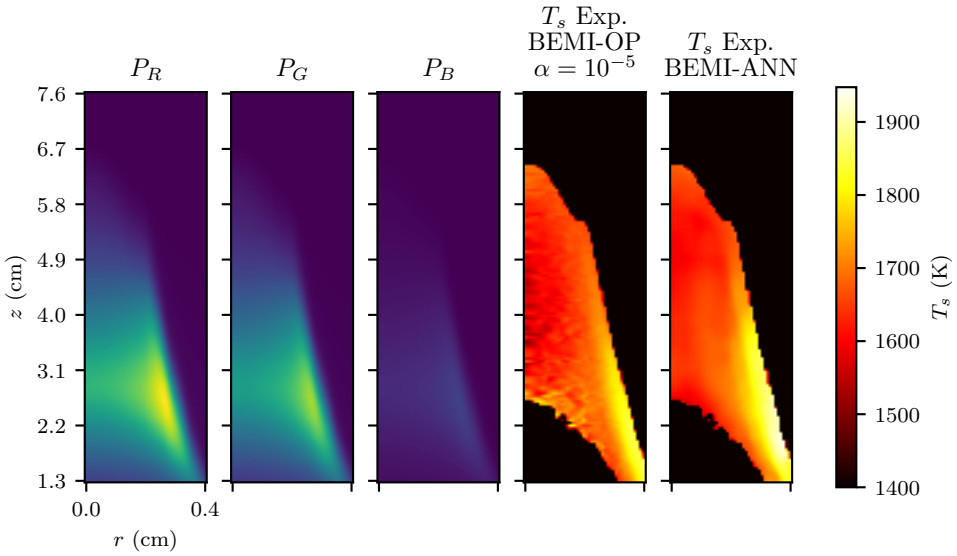


Figure 4.20: $P_{\{R,G,B\}}$ captured during an experimental campaign and predicted T_s field using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN.

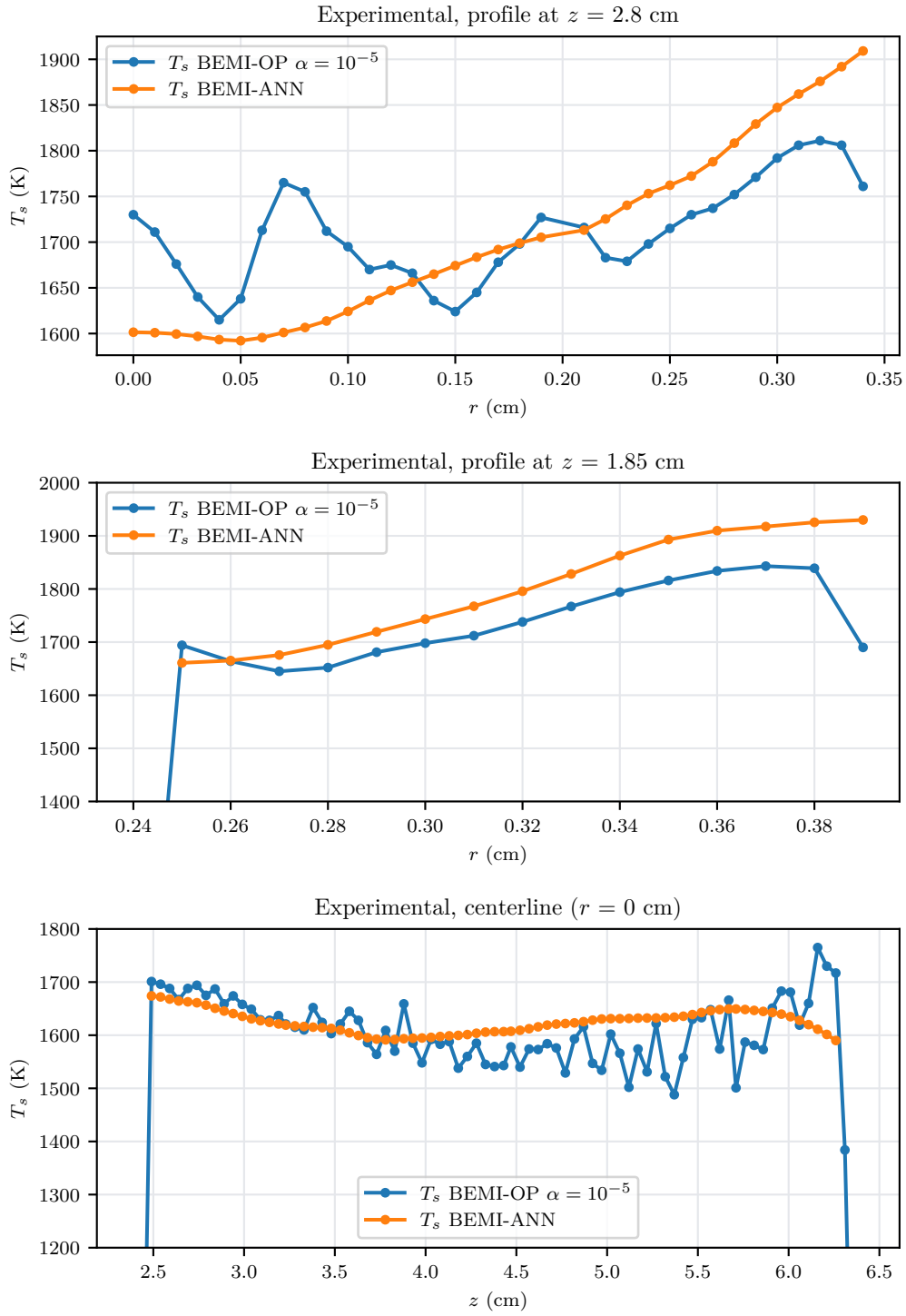


Figure 4.21: Predicted T_s profiles using BEMI-OP ($\alpha = 10^{-5}$) and our ANN-based method BEMI-ANN for data captured during an experimental campaign. Profiles at a HAB of $z = 2.8$ cm, HAB of $z = 1.85$ cm and along the flame centerline ($r = 0$ cm).

5 | Conclusions and Future Work

This thesis presented new techniques based on ANNs to reconstruct soot properties in CLAD flames from convoluted measurements. We analyzed the reconstruction of the soot volume fraction field f_s from LOSA measurements of the fraction of transmitted light τ_λ and the reconstruction of the soot temperature field T_s from the spectrally-integrated radiative light emission measurements captured with an RGB camera $P_{\{R,G,B\}}$.

We introduced the traditional methodologies used to reconstruct soot properties from convoluted measurements. The deconvolution step in these techniques requires the inversion of an Abel type integral, which we perform by using onion-peeling deconvolution with Tikhonov regularization. We also provided commentary on the disadvantages and pitfalls of these traditional methodologies.

We then presented a framework developed by members of the Energy Conversion and Combustion Group (EC^2G) and adapted for the systematic generation of physically-grounded synthetic soot fields from a converged solution generated using CoFlame. With this framework we created synthetic datasets used during the training of our ML models. We arrived to an ANN architecture based on U-Net to replace the traditional methodologies that require onion-peeling deconvolution with Tikhonov regularization. We separate our research in two sections: first using the synthetically-generated dataset containing soot volume fraction f_s and fraction of transmitted light τ_λ pairs we trained our ANN model to directly retrieve f_s from τ_λ without requiring explicit assumptions about the underlying physical process. Then, using the synthetically-generated dataset containing soot temperature T_s and emission measurements $P_{\{R,G,B\}}$, we trained our ANN model to directly retrieve T_s from $P_{\{R,G,B\}}$.

We evaluated the performance of the trained ANN models using entries of the synthetically-generated dataset and experimental campaigns in laboratory conditions. These results show that the trained ANN models outperform classical OP deconvolution when processing noisy synthetic images, delivering a smooth and consistent characterization of the 2D soot fields. Additionally, provided that the dataset of images used for the training of the ANN model is representative of the physical process, the ANN will infer an accurate model and provide estimates for the soot fields that are closer to the actual values than the ones obtained with traditional deconvolution techniques, which can be prone to measurement noise and over-regularization. In general, results over the experimental measurements show that both classical deconvolution and our ANN models produce similar results for low noise levels. However, the ANN demonstrates a higher resilience to noise in the measurements, making it a promising methodology to improve current inversion techniques in laboratories with optical equipment of different quality.

The combination of classical deconvolution methods with modern data-driven ML techniques could enable the development of new methods for combustion diagnostics. The new ML-based methodologies proposed in recent combustion literature represent a promising approach for developing non-intrusive and low-cost instrumentation for combustion diagnostics in laboratory environments. We expect that implementing these techniques using commercial off-the-shelf computing platforms will enable the development of low-cost instrumentation to improve the quality, reproducibility,

and scrutiny of experimental results in the characterization of soot properties for canonical CLAD flames. The results presented in this thesis are available as a Jupyter Notebook at `https://github.com/alonsorb/soot-ml`.

From our research, we identify the following topics as promising future work:

- We recommend the evaluation of different and more complex ANN architectures to characterize the soot particles emitted by CLAD flames. For example, Generative Adversarial Network (GAN) models have been successfully used for similar inversion problems in imaging. Additionally, similar problems in combustion could be modeled in a physics-informed machine learning framework, in which researchers incorporate differential equations associated with the physical process inside of the ML model.
- Given the difficulties of conducting experimental campaigns to validate the proposed models, most researchers evaluate their algorithms using a single set of flame conditions. A supervised learning model can only generalize over its training dataset, so a validation over multiple flame conditions could provide a better understanding of the generalization capabilities of the ML models.

Bibliography

[1] F. Escudero, ''Experimental and numerical contributions to soot production in laminar axisymmetric diffusion flames,'' Ph.D. dissertation, Aix-Marseille Université, 2019.

[2] J. E. Penner, H. Eddleman, and T. Novakov, ''Towards the development of a global inventory for black carbon emissions,'' *Atmospheric Environment. Part A. General Topics*, vol. 27, no. 8, pp. 1277--1295, 1993.

[3] M. Z. Jacobson, ''Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols,'' *Nature*, vol. 409, no. 6821, pp. 695--697, 2001.

[4] S. O. Baek, R. A. Field, M. E. Goldstone, P. W. Kirk, J. N. Lester, and R. Perry, ''A review of atmospheric polycyclic aromatic hydrocarbons: Sources, fate and behavior,'' *Water, Air, and Soil Pollution*, vol. 60, no. 3, pp. 279--300, 1991.

[5] D. Shindell, J. C. I. Kuylenstierna, E. Vignati, R. van Dingenen, M. Amann, Z. Klimont, S. C. Anenberg, N. Muller, G. Janssens-Maenhout, F. Raes, J. Schwartz, G. Faluvegi, L. Pozzoli, K. Kupiainen, L. Höglund-Isaksson, L. Emberson, D. Streets, V. Ramanathan, K. Hicks, N. T. K. Oanh, G. Milly, M. Williams, V. Demkine, and D. Fowler, ''Simultaneously Mitigating Near-Term Climate Change and Improving Human Health and Food Security,'' *Science*, vol. 335, no. 6065, pp. 183--189, 2012.

[6] Y. Wang and S. H. Chung, ''Soot formation in laminar counterflow flames,'' *Progress in Energy and Combustion Science*, vol. 74, pp. 152--238, 2019.

[7] Q. Zhang, ''Detailed Modeling of Soot Formation/Oxidation in Laminar Coflow Diffusion Flames,'' Ph.D. dissertation, University of Toronto, 2009.

[8] B. Franzelli, M. Roussillo, P. Scouflaire, J. Bonnety, R. Jalain, T. Dormieux, S. Candel, and G. Legros, ''Multi-diagnostic soot measurements in a laminar diffusion flame to assess the ISF database consistency,'' *Proceedings of the Combustion Institute*, vol. 37, no. 2, pp. 1355--1363, 2019.

[9] R. J. Santoro, H. G. Semerjian, and R. A. Dobbins, ''Soot particle measurements in diffusion flames,'' *Combustion and Flame*, vol. 51, no. C, pp. 203--218, jan 1983.

[10] R. J. Santoro, T. T. Yeh, J. J. Horvath, and H. G. Semerjian, ''The Transport and Growth of Soot Particles in Laminar Diffusion Flames,'' *Combustion Science and Technology*, vol. 53, no. 2-3, pp. 89--115, 1987.

[11] M. D. Smooke, C. S. McEnally, L. D. Pfefferle, R. J. Hall, and M. B. Colket, ''Computational and experimental study of soot formation in a coflow, laminar diffusion flame,'' *Combustion and Flame*, vol. 117, no. 1, pp. 117--139, 1999.

[12] C. Schulz, B. F. Kock, M. Hofmann, H. Michelsen, S. Will, B. Bougie, R. Suntz, and G. Smallwood, ''Laser-induced incandescence: recent trends and current questions,'' *Applied Physics B*, vol. 83, no. 3, p. 333, 2006.

[13] N. Peters, ''Laminar diffusion flamelet models in non-premixed turbulent combustion,'' *Progress in Energy and Combustion Science*, vol. 10, no. 3, pp. 319--339, 1984.

[14] R. Demarco, F. Nmira, and J.-L. Consalvi, ''Influence of thermal radiation on soot production in Laminar axisymmetric diffusion flames,'' *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 120, pp. 52--69, 2013.

[15] C. J. Dasch, ''One-dimensional tomography: a comparison of Abel, onion-peeling, and filtered backprojection methods,'' *Appl. Opt.*, vol. 31, no. 8, pp. 1146--1152, mar 1992.

[16] K. J. Daun, K. A. Thomson, F. Liu, and G. J. Smallwood, ''Deconvolution of axisymmetric flame properties using Tikhonov regularization,'' *Appl. Opt.*, vol. 45, no. 19, pp. 4638--4646, jul 2006.

[17] E. O. Åkesson and K. J. Daun, ''Parameter selection methods for axisymmetric flame tomography through Tikhonov regularization,'' *Appl. Opt.*, vol. 47, no. 3, pp. 407--416, jan 2008.

[18] P. C. Hansen and D. P. O'Leary, ''The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems,'' *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487--1503, 1993.

[19] K. A. Thomson, M. R. Johnson, D. R. Snelling, and G. J. Smallwood, ''Diffuse-light two-dimensional line-of-sight attenuation for soot concentration measurements,'' *Appl. Opt.*, vol. 47, no. 5, pp. 694--703, feb 2008.

[20] F. Escudero, A. Fuentes, J.-L. Consalvi, F. Liu, and R. Demarco, ''Unified behavior of soot production and radiative heat transfer in ethylene, propane and butane axisymmetric laminar diffusion flames at different oxygen indices,'' *Fuel*, vol. 183, pp. 668--679, nov 2016.

[21] M. F. Modest, *Radiative Heat Transfer*. Academic Press, 2013.

[22] W. H. Dalzell and A. F. Sarofim, ''Optical Constants of Soot and Their Application to Heat-Flux Calculations,'' *Journal of Heat Transfer*, vol. 91, no. 1, pp. 100--104, feb 1969.

[23] J. Yon, R. Lemaire, E. Therssen, P. Desgroux, A. Coppalle, and K. F. Ren, ''Examination of wavelength dependent soot optical properties of diesel and diesel/rapeseed methyl ester mixture by extinction spectra analysis and LII measurements,'' *Applied Physics B*, vol. 104, no. 2, pp. 253--271, 2011.

[24] A. R. Coderre, K. A. Thomson, D. R. Snelling, and M. R. Johnson, ''Spectrally resolved light absorption properties of cooled soot from a methane flame,'' *Applied Physics B*, vol. 104, no. 1, pp. 175--188, 2011.

[25] P. B. Kuhn, B. Ma, B. C. Connelly, M. D. Smooke, and M. B. Long, ''Soot and thin-filament pyrometry using a color digital camera,'' *Proceedings of the Combustion Institute*, vol. 33, no. 1, pp. 743--750, 2011.

[26] J. J. Cruz, L. F. F. da Silva, F. Escudero, F. Cepeda, J. C. Elicer-Cortés, and A. Fuentes, ''Soot Pyrometry by Emission Measurements at Different Wavelengths in Laminar Axisymmetric Flames,'' *Combustion Science and Technology*, vol. 0, no. 0, pp. 1--18, 2020.

[27] T. Ren, M. F. Modest, A. Fateev, G. Sutton, W. Zhao, and F. Rusu, ''Machine learning applied to retrieval of temperature and concentration distributions from infrared emission measurements,'' *Applied Energy*, vol. 252, p. 113448, 2019.

[28] T. Ren, H. Li, M. F. Modest, and C. Zhao, ''Efficient two-dimensional scalar fields reconstruction of laminar flames from infrared hyperspectral measurements with a machine learning approach,'' *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 271, p. 107724, 2021.

[29] ------, ''Machine learning applied to the retrieval of three-dimensional scalar fields of laminar flames from hyperspectral measurements,'' *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 279, p. 108047, 2022.

[30] T. Ren, Y. Zhou, Q. Wang, H. Liu, Z. Li, and C. Zhao, ''Machine learning-assisted soot temperature and volume fraction fields predictions in the ethylene laminar diffusion flames,'' *Opt. Express*, vol. 29, no. 2, pp. 1678--1693, jan 2021.

[31] Q. Wang, Z. Li, Z. Sun, H. Liu, W. Cai, and M. Yao, ''Simultaneous soot multi-parameter fields predictions in laminar sooting flames from neural network-based flame luminosity measurement I: methodology,'' *Opt. Lett.*, vol. 46, no. 16, pp. 3869--3872, aug 2021.

[32] L. Zhou, Y. Song, W. Ji, and H. Wei, ''Machine learning for combustion,'' *Energy and AI*, vol. 7, p. 100128, 2022.

[33] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, no. 7553, pp. 436--444, 2015.

[34] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ''Learning representations by back-propagating errors,'' *Nature*, vol. 323, no. 6088, pp. 533--536, 1986.

[36] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278--2324, 1998.

[37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[38] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional Networks for Biomedical Image Segmentation,'' in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234--241.

[39] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, ''Deep Convolutional Neural Network for Inverse Problems in Imaging,'' *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509--4522, 2017.

[40] M. T. McCann, K. H. Jin, and M. Unser, ''Convolutional Neural Networks for Inverse Problems in Imaging: A Review,'' *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85--95, 2017.

[41] V. Nair and G. E. Hinton, ''Rectified Linear Units Improve Restricted Boltzmann Machines,'' in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, pp. 807--814.

[42] S. Ioffe and C. Szegedy, ''Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,'' in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, pp. 448--456.

[43] N. A. Eaves, Q. Zhang, F. Liu, H. Guo, S. B. Dworkin, and M. J. Thomson, ''CoFlame: A refined and validated numerical algorithm for modeling sooting laminar coflow diffusion flames,'' *Computer Physics Communications*, vol. 207, pp. 464--477, 2016.

[44] N. A. Eaves, A. Veshkini, C. Riese, Q. Zhang, S. B. Dworkin, and M. J. Thomson, ''A numerical study of high pressure, laminar, sooting, ethane–air coflow diffusion flames,'' *Combustion and Flame*, vol. 159, no. 10, pp. 3179--3190, 2012.

[45] Q. Zhang, H. Guo, F. Liu, G. J. Smallwood, and M. J. Thomson, ''Modeling of soot aggregate formation and size distribution in a laminar ethylene/air coflow diffusion flame with detailed PAH chemistry and an advanced sectional aerosol dynamics model,'' *Proceedings of the Combustion Institute*, vol. 32, no. 1, pp. 761--768, 2009.

[46] M. Saffaripour, M. Kholghy, S. B. Dworkin, and M. J. Thomson, ''A numerical and experimental study of soot formation in a laminar coflow diffusion flame of a Jet A-1 surrogate,'' *Proceedings of the Combustion Institute*, vol. 34, no. 1, pp. 1057--1065, 2013.

[47] D. R. Snelling, K. A. Thomson, G. J. Smallwood, and Ö. L. Gülder, ''Two-dimensional imaging of soot volume fraction in laminar diffusion flames,'' *Appl. Opt.*, vol. 38, no. 12, pp. 2478--2485, apr 1999.

[48] F. Liu, H. Guo, and G. J. Smallwood, ''Effects of radiation model on the modeling of a laminar coflow methane/air diffusion flame,'' *Combustion and Flame*, vol. 138, no. 1, pp. 136--154, 2004.

[49] F. Liu, G. J. Smallwood, and O. L. Gülder, ''Band Lumping Strategy for Radiation Heat Transfer Calculations Using a Narrowband Model,'' *Journal of Thermophysics and Heat Transfer*, vol. 14, no. 2, pp. 278--281, 2000.

[50] F. Gelbard and J. H. Seinfeld, ''Simulation of multicomponent aerosol dynamics,'' *Journal of Colloid and Interface Science*, vol. 78, no. 2, pp. 485--501, 1980.

[51] S. H. Park and S. N. Rogak, ''A novel fixed-sectional model for the formation and growth of aerosol agglomerates,'' *Journal of Aerosol Science*, vol. 35, no. 11, pp. 1385--1404, 2004.

[52] N. A. Slavinskaya and P. Frank, ''A modelling study of aromatic soot precursors formation in laminar methane and ethene flames,'' *Combustion and Flame*, vol. 156, no. 9, pp. 1705--1722, 2009.

[53] S. B. Dworkin, Q. Zhang, M. J. Thomson, N. A. Slavinskaya, and U. Riedel, ''Application of an enhanced PAH growth model to soot formation in a laminar coflow ethylene/air diffusion flame,'' *Combustion and Flame*, vol. 158, no. 9, pp. 1682--1695, 2011.

[54] S. A. Eftekhari, ''A differential quadrature procedure with regularization of the Dirac-delta function for numerical solution of moving load problem,'' *Latin American Journal of Solids and Structures*, vol. 12, no. 7, pp. 1241--1265, 2015.

[55] F. Chollet and Others, ''Keras,'' https://keras.io, 2015.

[56] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, ''TensorFlow: A System for Large-Scale Machine Learning,'' in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, ser. OSDI'16. USA: USENIX Association, 2016, pp. 265--283.

[57] D. Kingma and J. Ba, ''Adam: A Method for Stochastic Optimization,'' *International Conference on Learning Representations*, dec 2014.

[58] J. A. H. Dreyer, R. I. Slavchov, E. J. Rees, J. Akroyd, M. Salamanca, S. Mosbach, and M. Kraft, ''Improved methodology for performing the inverse Abel transform of flame images for color ratio pyrometry,'' *Appl. Opt.*, vol. 58, no. 10, pp. 2662--2670, apr 2019.

[59] F. Liu, K. Thomson, and G. Smallwood, ''Soot temperature and volume fraction retrieval from spectrally resolved flame emission measurement in laminar axisymmetric coflow diffusion flames: Effect of self-absorption,'' *Combustion and Flame*, vol. 160, 2013.

[60] Z. Li, L. Zhang, and C. Lou, ''In-situ Measurement of Soot Volume Fraction and Temperature in Axisymmetric Soot-laden Flames Using TR-GSVD Algorithm,'' *IEEE Transactions on Instrumentation and Measurement*, vol. 70, p. 5001212, 2020.