



# UNIVERSITY OF LILLE

Doctorale School:

Matter, Radiation and Environmental Sciences

---

## Innovative fluorescence techniques in Mie-Raman Lidar for aerosol characterization and study of aerosol-clouds interactions

---

PhD Thesis prepared and defended publicly by

**Robin Miri**

on 08/11/2024, to obtain the doctoral degree in Physics

**Reviewers :**

Nicolas CEZARD	Research Director	ONERA, France
Lucas ALADOS ARBOLEDAS	Professor	University of Granada, Spain

**Examinators :**

Denis PETITPREZ (president of the jury)	Professor	Université de Lille, France
Michaël SICARD	Associated Professor	Université de La Réunion, France
Cyrielle DENJEAN	Researcher	Université de Toulouse, France

**Invited :**

Albert ANSMANN	Professor	Leibniz Institute for Tropospheric Research, Germany
Igor Veselovskii	Professor	Russian Academy of Sciences, Russia

**Directors :**

Philippe GOLOUB	Professor	Université de Lille, France
Olivier PUJOL	Associated Professor	Université de Lille, France

**Supervisor :**

Qiaoyun HU	Doctor	Université de Lille, France
------------	--------	-----------------------------

### Laboratoire d'Optique Atmosphérique

Département de Physique, Université de Lille, Sciences et technologies

59650 Villeneuve d'Ascq, France



# UNIVERSITE DE LILLE

Ecole doctorale :

Science de la Matière, du Rayonnement et de l'Environnement

---

## Utilisation de la fluorescence dans un lidar Mie-Raman pour la caractérisation des aérosols et l'étude de leurs interactions avec les nuages

---

Thèse de doctorat préparée et soutenue publiquement par

**Robin Miri**

Le 08/11/2024, pour obtenir le grade de docteur en Physique

**Rapporteurs :**

Nicolas CEZARD	Directeur de Recherches	ONERA, France
Lucas ALADOS ARBOLEDAS	Professeur	University of Granada, Espagne

**Examineurs :**

Denis PETITPREZ (président du jury)	Professeur	Université de Lille, France
Michaël SICARD	Maître de Conférences	Université de La Réunion, France
Cyrielle DENJEAN	Chercheure	Université de Toulouse, France

**Invités :**

Albert ANSMANN	Professeur	Leibniz Institute for Tropospheric Research, Allemagne
Igor Veselovskii	Professeur	Russian Academy of Sciences, Russie

**Directeurs :**

Philippe GOLOUB	Professeur	Université de Lille, France
Olivier PUJOL	Maître de Conférences	Université de Lille, France

**Encadrant :**

Qiaoyun HU	Docteure	Université de Lille, France
------------	----------	-----------------------------

### Laboratoire d'Optique Atmosphérique

Département de Physique, Université de Lille, Sciences et technologies

59650 Villeneuve d'Ascq, France



# Abstract

This thesis focuses on the analysis of atmospheric fluorescence measurements by lidar for characterizing aerosols and studying their interactions with water vapor and clouds. The LILAS (Lille Lidar Atmospheric Study) lidar, located on the ATOLL (ATmospheric Observations at LiLle) platform at the LOA (Laboratoire d'Optique Atmosphérique) in Lille, is a Mie-Raman-fluorescence lidar and a member of the European EARLINET-ACTRIS network (European Aerosol Research Lidar NETwork - Aerosol-Clouds and Trace gases Research InfraStructure). It is capable of profiling aerosol-induced fluorescence and has a high level of automation. These features were used for the development of FLARE-GMM (Fluorescence Lidar Aerosol REcognition based on Gaussian Mixture Model), an automatic aerosol classification algorithm based on a machine learning model. This algorithm identifies the type of aerosols detected by the lidar using their depolarization signature, fluorescence, and atmospheric humidity. Following validation, a statistical study of aerosols present in Lille from 2021 to 2023 was conducted using FLARE-GMM results.

The atmospheric fluorescence profile database was then used to improve the estimation of the hygroscopic properties of aerosols. Independent of the presence of water vapor, fluorescence allows for tracking aerosol concentration changes in a given layer, assuming the aerosols nature in that layer is homogeneous. This ability allows for compensating potential variations in aerosol concentration within the layer, enabling more accurate estimation of hygroscopic properties. This method, evaluated in different scenarios, demonstrated the advantages of using fluorescence. The hygroscopic properties of aerosols in Lille were subsequently estimated through an automatic detection method for hygroscopic growth events, revealing significant uncertainties primarily related to relative humidity estimation. The dependency of hygroscopic property estimation on relative humidity was modeled to better account for this phenomenon and understand how to minimize it.

Finally, fluorescence signals in clouds were studied. This part remains more exploratory, as many unknowns persist regarding the impact of clouds on fluorescence signals. Nevertheless, several hypotheses explaining how clouds might influence fluorescence signals were proposed and discussed. Additionally, a method for inverting the elastic backscatter of aerosols within the cloud layer was proposed. Although this approach currently has large uncertainties and relies on strong assumptions, it offers an original and promising perspective for studying aerosols in cloudy conditions through remote sensing.

**Key-words:** Remote sensing, Aerosols, Clouds, Lidar



---

# Résumé

Cette thèse se concentre sur l'analyse des mesures de fluorescence atmosphérique par lidar pour la caractérisation des aérosols et l'étude de leurs interactions avec la vapeur d'eau et les nuages. Le lidar LILAS (Lille Lidar Atmospheric Study), situé sur la plateforme ATOLL (ATmospheric Observations at LiLLe) du LOA (Laboratoire d'Optique Atmosphérique) à Lille, est un lidar Mie-Raman-fluorescence membre du réseau européen EARLINET-ACTRIS (European Aerosol Research Lidar NETwork-Aerosol-Clouds and Trace gases Research InfraStructure) capable de profiler la fluorescence induite des aérosols et disposant d'un haut niveau d'automatisation. Ces caractéristiques ont permis le développement de FLARE-GMM (Fluorescence Lidar Aerosol REcognition based on Gaussian Mixture Model), un algorithme de classification automatique des aérosols basé sur un modèle d'apprentissage automatique. Cet algorithme précise le type des aérosols détectés par le lidar à l'aide de leur signature de dépolarisation, de leur fluorescence et de l'humidité atmosphérique. Après validation, une étude statistique des aérosols présents à Lille sur la période 2021-2023 a été réalisée grâce aux résultats de FLARE-GMM.

La base de données de profils de fluorescence atmosphérique a ensuite été utilisée pour améliorer l'estimation des propriétés hygroscopiques des aérosols. Indépendante de la présence de vapeur d'eau, la fluorescence permet de suivre l'évolution de la concentration d'aérosols dans une couche donnée, en supposant que la nature des aérosols dans cette couche soit homogène. Cette capacité permet de compenser les variations éventuelles de concentration en aérosols dans la couche, permettant une estimation plus précise des propriétés hygroscopiques. Cette méthode, évaluée dans différentes situations, a montré les avantages apportés par la fluorescence. Les propriétés hygroscopiques des aérosols présents à Lille ont ensuite été estimées grâce à une méthode de détection automatique des cas de croissance hygroscopique, révélant des incertitudes importantes liées principalement à l'estimation de l'humidité relative. La dépendance de l'estimation des propriétés hygroscopiques à l'humidité relative a été modélisée pour mieux prendre en compte ce phénomène et comprendre comment le minimiser.

Enfin, les signaux de fluorescence dans les nuages ont été étudiés. Cette partie demeure plus prospective, car de nombreuses inconnues subsistent quant à l'impact des nuages sur les signaux de fluorescence. Néanmoins, plusieurs hypothèses expliquant comment les nuages peuvent influencer les signaux de fluorescence ont été avancées et discutées. Ensuite, une méthode d'inversion de la rétrodiffusion élastique des aérosols présents dans la couche nuageuse a été proposée. Bien que cette approche comporte, en l'état, de larges incertitudes et repose sur des hypothèses fortes, elle offre une perspective originale et prometteuse pour l'étude des aérosols en conditions nuageuses par télédétection.

**Mots-clés:** Télédétection, Aérosols, Nuages, Lidar



# Acknowledgement

I would like to thank all my supervisors, Philippe Goloub, Olivier Pujol and Qiaoyun Hu for their guidance, their support, their involvement in my work and their time. This study would not have been possible without their contribution. I am also grateful to everyone I have been able to interact with at the Laboratoire d'Optique Atmosphérique, every member of the laboratory has contributed to improving my work environment, allowing me to thrive throughout the thesis by putting me in the best conditions. I would like to thank specially William Boissière, Éric Bourriane, Gaël Dubois, Fabrice Ducos, Anne Priem, Marie-Lyse Liévin, Yuyang Chang, Maria Fernanda Sanchez and Élise Devigne, for enlightening my stay at the LOA.

I also would like to thank Igor Veselovskii for his unselfish help and guidance during my work. I have been able to learn a lot about lidar sciences from him and he greatly contributed to the scientific quality of this work. I would like to express my gratitude to Denis Petitprez and Albert Ansmann for being part of my Thesis Committee. Their encouragement and positive feedback throughout this process have been a constant source of motivation for me.

I would like to thank my family, my parents and my brother, for their presence and their support at all time. I also would like to thank my friends who have supported me during the writing of the manuscript. Their encouragement and assistance have enabled me to work in the best possible conditions, making this journey much more manageable.

Finally, I would like to thank the labex Cappa and the University of Lille, who provided financial support for this work.

# Contents

<b>ABSTRACT .....</b>	<b>2</b>
<b>RESUME .....</b>	<b>4</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>6</b>
<b>CONTENTS.....</b>	<b>7</b>
<b>LIST OF FIGURES .....</b>	<b>10</b>
<b>LIST OF TABLES .....</b>	<b>15</b>
<b>LIST OF IMPORTANT ABBREVIATIONS .....</b>	<b>16</b>
<b>1. INTRODUCTION.....</b>	<b>18</b>
1.1 CONTEXT .....	18
1.2 CURRENT STATE OF AEROSOL AND CLOUD STUDIES.....	21
1.2.1 <i>Observations</i> .....	21
1.2.2 <i>Models</i> .....	24
1.3 OVERVIEW AND LAYOUT OF THIS STUDY.....	25
<b>2. INSTRUMENTATION AND DATA PROCESSING .....</b>	<b>27</b>
2.1 LIGHT-MATTER INTERACTIONS .....	27
2.1.1 <i>Light properties, brightness temperature and lasers</i> .....	27
2.1.2 <i>Elastic scattering</i> .....	28
2.1.3 <i>Raman scattering and fluorescence</i> .....	29
2.2 LIDAR INSTRUMENT .....	31
2.2.1 <i>Lidar general description</i> .....	31
2.2.2 <i>The lidar equation</i> .....	33
2.2.3 <i>The LILAS system</i> .....	35
2.2.4 <i>LILAS profile inversion</i> .....	37
2.3 THE MICROWAVE RADIOMETER.....	40
2.3.1 <i>General presentation</i> .....	40
2.3.2 <i>The RPG-HATPRO G5</i> .....	40
2.4 DATA PREPARATION .....	42
2.4.1 <i>Dataset organization and temperature data</i> .....	42
2.4.2 <i>LILAS humidity calibration</i> .....	43
2.5 CHAPTER SUMMARY .....	45

---

<b>3. FLARE-GMM: AN AEROSOL TYPING MODEL BASED ON LILAS DATA, AEROSOL TYPING STATISTICAL ANALYSIS ON 3 YEARS OF DATA</b> .....	<b>47</b>
3.1 AEROSOL TYPING GENERALITIES .....	47
3.1.1 <i>Aerosol typing in the literature</i> .....	47
3.1.2 <i>Aerosol typing with LILAS</i> .....	49
3.2 FLARE-GMM PRESENTATION .....	52
3.2.1 <i>Generalities on machine learning</i> .....	52
3.2.2 <i>The Gaussian Mixture Model</i> .....	53
3.2.3 <i>FLARE-GMM: an automatic aerosol typing algorithm</i> .....	55
3.3 PREPARATION BEFORE TRAINING FLARE-GMM.....	57
3.3.1 <i>Filtering and preprocessing of the dataset</i> .....	57
3.3.2 <i>Hyperparameter selection</i> .....	59
3.4 SELECTION OF THE TRAINING SET, THE DIFFERENT VERSIONS OF FLARE-GMM .....	59
3.4.1 <i>Training with all available data</i> .....	60
3.4.2 <i>Training without the boundary layer</i> .....	66
3.4.3 <i>Training with selected cases</i> .....	71
3.4.4 <i>Use of RH and the treatment of hygroscopicity</i> .....	76
3.5 GENERALIZATION OF FLARE-GMM .....	80
3.5.1 <i>Classification of specific events</i> .....	80
3.5.2 <i>Comparison with NATALI aerosol typing</i> .....	85
3.6 AEROSOL TYPE ANALYSIS IN LILLE .....	87
3.7 CHAPTER SUMMARY .....	90
<b>4. HYGROSCOPIC GROWTH STUDY FROM MIE-RAMAN-FLUORESCENCE LIDAR</b> .....	<b>92</b>
4.1 AEROSOL HYGROSCOPIC GROWTH .....	92
4.1.1 <i>General description</i> .....	92
4.1.2 <i>Characterization of aerosol hygroscopic growth in the literature</i> .....	94
4.2 METHODOLOGY FOR ESTIMATING AEROSOL HYGROSCOPIC PROPERTIES WITH LILAS.....	97
4.3 CASE STUDIES .....	99
4.3.1 <i>Hygroscopic growth of urban aerosol during the night of 29 July 2021</i> .....	99
4.3.2 <i>Hygroscopic growth of biomass burning aerosols during the night of 9 March 2021.</i> .....	105
4.3.3 <i>Hygroscopic growth of urban aerosols during the night of 16 April 2021</i> .....	110
4.4 AUTOMATIC STUDY OF HYGROSCOPIC CASES.....	114
4.4.1 <i>Method</i> .....	114
4.4.2 <i>Results</i> .....	115
4.5 INFLUENCE OF RH UNCERTAINTY ON HYGROSCOPIC GROWTH PROPERTIES ESTIMATION.....	118
4.5.1 <i>Impact of a flat adjustment in RH</i> .....	118
4.5.2 <i>Modeling <math>\gamma</math> variations with RH and approximation of the slope</i> .....	120

---

4.6	CHAPTER SUMMARY .....	125
<b>5.</b>	<b>FLUORESCENCE MEASUREMENTS IN CLOUD LAYERS .....</b>	<b>126</b>
5.1	GENERALITIES ON CLOUD AND AEROSOL-CLOUD INTERACTION STUDIES WITH LIDARS .....	126
5.2	FLUORESCENCE SIGNALS IN CLOUDS .....	127
5.2.1	<i>Observations with LILAS</i> .....	127
5.2.2	<i>Fluorescence increase in cloud layers</i> .....	134
5.3	USE OF FLUORESCENCE TO RETRIEVE AEROSOL PROPERTIES IN CLOUD LAYERS .....	137
5.4	CHAPTER SUMMARY .....	139
<b>6.</b>	<b>CONCLUSION AND PERSPECTIVES .....</b>	<b>141</b>
6.1	CONCLUSION .....	141
6.2	PERSPECTIVES .....	142
	<b>REFERENCES .....</b>	<b>145</b>
	<b>APPENDIX A: SUPPLEMENTARY MATERIAL.....</b>	<b>159</b>
	<b>APPENDIX B: ARTICLE PUBLISHED IN AMT .....</b>	<b>166</b>

---

# List of Figures

<i>Figure 1.1: Change in effective radiative forcing (ERF) from 1750 to 2019 from different atmospheric components. Bars represent best estimates and uncertainties are given by error bars. Aerosol contribution is separated between contributions from aerosol-cloud interactions (ERFaci) and from aerosol-radiation interactions (ERFari) (cited from Intergovernmental Panel on Climate Change 2023).....</i>	<i>20</i>
<i>Figure 1.2: Active stations part of the network ACTRIS/EARLINET, 31 stations over Europe, the stations circled in red is the ATOLL platform operated by the LOA and IMT Douai (from <a href="https://www.earlinet.org/index.php?id=earlinet_homepage">https://www.earlinet.org/index.php?id=earlinet_homepage</a>).....</i>	<i>24</i>
<i>Figure 2.1: Schematic representation of Stokes and anti-Stokes Raman scattering between the ground state and a vibrational state. The blue arrow is the transition associated to the incident photon of wavelength <math>\lambda_0</math>, <math>\lambda_1</math> and <math>\lambda_2</math> are the wavelengths of the Stokes and anti-Stokes Raman scattering photon respectively, with <math>\lambda_2 &lt; \lambda_0 &lt; \lambda_1</math> .....</i>	<i>30</i>
<i>Figure 2.2: Schematic representation of fluorescence. <math>E_0</math> and <math>E_1</math> are the ground and excited electronic configuration energy levels respectively, <math>v_{0,1,2}</math> are the vibrational energy levels, <math>\lambda_0</math> is the incident photon wavelength, the red arrows are the non-radiative relaxation process (vibration), the green arrow is the fluorescence emission of a photon with wavelength <math>\lambda_f</math> higher than <math>\lambda_0</math> .....</i>	<i>31</i>
<i>Figure 2.3: Configuration types of lidars emitting and receiving parts, the arrows represent the emitted light by the emitting part, the dashed lines represent the optical axes of the various parts.....</i>	<i>32</i>
<i>Figure 2.4: schematic diagram illustrating overlap between the laser beam (Emission) and the field of view of the receiving telescope (Reception) for a simple mono-static bi-axial LiDAR system. The black dashed line marks the height above which full overlap (i.e., <math>O(R)</math> of 1) of the LiDAR transceiver is achieved and remains independent of height. ....</i>	<i>34</i>
<i>Figure 2.5: (a) Picture of LILAS (b) schematic representation of LILAS emission and the reception parts and their different components, more details on each element of the emitting and receiving parts can be found in part 2.2.3 .....</i>	<i>37</i>
<i>Figure 2.6: Picture of the RPG-HATPRO G5 Radiometer used in this study.....</i>	<i>41</i>
<i>Figure 2.7: spectral features used for microwave atmospheric profiling, from RPG-HATPRO G5 technical instrumental manual, the blue dots correspond to the measurement channels of the instrument: <a href="https://www.radiometer-physics.de/downloadftp/pub/PDF/Radiometers/General_documents/Manuals/2015/RPG_MWR_STD_Technical_Manual_2015.pdf">https://www.radiometer-physics.de/downloadftp/pub/PDF/Radiometers/General_documents/Manuals/2015/RPG_MWR_STD_Technical_Manual_2015.pdf</a>.....</i>	<i>41</i>
<i>Figure 2.8: Comparison between LILAS and Radiometer Absolute Humidity (a) and between Radiometer and ERA5 temperature (b) on 23 July, 2021 between 00:00 UTC and 01:00 UTC .....</i>	<i>44</i>
<i>Figure 3.1: Scattering and fluorescence spectra for an excitation at 355 nm (picture from Felidj 2016) .....</i>	<i>51</i>

---

Figure 3.2: Aerosol type with a depolarization/fluorescence capacity diagram. The ranges of the particle parameter variations for dust, pollen, smoke, and urban aerosol are given by rectangles, (adapted from Veselovskii et al. (2022)). ..... 52

Figure 3.3: Illustration of EM algorithm for GMM fitting to a dataset divided in two clusters. (a) corresponds to the initialization, the blue and red circles correspond to the Gaussian distributions centered around their mean and with radius corresponding to their variances (b) illustrates the expectation set, the data points are separated according to the maximum probability of belonging to either distribution (c) illustrates the maximization step, the new parameters of the Gaussians are computed according to the new distribution (d), (e) and (f) show the results of the EM algorithm after 2, 5 and 20 iterations respectively,  $L$  being the iteration number. (figure from Bishop (2006))..... 55

Figure 3.4: 2D histograms of the training set containing all filtered data from 2021, 2022 and 2023 (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR..... 61

Figure 3.5: Silhouette coefficient from K-Means partitions on the training set containing all filtered data from 2021, 2022 and 2023, for a number of clusters ranging from 3 to 8 ..... 62

Figure 3.6: Data repartition of the training set with FLARE-GMM according to (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR, each color is associated to a Gaussian distribution determined by FLARE-GMM..... 63

Figure 3.7: Data repartition of the training set with FLARE-GMM, with a training set containing data from 2021 and 2022 only, according to (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR, each color is associated to a Gaussian distribution determined by FLARE-GMM..... 66

Figure 3.8: 2D histograms of the training set containing all filtered data from 2021, 2022 and 2023, without the boundary layer and low aerosol load situations (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR..... 67

Figure 3.9: Data repartition of the training set with FLARE-GMM, trained on a dataset without boundary layer, according to (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR, each color is associated to a Gaussian distribution determined by FLARE-GMM ..... 69

Figure 3.10: Data repartition of the training set with FLARE-GMM, with a training set containing data from 2022 and 2023 only, without the boundary layer, according to (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR, each color is associated to a Gaussian distribution determined by FLARE-GMM ..... 71

Figure 3.11: (a) quick-looks of FLARE-GMM results (version from part 3.4.2) during the night between 6 and 7 September 2021 (b) Representation of classified data in function of the fluorescence capacity and the depolarization ..... 73

Figure 3.12: 2D histograms of the training set containing hand-selected data of pure cases (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR..... 74

Figure 3.13: Data repartition of the training set with FLARE-GMM, with a training set containing hand-selected data of pure cases, according to (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR, each color is associated to a Gaussian distribution determined by FLARE-GMM..... 75

Figure 3.14: 2D histograms of the training set containing hand-selected data of pure cases in function of humidity levels (a) below 60 % (b) between 60 % and 80 %, (c) over 80%, (d) all data ..... 77

---

Figure 3.15: Data repartition of the training set sections with the different versions of FLARE-GMM trained on these sections (a) section with $RH < 60\%$ , (b) $60\% < RH < 80\%$ , (c) $RH > 80\%$ , each color is associated to a Gaussian distribution determined by FLARE-GMM.....	78
Figure 3.16: Data repartition of the training set sections with the different versions of FLARE-GMM trained on these sections (a) section with $RH < 60\%$ , (b) $60\% < RH < 80\%$ , (c) $RH > 80\%$ , each color is associated to a Gaussian distribution determined by FLARE-GMM, with the representation of the negative log likelihood contour lines ( $-\ln(P(x))$ ).....	79
Figure 3.17: FLARE-GMM aerosol type estimation quick-look during the night between 15 and 16 March 2022 between 1000 m and 15000 m.....	81
Figure 3.18: FLARE-GMM aerosol type estimation quick-look during the night between 2 and 3 March 2021 between 1000 m and 15000 m.....	83
Figure 3.19: (a) Backward trajectory of 24 hours at 4000 m above ground level at 22:00 UTC on 19 July 2022 (b) Fire map from the Fire Information for Resource Management System (FIRMS) between 14 to and 19 July 2022 (source: <a href="https://firms.modaps.eosdis.nasa.gov/">https://firms.modaps.eosdis.nasa.gov/</a> , last access: 28 June 2024) (c) FLARE-GMM aerosol type estimation quick-look during the night of 19 July 2022 between 1000 m and 15000 m .....	84
Figure 3.20: Confusion matrix between FLARE-GMM and NATALI aerosol type estimation on 36 profiles from 2022 .....	86
Figure 3.21: (a) Violin plots and (b) Box plots of averaged altitudes, between the ground and 6 km above ground level in function of the aerosol type for all available data from 2021 to 2023 .....	88
Figure 3.22: (a) Histogram of aerosol type estimate from FLARE-GMM in function of the time and (b) share of each aerosol type in function of the season, with all data from 2021, 2022 and 2023 below 6000 m above ground level.....	89
Figure 4.1: Köhler curve (Saturation ratio vs particle diameter) of a sodium chloride (NaCl) particle with a dry size of $D_{dry} = 50\text{ nm}$ (at $T = 298\text{ K}$ ). The contributions of the Raoult effect (red curve) and the Kelvin effect (green curve) are shown. The critical saturation ratio $S_{crit}$ and the corresponding critical diameter $D_{crit}$ are indicated (blue bullet) (From P. C. Zieger (2011)). .....	94
Figure 4.2: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 29 July 2021.....	100
Figure 4.3: Profile of retrieved optical properties in function of altitude above ground level (a) water vapor mixing ratio [ $\text{g}/\text{kg}$ ] and potential temperature [ $\text{K}$ ], (b) elastic backscatter coefficient at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and RH, (c) fluorescence backscatter coefficient at 466 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and particular linear depolarization ratio at 532 nm, on 29 July 2021 from 22:00 to 23:00 UTC, the black dashed lines identify the area where hygroscopic growth is expected to occur .....	101
Figure 4.4: Evolution of the elastic backscatter coefficient at 532 nm in function of RH on 29 July 2021 from 22:00 to 23:00 UTC, between 1000 m and 1500 m above ground level, and results of the fit on the Hänel parameterization .....	102

---

---

Figure 4.5: Evolution of the normalized elastic backscatter coefficient at 532 nm (normalized with fluorescence backscatter coefficient) in function of RH on 29 July 2021 from 22:00 to 23:00 UTC, between 1000 m and 1500 m above ground level, and results of the fit to the Hänel parameterization .....	104
Figure 4.6: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $m^{-1}.sr^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 9 March 2021 .....	105
Figure 4.7: Profile of retrieved optical properties in function of altitude above ground level (a) water vapor mixing ratio [g/kg] and potential temperature [K], (b) elastic backscatter coefficient at 532 nm [ $m^{-1}.sr^{-1}$ ] and RH, (c) fluorescence backscatter coefficient at 466 nm [ $m^{-1}.sr^{-1}$ ] and particular linear depolarization ratio at 532 nm, on 9 March 2021 from 21:00 to 22:00 UTC, the black dashed lines identify the area where hygroscopic growth is expected to occur .....	106
Figure 4.8: Evolution of the elastic backscatter coefficient at 532 nm (a) without normalization and (b) with normalization with the fluorescence backscatter coefficient, in function of RH on 9 March 2021 from 21:00 to 22:00 UTC, between 850 m and 1280 m above ground level, and results of the fit to the Hänel parameterization .....	108
Figure 4.9: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $m^{-1}.sr^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 15 April 2021 .....	110
Figure 4.10: Profile of retrieved optical properties in function of altitude above ground level (a) water vapor mixing ratio [g/kg] and potential temperature [K], (b) elastic backscatter coefficient at 532 nm [ $m^{-1}.sr^{-1}$ ] and RH, (c) fluorescence backscatter coefficient at 466 nm [ $m^{-1}.sr^{-1}$ ] and particular linear depolarization ratio at 532 nm, on 16 April 2021 from 01:00 to 02:00 UTC, the black dashed lines identify the area where hygroscopic growth is expected to occur .....	112
Figure 4.11: Evolution of the elastic backscatter coefficient at 532 nm (a) without normalization and (b) with normalization with the fluorescence backscatter coefficient, in function of RH on 9 March 2021 from 21:00 to 22:00 UTC, between 1000 m and 1450 m above ground level, and results of the fit to the Hänel parameterization .....	113
Figure 4.12: Histograms of hygroscopic growth cases (a) time of occurrences (b) mean altitude above ground level (in meter) between February 2021 and December 2023 .....	116
Figure 4.13: boxplots of the different estimates of the enhancement factor at 85 % for $\beta_{532}$ , $f(85\%)$ for all aerosols and each aerosol type estimated from FLARE-GMM .....	117
Figure 4.14: Variation of the Hänel hygroscopic constant $\gamma$ (blue) and the enhancement factor at 85 % (red) at 532 nm, in function of the adjustment on the humidity ( $RH' = RH + \Delta RH$ ) for the 16 April 2021 between 01:00 and 02:00 UTC between 1000 m and 1450 m .....	119
Figure 4.15: Variation of the determination coefficient of the fit to the Hänel parameterization in function of the adjustment on the humidity ( $RH' = RH + \Delta RH$ ) for the 16 April 2021 between 01:00 and 02:00 UTC between 1000 m and 1450 m .....	120
Figure 4.16: $\ln\left(\frac{1-(RH+\Delta RH)}{1-(RH_{min}+\Delta RH)}\right)$ in function of $\ln\left(\frac{1-RH}{1-RH_{min}}\right)$ for different values of $\Delta RH$ (a) $\Delta RH = -0.08$ , (b) $\Delta RH = -0.04$ , (c) $\Delta RH = 0.04$ , (d) $\Delta RH = 0.08$ , the linear fit and the values of the slope .....	122

---

---

<i>Figure 4.17: <math>\ln\left(\frac{1-(RH+\Delta RH)}{1-(RH_{min}+\Delta RH)}\right)</math> in function of <math>\ln\left(\frac{1-RH}{1-RH_{min}}\right)</math> for different values of <math>\Delta RH</math> (a) <math>\Delta RH = -0.08</math>, (b) <math>\Delta RH = -0.04</math>, (c) <math>\Delta RH = 0.04</math>, (d) <math>\Delta RH = 0.08</math>, with the linear fit, and the approximation of the linear relationship expressed equation (4.21).....</i>	<i>123</i>
<i>Figure 4.18: Variation of the Hänel hygroscopic constant <math>\gamma</math> at 532 nm, in function of the adjustment on the humidity (<math>RH' = RH + \Delta RH</math>) for the 16 April 2021 between 01:00 and 02:00 UTC between 1000 m and 1450 m, and the approximation of the linear fit to this variation.....</i>	<i>124</i>
<i>Figure 5.1: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [<math>m^{-1}.sr^{-1}</math>] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 17 and 18 April 2021 .....</i>	<i>128</i>
<i>Figure 5.2: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [<math>m^{-1}.sr^{-1}</math>] and RH, (b) fluorescence backscatter coefficient at 466 nm [<math>m^{-1}.sr^{-1}</math>] and particular linear depolarization ratio at 532 nm on 18 April 2021 from 1:00 to 2:00 UTC .....</i>	<i>129</i>
<i>Figure 5.3: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [<math>m^{-1}.sr^{-1}</math>] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 3:00 UTC, on 2 and 3 July 2021.....</i>	<i>131</i>
<i>Figure 5.4: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [<math>m^{-1}.sr^{-1}</math>] and RH, (b) fluorescence backscatter coefficient at 466 nm [<math>m^{-1}.sr^{-1}</math>] and particular linear depolarization ratio at 532 nm on 2 July 2021 from 22:00 to 23:00 UTC.....</i>	<i>132</i>
<i>Figure 5.5: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [<math>m^{-1}.sr^{-1}</math>] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 29 and 30 July 2021.....</i>	<i>133</i>
<i>Figure 5.6: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [<math>m^{-1}.sr^{-1}</math>] and RH, (b) fluorescence backscatter coefficient at 466 nm [<math>m^{-1}.sr^{-1}</math>] and particular linear depolarization ratio at 532 nm on 29 and 30 July 2021 from 23:00 to 00:00 UTC.....</i>	<i>134</i>
<i>Figure 5.7: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [<math>m^{-1}.sr^{-1}</math>] and fluorescence backscatter coefficient at 466 nm [<math>m^{-1}.sr^{-1}</math>] (b) dry elastic backscatter coefficient of dry aerosol particles in the cloud layer at 532 nm [<math>m^{-1}.sr^{-1}</math>] and fluorescence backscatter coefficient at 466 nm [<math>m^{-1}.sr^{-1}</math>] on 18 April 2021 from 1:00 to 2:00 UTC.....</i>	<i>139</i>

---

# List of Tables

*Table 1.1: Overview aerosol characteristics, source type can be primary (emitted) and/or secondary (formed through multiple atmospheric mechanisms), the stated lifetime refers to tropospheric lifetime. Climate effect of increased aerosol is indicated as '+' for warming and '-' for cooling (adapted from Intergovernmental Panel on Climate Change 2023). ..... 18*

*Table 2.1: Summary table of the various important quantities used in this study with their characteristics ..... 43*

*Table 3.1: Summary of the various thresholds used to filter the dataset in the first time ..... 58*

---

# List of Important Abbreviations

<b>Abbreviation</b>	<b>Full name</b>
LILAS	Lille Lidar Atmospheric Study
LOA	Laboratoire d'Optique Atmosphérique
ATOLL	ATmospheric Observations at LiLLe
<i>PLDR</i>	Particle Linear Depolarization Ratio
<i>LR</i>	Lidar Ratio
<i>RH</i>	Relative Humidity
<i>IWV</i>	Integrated Water Vapor
FLARE-GMM	Fluorescence Lidar based Aerosol REcognition from Gaussian Mixture Model
GMM	Gaussian Mixture Model
NATALI	Neural Network Aerosol Typing Algorithm Based on Lidar Data
AUSTRAL	AUtomated Server for the TReatment of Atmospheric Lidars



# 1. Introduction

## 1.1 Context

An aerosol is a suspended particle in the air, in either liquid or solid form (Glickman et al., 2000). The sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) classifies aerosols as Short-Lived Climate Forcers (SLCFs) as their lifetime in the atmosphere is short, typically lasting no longer than weeks. Aerosols consist of sulfate, nitrate, ammonium, carbonaceous (black carbon, or organic aerosols), mineral dust and sea spray, with size ranging from nanometers to tens of micrometers (Intergovernmental Panel on Climate Change, 2023). Aerosols originate from various sources; anthropogenic activities, such as traffic and heating, contribute to a significant portion of aerosol emissions, particularly in urban areas. However, natural sources also play a role in aerosol presence in the atmosphere, such as mineral dust and sea sprays (Calvo et al., 2013; Mallet et al., 2018; Mascout et al., 2022, 2023).

Compounds	Source Type	Lifetime	Climate Forcing	Other Effects on Climate
<b>Sulphate aerosols</b>	Secondary	Minutes to weeks	-	Clouds, Ecosystem
<b>Nitrate aerosols</b>	Secondary	Minutes to weeks	-	Clouds, Ecosystem
<b>Carbonaceous aerosols</b>	Primary + secondary	Minutes to weeks	+/-	Cryo, Clouds, Ecosystem
<b>Sea Spray</b>	Primary	Day to weeks	-	Clouds, Ecosystem
<b>Mineral dust</b>	Primary	Minutes to weeks	+/-	Cryo, Clouds, Ecosystem

Table 1.1: Overview aerosol characteristics, source type can be primary (emitted) and/or secondary (formed through multiple atmospheric mechanisms), the stated lifetime refers to tropospheric lifetime. Climate effect of increased aerosol is indicated as '+' for warming and '-' for cooling (adapted from Intergovernmental Panel on Climate Change 2023).

Rising concerns on aerosols impact on both human health and climate have motivated research efforts to evaluate these effects. The presence of aerosols in the low atmosphere and their high concentration in city centers is responsible for the increase of cardiovascular and respiratory diseases. Depending on their size and composition, aerosols can penetrate the respiratory system and are responsible for oxidative stress due to their chemical interactions with the

lung's antioxidants, favoring the development of respiratory diseases (Shiraiwa et al., 2017). Smallest aerosols even penetrate the vascular system, and are responsible for inflammations which can have serious health impact (Rundell et al., 2007). Aerosols can also change visibility by fog formation. This phenomenon is especially noticeable in China, where aerosol concentrations can reach significant levels (Tie and Cao, 2009). It has also been observed in London, notably during the "great smog" of 1952. This event, attributed to specific meteorological conditions and coal combustion, led to an estimated 4000 deaths according to initial assessments (Polivka, 2018). Volcanic aerosols can damage plane engines and are therefore responsible for air traffic perturbation, as it was the case for the Eyjafjallajökull eruption in 2010, after which most European airports had to shut down for few days (Gislason et al., 2011).

Aerosols are also investigated for their impact on climate. Their presence in the atmosphere leads to changes in the scattering and absorption of incoming solar radiation. Aerosol effective radiative forcing (ERF) is defined as the impact of aerosols on the incoming and outgoing radiation at the top of the atmosphere. This effect considers the aerosol-radiation interactions contribution, affected by the aerosol optical properties, also known as ERF<sub>ari</sub>, as well as the contribution from the aerosol-cloud interactions, referred to as ERF<sub>aci</sub>.

Determining the radiative impact of aerosols is challenging due to their non-uniform distribution in the Earth's atmosphere. Their short lifetime in the air is responsible for significant disparities in their physical and chemical properties, which strongly depend on emissions. Moreover, the interaction processes involving aerosols are generally complex and challenging to characterize, adding further complexity to the assessment. However, a combined approach between observations and models allow for the improvement of the aerosol radiative forcing estimations (Liou, 2002). Aerosols can act as Cloud Condensation Nuclei (CCN) or ice nucleating particle (INP) on which water vapor condenses to form water droplets or ice crystals. This increases the cloud droplet and ice crystal number concentration and changes the clouds geometrical properties and albedo, impacting cloud radiative forcing (Twomey, 1959). Another outcome of aerosols influence on clouds is that the increase in droplet number concentration reduces cloud precipitation efficiency. The formed droplets are smaller and lighter, which reduces their ability to precipitate, increasing the cloud lifetime and consequently affecting Earth's radiation balance (Albrecht, 1989). Depending on cloud properties, they can exert either a positive or a negative radiative forcing. On the one hand, clouds having strong albedo, they reflect incoming solar radiation efficiently, thereby cooling the atmosphere. On the other hand, clouds contribute to the greenhouse effect, which balances their radiative forcing (Ramanathan et al., 1989). For this reason, predicting aerosol net impact

on cloud radiative forcing is challenging, as well as how this impact is expected to evolve in the near future.

The sixth Assessment Report of the IPCC compiles the latest findings and methods to estimate  $ER_{\text{Fari}}$  and  $ER_{\text{Faci}}$  from both models and observations. It concludes that aerosol effective radiative forcing is negative, indicating a cooling effect on the atmosphere, in contrast to greenhouse gasses such as carbon dioxide (Intergovernmental Panel on Climate Change, 2023). However, this report also highlights the existence of high uncertainties, especially in estimating  $ER_{\text{Faci}}$ . These uncertainties can be partly explained by the fact that some aerosol-cloud interaction processes are still missing from models. The total aerosol ERF for 2019 relative to 1750 is estimated at  $-1.1 \pm 0.7 \text{ W m}^{-2}$  with medium confidence.

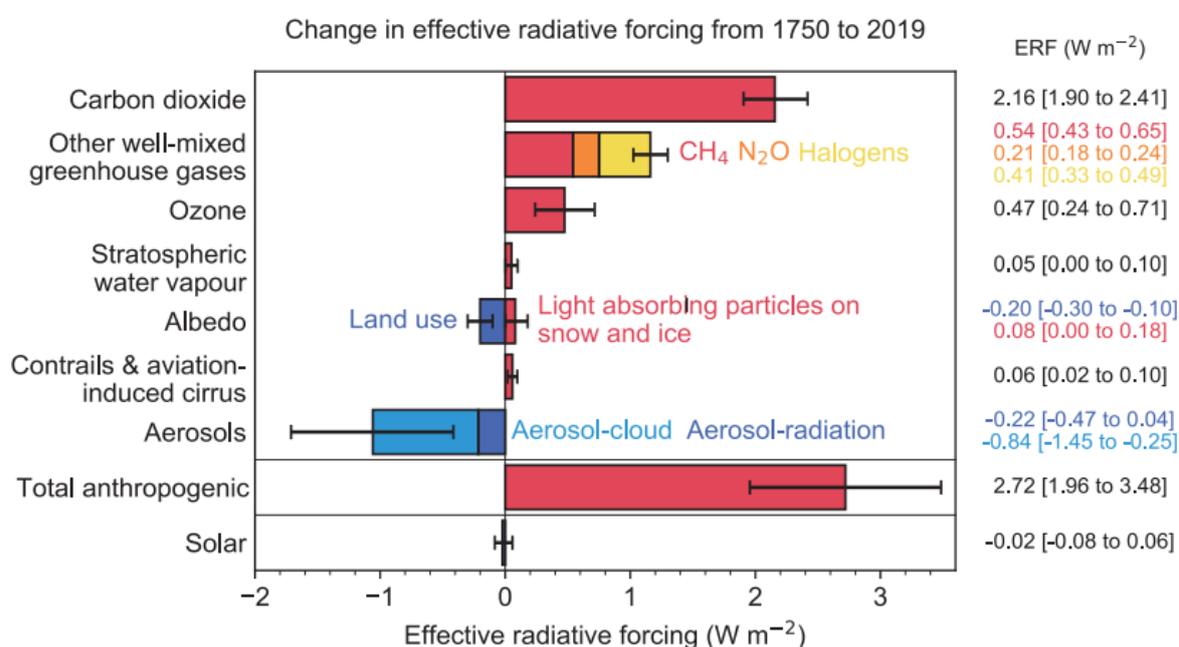


Figure 1.1: Change in effective radiative forcing (ERF) from 1750 to 2019 from different atmospheric components. Bars represent best estimates and uncertainties are given by error bars. Aerosol contribution is separated between contributions from aerosol-cloud interactions ( $ER_{\text{Faci}}$ ) and from aerosol-radiation interactions ( $ER_{\text{Fari}}$ ) (cited from Intergovernmental Panel on Climate Change 2023).

In this framework, the Laboratoire d'Optique Atmosphérique (LOA) located in Lille France, a research institute which is a mixed unit between the Centre National de la Recherche Scientifique (CNRS) and the University of Lille, is dedicated to the observation and modeling of both aerosols and clouds. The laboratory is divided between two teams, one is dedicated to the study of clouds, and the other to the study of aerosols. Both teams largely contribute to research, projects and measurement campaigns to improve the knowledge on atmospheric science from local to global scale (Cornet et al., 2010; Dubuisson et al., 2021; Hu et al., 2022; Waquet et al., 2013). This laboratory also features the ATmospheric Observations at LiLLe

(ATOLL) platform, an instrumentation platform dedicated to the observation of the atmosphere.

## **1.2 Current state of aerosol and cloud studies**

As mentioned above, more knowledge is needed on aerosols and their interactions with clouds. Scientific effort is devoted to the observation of these objects and a wide variety of instruments are dedicated to these tasks. Their use is essential to understand the complex interaction processes involving aerosols. Networks of instruments are developing in order to cover wide fractions of Earth, allowing to study these interactions with different types of aerosols and in various contexts. These processes can then be modeled to forecast climate and estimate how its evolution is affected by aerosols.

### **1.2.1 Observations**

Observations are essential to monitor aerosols and clouds, and understand the various physical phenomena surrounding them. Indeed, their presence in the atmosphere varies significantly across both time and space, requesting extensive spatial and temporal coverage. These observations can be either performed *in situ* or with remote sensing techniques.

*In situ* observations are performed when the instrument is co-located with the observable and collects samples to directly measure at the point of interest. In order to monitor aerosols, optical particle counters are used to measure the size distribution and concentration of particles. These instruments are generally cheap, allowing for their deployment in numerous locations (Aalto et al. 2005; Somsen et al. 2020), but they also come with a high degree of uncertainty (Hagan and Kroll, 2020). For this reason, scanning mobility particle sizers (SMPSs) are generally preferred for measuring aerosol size distributions and concentrations during field campaigns and for atmospheric studies. They provide valuable insight on local aerosol physical properties and are generally used during airborne campaign and for continuous monitoring at meteorological stations (Crumeyrolle et al., 2023; Harrison et al., 2019). Aerosol mass spectrometers are versatile instruments used to infer aerosol chemical compositions, allowing to identify aerosol sources during field campaigns or in controlled environments (Canagaratna et al., 2007). Aethalometers and nephelometers are used to measure optical properties, aethalometers allow to determine absorption properties of the aerosols, and nephelometers measure scattering properties. These instruments help understanding aerosol-radiation interactions, a crucial component for the assessment of aerosol impact on Earth radiation balance (Charlson et al., 1969; Anderson and Ogren, 1998; Arnott et al., 2005). Finally, Aerosol Chemical Speciation Monitor (ACSM) are instruments able to measure the chemical composition of aerosols, as well as their mass concentration, and are generally used to determine aerosol sources (Ng et al., 2011; Petit et al., 2014). For *in situ*

cloud monitoring, airborne probes are used to infer liquid and ice cloud properties such as droplet/crystal size distributions, or optical properties (McFarquhar et al., 2017).

*In situ* measurements have the drawback of measuring only at a certain point and time, making their temporal and spatial coverage limited. Their use can also be restricted to only certain aerosols, for example the ACSM can only be used on sub-micrometer aerosols. Moreover, their use during airborne campaigns incurs significant operational costs. Furthermore, these campaigns generally alter the environment in which the measurement is performed, enhancing the uncertainties (Spanu et al., 2020). For these reasons, remote sensing instruments, both ground-based and spaceborne, tend to be more and more used in atmospheric sciences. Spaceborne instruments have the advantage of covering the whole globe at the cost of important expense and technical constraints. On the other hand, ground-based instruments are limited to a single measurement spot on Earth but can be much more flexible. Active remote sensing instruments use their own source of light emission, the interaction between this source and the environment can then be analyzed to determine properties of the objects with which the emitted light interacted. On the other hand, passive remote sensing instruments do not operate their own light source and use other sources like sunlight.

Lidars (light detection and ranging) are active remote sensing instruments which can be used both ground-based or spaceborne, they measure the light backscattered by particles along the optical path of the emitted light. Lidars are widely used to determine aerosol optical properties along the atmospheric column, such as backscattering and extinction coefficient, as well as depolarization ratio and fluorescence coefficients (Ansmann et al., 1992a; Immler et al., 2005; Weitkamp, 2005; Sugimoto et al., 2012; Rao et al., 2018). But lidars are also able to identify the presence of clouds and to measure their thickness, height, and optical properties (Vaughan et al. 2004). Satellite lidars like Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), mounted on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO), launched in 2006 and stopped in 2023, or Earth Clouds, Aerosols and Radiation Explorer (EarthCARE), launched in May 2024, are used to continuously monitor aerosols and clouds around the globe (Hu et al., 2009; Kim et al., 2008). Details about this instrument are given part 2.2. Similarly to lidars, radars are used both on the ground and mounted on satellites to determine cloud height and properties (Stephens et al., 2008). Sun/sky photometers are passive remote sensing instruments, mostly used ground-based, that are able to infer the integrated aerosol optical depth (AOD) and the aerosol size distribution by pointing at the Sun and asserting atmospheric scattering at various wavelengths (Shaw, 1983; Holben et al., 1998). Finally, radiometers are used both ground-based and space-borne to monitor clouds and water vapor. These instruments measure the atmospheric brightness temperatures to assess profiles of atmospheric and cloud properties. Moderate-Resolution Imaging

Spectroradiometer (MODIS) or POLarization and Directionality of the Earth's Reflectances (POLDER), mounted on the Polarization & Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar (PARASOL) satellite are examples of space-borne instrument, giving valuable insight on Earth cloud coverage on a daily basis (Chepfer et al., 1998; Justice et al., 2002; Skou and Le Vine, 2006). More information on microwave radiometers is given part 2.3.

More and more efforts are invested in order to continuously monitor aerosols and clouds over the globe. Spatial missions dedicated to continuous observation of the atmosphere, such as the A-train, which CALIPSO is part of, launched by National Aeronautics Space Agency (NASA) and its national partners, have increasingly developed in recent years (Li et al., 2018; Schoeberl, 2002). On the other hand, networks of ground-based instruments have started to form such as AERosol RObotic NETwork (AERONET), first established by NASA and the Service National d'Observation (SNO) PHOTométrie pour le Traitement Opérationnel de Normalisation and Satellitaire (PHOTONS), a world-wide aerosol remote sensing network gathering sun photometer measurements. On the other hand, European Aerosol Research LIdar NETwork (EARLINET) gathers lidar data, for ground-based aerosol monitoring over Europe. EARLINET and AERONET are key components of the Aerosol-Clouds and Trace gases Research InfraStructure (ACTRIS) community, which develops observation networks of aerosols, clouds and trace gases. The ATOLL platform is a part of ACTRIS as a national facility.

Retrieval methods exploit these observations to derive quantitative properties on aerosols and clouds. These data can then be used to study physical and chemical processes, or implemented in atmospheric models in order to control their output and improve their performances.

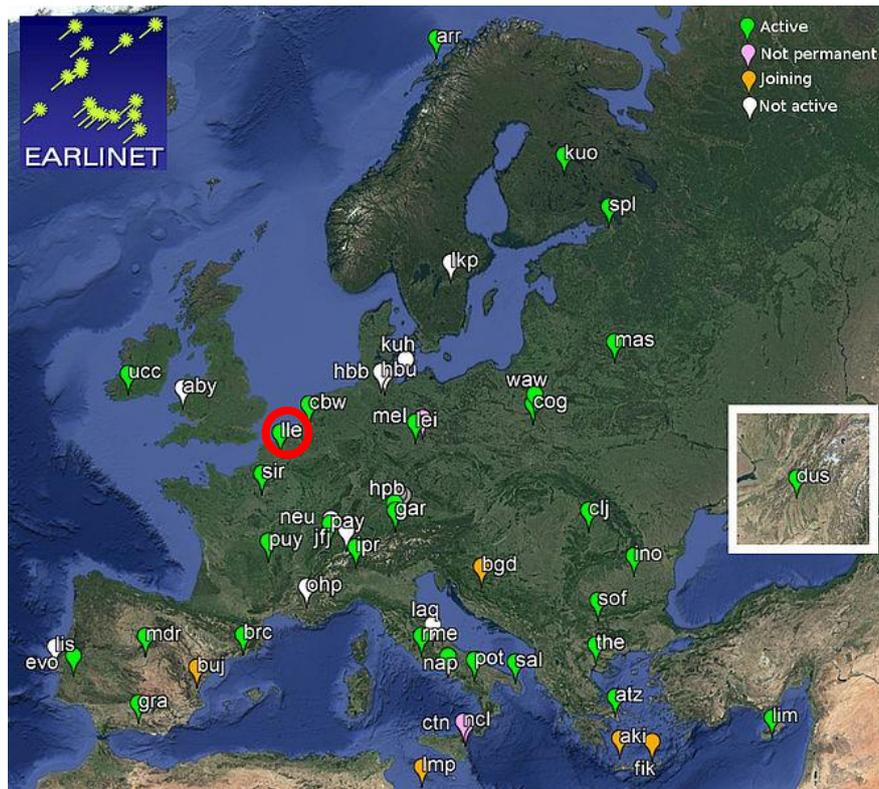


Figure 1.2: Active stations part of the network ACTRIS/EARLINET, 31 stations over Europe, the stations circled in red is the ATOLL platform operated by the LOA and IMT Douai (from [https://www.earlinet.org/index.php?id=earlinet\\_homepage](https://www.earlinet.org/index.php?id=earlinet_homepage))

### 1.2.2 Models

Models are an essential component of atmospheric research, providing valuable insights into the complex interactions and processes occurring within the atmosphere. They simulate thermodynamics and chemistry of the atmosphere with sets of equations and parameterizations to forecast atmospheric properties at various scales, from regional to global. They rely on observations to improve processes parameterization and correct their outputs. They are a valuable tool to understand and predict climate, and its evolution.

General circulation models (GCMs) rely on Navier-Stokes and thermodynamic equations to simulate Earth's atmosphere on a global scale. They are used for weather forecasting, to understand climate processes and forecast climate change. Coupled Model Intercomparison Project Phase 6 (CMIP6) is an international climate modeling project of the Working Group on Climate Modeling (WGCM), which compares the outputs of various climate models to understand which results are consistent across models and which are not [<https://wcrp-cmip.org/cmip-overview/#overview>, last access: May 2024]. Concerning the representation of clouds and aerosols, this group highlighted that while progress is being made, their interactions are in general challenging to implement in GCMs due to the scale at which these processes occur compared to the grid size at which GCMs work (Intergovernmental Panel on

Climate Change, 2023; Wang et al., 2021). Therefore, the understanding and parameterization of these processes is crucial to improve their implementation in climate models. These estimations are challenging to perform without collocated observations of clouds and aerosols, which is currently not possible with remote sensing instruments. More scientific efforts are therefore necessary to develop retrieval methods allowing to characterize particle properties inside of cloud layers. These efforts will contribute to better understanding and anticipating climate evolution.

### **1.3 Overview and layout of this study**

As pointed out previously, the presence of aerosols and clouds in the atmosphere greatly impacts climate, and more scientific work dedicated to understanding the interactions between them needs to be conducted. For this reason, the development of observations and retrieval methods of aerosols and clouds is necessary. In this context, the LOA, in the framework of Labex Chemical and Physical Properties of the Atmosphere (CaPPA) and ACTRIS infrastructure, has developed a Mie-Raman-Fluorescence lidar, Lille Lidar Atmospheric Study (LILAS). The singularity of this instrument is the detection channel dedicated to the measurement of laser-induced fluorescence signals emitted by the biogenic material present in the aerosols, and its high level of automation, allowing to maximize the observations. This kind of measurement technique is recent and more and more instruments are being equipped with it, such as the Multi-wavelength Atmospheric Raman lidar for Temperature, Humidity and Aerosol profiling (MARTHA) based at the Leibniz Institute for Tropospheric Research (TROPOS) (Gast et al., 2024). Fluorescence measurement opens new perspectives on aerosol remote sensing with lidars. In particular, the specificity of aerosol fluorescence is almost not influenced by the presence of water vapor in the air (Veselovskii et al., 2022b). The objective of this thesis is to exploit lidar fluorescence measurements to improve aerosol observations in humid and cloudy conditions. To begin with, after introducing the context, the main theoretical principles and the instruments that have been used, this study presents the development of an automatic aerosol typing method based on machine learning using lidar data, called Fluorescence Lidar Aerosol Recognition with Gaussian Mixture Model (FLARE-GMM). This model relies on the depolarization, fluorescence, and humidity measurements of LILAS to classify aerosols in function of their source, between urban aerosols, desert dust, and biomass burning smoke. This study also presents how fluorescence measurements allow to greatly improve the characterization of aerosol hygroscopic growth with lidars, defined as the growth of aerosols due to the uptake of water vapor. Using LILAS measurements and FLARE-GMM results, this method has been tested to estimate the hygroscopic growth factor of aerosols from different sources and compare them to literature values. These cases highlight the value brought by fluorescence measurement, as well as the limits of this method in the context of this

work. Eventually, perspectives are presented on how fluorescence signals in clouds could be used to retrieve information on particles in cloud layers. The fluorescence signal can be used to infer information on aerosols in cloud layers, providing information on potential CCN abundance in clouds. The potential limitations of this approach are also tackled.

This thesis consists of six chapters. In chapter 1, the context of this thesis is presented, with an overview of the different observations used in atmospheric sciences to characterize aerosols and clouds, how these observations are used in models and what are the current limitations which they face. The objectives of this thesis and its different parts are also presented. In chapter 2, fundamental background about basic theory and instrumentation is presented. The theory section presents fundamental bases on aerosol-radiation interactions, the instrumental section presents the lidar LILAS and the method to invert aerosol properties from its measurements in details, as well as the microwave radiometer used to calibrate LILAS water vapor channel. The calibration procedure and the data management processes are also laid out in this section. In chapter 3, bases on machine learning methods are introduced and the aerosol typing method FLARE-GMM is presented, with the analysis of its results on various cases and the comparison with another typing method. Chapter 4 presents the study of hygroscopic growth and how fluorescence can be used to improve the characterization of aerosol hygroscopic properties. Estimates of aerosol hygroscopic growth factors from this method are presented and compared to the literature. This chapter also presents the limits concerning the uncertainties of this retrieval method. In chapter 5, the question on the use of fluorescence in cloudy situations is raised, opportunities and limits brought by this question are also presented. Finally, in chapter 6, conclusions are drawn and perspectives are proposed.

## 2. Instrumentation and data processing

In this chapter, we provide the fundamental theoretical background on light-matter interactions for better understanding of this thesis. We then present atmospheric lidars, discussing their principles and general characteristics. The main instrument used in this study is the multi-wavelength Mie-Raman-fluorescence lidar LILAS (Lille Lidar Atmospheric Study), and this chapter outlines its technical specifications and the data processing protocols used to invert its measures for obtaining information on aerosols and water vapor. These descriptions are kept concise to provide a clear introduction to lidar specificities and operations. Finally, we discuss the water vapor calibration method, which involves a comparison with a reference instrument, in this case, the RPG-HATPRO G5 microwave radiometer, which is also described.

### 2.1 Light-matter interactions

#### 2.1.1 Light properties, brightness temperature and lasers

Light is classically characterized by its intensity, propagation direction, frequency and polarization. A blackbody is an idealized macroscopic condensed object in thermal equilibrium. It emits radiation according to the Planck law which describes the spectral energy density (per unit volume per unit energy):

$$u(\nu, T) = \frac{8\pi h\nu^3}{c^3} \frac{1}{\exp\left(\frac{h\nu}{k_B T}\right) - 1} \quad (2.1)$$

Where  $u(\nu, T)$  is the blackbody spectral energy density at frequency  $\nu$  and temperature  $T$ ,  $c$  is the speed of light,  $h$  is the Planck constant and  $k_B$  is the Boltzmann constant. This expression is used to estimate the brightness temperature, defined as the temperature at which the object would be if it was a blackbody.

The laser, standing for Light Amplification by Stimulated Emission Radiation, was invented in 1958 by Charles H. Townes and Arthur Leonard Schawlow and first built in 1960 by Theodore Maiman (Bromberg, 1988). This instrument relies on stimulated emission to emit light which is characterized by its high spatial and temporal coherence, allowing the light to be almost monochromatic and to produce short pulses of emission, and the light beam to be focused on

a tight spot while exhibiting small divergence. These characteristics allow the laser to have many applications.

### 2.1.2 Elastic scattering

Light interaction with matter can be separated in two categories: absorption and emission. Scattering, which is absorption followed by emission, can be seen classically as the oscillation of a dipole triggered by an incident electromagnetic field. The oscillation of this dipole then generates itself an electromagnetic wave, it is the scattered wave. Even though palpable matter appears mainly continuous and neutral, in reality, it is composed of discrete charges, resulting in light scattering being observable in numerous contexts (Bohren and Clothiaux, 2006).

The term elastic scattering is employed when the scattered light has the same frequency as the incident light, as opposed to inelastic scattering where the scattered light frequency is changed. Three regimes exist to characterize light scattering, depending on the size parameter of the scattering object, defined as  $\alpha = \pi D/\lambda$ , where  $\pi D$  is the circumference of the object and  $\lambda$  is the wavelength of the incident light. In function of the size parameter, approximation can be made to characterize the scattered electromagnetic wave.

If  $\alpha \ll 1$ , meaning that the object is small compared to the wavelength, then the Rayleigh scattering regime can be used (Rayleigh, 1871). It often applies for light scattering by molecules, for which the size is in the order of 1 nm, for light in the visible where the wavelength is in the order of 100 nm. The scattered light intensity can be expressed in this case as:

$$I(\lambda, R) = I_0 \frac{8\pi^2 p^2}{\lambda^4 R^2} (1 + \cos^2 \theta) \quad (2.2)$$

Where  $I(\lambda, R)$  is the scattered light intensity for an observer located at distance  $R$  from the scattering object,  $\lambda$  is the light wavelength,  $I_0$  is the incident light intensity,  $p$  is a parameter depending on the scattering object called polarizability, and  $\theta$  is the angle between the incident light direction and the observer. An important comment concerning this expression is that the scattered light intensity is proportional to  $\lambda^{-4}$ . This particularity explains why the sky is blue, as the shorter wavelengths are scattered more efficiently than the longer ones by atmospheric molecules.

If  $\alpha \gg 1$ , geometric scattering is used. These interactions are described by the laws of geometric optics. In the intermediate regime, when the particle size is comparable to the wavelength, the Mie theory applies. This theory introduces an exact solution to the Maxwell equations for the scattering of an harmonic plane wave by a homogeneous sphere under the form of an infinite series of multipole partial waves (Mie, 1908). This theory introduces the scattering, absorption and extinction efficiencies, characterizing the ability of the object to scatter or absorb incident radiation while the extinction corresponds to the decrease of light intensity in the forward

scattering direction and which combines scattering and absorption. The Mie theory, which can be applied only for spherical particles, is often used to describe light interaction with atmospheric particles, also called aerosols.

### **2.1.3 Raman scattering and fluorescence**

By opposition with elastic scattering, inelastic scattering means that the scattered light frequency is different from the incident light frequency. Raman scattering is an example of inelastic scattering discovered by C.V. Raman in 1928 after being predicted by Adolf Smekal in 1923. This scattering process is a much weaker phenomenon than elastic scattering, the ratio between the two intensities reaching up to  $10^6$  in favor of elastic scattering (Harris and Bertolucci, 1989).

Raman scattering by a molecule is explained by its energy levels. A molecule exhibits discrete energy levels. These levels correspond to a rotation or a vibration mode of the molecule, or a change in its electronic configuration. These mechanisms are associated with discrete energy states, with the highest energy being the one for the change of electronic configuration, followed by the vibration, the rotation having the lowest energy transition, the values of the transitional energies depending on the type of the molecule, and the conditions. In general, Raman scattering involves vibrational or rotational energy states.

When a photon of a certain energy collides with a molecule in its ground state, the molecule goes to a higher energy level, the energy difference being equal to the photon energy. In case of scattering, the molecule does not stay in this high energy state, a photon is emitted to lower the molecule energy. It can be of the same energy as the incident one, in that case the molecule comes back to its ground state, this is elastic scattering. But the emitted photon can also be less energetic if a vibrational energy level exists between the high energy state and the ground state. In that case, the molecule can reach the vibrational state instead of the ground state. Then, it emits a photon which is less energetic, this is Stokes Raman scattering. In this case, the emitted light has a higher wavelength than the received one, according to the Planck-Einstein relationship:  $E = h\nu$ , where  $E$  is the photon energy,  $h$  is the Planck constant, and  $\nu$  is the frequency. The anti-Stokes Raman scattering corresponds to the situation in which the molecule is initially in its vibrational state, and finishes in its ground state after the scattering process. The emitted photon is then more energetic and has a shorter wavelength than the received one. In these cases, the energy difference between the incident and emitted photons is characteristic of the molecule, and can be used to perform spectroscopy, or to measure the concentration of specific constituents.

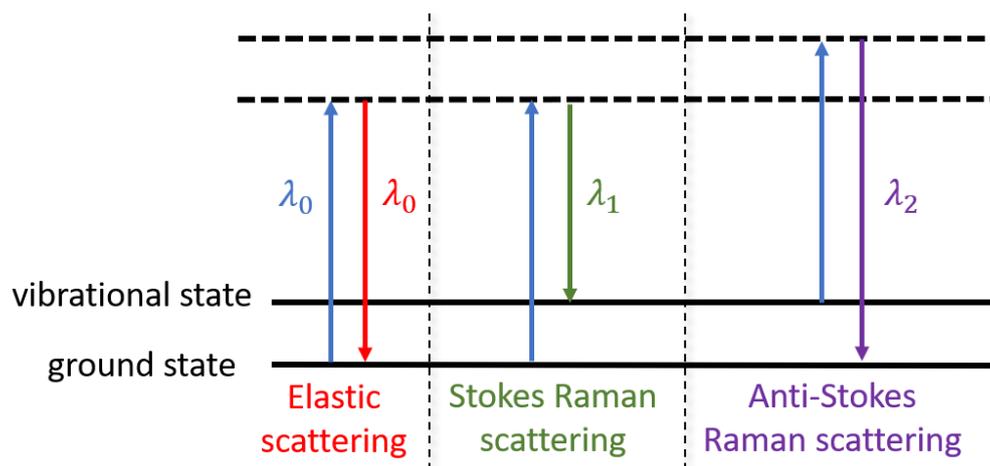


Figure 2.1: Schematic representation of Stokes and anti-Stokes Raman scattering between the ground state and a vibrational state. The blue arrow is the transition associated to the incident photon of wavelength  $\lambda_0$ ,  $\lambda_1$  and  $\lambda_2$  are the wavelengths of the Stokes and anti-Stokes Raman scattering photon respectively, with  $\lambda_2 < \lambda_0 < \lambda_1$

First described in the XVI century, fluorescence is also an inelastic scattering process. In general, a photon emitted by fluorescence is less energetic than the absorbed one and has therefore a longer wavelength. Fluorescence can be witnessed in nature by molecules like chlorophyll or quinine which absorb ultra-violet photons to emit in the visible.

This process is different from Raman scattering as it is much slower. First, the molecule absorbs the photon and reaches a higher energy state. This differs from scattering, as in the scattering process, this higher energy level is a transitional state that is unobservable. During fluorescence however, the molecule absorbs the photon and can be observed in the higher energy state in which it remains for a longer time. When a photon is absorbed in the UV, this is often followed by a change in the electronic configuration of the molecule. Then, the molecule generally relaxes through vibration to get to a lower energy state. Finally, fluorescence occurs when a photon is emitted, enabling the molecule to return to its initial electronic configuration. The energy state the molecule reaches after the photon emission can be the ground state, but is often a vibrational state.

The fluorescence lifetime is the average time during which the molecule remains in its excited state. It conforms to a first-order decay law and typically, it results into lifetimes on the order of nanoseconds for photon emission within the visible energy range.

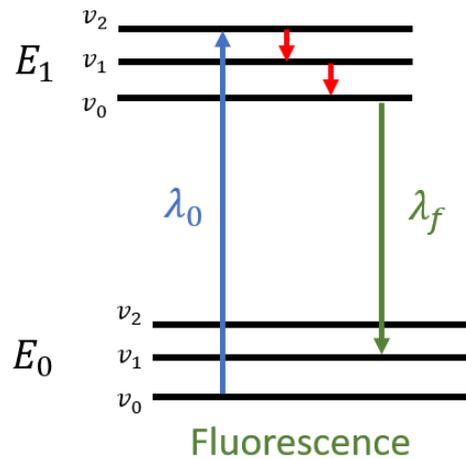


Figure 2.2: Schematic representation of fluorescence.  $E_0$  and  $E_1$  are the ground and excited electronic configuration energy levels respectively,  $v_{0,1,2}$  are the vibrational energy levels,  $\lambda_0$  is the incident photon wavelength, the red arrows are the non-radiative relaxation process (vibration), the green arrow is the fluorescence emission of a photon with wavelength  $\lambda_f$  higher than  $\lambda_0$ .

## 2.2 Lidar instrument

### 2.2.1 Lidar general description

Light detection and ranging (lidar), is an active remote sensing technology whose principle dates back from pre-laser times, and which saw one of its first applications in meteorology by Fiocco and Smullin in 1963, shortly after the development of lasers. Comparable to radar (radio detection and ranging) or sodar (sound navigation and ranging), lidar utilizes light, with wavelength varying from 300 nm to 12 000 nm typically, and finds applications in many fields, from geology to autonomous cars, or atmospheric physics. Lidar instruments are generally separated in two categories; hard target lidars are used to measure distances and surface properties, while atmospheric lidars study the medium in which light travels, which most of the time is the atmosphere. In meteorology, lidars can profile atmospheric parameters such as temperature, pressure, humidity, wind, clouds, trace gases or aerosols in function of the characteristics of the instrument (Wandinger et al., 2015).

Lidar setup consists of an emitting and a receiving part organized in different configurations in function of the instrument; crossed bistatic, paraxial monostatic, coaxial bistatic or monostatic, described Figure 2.1. The emission part is made of a laser source emitting pulses ranging from a few to several hundred nanoseconds with specific spectral properties. Nd:YAG (Neodymium-doped Yttrium Aluminum Garnet) lasers have seen their use more and more democratized in the lidar community since the 1980s. The Nd:YAG crystal emits in the infrared at 1064 nm and can be doubled or tripled by non-linear optics to emit in the green at 532 nm and in the ultra-violet at 355 nm. Typically, a beam expander is integrated to diminish the

divergence of the source light beam prior to its release into the atmosphere, aiming to prevent significant loss of light intensity at long distance from the instrument.

On the reception end, a telescope gathers the photons backscattered by the atmosphere. The detection mode can be either direct or involve heterodyne detection, where the collected light is mixed with the source, generally found in wind lidars. The light is then analyzed, to estimate its intensity, wavelength, polarization state or other properties, depending on the application of the instrument. This can be achieved generally by dichroic mirrors, interference filters and polarization beam splitters. The backscattered wave is then collected on a detector. The detection is performed by Photomultiplier tubes (PMTs), or photodiodes. The photon-counting technique allows detection of individual photons and is therefore highly sensitive, however for strong backscattered signals, saturations can be observed. For this reason, analog detection is used in these conditions. This technique consists in measuring the average current generated by the incident photons, and converting it to digital signal.

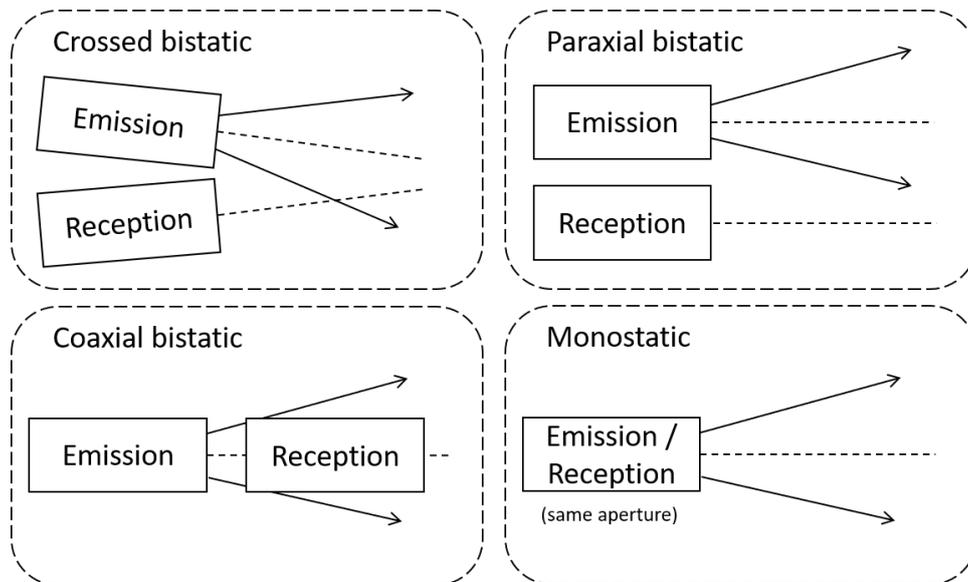


Figure 2.3: Configuration types of lidars emitting and receiving parts, the arrows represent the emitted light by the emitting part, the dashed lines represent the optical axes of the various parts

Considering the speed of light, and the time between the laser pulse emission, and the reception, it is possible to estimate the distance of the object which scattered the light back toward the instrument:  $D = (c t)/2$ , where  $D$  is the distance between the object and the instrument and  $t$  is the duration between the pulse emission, and the reception. Considering this expression and the time resolution of the detection system, it is possible to estimate the range resolution of the instrument:  $\Delta R = (c \Delta t)/2$  where  $\Delta R$  is the range resolution of the instrument and  $\Delta t$  is the time resolution of the detection system. A time resolution of the detection system of 50 ns thus results in a range resolution of 7.5 m, considering the laser pulse

being shorter. Typically, a few to several thousand laser pulses are emitted per second, and lidar signals are averaged over time intervals ranging from a few seconds to a few minutes. This helps in reducing the volume of stored data since high time resolution is not significant for atmospheric observations with lidars.

### 2.2.2 The lidar equation

The lidar equation defines the power of the signal detected by the lidar system, which originates from the backscattering of the emitted laser beam by an object located at a distance  $R$  from the instrument.

This equation considers the emission of a laser pulse which propagate through the atmosphere on the distance  $R$ . As the laser propagates, it is attenuated by the atmosphere, this attenuation is expressed by the atmospheric transmission function  $T(\lambda, R)$ , which can be expressed as:

$$T(\lambda, R) = \exp \left[ \int_0^R -\alpha(\lambda, r) dr \right] \quad (2.3)$$

Where  $\alpha(\lambda, r)$  is the atmospheric extinction coefficient, which quantifies the extinction properties of the atmosphere. Since the light travels both in the forward and backward direction, after being backscattered by the object, the transmission factor is squared in the lidar equation. Once the light reaches the object, it is scattered in every direction, and a certain proportion of the light goes back towards the instrument. This quantity is described by the atmospheric backscattering coefficient,  $\beta(\lambda, R)$  which quantifies the proportion of light backscattered compared to the incoming light in function of the wavelength. The backscatter coefficient can be expressed as  $\beta(\lambda) = \sigma_s(\lambda)P(\pi)/4\pi$ , where  $\sigma_s(\lambda)$  is the object scattering cross section at wavelength  $\lambda$ , expressing the object ability of scatter light, and  $P(\pi)$  is the phase function, describing the angular dependency of the object scattering efficiency, evaluated in the backward direction.

In the approximation of single scattering, the lidar equation can therefore be expressed as:

$$P(\lambda, R) = A_0 \frac{O(R)}{R^2} \beta(\lambda, R) T^2(\lambda, R) \quad (2.4)$$

Where  $P(\lambda, R)$  is the received light power at wavelength  $\lambda$ , that was backscattered by an object at a distance  $R$  from the instrument.  $A_0$  is an instrumental constant which takes into account the laser source properties as well as the various optics and detector efficiencies. Eventually,  $O(R)$  is the overlap function. It expresses the fraction of the backscattered light going in the receiver field of view. It is null at  $R = 0$  and reaches 1 at a certain range. At mid-range, its expression depends on the characteristics of both the emission and reception parts, as well as the configuration of their optical axes. Figure 2.4 shows a schematic illustration of the overlap

---

function and its value in function of the distance, the light beam and the receiver field of view geometry. The instrumental constant and the overlap function are both variables that depend only on the system and that need to be estimated to obtain information about the atmosphere, which are contained in the backscattering coefficient and the extinction coefficient.

In the atmosphere, the scattering object is an air parcel containing both atmospheric molecules and particles, both responsible for the backscattering of light. Therefore, the backscatter coefficient can be separated in two terms:  $\beta(\lambda, R) = \beta_{mol}(\lambda, R) + \beta_{part}(\lambda, R)$  where  $\beta_{mol}(\lambda, R)$  is the contribution from the molecules to the backscattering coefficient, and  $\beta_{part}(\lambda, R)$  is the contribution from the particles to the backscattering coefficient. In the case where polarization is investigated, it is also possible to separate the backscattering coefficient between the contribution to the parallel and cross polarization, compared to the laser polarization state:  $\beta_{mol/part}(\lambda, R) = \beta_{mol/part}^{\parallel}(\lambda, R) + \beta_{mol/part}^{\perp}(\lambda, R)$ , where  $\beta_{mol/part}^{\parallel}(\lambda, R)$  correspond the parallel backscattering coefficient, meaning the capacity of the object to scatter light while keeping its polarization state, and  $\beta_{mol/part}^{\perp}(\lambda, R)$  is the cross backscattering coefficient which is the capacity of the object to backscatter light and change its polarization state.

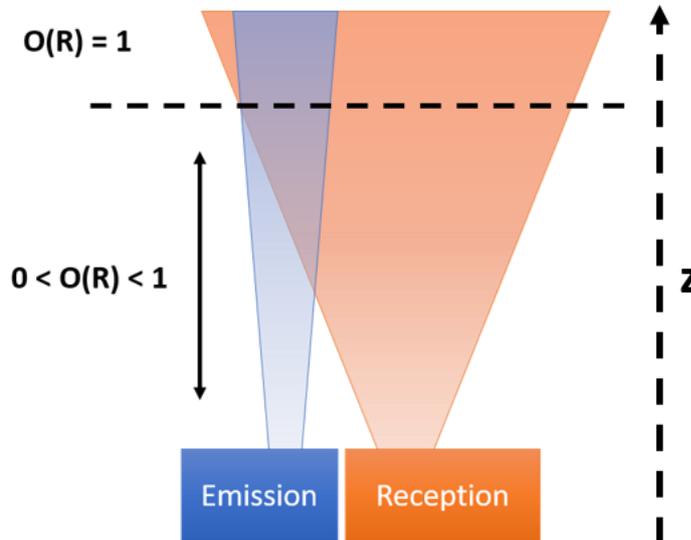


Figure 2.4: schematic diagram illustrating overlap between the laser beam (Emission) and the field of view of the receiving telescope (Reception) for a simple mono-static bi-axial LiDAR system. The black dashed line marks the height above which full overlap (i.e.,  $O(R)$  of 1) of the LiDAR transceiver is achieved and remains independent of height.

Similarly to the backscatter coefficient, the extinction coefficient can be separated between the contribution from the molecules, and the one from the particles:  $\alpha(\lambda, r) = \alpha_{mol}(\lambda, r) + \alpha_{part}(\lambda, r)$  where  $\alpha_{mol}(\lambda, r)$  and  $\alpha_{part}(\lambda, r)$  are the contributions from the molecules and the particles respectively. However, contrary to the backscatter coefficient, in general it is

estimated that the extinction coefficient is independent with regard to the polarization state of the light.

The lidar equation can now be written in its common form as:

$$P(\lambda, R) = A_0 \frac{O(R)}{R^2} \left( \beta_{mol}(\lambda, R) + \beta_{part}(\lambda, R) \right) \exp \left[ \int_0^R -2 \left( \alpha_{mol}(\lambda, r) + \alpha_{part}(\lambda, r) \right) dr \right] \quad (2.5)$$

This equation applies for elastic scattering (Weitkamp, 2006). In the case of inelastic scattering or fluorescence, the light wavelength changes between the forward and the backward trajectory. Therefore, the transmission term changes in the lidar equation, but the other terms are not impacted. The lidar inelastic equation can then be expressed as:

$$P(\lambda_1, R) = A_1 \frac{O(R)}{R^2} \beta_1(\lambda_0, R) \exp \left[ \int_0^R -(\alpha(\lambda_0, r) + \alpha(\lambda_1, r)) dr \right] \quad (2.6)$$

$$P(\lambda_1, R) = A_1 \frac{O(R)}{R^2} \beta_1(\lambda_0, R) T(\lambda_0, R) T(\lambda_1, R) \quad (2.7)$$

Where  $A_1$  is the instrumental constant associated to the considered detection channel,  $\lambda_0$  is the wavelength of the emitted light,  $\lambda_1$  is the received light wavelength, and  $\beta_1(\lambda_0, R)$  is either the Raman backscattering coefficient in case of Raman scattering, or the fluorescence backscattering coefficient in case of fluorescence. The fluorescence backscattering coefficient is defined by analogy with the Raman lidar equation, and describes the ability of the object to produce fluorescence. Inverting the elastic and inelastic lidar equations provides insights into the scattering and extinction properties of the atmosphere, offering a valuable approach for atmospheric research.

### 2.2.3 The LILAS system

Lille Lidar Atmospheric Study (LILAS), is a Mie-Raman-Fluorescence atmospheric lidar in coaxial bistatic configuration present on the ATOLL observation platform in Lille. Operating since 2013, it is an instrument designed to investigate both elastic and inelastic scattering, as well as fluorescence and depolarization with a high automation level. As of 2021, all detection channels have been available simultaneously on the instrument. LILAS is currently operating in the frame of the SNO PHOTONS and ACTRIS-CARS (Centre for Aerosol Remote Sensing), as well as EARLINET since 2015 (Veselovskii et al., 2016, 2020).

The laser of LILAS is a Nd:YAG with frequency doubling and tripling, emitting therefore at 1064 nm, 532 nm and 355 nm with a power of 100 mJ, 100 mJ and 90 mJ for each wavelength

respectively. The laser pulse rate is at 20 Hz, its beam width is at 3 mm with a divergence of 0.5 mrad. A beam expander is used to decrease the divergence of the laser beam by a factor 5.

The receiving part comprises a telescope with a diameter of 40 cm which sends the light toward the receiving optics. The collected light is then separated by dichroic mirrors and polarization beam splitters into the different detection channels. A schematic illustration of LILAS architecture is displayed Figure 2.5.

There are currently 18 detection channels operating on LILAS. 10 are dedicated to elastic scattering detection: for each emitted wavelength, the received light is separated between parallel and cross polarization to compute the depolarization ratio. LILAS also features three Raman channels; at 387 nm, 530 nm and 408 nm. The first correspond to anti-Stokes vibrational Raman scattering by N<sub>2</sub>, the channel at 530 nm corresponds to the Stokes rotational Raman lines of the diatomic molecules N<sub>2</sub> and O<sub>2</sub>, and the channel at 408 nm to anti-Stokes vibrational Raman scattering by H<sub>2</sub>O. For each Raman and elastic detection channel, an interference filter is used prior to the detector to accurately select the detected light wavelength. Finally, LILAS also exhibits a fluorescence channel with a filter centered around 466 nm and with a width of 44 nm. For each collected channel, PMTs are used to acquire the signal, which is afterward processed both in analog and photon-counting modes, except for the elastic channels at 1064 nm where avalanche photodiodes are used instead, and only analog detection is performed due to the high intensity of the signal. The use of both analog and photon-counting modes allows to perform gluing in order to improve the coverage of the instrument, considering the superior performance of analog processing at lower altitudes, where the backscattered signal is more important, and photon-counting at higher altitudes, where the backscattered light is weaker. The detection system temporal resolution is 50 ns which corresponds to a vertical resolution of 7.5 m for the vertical profiles. Moreover, the fluorescence lifetime falling in the nanosecond range, being shorter than the detector resolution, the fluorescence process is therefore seen as instantaneous by the instrument, and the measurement is not impacted by the non-instantaneous aspect of this phenomenon. EARLINET checking procedures are regularly performed in order to certify lidar measurement quality. These procedures are the Rayleigh fit, the telecover test, and  $\Delta 90^\circ$  calibration (Freudenthaler et al., 2018). More details on the instrument can be found in Hu 2018.

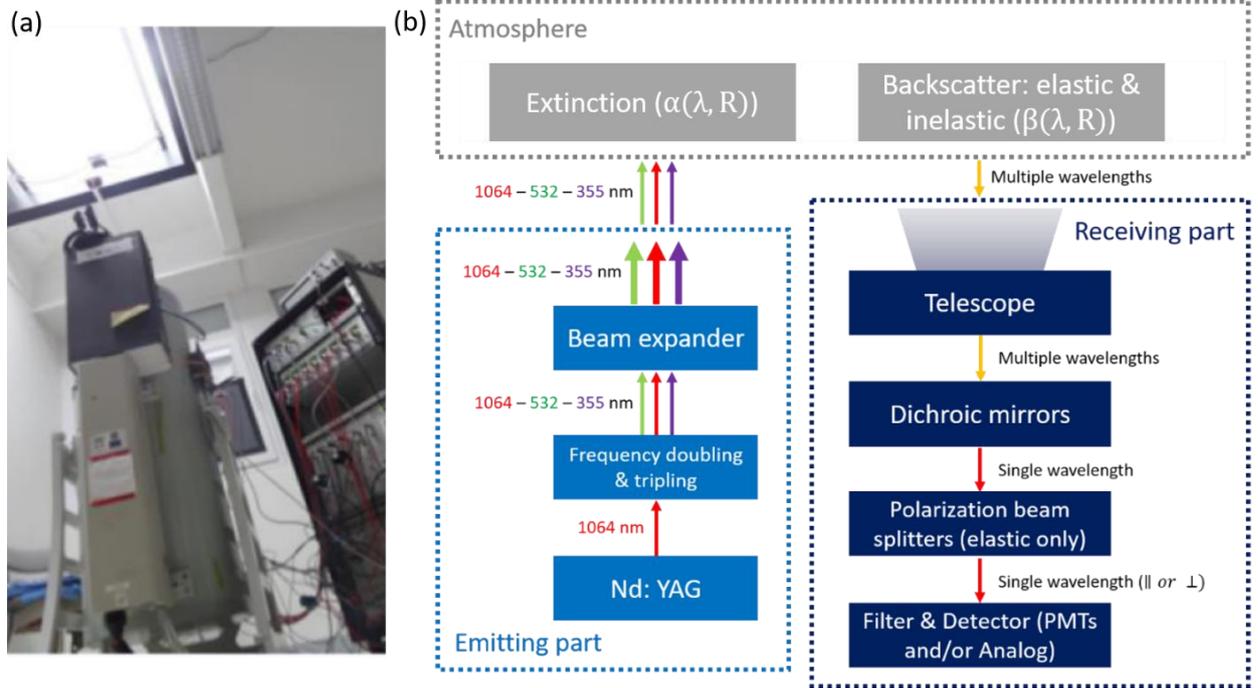


Figure 2.5: (a) Picture of LILAS (b) schematic representation of LILAS emission and the reception parts and their different components, more details on each element of the emitting and receiving parts can be found in part 2.2.3

#### 2.2.4 LILAS profile inversion

The objective of this section is to describe LILAS raw data (level 0) inversion, to derive profiles of humidity and aerosol backscatter coefficients (level 1.5). Pre-processing steps are applied before treating the data, these are trigger delay correction, dead-time correction and electric noise subtraction and are described in Hu 2018. The inversion is managed by AUSTRAL (AUtomatic Server for the TReatment of Atmospheric Lidars), a processing server developed internally at the laboratory (Ducos et al., 2022).

In order to invert lidar profiles, different methods exist. The one chosen in this work is the modified Raman technique described in Veselovskii et al. 2022, which is based on the classic Raman inversion method (Ansmann et al., 1992a). This method includes the choice of a reference height in which the contribution of aerosols to the backscatter signal is negligible.

If the overlap function is assumed to be equal to 1, which is true at a certain distance from the instrument, the ratio between (2.5) and (2.7) for rotational Raman scattering at 530 nm allows to estimate  $\beta_{part}(\lambda_0, R)$ :

$$\frac{P(\lambda_0, R)}{P(\lambda_1, R)} = \frac{A_0 \beta_{part}(\lambda_0, R) + \beta_{mol}(\lambda_0, R) T(\lambda_0, R)}{A_1 \beta_1(\lambda_0, R) T(\lambda_1, R)} \quad (2.8)$$

$$\beta_{part}(\lambda_0, R) = \frac{P(\lambda_0, R) A_1}{P(\lambda_1, R) A_0} \beta_1(\lambda_0, R) \frac{T(\lambda_0, R)}{T(\lambda_1, R)} - \beta_{mol}(\lambda_0, R) \quad (2.9)$$

For rotational Raman scattering,  $\beta_1(\lambda_0, R)$  depends on the scattering molecule properties and its quantity in the atmosphere. Therefore, it can be expressed as  $\beta_1(\lambda_0, R) = \sigma_1(\lambda_0)\rho(R)$ , where  $\sigma_1(\lambda_0)$  is the Raman differential scattering cross section in the backward direction (in  $\text{m}^2 \cdot \text{rad}^{-1}$ ), and  $\rho(R)$  is the density of Raman scatters (in  $\text{m}^{-3}$ ).

By introducing the instrumental constant  $K$  as  $K = (A_1/A_0) \sigma_1$ ,  $\beta_{part}(\lambda_0, R)$  can be expressed as:

$$\beta_{part}(\lambda_0, R) = K \frac{P(\lambda_0, R)}{P(\lambda_1, R)} \rho(R) \frac{T(\lambda_0, R)}{T(\lambda_1, R)} - \beta_{mol}(\lambda_0, R) \quad (2.10)$$

To compute  $\beta_{part}(\lambda_0, R)$ ;  $K$ ,  $\rho(R)$ ,  $T(\lambda_0, R)/T(\lambda_1, R)$  and  $\beta_{mol}(\lambda_0, R)$  need to be estimated. In order to estimate  $K$ , a reference height  $R_{ref}$  needs to be selected, like in the traditional Raman inversion. This reference height should almost be aerosol free, the reason why it is often chosen at high altitude.  $K$  can then be computed as:

$$K = \frac{P(\lambda_1, R_{ref})}{P(\lambda_0, R_{ref})} \frac{1}{\rho(R_{ref})} \frac{T(\lambda_1, R_{ref})}{T(\lambda_0, R_{ref})} \beta_{mol}(\lambda_0, R_{ref}) \quad (2.11)$$

Where  $R_{ref}$  is the reference height, at which  $\beta_{part}(\lambda_0, R_{ref}) = 0$ . It is possible to estimate  $\beta_{mol}(\lambda_0, R)$  and  $\rho(R)$  from standard atmospheric profiles, which is automatically managed by austral. The ratio of atmospheric transmissions between the elastic channel at 532 nm and the rotational Raman channel at 530 nm can be approximated as being equal to 1. This approximation can be performed considering the difference between the emission and reception wavelength being small. These assessments allow to determine the aerosol backscattering coefficient at 532 nm,  $\beta_{part}(\lambda_0, R)$ .

The benefit of this approach is that  $K$  is expected to be stable throughout the day if no change is made on the instrument. Therefore, it becomes possible to invert profiles that do not exhibit any reference zone (because of the presence of low clouds for example), by using the constant  $K$  of the nearest clear sky profile. This approach is not possible with the classical Raman inversion.

The particle linear depolarization ratio ( $PLDR$ ) is defined as:

$$PLDR(\lambda_0, R) = \frac{\beta_{part}^{\perp}(\lambda_0, R)}{\beta_{part}^{\parallel}(\lambda_0, R)} \quad (2.12)$$

This ratio can be computed as the elastic channel is in fact separated with a polarization beam splitter, between the light polarized parallel to the laser polarization ( $P^{\parallel}(\lambda_0, R)$ ) and the light polarized perpendicularly to the laser polarization ( $P^{\perp}(\lambda_0, R)$ ) which are used to estimate  $\beta_{part}^{\parallel}(\lambda_0, R)$  and  $\beta_{part}^{\perp}(\lambda_0, R)$ . This can be done after performing calibration procedures

accounting for the efficiencies of each channel (Freudenthaler et al., 2009; Hu, 2018), and the estimation of the molecular and particular contributions to the total backscatter coefficient. The investigation of the *PLDR* is essential as it allows us to obtain information on the particle's geometry, as perfect spheres tend to keep the light polarization state after scattering, while more complex structures exhibit higher depolarization (Sassen, 2000).

In order to compute the water vapor mixing ratio, the N<sub>2</sub> vibrational Raman channel at 387 nm is used as a reference. The ratio between the lidar detection channel at 408 nm, and the reference channel allows to express the Raman backscatter coefficient:

$$\beta_{H_2O}(\lambda_0, R) = \frac{P(408 \text{ nm}, R) A_{387}}{P(387 \text{ nm}, R) A_{408}} \beta_{N_2}(\lambda_0, R) \frac{T(408, R)}{T(387, R)} \quad (2.13)$$

$\beta_{H_2O}(\lambda_0, R)$  being the Raman backscatter coefficient of the Raman vibrational line of H<sub>2</sub>O, which can be expressed as  $\beta_{H_2O}(\lambda_0, R) = \sigma_{H_2O}(\lambda_0) \rho_{H_2O}(R)$ . Where  $\sigma_{H_2O}(\lambda_0)$  is the Raman differential scattering cross section in the backward direction of water vapor for the Raman line at 408 nm which is known, and  $\rho_{H_2O}(R)$  is the water vapor density.  $\beta_{N_2}(\lambda_0, R)$  is the Raman vibrational line of N<sub>2</sub> which can be estimated from standard atmosphere properties. The ratio of atmospheric transmissions can be approximated using Rayleigh computation for the molecular transmission ratio, and by assuming an Ångström exponent for the particle transmission ratio. The Ångström exponent being defined as:

$$\frac{\alpha(\lambda_1)}{\alpha(\lambda_2)} = \left(\frac{\lambda_1}{\lambda_2}\right)^{-a} \quad (2.14)$$

Where  $\alpha(\lambda_1)$  and  $\alpha(\lambda_2)$  are the aerosol extinction coefficients at wavelengths  $\lambda_1$  and  $\lambda_2$  respectively, and  $a$  is the Ångström exponent.

Finally, to estimate the water vapor density, which can be converted to mixing ratio using standard atmospheric properties, the ratio  $A_{387}/A_{408}$  needs to be known. To do so, the lidar measurement is compared to an external reference. This process is described in part 2.4.2 in this work.

Finally, in order to compute the fluorescence backscatter coefficient, like for the water vapor computation, the N<sub>2</sub> vibrational Raman line at 387 nm is also used as a reference. Equation (2.13) then becomes:

$$\beta_{fluor}(\lambda_0, R) = \frac{P(466 \text{ nm}, R) A_{387}}{P(387 \text{ nm}, R) A_{466}} \beta_{N_2}(\lambda_0, R) \frac{T(466, R)}{T(387, R)} \quad (2.15)$$

The only difference with water vapor inversion however is that there is no reference for aerosol fluorescence to estimate  $\beta_{fluor}(\lambda_0, R)$ . Therefore, the characteristics of the dichroic mirror, and of the interference filter need to be considered. Furthermore, the ratio between the gain of the

detectors of 466 nm and 387 nm LILAS channels needs to be estimated regularly by calibration procedures.

The constant  $K$  is in general computed using a 4-hour time window. All these inversions are then performed on each lidar raw data profile, giving an inverted profile every 1 to 3 minutes, depending on the configuration of the raw time resolution of each profile.

## **2.3 The microwave radiometer**

### **2.3.1 General presentation**

The radiometer is a passive remote sensing instrument measuring incoming radiant flux of electromagnetic radiation, often expressed as brightness temperatures, within a specific wavelength range. First introduced in Dicke et al. (1946), microwave radiometers are instruments designed to estimate the atmospheric temperature and humidity profiles as well as integrated information such as Integrated Water Vapor (*IWV*) and Liquid Water Path (*LWP*). They measure in the microwave domain, defined as electromagnetic waves featuring wavelength ranging from the millimeter to the meter. In general, electromagnetic waves in this domain are not characterized by their wavelength but by their frequency, ranging from 1 GHz to 300 GHz.

In this domain, both oxygen and water molecules exhibit absorption lines. They are located around 60 GHz and 118.75 GHz for the oxygen, and 22.235 GHz and 183.31 GHz for water. Furthermore, liquid water continuously absorbs light in the microwave region.

Microwave radiometer measures along these oxygen and water vapor absorption lines to obtain information on atmospheric properties. The shapes of the absorption lines depend on the molecule concentration, temperature, and pressure. Due to pressure broadening, which can be described as the broadening of absorption lines due to pressure, the shapes of the absorption lines measured by the radiometer are complex and vary with the vertical distribution of the molecule. Since this distribution is known in the case of oxygen, it allows inverting the atmospheric temperature profile. For water vapor, the same method is applied, the information on temperature given by the oxygen line allows to invert water vapor concentration (Skou and Le Vine, 2006). These inversions are not straightforward and usually involve statistical methods such as optimal estimation approaches or artificial intelligence methods like machine learning and deep learning.

### **2.3.2 The RPG-HATPRO G5**

Commercialized by “Radiometer Physics, a Rohde & Schwarz company”, the RPG-HATPRO G5 is a microwave radiometer measuring water vapor and temperature profiles. Installed on

the ATOLL platform since 2021, the instrument also features an infrared extension for cloud base height and ice cloud detection, as well as a weather station allowing it to retrieve atmospheric properties at ground level.



Figure 2.6: Picture of the RPG-HATPRO G5 Radiometer used in this study

The detection channels of this radiometer are around 22.235 GHz for water vapor retrieval, between 51 to 58 GHz for temperature retrieval, and at 31.4 GHz for liquid water retrieval. The absorption spectrum as well as the detection channels can be found Figure 2.7. It exhibits the fact that only one side of the absorption lines is investigated, as their profile is symmetrical. 14 channels are simultaneously acquired, providing profiles every second.

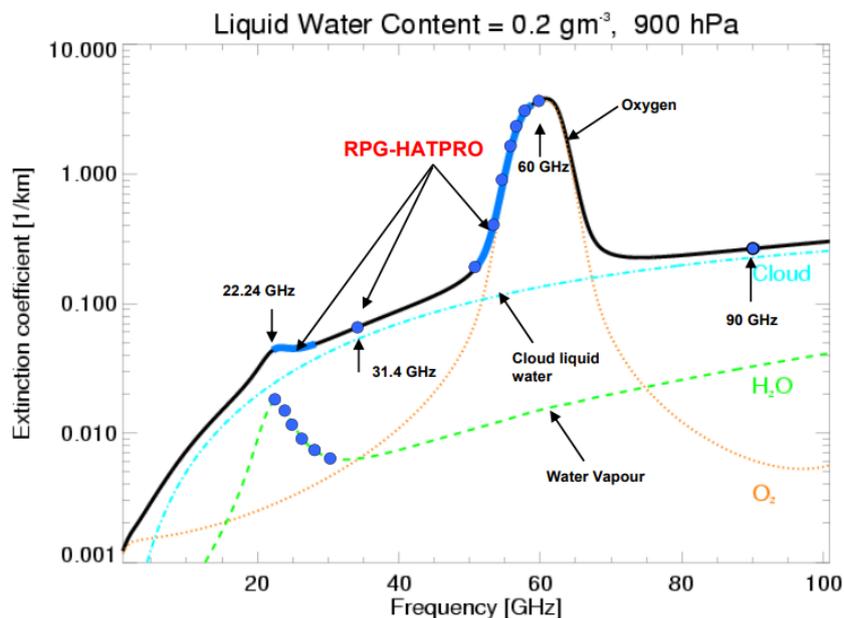


Figure 2.7: spectral features used for microwave atmospheric profiling, from RPG-HATPRO G5 technical instrumental manual, the blue dots correspond to the measurement channels of the instrument: [https://www.radiometer-physics.de/downloadftp/pub/PDF/Radiometers/General\\_documents/Manuals/2015/RPG\\_MWR\\_STD\\_Technical\\_Manual\\_2015.pdf](https://www.radiometer-physics.de/downloadftp/pub/PDF/Radiometers/General_documents/Manuals/2015/RPG_MWR_STD_Technical_Manual_2015.pdf)

The retrieval algorithm of the RPG-HATPRO G5 is based on a neural network approach specifically trained and adapted to the conditions of the measurement site. The instrument

present at the ATOLL platform has been trained with a generated dataset representing over 20 000 possible atmospheric situations gathered by radio sounding measurements, covering a wide range of atmospheric situations, enhancing the performance of the retrieval. On each radiosounding profile, a forward model is used to simulate the measurement the radiometer would perform. The different accuracies of retrievals are of  $0.4 \text{ g. m}^{-3}$  for the humidity profile,  $0.25 \text{ K}$  to  $1 \text{ K}$ , in function of the altitude for the temperature profile, and  $0.12 \text{ kg. m}^{-2}$  for the IWV measurement. The vertical resolution of the retrieved profiles evolves with the altitude, from  $25 \text{ m}$  near the ground to  $300 \text{ m}$  up to  $10 \text{ km}$ , the maximum altitude.

## **2.4 Data preparation**

### **2.4.1 Dataset organization and temperature data**

The dataset spans from February 2021 to the end of 2023, a period during which the radiometer as well as all LILAS detection channels were simultaneously available. The lidar profiles have been inverted a first time to estimate the different instrumental constants for each day. Since the determination of the instrumental constant is not accurate in the presence of clouds, the instrumental constant of the closest clear sky situation has been used to invert cloudy profiles.

In order to reduce the amount of data, and to improve the signal to noise ratio, both the radiometer and lidar profiles have been averaged hourly. Due to the low intensity of the fluorescence emission and of Raman scattering, LILAS measures fluorescence and humidity only during nighttime, daytime data being contaminated by background daylight. Therefore, only profiles measured between 8 p.m. to 4 a.m. were used in this work.

In order to use both lidar and radiometer data, all profile ranges have been aligned on the lidar vertical distribution; ranging from the ground to  $15 \text{ km}$ , with a vertical resolution of  $7.5 \text{ m}$ . Since the radiometer measurement stops at  $10 \text{ km}$ , its profiles were not considered above this altitude. Linear interpolation has been used on the radiometer profiles in order to align the measures on the same ranges.

To provide a temperature reference for comparison with radiometer retrieval, temperature data estimated by the ERA5 reanalysis has also been imported. ERA5 is an ECMWF (European Center for Medium-range Weather Forecast) atmospheric reanalysis. It relies on general circulation models and data assimilation of a vast amount of historical observation. The ERA5 temperature data have been gathered using the ICARE data and services center (<https://www.icare.univ-lille.fr/>), one of the 4 data centers of AERIS, the French data infrastructure for the atmosphere created in 2014. The closest grid point to Lille has been isolated in order to derive the hourly averaged temperature profile in Lille. Similarly to the

radiometer data, temperature data ranges have been aligned on LILAS ranges using linear interpolation.

<b>Quantity</b>	$\beta_{532}$	$PLDR_{532}$	$\beta_{fluo}$	$x_{H_2O}$	AH	IWV	Temperature
<b>Units</b>	$sr^{-1} \cdot m^{-1}$	$\emptyset$	$sr^{-1} \cdot m^{-1}$	$g \cdot kg^{-1}$	$g \cdot m^{-3}$	$g \cdot m^{-2}$	K
<b>Availability</b>	All	All	Night	Night	All	All	All
<b>Source</b>	LILAS	LILAS	LILAS	LILAS	Radiometer	Radiometer	ERA-5 Reanalysis
<b>Time Resolution (initial)</b>	3 min	3 min	3 min	3 min	3 min	10 min	1 H
<b>Time Resolution (final)</b>	1 H	1 H	1 H	1 H	1 H	1H	1 H
<b>Range Resolution (initial)</b>	7.5 m	7.5 m	7.5 m	7.5 m	Variable	Variable	Variable
<b>Range Resolution (final)</b>	7.5 m	7.5 m	7.5 m	7.5 m	7.5 m	7.5 m	7.5 m

Table 2.1: Summary table of the various important quantities used in this study with their characteristics

#### 2.4.2 LILAS humidity calibration

As mentioned in part 2.2.4, unlike other lidar inversions, the retrieval of water vapor with lidar requires the use of an external reference in order to estimate the instrumental constant necessary to invert lidar measurement at 408 nm. Radiosounding data are usually used to perform this calibration as they provide an accurate water vapor profile twice a day (Foth et al., 2015; Mattis et al., 2002). Unfortunately, the proximity of the ATOLL platform to an airport prevents the use of radiosounding. Before the radiometer installation, radiosoundings from Brussels, in Belgium (150 km away from Lille) were used to calibrate the lidar. This process could introduce a lot of biases and therefore, a calibration method based on the radiometer has been developed.

The first approach consisted in using the part of the profile where the lidar water vapor retrieval was the most accurate. This part is approximately located between 2 km and 3.5 km. To make sure that the profiles given by the radiometer were accurate, the AH (absolute humidity) profiles of the radiometer have been compared to LILAS uncalibrated measurement. The water vapor mixing ratio measured by LILAS has been converted to absolute humidity by using the ERA5 temperature.

$$AH = x_{H_2O} \frac{P}{R_a T}, \text{ with } P = P_0 \exp\left(-\frac{gz}{R_a T}\right) \quad (2.16)$$

Where  $AH$  is the absolute humidity,  $x_{H_2O}$  is the water vapor mixing ratio,  $P$  is the atmospheric pressure,  $R_a = 287.05 \text{ J kg}^{-1}\text{K}^{-1}$  is the specific gas constant for dry air,  $T$  is the temperature,  $P_0$  is the reference atmospheric pressure, set at 1013.25 hPa,  $g$  is the acceleration due to Earth gravity and  $z$  is the altitude above sea level.

In order to compare radiometer profiles to lidar measures which have not been calibrated, both profiles have been normalized, meaning that they have been divided by the value of their integral between the ground to 6 km, which is the maximum detection range for lidar water vapor. This process enabled the examination of the profile shapes to determine their agreement. The radiometer temperature has also been compared to the ERA5 temperature.

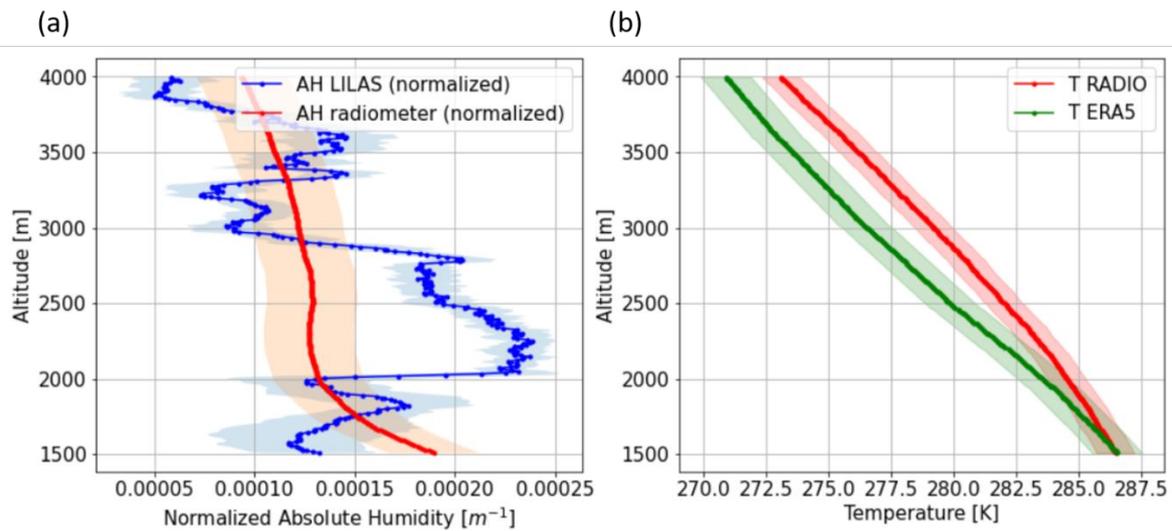


Figure 2.8: Comparison between LILAS and Radiometer Absolute Humidity (a) and between Radiometer and ERA5 temperature (b) on 23 July, 2021 between 00:00 UTC and 01:00 UTC

Figure 2.8 shows an example of comparison between LILAS and radiometer AH, and between radiometer and ERA5 temperature profiles during 23 July 2021, averaged between 00:00 UTC and 01:00 UTC. The comparison between the radiometer and LILAS AH profiles shows that there are noticeable differences between the two estimations which are more important than measurement noise. The profile measured by LILAS is more directly obtained than the one measured by the radiometer and therefore seems more realistic, while the profile determined by the radiometer is smoother. This difference is due to the retrieval method employed to invert the radiometer humidity profile. Indeed, this method, which uses artificial intelligence, relies on little information to rebuild the whole profile, while LILAS is able to obtain information at each point of the profile, making this retrieval method more accurate. The same comment can be made on the difference between radiometer and ERA5 temperature profiles. The difference between the two values reaching up to 3 K, even though ERA5 temperature is not directly

measured, it is the result of models gathering much more information than the radiometer retrieval, and therefore, the ERA5 temperature profile is likely to be more realistic than the radiometer one.

After analyzing the comparison between the two instrument results, it has been concluded that relying solely on the segment of the radiometer profile between 2 km and 3.5 km would not provide a reliable estimation of the humidity calibration constant due the inaccuracies of the retrieval method in comparison to the lidar measurement. Instead, it has been decided to use the radiometer integrated information. Due to the nature of the radiometer retrieval process, this measurement is expected to be much more reliable than the retrieved profile. Following the method presented in Foth et al. 2015, by supposing that most water in the atmosphere is contained between the ground and 6 km, it is possible to use the *IWV* measurement of the radiometer to compute the lidar instrumental constant at 408 nm following equation (2.17).

$$K_{408} = \frac{1}{IWV} \int_0^{h_{max}} x'_{H_2O}(z) \frac{P(z)}{R_a T(z)} dz \quad (2.17)$$

Where  $K_{408}$  is the lidar instrumental constant for humidity measurement, *IWV* is the radiometer Integrated Water Vapor measure, and  $x'_{H_2O}$  is the uncalibrated lidar water vapor mixing ratio measure, and  $h_{max}$  is the maximum considered altitude, here equal to 6 km. This method implies another approximation which concerns the overlap function  $O(R)$  presented equation (2.5). Indeed, in the inversion methods presented in section 2.2.4, this overlap function is approximated as being equal to 1, which is true for LILAS at a distance of approximately 1 km from the instrument, but which does not hold true in lower altitudes, and specifically near to the ground. However, the computation of humidity from lidar data involves the ratio between two detection channels (408 nm and 387 nm). By supposing that the overlap function of these two detection channels is the same, the computation of water vapor as expressed equation (2.13) is still valid.

This method has been used to compute the lidar instrumental constant in clear sky conditions. For cloudy sky situations, it is possible to compute the relative humidity using the lidar water vapor mixing ratio, ERA5 temperature, and the pressure. The relative humidity being defined as the ratio between the water vapor partial pressure and the equilibrium vapor pressure, its value is supposed to be at 100 % at cloud base. This gives a reference point in the profile which allows an estimate of the calibration coefficient.

## 2.5 Chapter summary

This chapter primarily focuses on the instrument used in this study, the Mie-Raman-fluorescence lidar LILAS. It covers the technical characteristics of LILAS and the data

---

processing protocol applied. The data processing in this work utilizes a modified Raman inversion method, which introduces an instrumental constant,  $K$ , allowing for profile inversion without the need for a reference zone in every case. This method is particularly advantageous for processing profiles that contain low clouds or lack a reference zone. For water vapor inversion with LILAS, a reference instrument, the RPG-HATPRO G5 Radiometer, is required, and its details are also provided in this chapter. The calibration method for LILAS assumes that most water vapor is concentrated between the ground and 6 km above ground level, enabling the use of  $IWV$  measurements from the radiometer, as the water vapor profiles from this instrument were not sufficiently precise for direct LILAS calibration.

# **3. FLARE-GMM: an aerosol typing model based on LILAS data, aerosol typing statistical analysis on 3 years of data**

This chapter focuses on FLARE-GMM (Fluorescence Lidar based Aerosol Recognition from Gaussian Mixture Model), an automated aerosol typing algorithm based on LILAS data that uses the Gaussian Mixture Model (GMM). The chapter begins with a discussion on general concepts related to aerosol typing, followed by a brief introduction to machine learning, highlighting its key principles, benefits, and challenges. Next, the development and functioning of FLARE-GMM are presented. Throughout this work, various versions of the algorithm have been trained using different datasets. This chapter details the evolution of FLARE-GMM, addressing the limitations encountered and the improvements made leading to the final version. The final version is then compared with other classification methods to assess its performance. Finally, a brief statistical analysis of aerosol types presents in the Lille atmosphere from 2021 to 2023 is conducted using the LILAS dataset, applying the FLARE-GMM algorithm.

## **3.1 Aerosol typing generalities**

### **3.1.1 Aerosol typing in the literature**

As stated in the first chapter, atmospheric aerosols can originate from various sources. These sources generally fall into two categories: natural and anthropogenic. Aerosols can be directly emitted into the atmosphere, such as desert dust, which are known as primary aerosols. Alternatively, they can form in the atmosphere from gaseous precursors, like Secondary Organic Aerosols (SOA), they are then referred to as secondary aerosols. The physical and chemical properties of aerosols can vary based on their origin. The objective of aerosol typing methods is to use these properties in order to determine the aerosols source type.

Aerosol chemical composition can be estimated *in situ* with instruments like ACSM, which continuously measure the mass and chemical composition of sub-micron aerosols (Ng et al., 2011; Budisulistiorini et al., 2014; Parworth et al., 2015). Filters can also be used to collect both fine and coarse particles, and the samples can then be analyzed to determine the aerosols chemical composition (Edgerton et al., 2005; Yli-Tuomi et al., 2003; Yttri et al., 2015). The Positive Matrix Factorization (PMF) is generally the most commonly used method to estimate aerosol sources from the chemical composition. This method relies on prior knowledge of the specific chemical characteristics of particles emitted by different sources. Matrix factorization allows for the deconvolution of the chemical compositions of observed aerosols using this *a priori* information. The contribution of each source type to ambient aerosols can then be estimated with this method (Paatero and Tapper, 1994; Reff et al., 2007). However, *in situ* methods suffer from numerous limitations to determine aerosol type. The first one is that *in situ* measurement is limited to the instrument location only, which reduces the amount of data and of situations which can be analyzed. Furthermore, the instruments can also show strong constraints. The ACSM for example is limited to analyzing sub-micron aerosols, making it unsuitable for studying coarse aerosols. In contrast, filters can be used to study both fine and coarse particles. However, the chemical composition analysis is typically conducted off-line. As a result, the amount of data obtained is usually much smaller compared to other methods, and the sample acquisition time is very long (often around 24 hours), reducing the opportunities for interpreting and analyzing the results. Moreover, the PMF method also suffers from drawbacks. Namely, the *a priori* information on the source chemical characteristics is generally difficult to accurately estimate and can be subjectively set. Therefore, the analyses can be challenging to interpret and the results can be contested.

Aerosol typing using remote sensing instruments has greatly advanced in recent years. The preferred instrument for conducting these classifications is generally the lidar, as it allows for the measurement of numerous optical properties of aerosols in the atmosphere. These optical properties vary in function of the particle origins, and different methods have been developed to estimate the aerosol sources from them. Various quantities estimated from lidar measurements have traditionally been used for this purpose. The Lidar Ratio ( $LR$ ) is the ratio between the aerosol elastic backscatter coefficient and the extinction coefficient, as the elastic backscatter coefficient can be expressed as  $\beta(\lambda) = \sigma_s(\lambda)P(\pi)/4\pi$ , with  $\sigma_s(\lambda) = \alpha(\lambda)\omega$ , where  $\omega$  is the aerosol single scattering albedo, defined as the ratio of the scattering coefficient to the extinction coefficient. It represents the fraction of the total light attenuation due to scattering alone. Therefore,  $LR$  can be expressed as  $LR = 4\pi/\omega P(\pi)$ , and is therefore an intensive parameter characteristic of the particle, widely used for aerosol characterization. The backscatter color ratio is the ratio between elastic backscatter coefficients measured at different wavelengths. The spectral depolarization ratio describes the dependency of the  $PLDR$

with the wavelength. All these intensive quantities, alongside with the Ångström exponent and the *PLDR* already defined, are generally used to estimate the aerosol type (Ansmann et al., 2011; Baars et al., 2016; Burton et al., 2012; Hofer et al., 2020; Müller et al., 2007). In the difference from *in situ*, deconvolution methods from remote sensing measurements are much more difficult to perform. The main reasons involve the uncertainty levels which are more important, as well as the quantity of information accessible which is more limited than the aerosol chemical composition information obtained with filters. Therefore, aerosol typing from remote sensing instruments generally discriminate aerosols between sources with a wider nature (urban for example, gathering all anthropogenic emissions from city activities), and attempts to identify the main aerosol source a layer is composed of, instead of estimating the contribution of each source type. However, as lidars are able to cover more situations in the atmosphere by measuring along the atmospheric column, and gather more data, advanced statistical classification methods such as machine learning and deep learning can be used to generate automatic classification tools able to perform aerosol typing on large amount of data (Choi et al., 2021; Nicolae et al., 2018a; Papagiannopoulos et al., 2018; Voudouri et al., 2019). These algorithms need to be trained on specific dataset in order to operate, these methods can then be limited by the measurement uncertainties of the optical properties used for the aerosol typing. This mainly concerns the lidar ratio and the Ångström exponent which are computed from the aerosol extinction coefficient which can exhibit high uncertainty levels due to its retrieval method (Ansmann et al., 1992a). In order to overcome this challenge, models can be trained on simulated data. However, this method also shows limitations as simulated data are not expected to perfectly represent real conditions and are subject to biases from the models used for the simulation, which can therefore impact the accuracy of aerosol typing. Simulated data may also not capture all the complexities and variabilities of real-world scenarios, leading to discrepancies in aerosol classification.

### **3.1.2 Aerosol typing with LILAS**

In this work, LILAS data have been used to perform aerosol typing. This instrument features the configuration to estimate the various quantities traditionally used for aerosol typing that has been previously presented. However, LILAS has the particularity of exhibiting a high automation degree, and in order to estimate the intensive properties mentioned earlier, a Raman or Klett inversion of the lidar profiles needs to be performed. These inversions are not straightforward as they require to select a reference zone almost aerosol free in the case of the Raman inversion, or a fixed parameter which is in general either the lidar ratio or the aerosol optical depth (AOD), in the case of the Klett inversion (Ansmann et al., 1992a). Instead, the inversion method presented in chapter 2.2.4 allows to automatically perform the inversion by estimating the lidar instrumental constant for optimal situations and using it in more complex

cases such as cloudy situations. This way, it is possible to automatically inverse the available lidar profiles and work on large amounts of data. However, this method only allows to invert the elastic backscatter coefficient and the *PLDR* at one wavelength, 532 nm, which has been chosen as it is the one exhibiting the best signal to noise ratio out of the three LILAS elastic wavelengths. Therefore, without the extinction coefficient, the *LR*, Ångström exponent, backscatter color ratio and spectral depolarization ratio cannot be estimated in this configuration. Nevertheless, it is still possible to perform relevant aerosol typing with LILAS.

The other particularity of this instrument is its capacity to measure aerosol fluorescence. The fluorescence signal at 466 nm, for an emission at 355 nm, is commonly emitted by bioaerosols. Fluorophore commonly found in these aerosols are phenylalanine, tyrosine, tryptophan, nicotinamide adenine dinucleotide compounds (NADH), flavins and chlorophylls which are generally found in vegetation (Hill et al., 1999; Hoekstra et al., 2007; Pan et al., 1999, 2012). For this reason, smoke aerosol from biomass burning and pollen are expected to exhibit strong fluorescence backscatter coefficients.

However, the fluorescence backscatter coefficient is an extensive quantity which therefore depends on the aerosol concentration. Therefore, it cannot be used for aerosol typing. Instead, in order to use an intensive information on fluorescence, the fluorescence capacity is as:

$$G_{fluor} \equiv \frac{\beta_{fluor}}{\beta_{532}} \quad (3.1)$$

Where  $G_{fluor}$  is the fluorescence capacity, an intensive quantity, which can therefore be used to perform aerosol typing. Veselovskii et al. (2022) proposed a classification method based on two features: the depolarization ratio at 532 nm, and the fluorescence capacity. The depolarization ratio allows to discriminate between smooth aerosols which do not depolarize much, such as biomass burning aerosols, and coarse particles, like desert dust, which strongly depolarize the incident light (Freudenthaler et al., 2009; Gobbi, 1998; McNeil and Carswell, 1975; Murayama et al., 1996). On the other hand, the fluorescence capacity allows to distinguish aerosols containing important traces of bioaerosols such as pollen or smoke, from aerosols containing less bioaerosols like urban aerosols or desert dust.

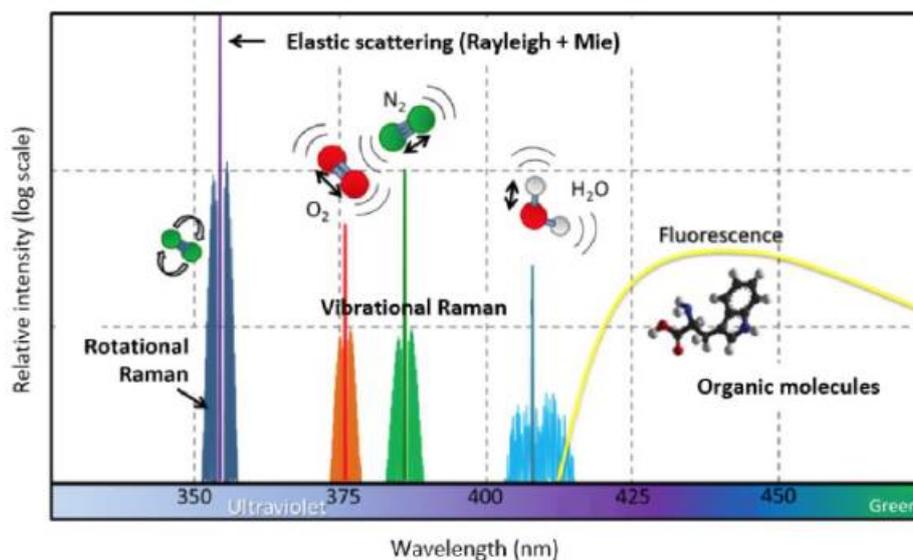


Figure 3.1: Scattering and fluorescence spectra for an excitation at 355 nm (picture from Felidj 2016)

With this approach and in the context of Lille environmental conditions, Veselovskii et al. (2022) demonstrated that it is possible to discriminate between four aerosol types using these two quantities measured by LILAS. Urban aerosols correspond to aerosols with low fluorescence and low depolarization. Desert dust aerosols exhibit low fluorescence as well but high depolarization. Biomass burning aerosols have strong fluorescence but low depolarization. Eventually, pollen show strong fluorescence and strong depolarization. These results are summarized in Figure 3.2.

The main limitation of this method raised in Veselovskii et al. (2022) concerns the treatment of hygroscopicity. Hygroscopic growth, which is treated more in-depth in chapter 4, corresponds to the absorption of water vapor by aerosols, resulting in a potential change of physical and chemical properties. At high relative humidity ( $RH$ ), the aerosol size increases when the aerosol absorbs water vapor, and therefore, the elastic backscatter coefficient increases. However, the fluorescence which is emitted only by the fluorophores present in the aerosols and not by water molecules, is not expected to be impacted by hygroscopic growth. Consequently, at high humidity conditions, the fluorescence capacity is expected to decrease for hygroscopic aerosols, as  $\beta_{fluo}$  remains constant but  $\beta_{532}$  increases, affecting the efficiency of the presented classification method in high humidity conditions. Moreover, hygroscopic growth generally also smoothens aerosols, consequently decreasing the depolarization ratio, further impacting the classification accuracy (Cooper et al., 1974; Ferrare et al., 2023; Haarig et al., 2017).

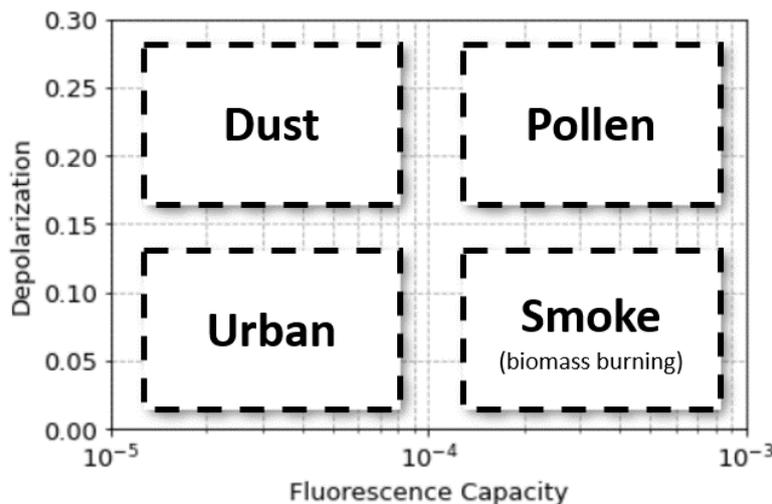


Figure 3.2: Aerosol type with a depolarization/fluorescence capacity diagram. The ranges of the particle parameter variations for dust, pollen, smoke, and urban aerosol are given by rectangles, (adapted from Veselovskii et al. (2022)).

In order to improve the aerosol typing method developed in Veselovskii et al. (2022), a statistical approach has been used to automatically estimate the clusters associated to each aerosol type, and the  $RH$  value measured with LILAS has been used to take hygroscopic growth into account. The next part is dedicated to the presentation of the machine learning model that has been used to perform the automatic classification.

## 3.2 FLARE-GMM presentation

### 3.2.1 Generalities on machine learning

Machine learning is a subset of artificial intelligence that focuses on the development of algorithms and statistical models enabling computers to perform tasks without explicit instructions. The objective of machine learning algorithms is to automatically identify patterns in data, and to take actions such as classification in different categories, or regression. A machine learning model is an algorithm featuring tunable parameters that are estimated during the training phase, also known as the learning phase. The training set is the dataset used for this training phase, it is ideally representative of the different conditions in which the algorithm is supposed to operate. In some occurrences, the training set needs to be preprocessed, either to accelerate computation time, or to form new variables more relevant for the considered problem. The objective of the training phase is to tune the parameters of the model, generally by minimizing a cost function, in order to be able to predict the desired output  $y(x)$  for unseen inputs  $x$ , not contained in the training set. The ability of the model to correctly predict  $y(x)$  for unseen data is called generalization, and is a central aspect of machine learning. Examples of machine learning algorithms are least-squares regression, or hand digit recognition.

Depending on the output nature and of the training set, machine learning algorithms fall in different categories. First, models such as least-square regression, which predict continuous outputs are referred to as regression models. On the other hand, models such as hand digit recognition, which predict discrete outputs are called classification models. When the training set contains a set of inputs as well as the corresponding expected outputs, the model is qualified as a supervised machine learning algorithm, while if the training set does not contain outputs, we talk about unsupervised algorithms.

Probability, decision and information theories are at the heart of the machine learning field, and allow us to work with real data containing uncertainties to solve complex physical problems from a data-driven statistical approach. Machine learning is a field that has significantly developed in recent years, especially with the explosion of modern computing power. Machine learning algorithms are now being used in various domains such as healthcare, finance, and of course, physics (Bishop, 2006).

### **3.2.2 The Gaussian Mixture Model**

Gaussian Mixture Model (GMM) is an unsupervised classification model, also referred to as clustering algorithm. The objective of this model is to recognize patterns in the dataset to identify latent variables and separate the dataset between classes. This algorithm operates under the assumption that the dataset is generated from a finite mixture of Gaussian distributions with unknown parameters. Each Gaussian distribution represents a cluster within the data, and the overall model is a weighted sum of these Gaussian distributions (Bishop, 2006).

Before the training phase begins, the number of Gaussian distributions must be specified. These kinds of specified values are called hyperparameters, defined as parameters with a fixed value, used to shape the model's structure and guide the training process. The training phase consists then in estimating the Gaussians means, variances and weights that best fit the dataset according to the maximum likelihood criteria. This phase can be performed with different methods but the most common one is the Expectation Maximization (EM), an iterative algorithm used to find maximum likelihood estimates of parameters in statistical models with latent variables. This algorithm alternates between two main steps: the expectation step, which consists in computing the expected values of the latent variables given the current estimates of the parameters, and the maximization step, during which the parameters are updated to maximize the likelihood based on these expected values of latent variables (Dempster et al., 1977; Gentle et al., 2012).

In the case of GMM, after the initialization, the expectation step corresponds to the computation of the probability fields of being associated to each distribution. The data points

can then be attributed to a certain Gaussian component, or a certain class, according to the maximum probability. During the maximization step, new Gaussian parameters (means, variances and weights) are computed according to the data repartition issued from the expectation step. By repeating this process, it is possible to lead to a convergence of the GMM to a local maximum of the likelihood function.

This method however shows some limitations. First, by its sensitivity to the initialization, as the algorithm converges toward a local optimum which might differ from the global solution. Moreover, for GMM, computation times of EM algorithms can be very important. For these reasons, the initialization is often selected from the result of a K-means clustering algorithm, a simpler clustering algorithm, which aims at initially getting closer to the optimal solution, reducing computation time and avoiding converging toward local optimums. Moreover, the selection of the a priori number of clusters is also constraining, as a number of clusters not adapted to the real data representation would result in suboptimal solutions. Eventually, preprocessing steps are essential before running the EM algorithm effectively on the training set. Specifically, rescaling the dataset ensures each variable contributes equally to the likelihood estimation. This involves centering the data by subtracting the mean of the overall distribution and scaling it by dividing by the variance. These steps normalize the variables, optimizing the performance and convergence of the algorithm. Indeed, when variables are on disparate scales, the dominant variable exerts more influence on the likelihood estimation, potentially biasing the results (Bishop, 2006). An illustration of the EM algorithm is represented in Figure 3.3.

Once the algorithm is trained, the dataset can then be represented as a sum of Gaussian distributions, and it is possible to express the probability of each data point to have been generated by a specific Gaussian component:

$$p(k|x) = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)} \quad (3.2)$$

Where  $p(k|x)$  is the probability for a data point  $x$  to have been generated by the Gaussian component  $k$  of weight  $\pi_k$ , mean  $\mu_k$  and variance  $\Sigma_k$ ,  $\mathcal{N}(x|\mu_k, \Sigma_k)$  is the multivariate Gaussian distribution evaluated in  $x$ , and  $K$  is the total number of Gaussian distributions specified before the training phase. Being able to compute  $p(k|x)$  for each Gaussian distribution allows classification of the dataset by identifying the maximum probability among the different components.

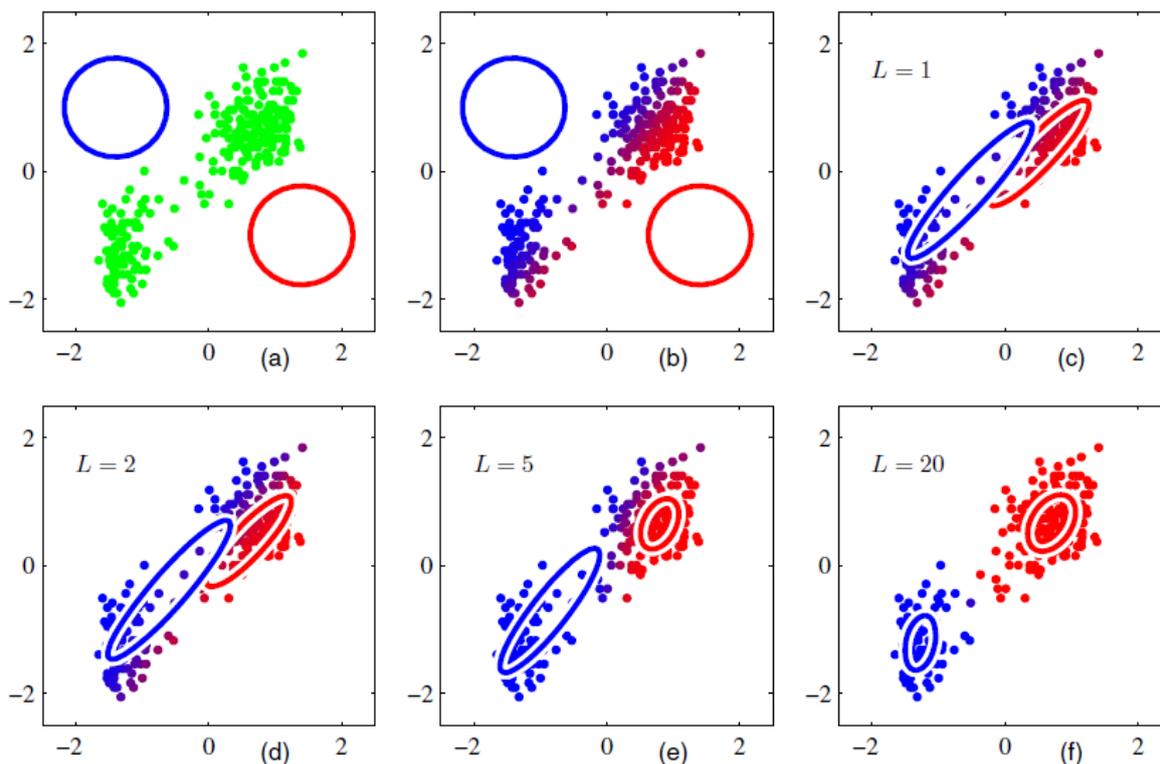


Figure 3.3: Illustration of EM algorithm for GMM fitting to a dataset divided in two clusters.

(a) corresponds to the initialization, the blue and red circles correspond to the Gaussian distributions centered around their mean and with radius corresponding to their variances (b) illustrates the expectation set, the data points are separated according to the maximum probability of belonging to either distribution (c) illustrates the maximization step, the new parameters of the Gaussians are computed according to the new distribution (d), (e) and (f) show the results of the EM algorithm after 2, 5 and 20 iterations respectively,  $L$  being the iteration number. (figure from Bishop (2006))

### 3.2.3 FLARE-GMM: an automatic aerosol typing algorithm

Fluorescence Lidar based Aerosol REcognition from Gaussian Mixture Model (FLARE-GMM) is an automatic aerosol typing algorithm based on LILAS data and the GMM algorithm introduced earlier. Different versions of this model have been tested and are going to be presented. The model takes as features the fluorescence capacity, the  $PLDR$  at 532 nm, as well as  $RH$  depending on the version. The objective of this model is to perform automatic aerosol typing between urban aerosols, desert dust, biomass burning aerosols, and pollen, from LILAS data, according to the repartition represented Figure 3.2.

The advantages of using GMM for aerosol typing lies in several aspects. First, compared to a manual subjective approach, the automatic and statistical nature of this algorithm allows for more rigorous determinations of the different clusters composing the dataset, and of the decision boundaries used for the automatic identification of aerosol type. Moreover, GMM allows us to deal with multi-dimensional datasets. Although it is preferable to keep the dataset dimension as low as possible to reduce computation times and the risk of underfitting, machine

learning models are capable of handling multi-dimensional problems that are much more complex to analyze manually. In our case, this means that incorporating information on *RH* into our model is straightforward, as well as the inclusion of further aerosol optical properties such as Lidar Ratios, Ångström Exponents, or spectral depolarization ratio as the profile inversion method evolves.

Then, compared to other machine learning algorithms, GMM shows many benefits. Compared to K-means, a simpler clustering algorithm, GMM algorithms provide more detailed information and enable finer classification, at the cost of longer computation time. K-means is a non-probabilistic model that uses a hard-clustering approach, whereas GMMs are probabilistic models. This distinction allows GMMs to handle cluster overlap more efficiently than K-means. Additionally, GMMs facilitate the treatment of uncertainty because clusters are represented by Gaussian distributions. This representation allows for better identification of outliers and data points near decision boundaries, a capability that K-means lacks since it only indicates the class to which each data point belongs (Bishop, 2006; Patel and Kushwaha, 2020).

Eventually, compared to neural network methods, GMM also has many advantages. Neural networks are popular algorithms, allowing to solve very complex problems, and have been used in various occasions for aerosol typing (Nicolae et al., 2018a; Papagiannopoulos et al., 2018; Voudouri et al., 2019). The ability of neural networks to solve complex problems lies in the number of their parameters, which can be important depending on the network complexity. However, the high number of parameters means that the computation time for training neural networks can be significant. Additionally, a very large training set is necessary to prevent overfitting, a strong limitation of deep learning algorithms. Finally, neural networks are supervised models, implying that the training set already contains the expected aerosol type. Constituting the training set can therefore represent an important challenge if the classification of aerosols is performed manually, moreover considering that the training set size needs to be large in order to correctly train the neural network. This aspect has motivated researchers to work with simulated data, as it is the case for Nicolae et al. (2018), in order to work with a large training set containing the aerosol type of each data point. The advantage of GMM compared to neural networks lies in the ability to work with real, unclassified data, whereas neural networks require simulated labeled data for training.

The main drawback of GMM however is that it supposes that the data points are generated by Gaussians, making it challenging to identify clusters if the data distribution is not represented well by this distribution. Moreover, in the case of mixtures of different aerosol types, the lidar performs a convolution of the aerosol optical properties due to the fact that the instrument measures the average optical properties of a volume in which different aerosols are present. If

aerosols from different sources are present in this volume, the resulting optical property measured by LILAS would then be a convolution of each source aerosol properties. This aspect is not modeled well by GMM, especially during the training phase, leading to biases in the cluster determination as it will be illustrated later.

### 3.3 Preparation before training FLARE-GMM

#### 3.3.1 Filtering and preprocessing of the dataset

As mentioned in the former part, in order to correctly train GMM algorithms, the training set needs to be preprocessed. Before rescaling the dataset to make sure each variable contributes equally to the cluster identification, it needs to be filtered, as the presence of outliers in the training set can strongly impact the model.

In order to filter LILAS data, the *PLDR* at 532 nm, the elastic backscatter coefficient at 532 nm, the fluorescence capacity, *RH* and the altitude have been considered. First concerning the *PLDR*, situations in which the depolarization ratio is negative have been filtered out as negative values of *PLDR* are non-physical. This means that a problem occurred during the profile inversion, which concerns about 5% of the dataset. Then, regarding the upper limit, it has been chosen to fix it at 40%. This choice has been motivated by the maximum values reached by desert dust aerosols, which are in this range at 532 nm (Haarig et al., 2022). Pollen aerosols may exhibit higher *PLDR* values at 532 nm, up to 80% (Bohlmann et al., 2021), however as it will be shown later, occurrences of pure pollen aerosols have not been observed in the dataset. Instead, pollen is often mixed with urban aerosols in the atmospheric boundary layer, making it difficult to reach such levels of depolarization. Situations for which the *PLDR* value reaches higher values than 40% correspond then to ice clouds, which can be misclassified as desert dust aerosols otherwise.

Then, the elastic backscatter coefficient at 532 nm has also been used to filter the dataset. The objective here is to filter out cloudy cases which can alter the data, either by impacting the profile inversion, or due to the screen effect of optically thick clouds. Therefore, after some tests, it has been chosen that profiles in which  $\beta_{532}$  reaches over  $10 \text{ Mm}^{-1}\text{sr}^{-1}$  have been filtered out to avoid cloudy situations. Moreover, the elastic backscatter coefficient has also been used to filter situations with low aerosol load. It is important to filter these situations as performing aerosol typing if there is little to no aerosol is not relevant. But moreover, if the backscatter coefficient is low, the *PLDR* becomes then very sensitive to measurement noise. Indeed, the depolarization is computed from a ratio between the elastic backscatter coefficients measured in parallel and cross polarization states. Therefore, as these values decrease, the sensitivity of *PLDR* to measurement noise increases and can reach outlier values under low aerosol load conditions. Therefore, if the  $\beta_{532}$  value is below  $0.7 \text{ Mm}^{-1}\text{sr}^{-1}$ , the case has been

filtered out from the training set. This threshold has also been applied to the result of FLARE-GMM, and these situations have been classified as “background”.

Regarding the fluorescence capacity, the results from Veselovskii et al. (2022) as well as the observation of the LILAS data have been used to determine the various thresholds to filter outliers. In the dataset used for this study, smoke aerosols can have fluorescence capacity reaching up to  $10^{-3}$ . This value has then been chosen as the upper threshold for fluorescence capacity, and cases exhibiting higher values have been filtered out. For the lower threshold, a value of  $9.10^{-6}$  has been selected. This choice has been motivated by observations of the fluorescence capacity of urban and dust cases, which rarely fall below this value. Lower fluorescence capacities are typically observed only in clouds. This lower threshold is then a secondary filtering process enabling to exclude such conditions.

Then, concerning  $RH$ , filters have been applied on the dataset to exclude physically unrealistic values. Specifically, situations with negative  $RH$  values or values above 100% have been filtered out, which concerns about 0.5% of the dataset. And finally, only altitudes above 1000 m above ground level have been considered to ensure that the overlap function is equal to 1, guaranteeing the quality of the LILAS profile inversion. Similarly, only altitudes below 6000 m have been considered in the first place to maintain an acceptable signal-to-noise ratio for the  $RH$  data.

After the filtering, preprocessing of the data has been applied. The fluorescence capacity can vary on very large ranges (between  $9.10^{-6}$  to  $10^{-3}$ ), compared to  $PLDR$  and  $RH$ , which typically vary within a single order of magnitude. This wide variation implies that some information may be lost during the rescaling process. To mitigate this effect, the logarithm of the fluorescence capacity has been used as a feature for the GMM training. By applying the logarithm, the range of fluorescence capacity is compressed, reducing the disparity with other variables. Afterward, the training set has been rescaled by retracting the mean of the dataset, and dividing by the variance, to ensure that each variable contributes equally to cluster identification. This rescaling step is crucial as it balances the influence of each variable in GMM, allowing the model to more effectively identify clusters without being dominated by any single variable.

Quantity	$\beta_{532}$	$PLDR$	$G_{fluo}$	$RH$	Altitude
Maximum threshold	$10 \text{ sr}^{-1}\text{Mm}^{-1}$ (profile removed)	40%	$10^{-3}$	100%	6 km
Minimum threshold	$0.7 \text{ sr}^{-1}\text{Mm}^{-1}$	0%	$9.10^{-6}$	0%	1 km

Table 3.1: Summary of the various thresholds used to filter the dataset in the first time

### 3.3.2 Hyperparameter selection

As depicted in the presentation of GMM part 3.2.2, the expected number of clusters has to be selected before training the model. The selection of the appropriate number of expected clusters is essential to optimize GMM outcome, and efficiently represent the dataset. In order to correctly select this hyperparameter, different methods exist. In this work, it has been decided to use the silhouette method (Dinh et al., 2019; Zhou and Gao, 2014).

This method relies on the silhouette coefficient which is a metric used to evaluate the quality of the model in a clustering analysis. It measures how similar a data point is to its own cluster compared to other clusters. The silhouette coefficient of a data point can be expressed as:

$$s = \frac{B - A}{\max(A, B)} \quad (3.3)$$

Where  $s$  is the silhouette coefficient,  $A$  is the average distance between the point and all other points in the same cluster, and  $B$  is the average distance between the point and all points in the nearest cluster that it is not a part of (Rousseeuw, 1987). The silhouette coefficient varies between -1 and 1. If it is equal to 1 it means that the point is well classified, if it is equal to 0, it means that the data point is on or very close to the decision boundary between two neighboring clusters, and a negative value indicates that the data point may have been assigned to the wrong cluster. By considering the average silhouette coefficient of all data points, it is possible to estimate an overall assessment of the clustering quality.

Therefore, a method to select the most appropriate number of clusters that best fit the dataset is to train the model for different values of cluster numbers, compute the average silhouette coefficient of each model, and select the number of clusters which exhibit the highest silhouette coefficient value. This method ensures to choose the number of clusters which best fits the dataset (Dinh et al., 2019; Zhou and Gao, 2014).

In order to reduce computation time, K-means has been used instead of GMM to run this method and estimate the silhouette coefficient. This choice could be disputed as K-means and GMM might have different outcomes in terms of cluster determination. However, the ideal number of clusters depends much more on the dataset than on the model used. Therefore, after conducting some tests, it has been determined that the overall results in terms of choice of cluster number have been consistent between the two models, allowing to run K-means to reduce computation time.

## 3.4 Selection of the training set, the different versions of FLARE-GMM

During the development of FLARE-GMM, several training sets have been tested. In this part, the different training sets used to train the GMM algorithm are going to be presented, as well as the results of FLARE-GMM clusters identification, the conclusions that have been drawn for each version and the evolutions of the classification method.

### 3.4.1 Training with all available data

The first tested version of FLARE-GMM has been trained on all data from LILAS measurements from 2021 to 2023. The objective of this first step is to initially evaluate the performances of the training on all available data. After the filtering, this training set contains 1141 profiles, for a total of 106979 points, which is a really high number, ideal to avoid overfitting issues that might occur in machine learning problems.

Before proceeding to the silhouette method to select the number of clusters, and to the training phase, it can be interesting to investigate the data repartition to check if clusters can clearly be identified in the dataset. Figure 3.4 shows 2D histograms of the training set containing all filtered data from LILAS, in function of the different features used to train FLARE-GMM.

Figure 3.4 (a) shows the repartition of the dataset in function of the fluorescence capacity and the depolarization ratio. From this figure, it is possible to identify a cluster with high depolarization (over 30 %), and low fluorescence capacity (below  $10^{-4}$ ). According to the repartition proposed by Veselovskii et al. (2022), this cluster is supposed to be associated with desert dust. Other than this one, it is possible to identify another cluster, with low depolarization (below 15 %) and relatively low fluorescence (below  $2 \cdot 10^{-4}$ ). This cluster could be attributed to urban aerosols according to Veselovskii et al. (2022), which is consistent with the fact that this cluster seems to gather a large proportion of the data points, which is expected as in the situation of Lille, the boundary layer is most of the time filled with urban aerosols due to the city activity. Then, it is possible to distinguish another cluster at low depolarization (below 10 %) but with high fluorescence capacity (over  $2 \cdot 10^{-4}$ ). These optical properties are characteristics of smoke, or biomass burning aerosols (Veselovskii et al., 2022b). Eventually, one last cluster can be identified in the center of the graph. This cluster probably gathers data points from mixtures. As the lidar measures the average optical properties of the aerosols contained in a defined volume, these points are the result of convolutions between different distributions, which can be challenging to tackle with GMM, and will be discussed later.

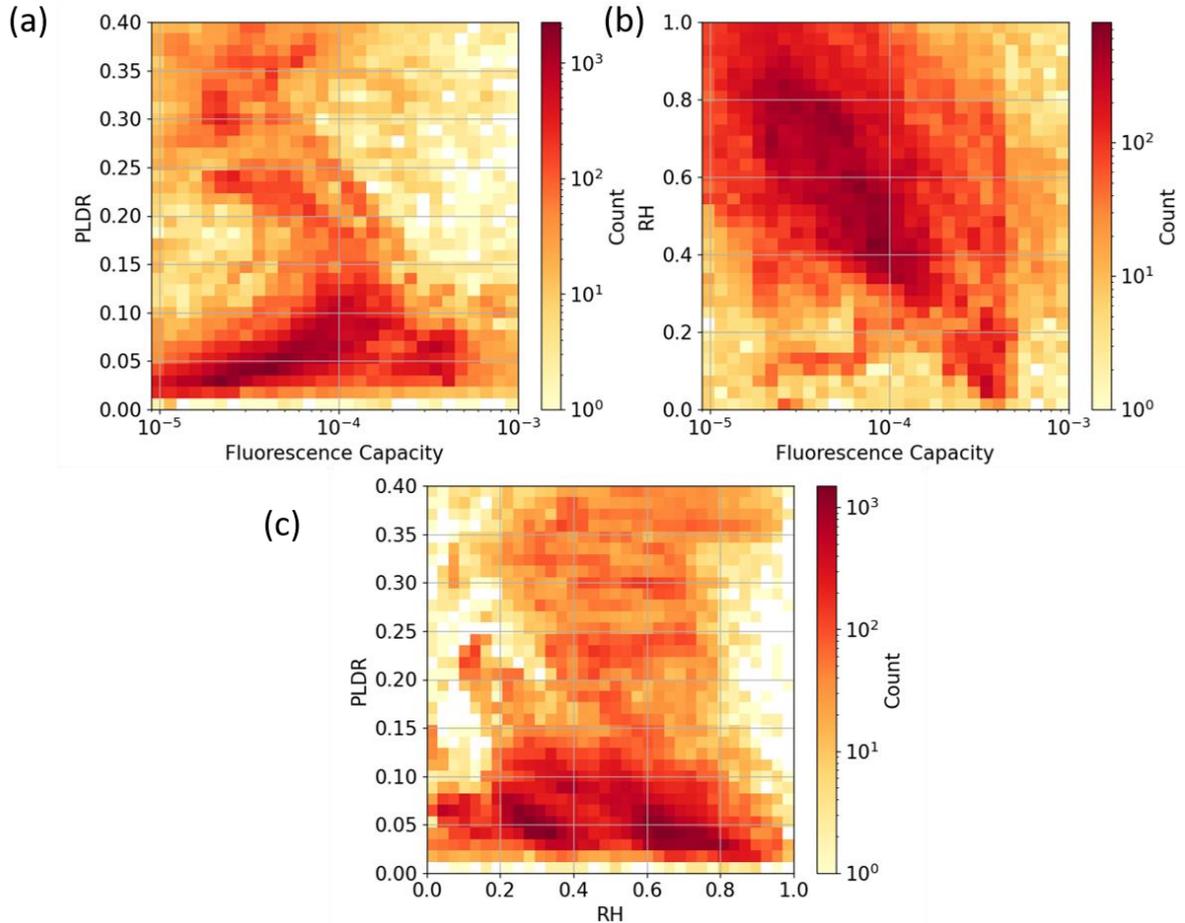


Figure 3.4: 2D histograms of the training set containing all filtered data from 2021, 2022 and 2023 (a)  $G_{fluo}$  vs  $PLDR$ , (b)  $G_{fluo}$  vs  $RH$ , (c)  $RH$  vs  $PLDR$

Figure 3.4 (b) shows the data repartition in a 2D histogram in function of the fluorescence capacity and  $RH$ . Only one cluster seems to appear by looking at this figure, but it can be interesting to analyze the location of this cluster. Indeed, this cluster can be seen as an illustration of hygroscopic growth impact on aerosol optical properties, which has been mentioned in part 3.1.2 and is dealt with more in depth in chapter 4. This impact is characterized by the fact that the aerosol size increases with  $RH$ , thus increasing the elastic backscatter coefficient of the aerosol. As the fluorescence backscatter coefficient remains constant in high humidity conditions, the fluorescence capacity consequently decreases at high humidity. This can be seen in Figure 3.4 (b) as the data distribution mostly ranges from high fluorescence capacity and low  $RH$ , to high  $RH$  and low fluorescence capacity.

Figure 3.4 (c) is more difficult to analyze. It is expected that the  $PLDR$  decreases with  $RH$  due to hygroscopic growth, as the aerosol is expected to smoothen at high humidity (Cooper et al., 1974; Ferrare et al., 2023; Haerig et al., 2017). But this aspect is more difficult to identify than in the case of fluorescence capacity. Two clusters can still be identified at low  $PLDR$  (below 10 %), one at high  $RH$  (above 50 %) and the other at low  $RH$  (below 50%). Among these clusters,

it could be possible to observe a decrease of  $PLDR$  with  $RH$ , however this decrease is not obvious and difficult to evidence.

One comment that can be made, however, is that almost the entire domain represented in the various graphs contains data. This could be due to various factors, such as measurement uncertainty, which increases the size of the different clusters. But it can also be explained by the presence of numerous cases of mixing between different aerosol types. Due to the measurement method, these cases are represented by convolutions between different distributions of these aerosol optical properties, creating data centered between these different distributions.

However, Figure 3.4 (a) shows that aerosols having high depolarization ratio and high fluorescence capacity are rarely observed. Pollen is expected to meet these criteria (Veselovskii et al., 2022b), indicating that pure pollen conditions are almost never observed in the atmosphere around Lille. This is likely because pollen is generally found in the atmospheric boundary layer, where they often mix with urban aerosols. This mixing reduces the fluorescence capacity and the  $PLDR$  of the mixture, making the detection of pollen in this context particularly challenging.

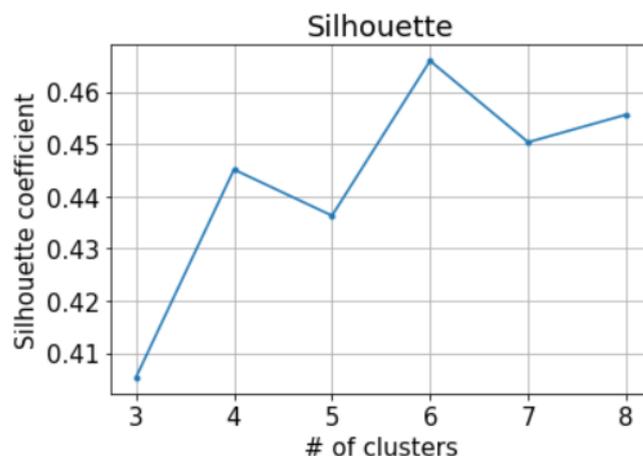


Figure 3.5: Silhouette coefficient from K-Means partitions on the training set containing all filtered data from 2021, 2022 and 2023, for a number of clusters ranging from 3 to 8

Before proceeding to FLARE-GMM training, the expected number of clusters has to be chosen. The silhouette coefficient has been computed for K-Means partitions performed on this training set, for a number of clusters ranging from 3 to 8. The different silhouette coefficients are shown Figure 3.5.

According to the silhouette coefficient method, the number of clusters that is most appropriate to represent the dataset is therefore 6 clusters. This number could have been anticipated as

four different clusters have been identified in Figure 3.4 (a), and two in Figure 3.4 (c) at low  $PLDR$ . The clusters for smoke and urban aerosols can then probably be separated between dry conditions and humid conditions.

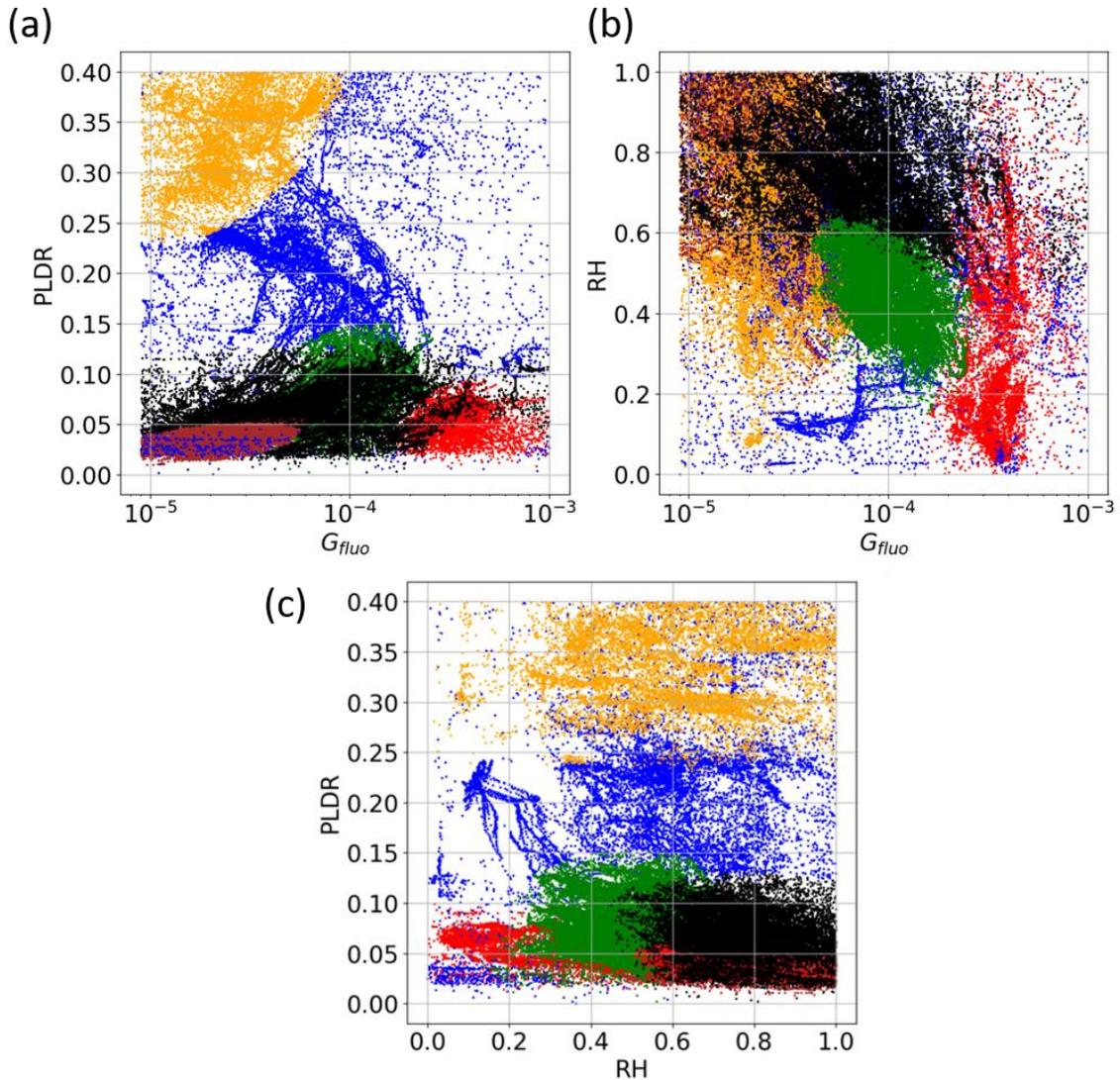


Figure 3.6: Data repartition of the training set with FLARE-GMM according to (a)  $G_{fluo}$  vs  $PLDR$ , (b)  $G_{fluo}$  vs  $RH$ , (c)  $RH$  vs  $PLDR$ , each color is associated to a Gaussian distribution determined by FLARE-GMM

Once these steps are completed, the training of FLARE-GMM can be performed with 6 clusters on the dataset. In order to represent the resulting trained algorithm, the data repartition of the training set can be investigated. The analysis of this repartition allows to observe a representation of the  $p(k|x)$  functions defined equation (3.2) which represent the various Gaussian distributions of each component. By comparing this analysis to the distribution determined by (Veselovskii et al., 2022b), it is first possible to observe if the obtained repartition is relevant, and then to identify each cluster to associate it to its corresponding aerosol type.

Figure 3.6 shows the data repartition between the different clusters determined by FLARE-GMM on the training set according to the different features used for the classification. The objective of this section is to analyze the different figures to identify the aerosol types that could be represented by each cluster. Figure 3.6 (a) shows the data repartition according to the fluorescence capacity and *PLDR*. This figure can be used in comparison to Figure 3.2 to identify the different clusters. First, concerning the orange cluster, it is characterized by an important depolarization (over 20 %), and a low fluorescence capacity (below  $10^{-4}$ ), for this reason, this cluster most certainly corresponds to desert dust aerosols. This can be confirmed by the observation of Figure 3.6 (c), which shows that the depolarization ratio of this aerosol type remains constant with humidity variations. This aspect can be explained by the fact that desert dust exhibit little to no hygroscopic growth (Chen et al., 2020; Herich et al., 2009). Therefore, the fact that the depolarization ratio of these aerosols does not decrease in high humidity conditions is consistent with the suggestion that this cluster corresponds to desert dust. Regarding the blue cluster, it spans a wide range of aerosol optical properties and is centered around medium *PLDR* and fluorescence capacity. For this reason, it is probably associated with the different mixture cases. The clusters showing low depolarization are more difficult to analyze as they are partly superposed in Figure 3.6 (a), making it difficult to distinguish between them. Nevertheless, it is possible to identify that the brown cluster shows low fluorescence capacity (below  $7.10^{-5}$ ) and the red cluster shows high fluorescence capacity (over  $2.10^{-4}$ ). For this reason, the first corresponds probably to urban aerosols, and the second to smoke. The analysis of the black and green cluster is more complicated as they both share similar optical properties. In order to distinguish between them, it is possible to analyze the distributions represented according to *RH* Figure 3.6 (b) and Figure 3.6 (c). We can observe from these figures that the green cluster corresponds to dryer conditions than the black cluster. Since the aerosol fluorescence capacity decreases in high humidity conditions due to hygroscopic growth, the aerosols represented by the green cluster are therefore less fluorescent than the ones represented by the black cluster. For this reason, it is possible to associate the black cluster to humid smoke aerosols, while the red cluster corresponds to dry smoke aerosols. The green cluster can be associated with dry urban aerosols, while the brown cluster corresponds to humid urban aerosols showing lower depolarization and lower fluorescence capacity.

However, this repartition can be criticized on numerous aspects. Mainly, concerning the distributions associated with the smoke aerosols. Concerning the red cluster, associated with the dry smoke aerosols, we can observe from Figure 3.6 (b) and Figure 3.6(c) that it gathers data points with *RH* reaching up to 100 %, which contradicts the fact that this distribution corresponds to dry conditions. Furthermore, concerning the black cluster, it gathers levels of

fluorescence capacity going all the way down to  $10^{-5}$  which is extremely low for smoke aerosols, even at high humidity conditions, showing that the distribution repartition determined by FLARE-GMM is still ambiguous with this training set. Eventually, an interesting aspect is the attribution of data points with low depolarization and low fluorescence capacity associated with the blue cluster. These optical properties correspond to urban aerosols but are associated with a cluster identified as representing the mixtures. This aspect can be explained by the variance of the Gaussian distribution and the probabilistic nature of the GMM model. Since the blue distribution has a higher variance than the others, outliers or data points near the decision boundaries are more likely to be attributed to the blue cluster. This is because the higher variance increases  $p(k|x)$ , the Gaussian probability of association with this distribution, far from the means of each cluster. In order to mitigate this effect, the likelihood has been used to deal with outliers and identify them as unclassified data. This method has been applied in the final version of FLARE-GMM which is presented in part 3.4.3.

Eventually, to evaluate the robustness of FLARE-GMM in this context, it has been tested to train the algorithm with only a fraction of the dataset. The objective here is to evaluate if the distributions remain stable when a part of the training set is removed. To evaluate this robustness, the data from 2023 have been removed from the training set and the training method on FLARE-GMM has been performed with this reduced dataset.

The silhouette coefficient method (figure in the Appendix), shows that the ideal number of clusters for this reduced dataset is not 6 anymore but 5, already suggesting that the distribution might be impacted by this change.

Figure 3.7 shows the repartition of the training set data points in the different clusters from FLARE-GMM when it is trained on data only from 2021 and 2022. By comparing this figure with Figure 3.6, it is possible to observe that some clusters have been strongly impacted by the change of dataset. Figure 3.7 (a) demonstrates that the orange and blue clusters, previously associated with desert dust aerosols and mixtures, now exhibit similar optical properties. This similarity makes it difficult to distinguish between pure desert dust cases and mixtures in this case. Furthermore, the red cluster, previously identified as smoke aerosols now exhibits a much higher variance. Consequently, data points with *PLDR* going up to 17 % are associated with this cluster, which cannot be possible in the case of pure smoke.

These differences demonstrate the instability of the distributions previously presented. An explanation of this instability is the presence of mixture cases. As these cases result from convolutions between different distributions, their presence can therefore be responsible for the distortion of the clusters. Indeed, the GMM algorithm assumes that the entire dataset is generated by a finite number of Gaussians, which in this case represent the different types of

aerosols. However, this algorithm does not account for the presence of data points resulting from the convolution of different distributions. Consequently, the working assumptions are not met by the dataset, which explains this instability.

Different training sets have then been formed to reduce as much as possible mixture cases occurrences. The results of FLARE-GMM training on these different datasets are presented in the following parts.

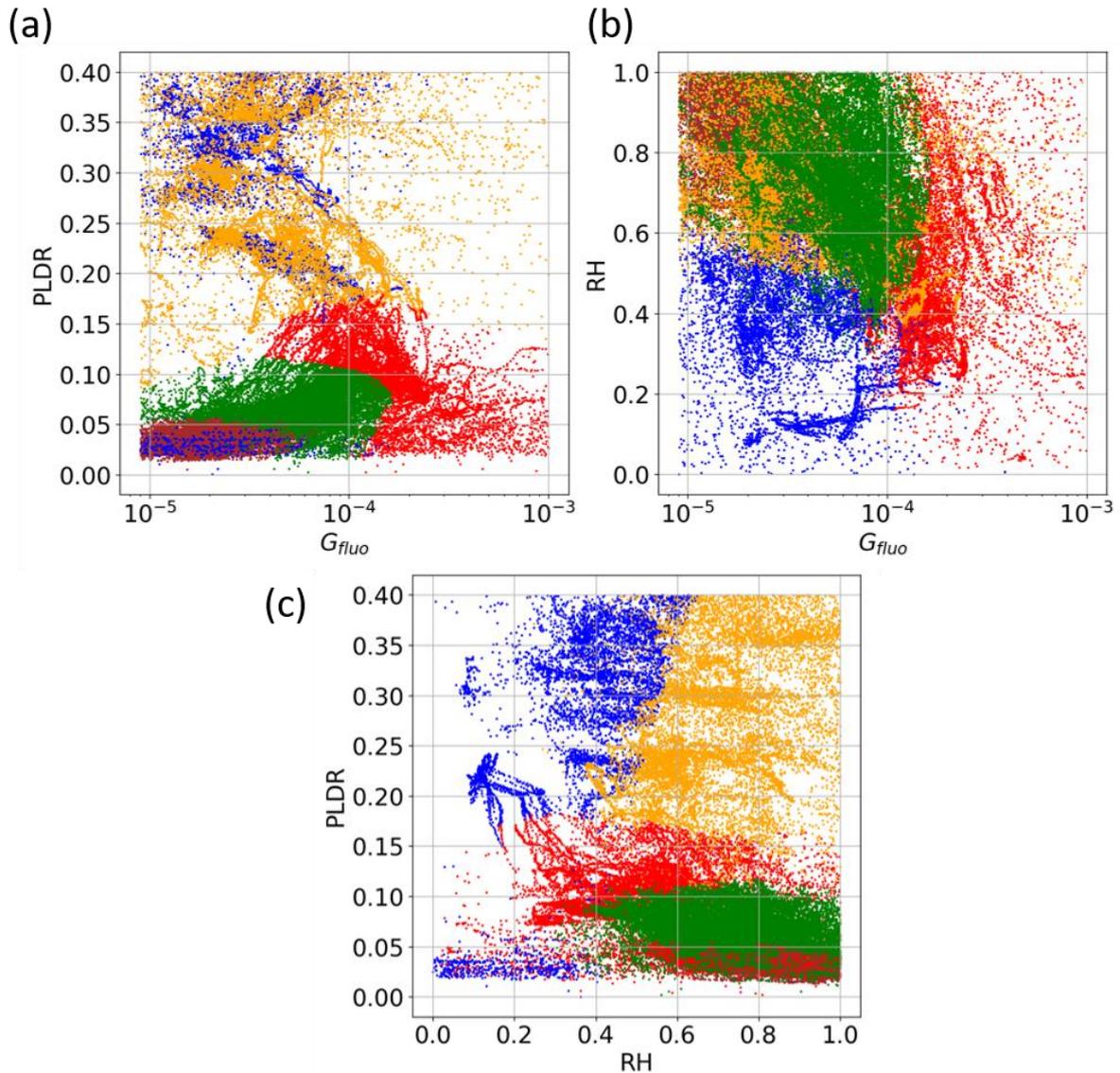


Figure 3.7: Data repartition of the training set with FLARE-GMM, with a training set containing data from 2021 and 2022 only, according to (a)  $G_{fluo}$  vs PLDR, (b)  $G_{fluo}$  vs RH, (c) RH vs PLDR, each color is associated to a Gaussian distribution determined by FLARE-GMM

### 3.4.2 Training without the boundary layer

In order to improve the performance of FLARE-GMM, it has been shown that it is necessary to reduce the amount of mixture cases in the dataset, as they are responsible for the model

instability. For this reason, it has been tested to remove the boundary layer data from the training set, as well as situations with medium to low aerosol loads. The objective of this process is to focus the training set on strong events during which the aerosol properties are specific to one aerosol type, therefore improving the ability of FLARE-GMM to correctly identify the clusters.

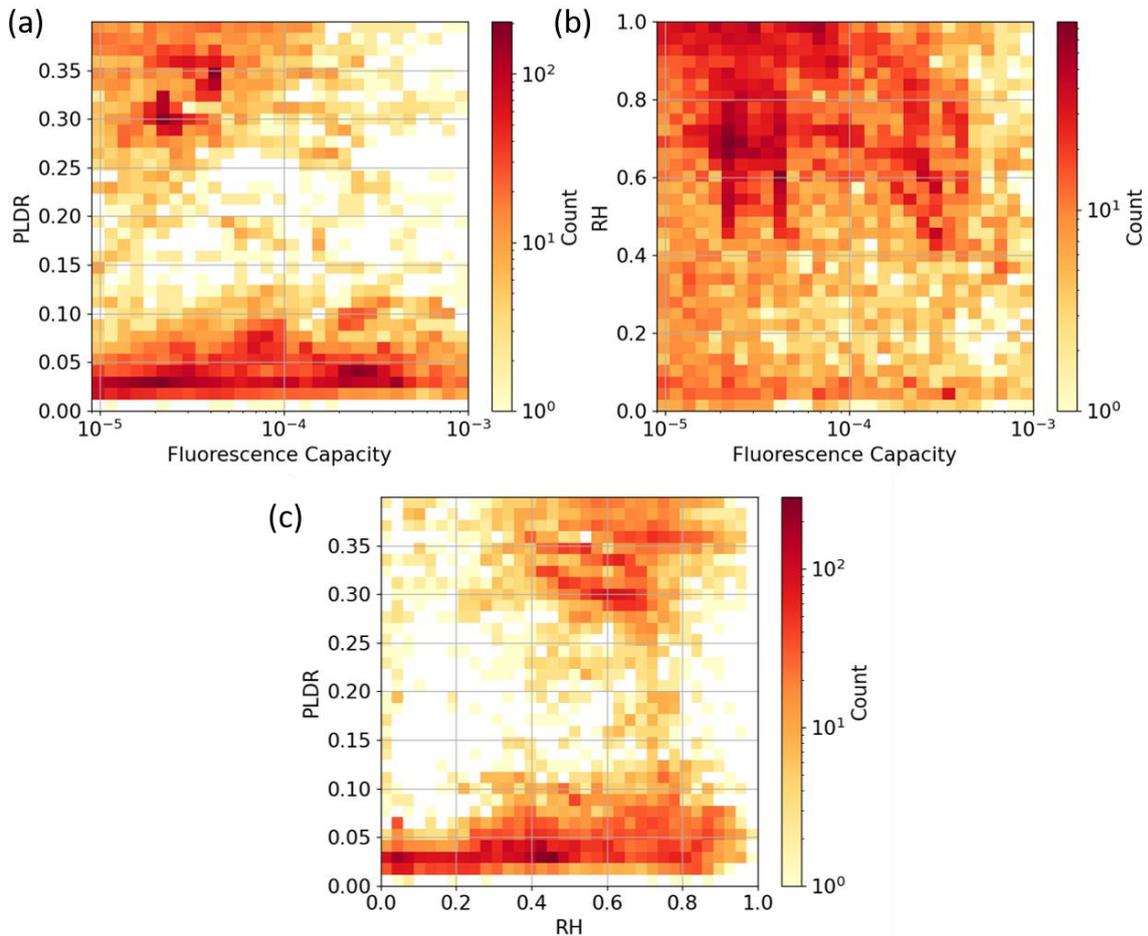


Figure 3.8: 2D histograms of the training set containing all filtered data from 2021, 2022 and 2023, without the boundary layer and low aerosol load situations (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR

The lower threshold on the elastic backscatter coefficient has then been increased from  $0.7 \text{ Mm}^{-1}\text{sr}^{-1}$ , to  $2 \text{ Mm}^{-1}\text{sr}^{-1}$ , and only altitudes above 2000 m above ground level have been considered, instead of 1000 m. These thresholds have been selected after testing to balance between having sufficient data to accurately represent each aerosol type and filtering out as many mixture cases as possible. The dataset used to train FLARE-GMM in this section contains 321 profiles, for a total of 11389 data points. Although this is significantly smaller than the previous training set, it is still sufficient to effectively train FLARE-GMM without causing overfitting, and this training set still covers a wide variety of situations.

Figure 3.8 shows the 2D histogram of the training set used in this section. From Figure 3.8 (a), it is possible to observe that a cluster with low fluorescence and high depolarization, attributable to desert dust, can still be identified. Similarly, a cluster with low depolarization, containing both urban aerosols and smoke, can be observed, making it difficult to distinguish between the two. But the most significant observation is that Figure 3.8 (a) illustrates how filtering the boundary layer and low aerosol load situations effectively eliminates the majority of mixture cases. This is evidenced by the absence of the data points located at the center of Figure 3.8 (a) which can be identified Figure 3.4 (a). Even though some residual points can be observed, a large portion of these cases has been filtered out. This helps minimize their influence on cluster identification and the determination of decision boundaries. Finally, it is important to note that it is difficult to assess whether the mixtures of urban aerosols and smoke have been correctly filtered. Unlike the previously mentioned data, which can be easily identified, these mixtures have physical properties that are more difficult to distinguish from pure cases of urban aerosols and smoke. Therefore, this factor should be kept in mind when analyzing the classification results.

It has been determined with the silhouette method (in the Appendix) that the number of clusters best representing the dataset is 5. FLARE-GMM has been trained using this parameter on the training set presented in this section.

Figure 3.9 shows FLARE-GMM class repartition on the training set, in function of the different features used to train the algorithm. Figure 3.9 (a) can be used to identify the different clusters and to attempt to associate them with the various aerosol types. First, the blue cluster is characterized by a high depolarization ratio (over 25 %) and low fluorescence capacity (below  $2.10^{-4}$ ). These properties are specific to desert dust aerosols, and similarly to the situation analyzed in part 3.4.1, they are almost independent of humidity levels. This confirms that this cluster is attributed to desert dust, as it indicates that these aerosols are almost not hygroscopic. The yellow cluster appears to exhibit a large variance, gathering data points with highly variable optical properties. This cluster appears to be located at the center of Figure 3.9 (a), indicating that it could be associated with mixtures. It is interesting to notice that despite the fact that most mixture cases have been filtered out, as it is suggested from the analysis of Figure 3.8, a cluster gathering mixed cases is still identified during the training of FLARE-GMM, meaning that the filtering process is probably not enough to completely mitigate the impact of these situation on the identification of the decision boundaries. Concerning the clusters with low depolarization, analyzing Figure 3.9 (a) allows to easily distinguish the cluster associated with smoke aerosols from the one linked to urban aerosols. The green cluster consists of cases with very low fluorescence capacity (below  $10^{-4}$ ), whereas the red cluster includes cases with significantly higher fluorescence capacity (over  $10^{-4}$ ). Therefore, the green

cluster can be attributed to urban aerosols, while the red cluster can be associated with smoke aerosols. It is more complicated to correctly analyze the brown cluster. It gathers cases with highly variable fluorescence capacity, low  $PLDR$ , and low  $RH$ . It is therefore probably not associated with a specific aerosol type but rather gathers dry conditions.

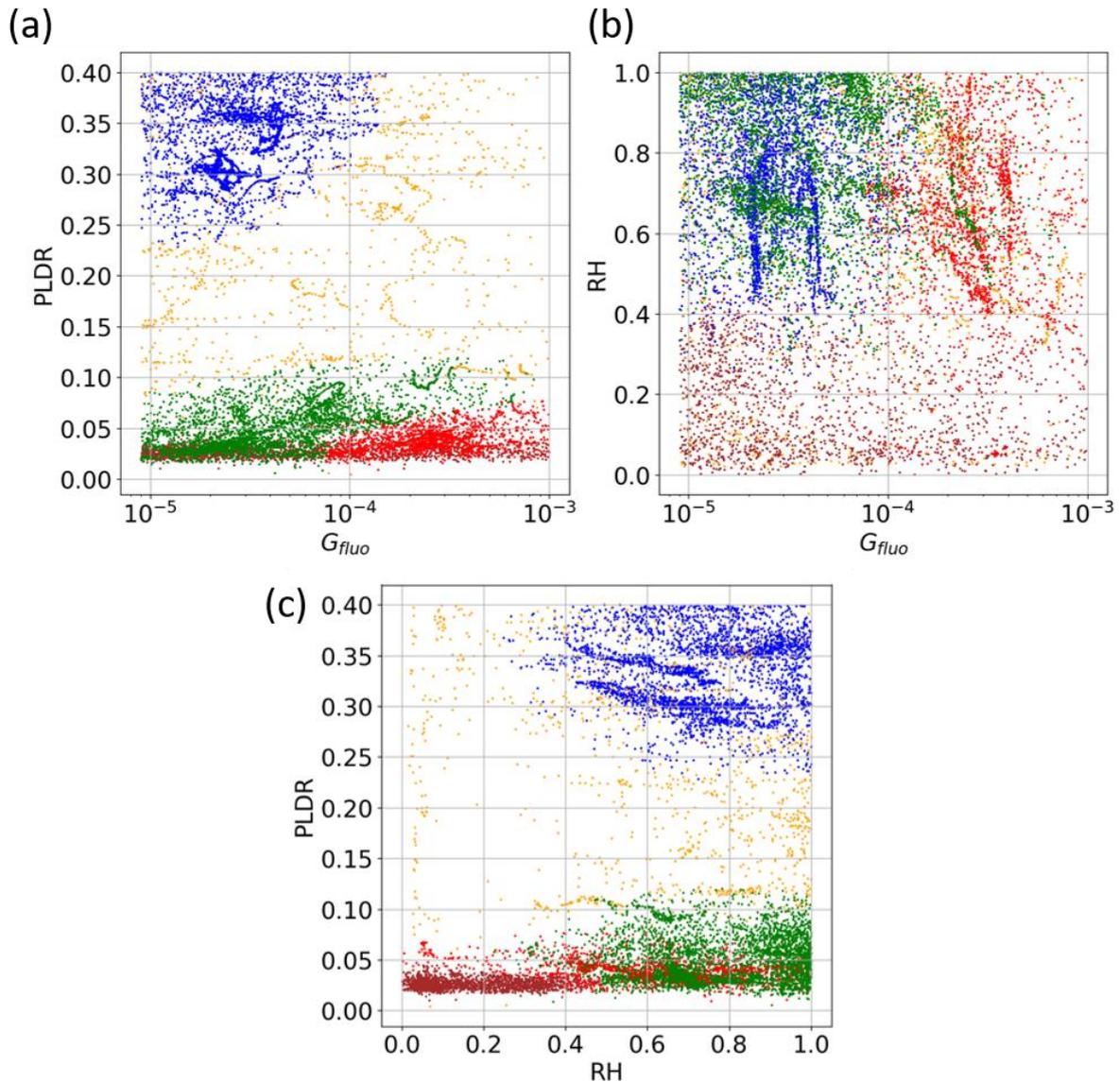


Figure 3.9: Data repartition of the training set with FLARE-GMM, trained on a dataset without boundary layer, according to (a)  $G_{fluo}$  vs  $PLDR$ , (b)  $G_{fluo}$  vs  $RH$ , (c)  $RH$  vs  $PLDR$ , each color is associated to a Gaussian distribution determined by FLARE-GMM

This version of FLARE-GMM seems more reliable than the previous one presented in part 3.4.1. It is easier to distinguish the different clusters, and apart from the brown one, each type of aerosol (dust, urban, smoke) can be easily associated with a specific cluster that stands out from the others. However, it is important to keep in mind that this distribution is still influenced by mixture cases grouped into one cluster, which poses a problem concerning the model working assumptions. Furthermore, the presence of the brown cluster which gathers

dry data from various aerosol types also raises questions. To ensure the robustness of this distribution, the same test as for the previous version has been conducted. One year of data has been removed from the dataset, and the results have been analyzed to ensure that the shape of the distribution is not impacted.

The results after removing the cases that occurred in 2021 are presented. For this training set, the silhouette method (in the Appendix) estimates that the optimal number of clusters for representing the dataset is 6, which already marks a difference with the previous situation. Figure 3.10 shows the data repartition of the training set with FLARE-GMM when it is trained without 2021 data. It is possible to observe from Figure 3.10 (a) that the clusters identified with FLARE-GMM in this configuration are different from the ones presented in Figure 3.9. This mainly concerns the cluster attributed to the desert dust aerosols, which is here separated between the blue and the green cluster, and with the decision boundaries significantly altered compared to those shown Figure 3.9. Furthermore, while the clusters associated with urban and smoke aerosols can still be identified as brown and red, respectively, based on their optical properties, the presence of a black cluster containing a substantial amount of data raises questions. This cluster is similar to the brown cluster represented Figure 3.9, and attributing a specific aerosol type to it is difficult.

Therefore, while it is possible to identify similar clusters, it is also possible to observe that their decision boundaries and characteristics have been largely impacted by the change of the dataset. Therefore, this aspect demonstrates that this method is not robust enough and probably still not adapted to the problem. Once again, the presence of mixed cases can explain why cluster determination is highly sensitive and can change drastically with variations in the training set.

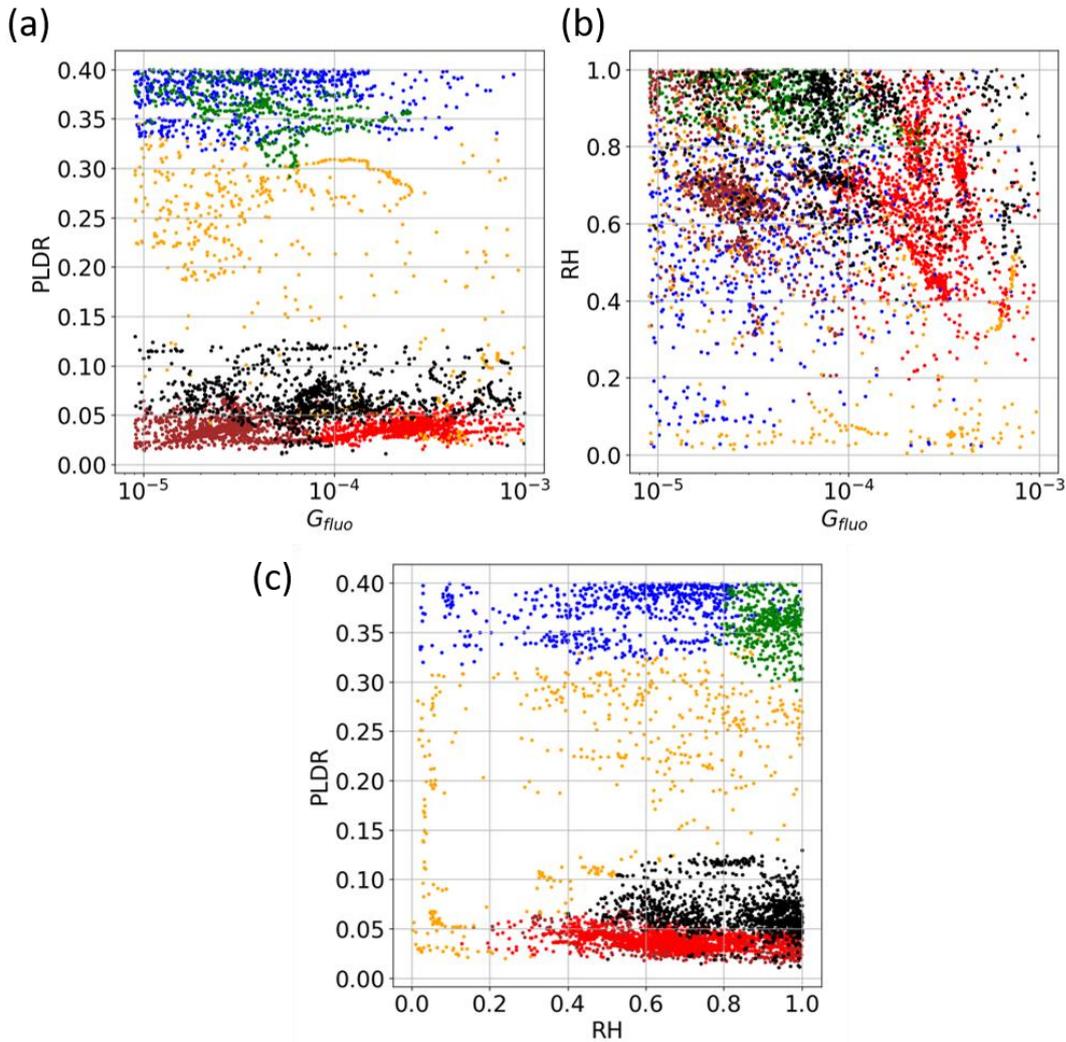


Figure 3.10: Data repartition of the training set with FLARE-GMM, with a training set containing data from 2022 and 2023 only, without the boundary layer, according to (a)  $G_{fluo}$  vs  $PLDR$ , (b)  $G_{fluo}$  vs  $RH$ , (c)  $RH$  vs  $PLDR$ , each color is associated to a Gaussian distribution determined by FLARE-GMM

### 3.4.3 Training with selected cases

In this final version of the training set for FLARE-GMM, the approach significantly changed. The former versions of the algorithm presented in previous parts evidenced the fact that the presence of mixed cases can strongly impact the training of FLARE-GMM and the determination of the clusters.

In order to solve this issue, it has been decided to train FLARE-GMM on a dataset made of hand-selected cases, gathering, as much as possible, pure cases only. In order to select the different cases used for the training of FLARE-GMM, the second version of the classification algorithm presented in part 3.4.2 has been used on LILAS quick-looks. These products are inverted profiles on which the 2 hours average has not been performed. These data, with a higher temporal frequency (approximately one profile every 3 minutes), are used to create

quick-looks that help better identify aerosol layers and cloud structures, at the cost of higher uncertainty. The use of FLARE-GMM on these products allows to identify pure aerosol layers, on which FLARE-GMM aerosol typing is constant. This solution is not ideal since the objective is to work with as much data as possible to cover every scenario. However, it is more important to us to improve the reliability of FLARE-GMM. Indeed, the objective of this algorithm is to work with a large volume of data and being able to quickly analyze estimations of aerosol typing over long periods, it is therefore important to assure the quality of this estimation.

Figure 3.11 shows an example of quick-look that has been used to identify profiles that should be selected to train FLARE-GMM. It is possible to observe from Figure 3.11 (a) that a dust aerosol layer is present in the atmosphere above Lille during the night between 6 and 7 September 2021, located around 2500 m above ground level. Both Figure 3.11 (a) and Figure 3.11 (b), which represents the classified data in function of the fluorescence capacity and the *PLDR*, allow us to estimate that in this situation, the aerosol layer is composed of pure desert dust. The optical properties of this layer represented in Figure 3.11 (b) are indeed characteristic of this aerosol type. This case has therefore been used for the new dataset used to train FLARE-GMM.

With this method, by analyzing the quick-looks of classifications and the scattering plots of the classified data in function of the fluorescence capacity and the *PLDR*, it has been possible to create a training set composed almost only of pure cases. These cases have been filtered with the method detailed part 3.3.1, which allowed to form a training set gathering 138 profiles, for a total of 17078 data points.

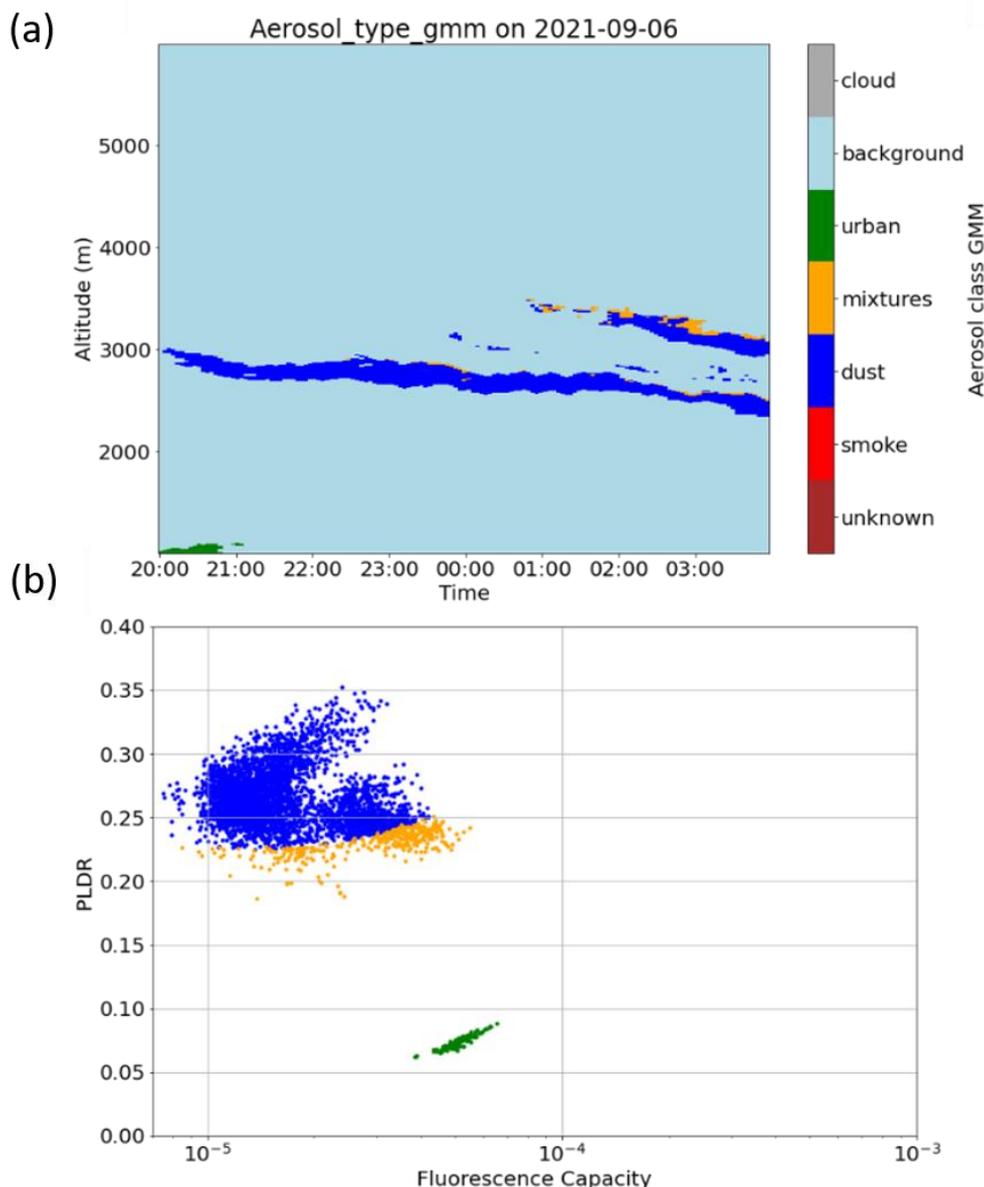


Figure 3.11: (a) quick-looks of FLARE-GMM results (version from part 3.4.2) during the night between 6 and 7 September 2021 (b) Representation of classified data in function of the fluorescence capacity and the depolarization

By observing Figure 3.12 (a), it is possible to notice that no mixture case can be observed anymore in comparison with Figure 3.8 (a) and Figure 3.4 (a). The cluster associated with dust cases with low fluorescence and high depolarization is now clearly identified. Furthermore, concerning the smoke and urban clusters, Figure 3.8 (b) shows that there seems to be a separation between a cluster with high fluorescence capacity and another cluster with lower fluorescence capacity. This observation suggests that even though the clusters are close in terms of optical properties, the use of this training set allows us to differentiate more easily between smoke and urban cases.

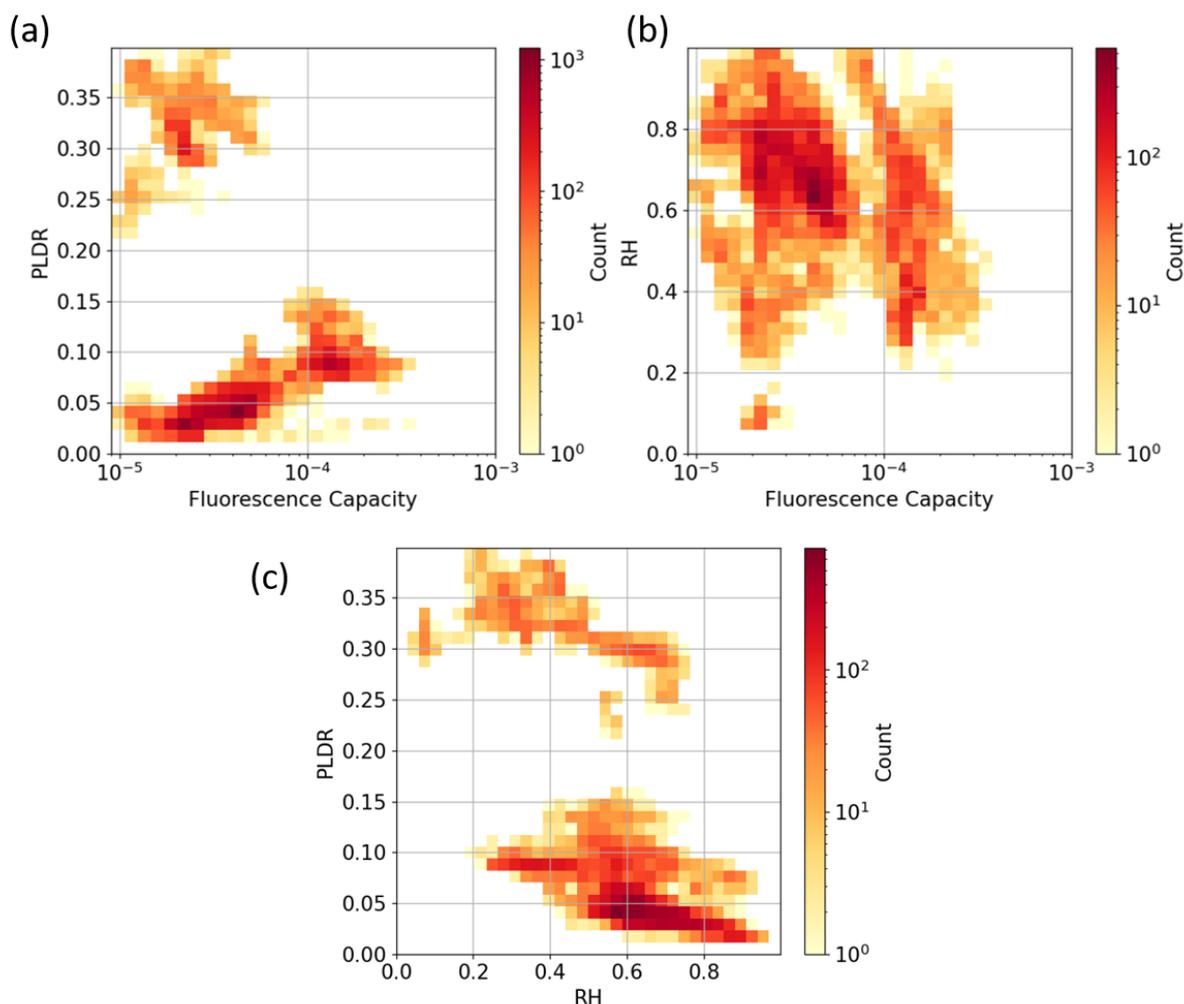


Figure 3.12: 2D histograms of the training set containing hand-selected data of pure cases (a)  $G_{flu0}$  vs PLDR, (b)  $G_{flu0}$  vs RH, (c) RH vs PLDR

The silhouette method has been applied on this dataset to estimate the optimal number of clusters represented in the data. While it has been anticipated that the dataset would be represented by 3 clusters, one attributed to urban aerosols, one to smoke, and one to dust, the silhouette coefficient method has actually estimated that the dataset is composed of 5 clusters, as represented in the Appendix.

FLARE-GMM has therefore been trained with 5 clusters, and the results of the training data classification is shown Figure 3.13. Figure 3.13 (a) shows that the blue cluster can be associated with desert dust as it gathers data with high depolarization (above 20 %) and low fluorescence (below  $10^{-4}$ ). Regarding the discrimination between urban and smoke clusters, it is possible to observe that both red and green clusters exhibit higher fluorescence capacity than the orange and brown clusters. Figure 3.13 (b) also shows that the green and orange clusters gather data in higher humidity conditions than the brown and red clusters. Moreover, the brown and red clusters seem to exhibit higher fluorescence capacity than the orange and green clusters

respectively, suggesting that hygroscopic growth occurred, thus reducing the fluorescence capacity. For these reasons, it is possible to attribute the green cluster to humid smoke, while the red cluster can be associated with dry smoke. On the other hand, the brown and orange clusters can be attributed to dry and humid urban aerosols respectively.

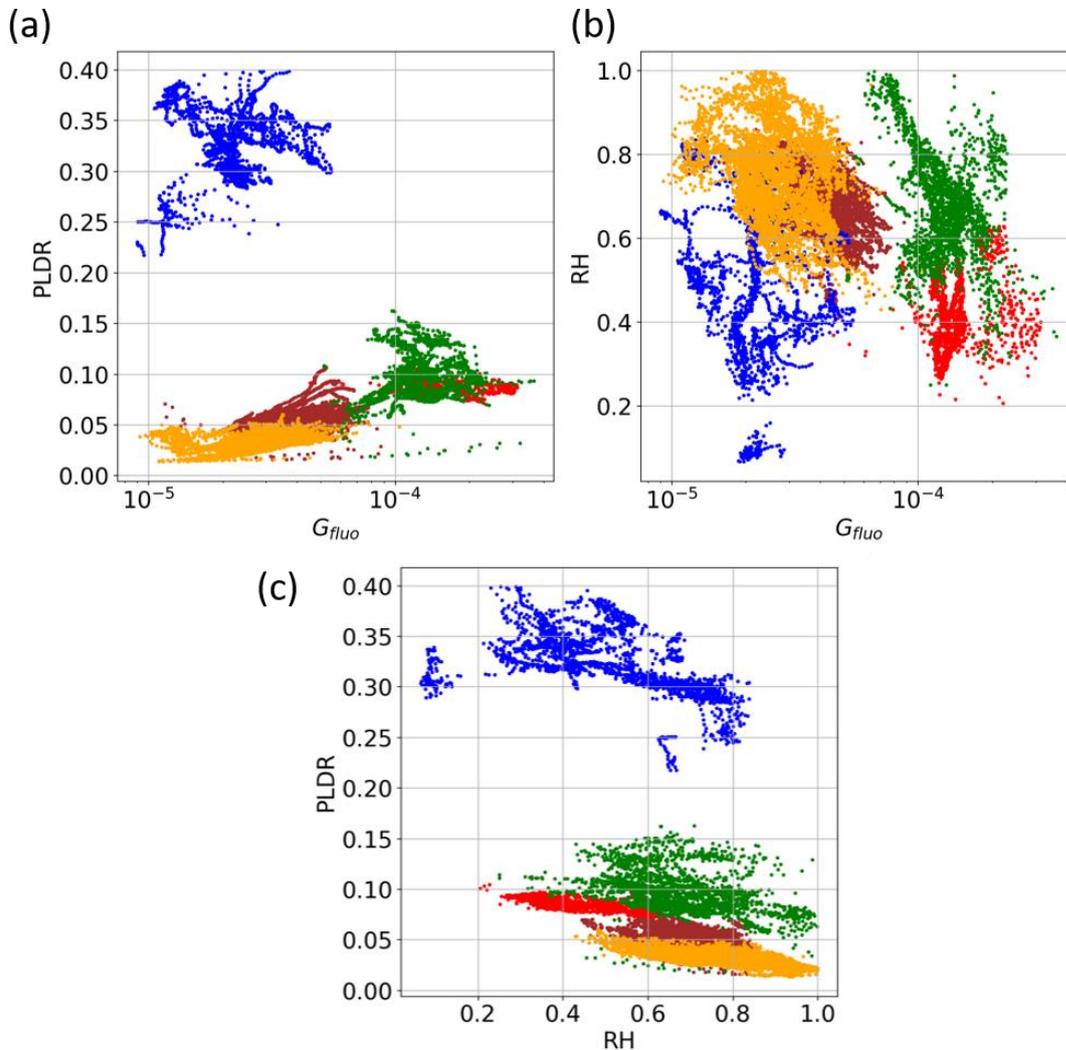


Figure 3.13: Data repartition of the training set with FLARE-GMM, with a training set containing hand-selected data of pure cases, according to (a)  $G_{fluo}$  vs  $PLDR$ , (b)  $G_{fluo}$  vs  $RH$ , (c)  $RH$  vs  $PLDR$ , each color is associated to a Gaussian distribution determined by FLARE-GMM

This data repartition appears ideal, it is possible to differentiate between various aerosol types and to identify if an aerosol layer contains either urban, smoke or dust aerosols accurately with this method. However, questions arise on the separation of the urban and smoke cases between different Gaussian components. It can be interesting to notice that it is not the case for dust aerosols. This aspect can be interpreted by the fact that these aerosols do not exhibit hygroscopic growth, therefore, their optical properties are not impacted by humidity.

However, it may seem ambiguous that the clusters for smoke and urban aerosols are separated into two instead of being combined into a single cluster that accounts for the effects of hygroscopic growth on optical properties. One explanation for this is that hygroscopic growth is not well represented by a Gaussian model. In high humidity conditions, hygroscopic growth decreases fluorescence capacity and depolarization. As a result, if we consider a cluster represented in the  $[G_{fluo}; PLDR]$  plane, the effect of hygroscopic growth can be seen as a translational movement in this plane. Indeed, at low humidity,  $G_{fluo}$  and  $PLDR$  will be high, and they will decrease, along a specific trajectory as  $RH$  increases. Therefore, it can be estimated that the data repartition does not follow a Gaussian distribution according to  $RH$ , as this movement does not align with Gaussian modeling in 3 dimensions. This aspect could explain why the smoke and urban aerosol clusters are separated, and indicated that another representation of this phenomenon should be used.

#### 3.4.4 Use of $RH$ and the treatment of hygroscopicity

In order to solve the issues raised in the former part, and improve the treatment of hygroscopic growth by FLARE-GMM, it has been decided that  $RH$  should not be considered as a feature to classify the data. This choice has been motivated by the nature of the hygroscopic growth phenomenon, which cannot be represented by a Gaussian distribution, as mentioned previously.

Instead, it has been decided to train three different models, one for dry conditions, one for very humid conditions, and one corresponding to medium conditions. To do so, the training set has been divided into three groups based on the  $RH$  value. Dry data points have been selected for  $RH$  values below 60%, very humid data points for  $RH$  values over 80%, and data with  $RH$  between 60% and 80% have been attributed to the dataset corresponding to medium conditions. This repartition has been decided after running some tests. The challenge is to balance between having as many sections as possible to correctly consider the impact of hygroscopic growth on the aerosol optical properties, while also providing that each aerosol class is well represented with enough data points in each section, to correctly train the model. The other benefit of this method is that only two features are used, which allows to visualize the data much more easily.

Figure 3.14 shows the histograms of the different sections of the training set in function of the fluorescence capacity and the depolarization ratio. This figure shows that in each section, the different aerosol types, urban, smoke and desert dust, are represented. The only exception being for dust cases when  $RH$  is over 80 %, with a very low number of data points. However, this low amount of data does not impact the training of FLARE-GMM, as it will be evidenced later. Each section contains 4719, 9970 and 2475 for the dry, medium and wet sections

respectively, which is enough to correctly train FLARE-GMM without risking overfitting the model.

Figure 3.14 also shows an interesting phenomenon, showing a clear shift of the urban and smoke clusters in the  $[G_{flu}; PLDR]$  plane towards lower values of fluorescence and depolarization as  $RH$  increases. This aspect is a direct illustration of the hygroscopic growth impact on aerosol optical properties, which can be clearly observed in this situation.

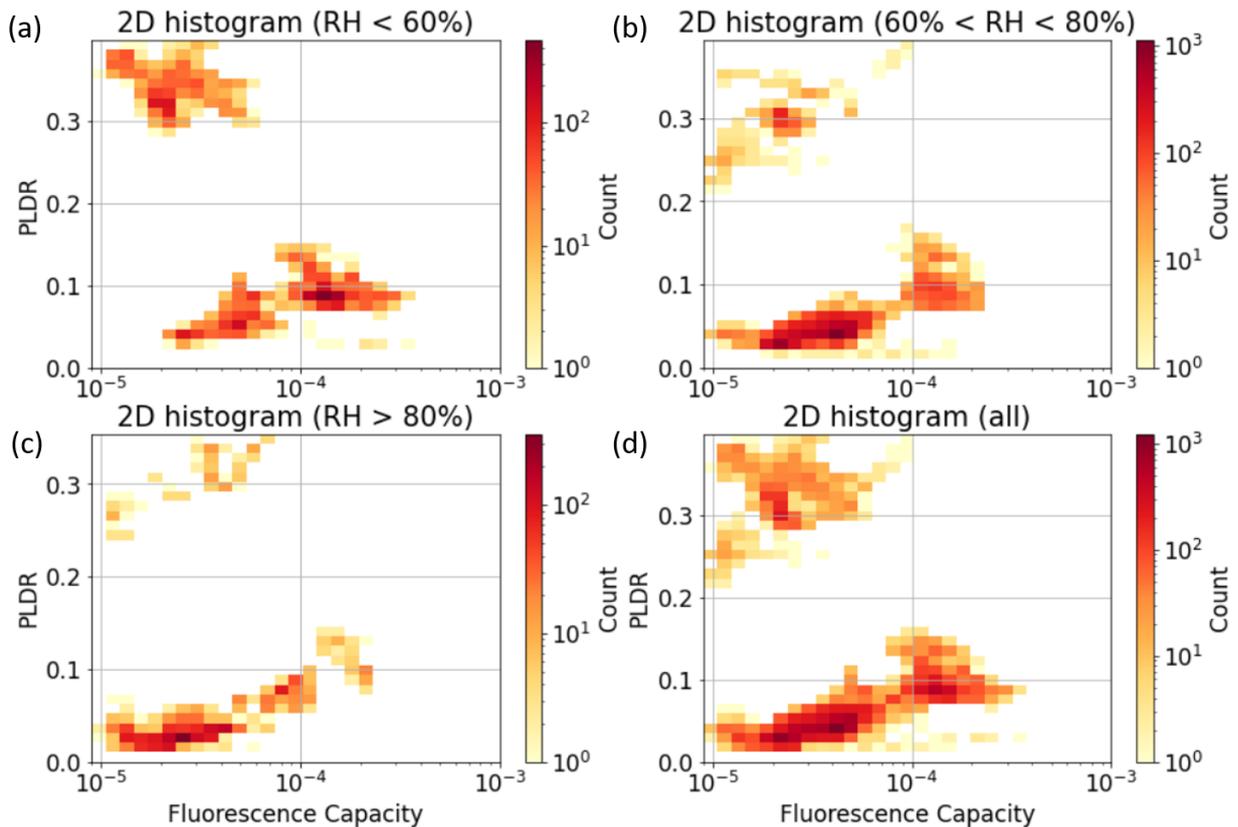


Figure 3.14: 2D histograms of the training set containing hand-selected data of pure cases in function of humidity levels (a) below 60 % (b) between 60 % and 80 %, (c) over 80%, (d) all data

The silhouette coefficient method, shown in the Appendix has been used on each section of the dataset, and in each case, the optimal cluster number has been determined as 3, which is the result that has been expected by the observation of the dataset 2D histograms. This result supports the conclusion made in part 3.4.3, that hygroscopic growth has been responsible for the separation of the clusters in the model presented in this part.

FLARE-GMM models have been trained on each section of the training set, and the results of the training set data repartition are displayed Figure 3.15. On this figure, it is possible to observe that the different clusters are well defined and separated in each case. The association of each cluster to its aerosol type, urban, smoke and dust, is not ambiguous and can be

performed straightforwardly. For this reason, this model is the one that has been chosen as the best aerosol typing method.

However, limitations of this method can be mentioned. First concerning its ability to perform on unseen situations (generalization). Given that the training set contains a limited number of cases, due to the selection process, it can be more challenging for the model to accurately classify an aerosol with properties that are similar to, but distinct from, those on which the model has been trained. Furthermore, the treatment of mixture cases can also be problematic. Unlike the models presented in parts 3.4.1 and 3.4.2, FLARE-GMM in this instance does not have a cluster dedicated to identifying mixtures, making it challenging to accurately classify them.

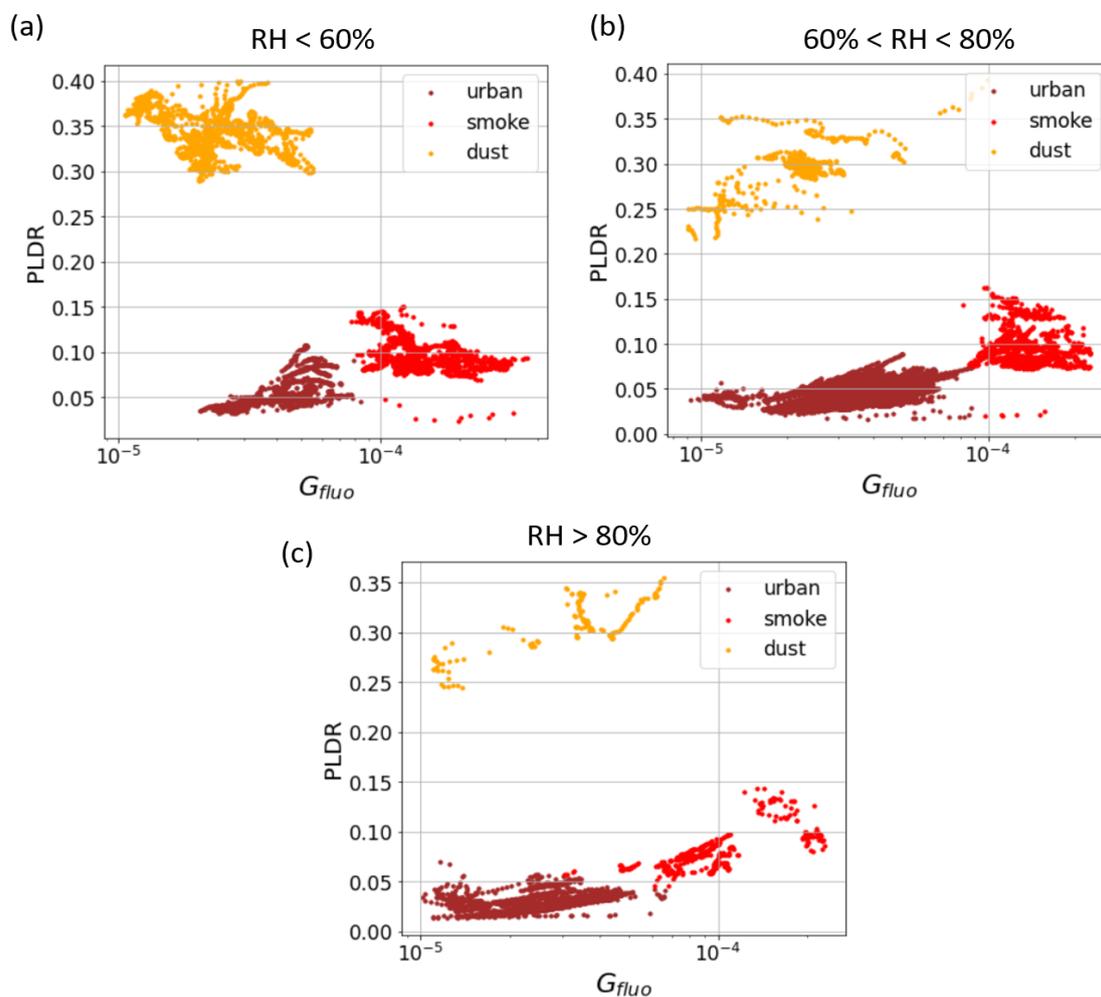


Figure 3.15: Data repartition of the training set sections with the different versions of FLARE-GMM trained on these sections (a) section with  $RH < 60\%$ , (b)  $60\% < RH < 80\%$ , (c)  $RH > 80\%$ , each color is associated to a Gaussian distribution determined by FLARE-GMM

In order to address these issues, the likelihood function can be used. It is defined as:

$$P(x) = \sum_{j=1}^K \pi_j \mathcal{N}(x|\mu_j, \Sigma_j) \quad (3.4)$$

This function can be interpreted as the probability that the point  $x$  has been generated by a Gaussian distribution of the model. Therefore, by choosing a threshold on the likelihood value, it is possible to balance between enlarging the clusters to mitigate the impact of the limited training set, and excluding mixture cases and outliers from the repartition process.

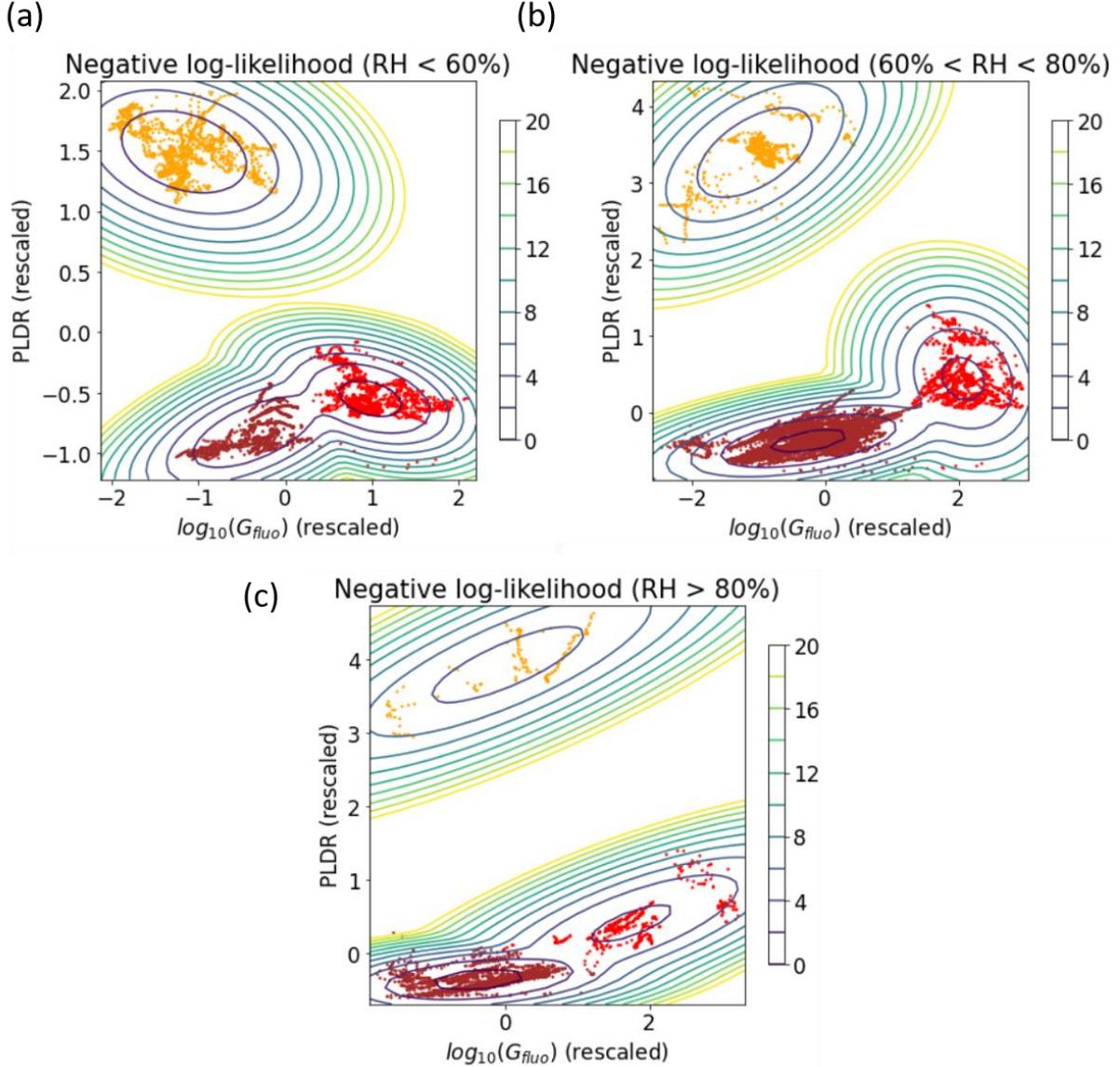


Figure 3.16: Data repartition of the training set sections with the different versions of FLARE-GMM trained on these sections (a) section with  $RH < 60\%$ , (b)  $60\% < RH < 80\%$ , (c)  $RH > 80\%$ , each color is associated to a Gaussian distribution determined by FLARE-GMM, with the representation of the negative log likelihood contour lines ( $-\ln(P(x))$ )

Figure 3.16 shows the contour lines of the negative log likelihood field, which corresponds to the negative of the logarithm of equation (3.4), in function of the rescaled features. This figure

has been analyzed in order to select the optimal threshold of the negative log likelihood, in order to correctly filter out mixture cases and outliers, while also widening the clusters to consider the limits of the training set. Eventually, it has been chosen to fix the threshold at 8. This value has been selected as it allows for unseen cases, for which the negative log likelihood value is generally below 8, to be correctly classified, while also excluding mixture cases and outliers, for which the negative log likelihood value is generally over 8.

Another benefit of this approach is that it can be used at high altitudes, where  $RH$  is much more difficult to obtain accurately. In order to use FLARE-GMM above 6 km, altitude above which  $RH$  cannot be measured by LILAS, it has been decided to use the driest model to classify cases. The motivation behind this approach is that lidar can measure fluorescence and depolarization at very high altitudes, and in clear sky conditions, the  $RH$  level is typically lower at these altitudes. Consequently, hygroscopic growth cases are rarely detected at high altitudes, which is presented in chapter 4. This supports using a dry aerosol model for aerosol typing in such cases. Allowing aerosol typing up to 15 km during the night.

Eventually, it is possible to observe that the cluster associated with urban aerosols and the one associated with smoke particles are very close to each other. Moreover, there is no separation between them once the negative log likelihood criterion is applied. This makes it currently very difficult to differentiate layers containing a mixture of urban and smoke cases from a layer containing pure particles of one aerosol type. Instead, FLARE-GMM can estimate which aerosol has the higher contribution to the mixture's optical properties. In our case, it is not possible to improve this result, as the clusters are too close. The use of other optical properties in the future, such as the  $LR$  or the Ångström exponent, could be essential to help solve this issue.

### **3.5 Generalization of FLARE-GMM**

The objective of this section is to assure the generalization of FLARE-GMM. To do so, the results of FLARE-GMM have been compared to references for unseen cases. In the first part, this concerns cases of extreme events in which the aerosol type is not ambiguous and that has already been documented by other studies. In the second part, FLARE-GMM is compared to NATALI, another aerosol typing model based on lidar data, which uses a neural network.

#### **3.5.1 Classification of specific events**

Assessing the accuracy of clustering models such as GMM can be challenging in the absence of definitive reference. In this section, an analysis of specific events has been performed to have an idea of the algorithm performances to identify aerosol types from LILAS data.

Our initial approach involves exploring FLARE-GMM aerosol typing estimation in instances where aerosol types are not ambiguous and easily identified. These scenarios mainly manifest during specific events of dust or smoke occurrences. Fortunately, the region of Lille frequently experiences such events, which are consistently documented and analyzed by the LOA, and which origins can be checked from backward trajectories (Baars et al., 2019; Draxler et al., 2023; Stein et al., 2015).

The first event analyzed in this section occurred during the night between 15 and 16 March 2022. In this period, strong manifestations of Sirocco winds have been experienced. These phenomena are responsible for the advection of Saharan desert dust over Europe. Consequently, desert dust can be observed in Lille during these events (Husar, 2004; Stohl et al., 2004). The backward trajectory for this night (in the Appendix) confirms that the air present in the atmosphere above Lille in this situation comes from the Saharan region, thus supporting the fact that desert dust is expected to be present in Lille.

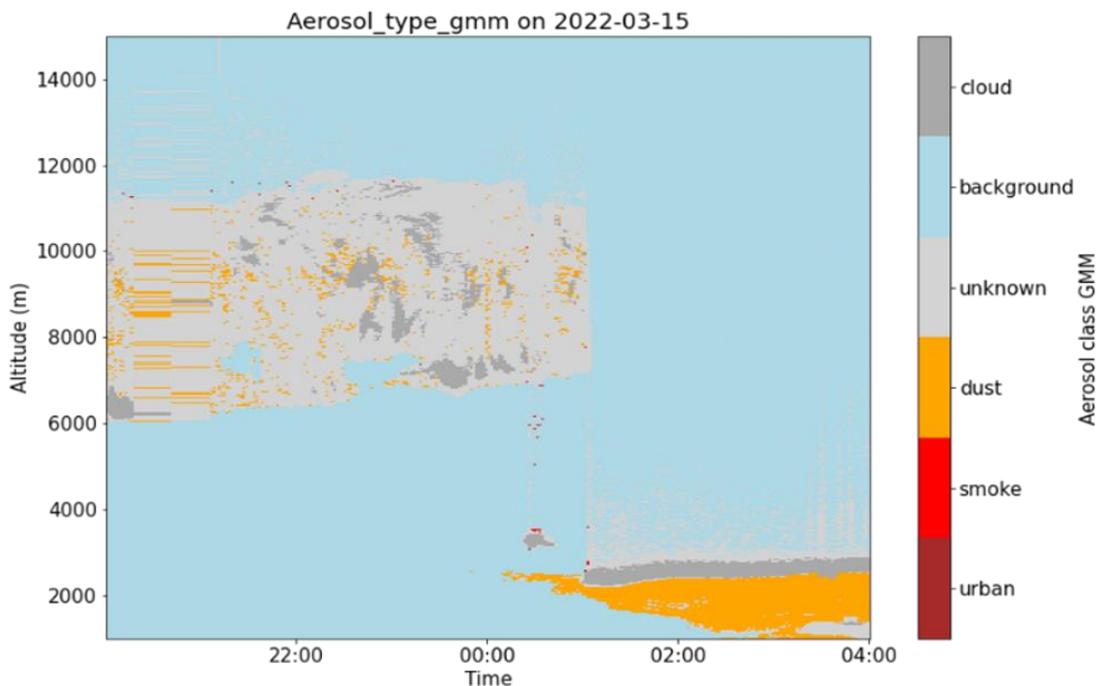


Figure 3.17: FLARE-GMM aerosol type estimation quick-look during the night between 15 and 16 March 2022 between 1000 m and 15000 m

On the other hand, Figure 3.17 shows the quick-look of FLARE-GMM aerosol type estimation during the night between 15 and 16 March 2022. Background and cloud classes are automatically attributed when the elastic backscatter coefficient is below  $0.5 \text{ Mm}^{-1} \cdot \text{sr}^{-1}$  or over  $15 \text{ Mm}^{-1} \cdot \text{sr}^{-1}$  respectively, while the unknown class gathers the outliers and mixture cases for which the negative log likelihood of FLARE-GMM is over 8. Figure 3.17 shows that during the second part of the night, after 00:00 UTC on 16 April 2022, an aerosol layer has

been present below 3000 m, with clouds at the top of the layer. FLARE-GMM estimates that these aerosols are certainly desert dust aerosols, which is consistent with the backward trajectories as well as the analyses and the different reports made on this particular situation (Bouteiller, 2022). FLARE-GMM aerosol typing estimation concurs with the expected result in this case, supporting the fact that FLARE-GMM is able to identify desert dust aerosols in such events.

The second case used in this study occurred during the night between 2 and 3 March 2021. Similarly to the previous case, the presence of strong Sirocco winds has been responsible for the transportation of Saharan desert dust over Europe. The backward trajectory for this case (in the Appendix) shows that the air present above Lille during this night originated from the North African region, supporting the presence of desert dust in the atmosphere. Figure 3.18 shows the quick-look of FLARE-GMM aerosol type estimation for this case. During this night, FLARE-GMM identifies the presence of desert dust aerosols in a layer spanning from 2000 m to almost 7000 m. Clouds can also be observed after 00:00 UTC, ranging from 3000 m to 10000 m, while the lowest part of the atmosphere is associated with unknown aerosol type, which could correspond to a mixture between desert dust and urban aerosols in the boundary layer. Such as the previous studied case, in this situation, the ability of FLARE-GMM to correctly identify desert dust aerosol layers is illustrated here, as FLARE-GMM aerosol typing estimation corresponds to the aerosol type expected from backward trajectory and previous analyses (Veselovskii et al., 2022b).

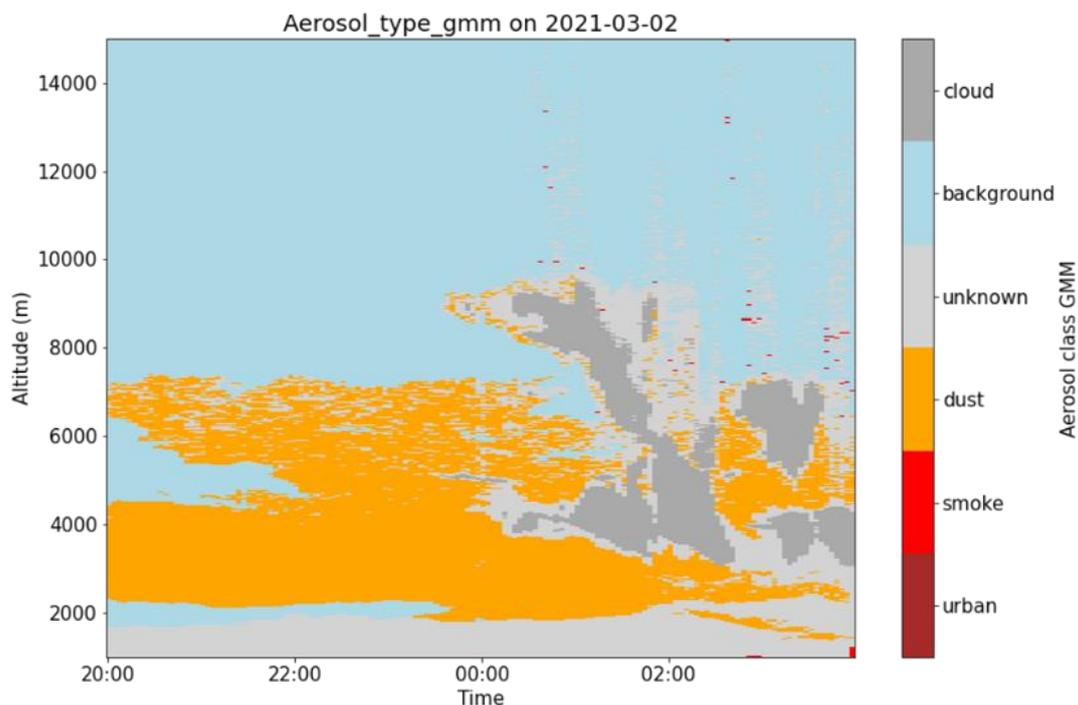


Figure 3.18: FLARE-GMM aerosol type estimation quick-look during the night between 2 and 3 March 2021 between 1000 m and 15000 m

Eventually, the last case investigated in this part occurred during the night of 19 July 2022. During this period, significant forest fires occurred in the Gascogne region in southeastern France. The winds blowing northward during this event transported the biomass burning aerosols all the way to Lille. This can be clearly observed in Figure 3.19 where both the fire map and the backward trajectory are represented. Figure 3.19 (a) shows the backward trajectory in this situation and Figure 3.19 (b) shows the fire map between 14 and 19 July 2022, obtained from the Fire Information for Resource Management System (FIRMS), which uses data from MODIS and the Visible Infrared Imaging Radiometer Suite (VIIRS), and is managed by NASA (source: <https://firms.modaps.eosdis.nasa.gov/>, last access: 28 June 2024). On this map is also highlighted the fire data corresponding to the forest fires which occurred in the Gascogne region at this period. By analyzing both these figures, it is possible to observe that biomass burning aerosols emitted from the forest fires, have been transported to Lille during this period. On the other hand, Figure 3.19 (c) shows the aerosol typing estimation from FLARE-GMM during the night of 19 July 2022. This figure shows that FLARE-GMM correctly recognizes the presence of a smoke layer ranging from 2000 m to 6000 m with the presence of unknown aerosols in the lower part of the atmosphere, which could correspond to a mixture with another aerosol type like urbans, or to outliers. Nevertheless, this case illustrates well the ability of FLARE-GMM to identify smoke layers in such conditions, supporting its efficiency and generalization.

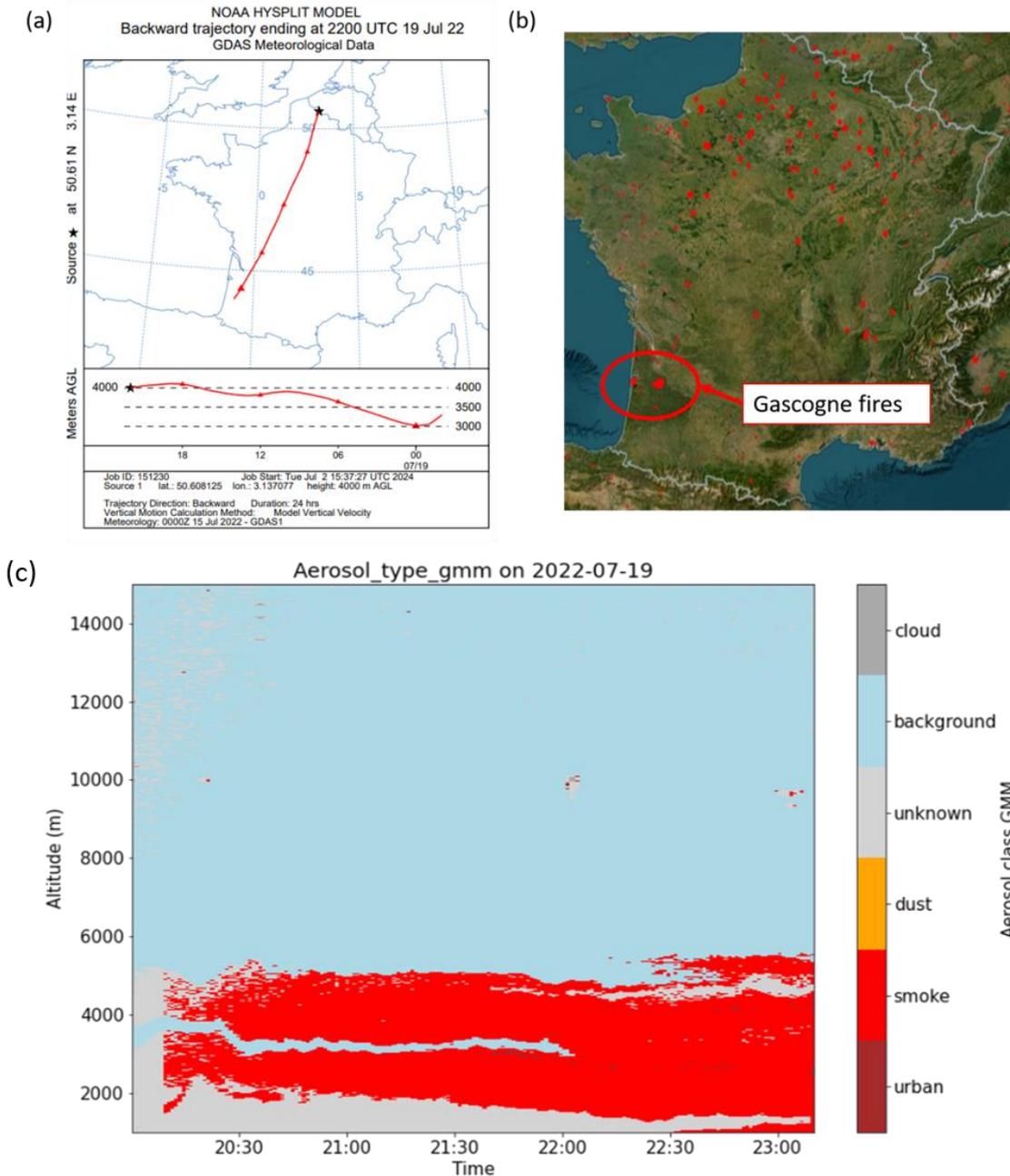


Figure 3.19: (a) Backward trajectory of 24 hours at 4000 m above ground level at 22:00 UTC on 19 July 2022 (b) Fire map from the Fire Information for Resource Management System (FIRMS) between 14 to and 19 July 2022 (source: <https://firms.modaps.eosdis.nasa.gov/>, last access: 28 June 2024) (c) FLARE-GMM aerosol type estimation quick-look during the night of 19 July 2022 between 1000 m and 15000 m

These three presented cases allow for the evaluation of the performance of FLARE-GMM in occurrences of strong events. In these situations, the aerosol type estimated by the algorithm is consistent with the expected aerosol type present in the atmosphere. The ability of FLARE-GMM to correctly identify aerosol types in such cases is thus supported by these examples. However, this approach is limited since it uses a low number of specific situations, which is

therefore not ideal to evaluate the algorithm performance in general. In order to do so, as previously mentioned, a reference is necessary to compare with FLARE-GMM results.

### 3.5.2 Comparison with NATALI aerosol typing

Neural Network Aerosol Typing Algorithm Based on Lidar Data (NATALI) is a deep learning algorithm developed to estimate the most probable aerosol type from lidar data. This algorithm uses the EARLINET  $3\beta + 2\alpha (+1\delta)$  profiles, which are multispectral profiles that can be obtained from LILAS and are regularly inverted. However, these profiles differ from the data used in this study, as both the inverted variables and the inversion method are different. NATALI has been trained on synthetic data, using the aerosol Ångström exponent, color index, color ratios,  $LR$  and  $PLDR$  of the aerosols as features to perform the classification. From these properties, the algorithm is able to determine the aerosol type among continental, continental polluted, smoke, dust, marine and volcanic (Nicolae et al., 2018a).

Thus, the use of NATALI allows us to have a second estimate of the aerosol type, which can be used as a relative reference to evaluate FLARE-GMM performance. 36 cloud-free profiles from 2022, covering different situations and aerosol types have been selected randomly to compare FLARE-GMM estimates with NATALI. Figure 3.20 shows the confusion matrix between the two aerosol type estimates. Confusion matrices are usually used to evaluate the performance of a classification algorithm. They display the counts of true positives, true negatives, false positives, and false negatives, helping to assess the accuracy, precision and overall effectiveness of the model. In this case, as the reference is not absolute, this matrix can then be used to compare the results from the two models, analyze their agreements and disagreement, in order to evaluate their performances, and biases.

The confusion matrix indicates that the agreement rate between the two models is at 38 %. This rate is acceptable given that the two models utilize different features for classification and differ in their algorithmic nature. Indeed, NATALI is a supervised learning algorithm, while FLARE-GMM is unsupervised. Regarding the disagreements between FLARE-GMM and NATALI, we can first evidence the confusion that exists between smoke and urban, or continental aerosols. Indeed, in a substantial number of cases, NATALI and FLARE-GMM disagree between smoke and urban aerosols. This confusion has been expected since the optical properties of smoke and urban aerosol are close, as evidenced part 3.4.4. Furthermore, as evidenced in the previous sections, the investigation of humidity is essential to discriminate between smoke and urban aerosols in dry and humid conditions. However, NATALI does not consider hygroscopic growth as it does not use  $RH$  as a feature, which is essential to classify aerosols. Therefore, the discrimination between these aerosol types constitutes a limitation of

NATALI (Nicolae et al., 2018b). Thus, it is challenging to conclude on the disagreements between the two models in this case.

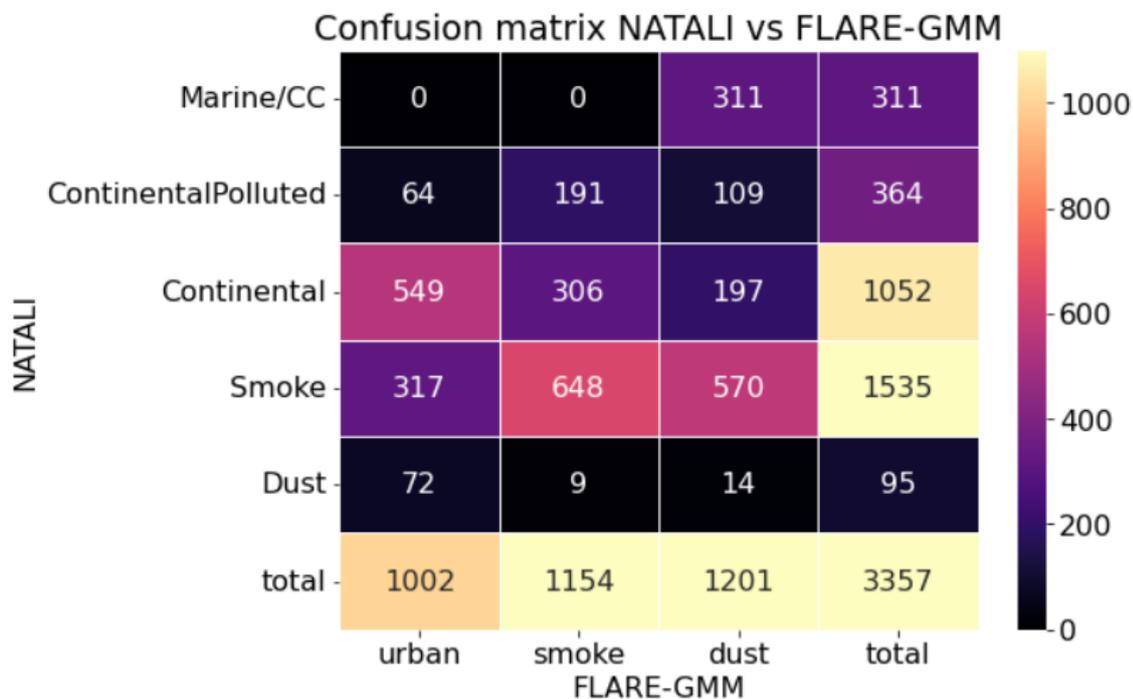


Figure 3.20: Confusion matrix between FLARE-GMM and NATALI aerosol type estimation on 36 profiles from 2022

Moreover, we can notice that most cases identified as desert dust by FLARE-GMM are classified differently by NATALI, either as smoke or as marine aerosols. This confusion is more surprising since desert dust optical properties are supposed to be different from these aerosol types. In particular, desert dust *PLDR* is expected to range close to 30 %, while smoke and marine aerosols are expected to have much lower *PLDR*. This difference can be explained by several factors. First, the methods used to invert the profiles used by FLARE-GMM and NATALI are different. Therefore, inversion errors might appear on one dataset and not the other, introducing disagreements. Then, the different natures of the two algorithms might as well explain these differences. As it has been mentioned, NATALI is a supervised learning model that has been trained on synthetic data, as opposed to FLARE-GMM that is an unsupervised learning model which has been trained using data from LILAS instrument specifically. This aspect is an advantage for FLARE-GMM as the specificities of the site in Lille, as well as the specificities of the instrument are therefore considered by the model. On the other hand, NATALI, which has been trained on synthetic data, might contain biases from the model used to simulate the aerosol optical properties. Moreover, the features used by NATALI might as well explain these differences with FLARE-GMM estimates. Indeed, the *LR* and the Ångström exponent, both relying on the extinction coefficient estimation, are used by NATALI. However, these properties are difficult to estimate accurately with the Raman inversion,

performed on the EARLINET profiles. These quantities often exhibit high uncertainties (Ansmann et al., 1992a) thus impacting NATALI estimation quality. This might explain why NATALI predicts the presence of marine aerosols while the measurement site is located 70 km away from the coast, thus making it unlikely to observe aerosol layers mainly composed of marine particles. Eventually, the treatment of hygroscopicity can also be responsible for the differences between NATALI and FLARE-GMM aerosol typing estimates. This phenomenon, which significantly alters aerosol properties, is taken into account by NATALI in the training set, as it is modeled according to different  $RH$  levels, but  $RH$  is not used as an input to determine the aerosol type (Nicolae et al., 2018a). On the other hand, as FLARE-GMM uses real data in the training set, it naturally covers a wide range of humidity levels, and furthermore, even if  $RH$  is not used as a feature, its influence on aerosol optical properties is considered with the use of different trained modeled in function of the  $RH$  levels. These differences of treatment can be responsible for important differences between the two estimates, especially between urban and smoke aerosols, as their optical properties can be difficult to distinguish if humidity is not considered.

Nevertheless, the comparison between FLARE-GMM and NATALI allows us to evaluate FLARE-GMM performance, compared to another automatic aerosol typing model. While the comparison suffers from limitations that have been raised, making it challenging to formulate quantitative conclusions, it is still providing encouraging results. Indeed, despite their differences, in terms of architecture, training methods and datasets used to perform the classification, the agreement rate between the two models is almost at 40 %. Moreover, disagreements between NATALI and FLARE-GMM can find explanations in many factors that have been mentioned. Therefore, this comparison, with the analysis of extreme events performed part 3.5.1 are promising for the model performance. They also provide a positive outlook for its potential future improvements and advancements, indicating that the model has a solid foundation and can continue to evolve successfully.

### **3.6 Aerosol type analysis in Lille**

In this section, FLARE-GMM is used to analyze aerosol type estimates in Lille on all the available dataset. The advantage of developing an automatic aerosol typing method is that such analyses are easily performed quickly on a very large amount of data. The results can then be analyzed to study aerosol properties in Lille, and evaluate potential trends.

By using FLARE-GMM on the LILAS dataset from 2021 to 2023, we can investigate the aerosol type repartition, as well as the seasonality of the aerosols in the Lille region. To do so, the aerosol type has been estimated by FLARE-GMM for each available profile between 2021 and 2023. Each profile containing more than 15 data points classified as a specific aerosol type have

been considered in order to avoid treating outliers. Eventually, the considered altitudes for this analysis have been selected below 6 km. This choice has been motivated to avoid taking cirrus clouds into account, which might be classified as dust aerosols by FLARE-GMM due to their high depolarization and low fluorescence. Indeed, cirrus clouds often show low optical thickness and are therefore more complicated to differentiate from aerosols by using a criterion on the elastic backscatter coefficient. By considering only the data below 6 km, it is possible to mitigate the impact of cirrus clouds in the statistics while also considering most of the aerosol cases, which are in majority present in low altitudes.

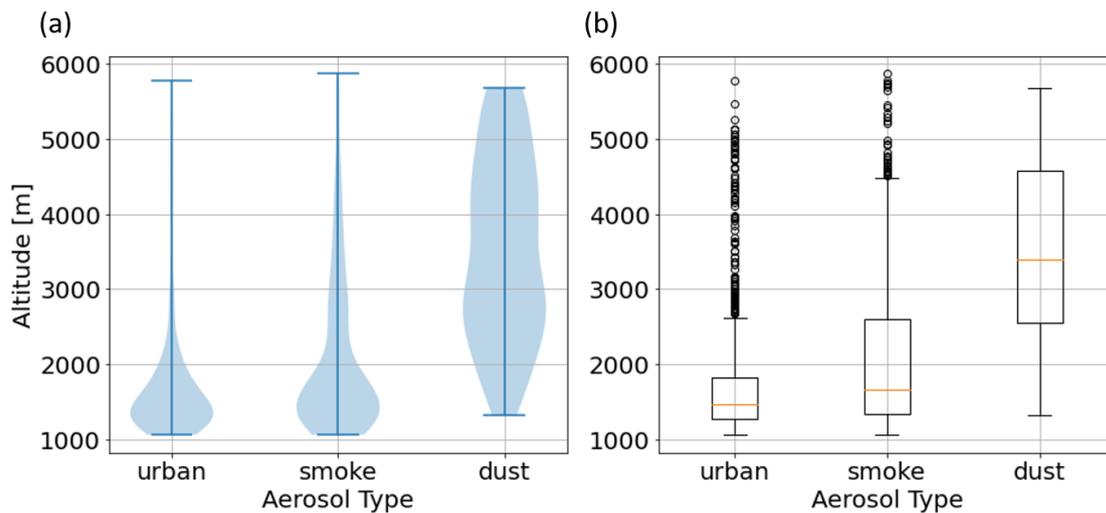


Figure 3.21: (a) Violin plots and (b) Box plots of averaged altitudes, between the ground and 6 km above ground level in function of the aerosol type for all available data from 2021 to 2023

First, it can be interesting to investigate the altitude distribution for each aerosol type. Figure 3.21 shows the violin plots (a) and the box plots (b) of the averaged altitudes above the ground of the identified aerosol layers in function of the aerosol type. This plot indicates that both urban and smoke aerosols are predominantly detected at low altitudes, mainly within the boundary layer. The distribution for smoke aerosols exhibits a longer tail compared to urban aerosols. This is expected, as smoke aerosols, generally originating from fires, are emitted at high temperatures and can be injected into higher altitudes. In contrast, urban aerosols usually remain confined to the boundary layer and rarely reach higher altitudes. Regarding dust aerosols, their distribution shows that they can be present at much higher altitudes. This result can be interpreted in different ways. First, it could be a consequence of ice cloud detection, however, below 6 km, the presence of ice clouds is less probable in the Lille atmosphere. This could also be due to the fact that the primary sources of dust in Lille are not local. Unlike urban and smoke aerosols, which may be emitted locally, dust often originates from the Sahara or other deserts and is carried to Lille by the wind. As a result, dust particles are found at more dispersed altitudes compared to urban and smoke aerosols.

Figure 3.22 (a) shows the histogram of FLARE-GMM aerosol type estimations on all available data below 6000 m in function of the time. This figure shows that in general, more aerosol layers are identified during spring and summer. This situation is influenced by Lille's meteorology, as it often rains in this region during winter. LILAS does not measure during rain, which reduces the amount of data available for analysis. Additionally, during winter, the boundary layer is generally lower due to decreased temperatures. Given that the minimum considered altitude is 1000 m, aerosol layers may not be detected, further limiting the data.

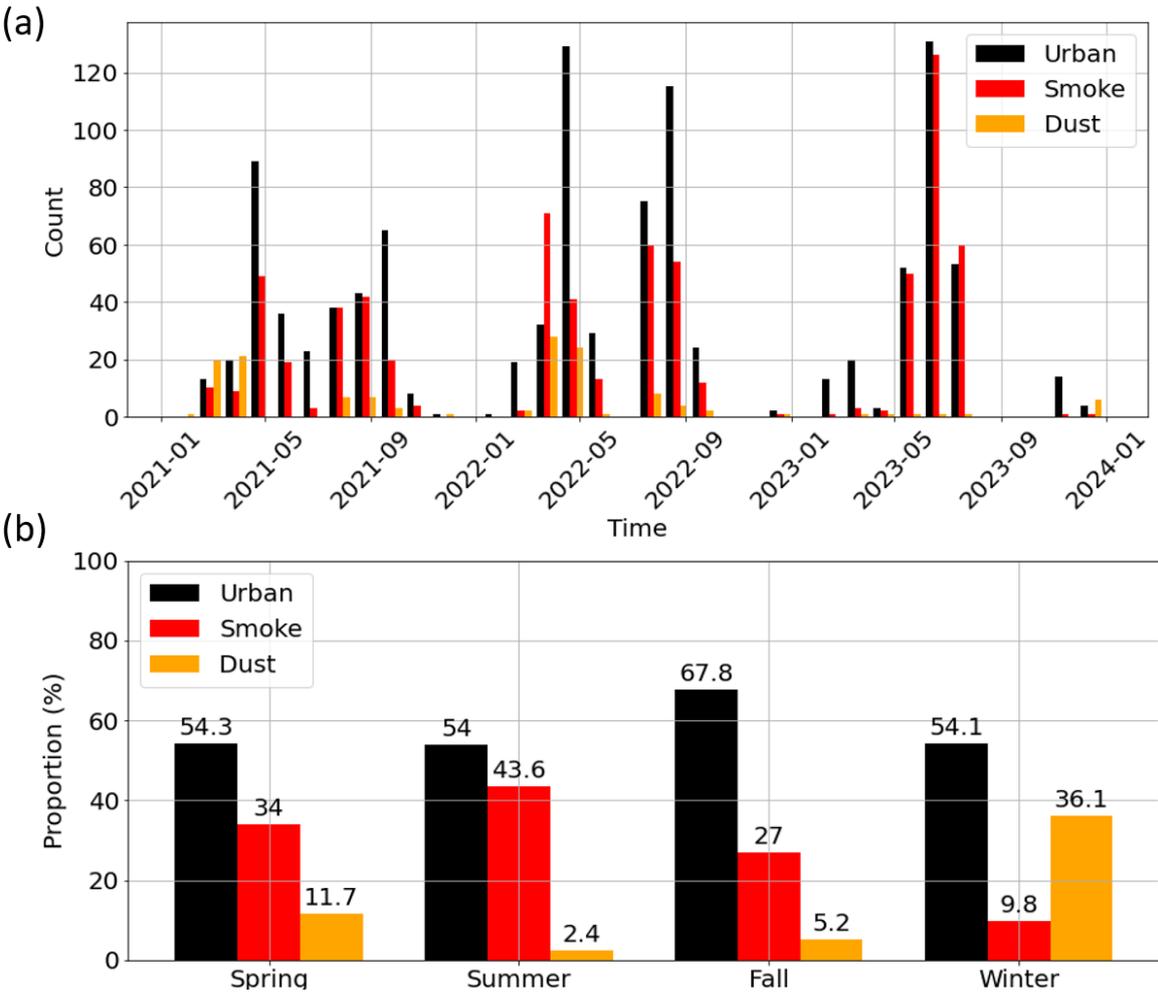


Figure 3.22: (a) Histogram of aerosol type estimate from FLARE-GMM in function of the time and (b) share of each aerosol type in function of the season, with all data from 2021, 2022 and 2023 below 6000 m above ground level

Figure 3.22 (b) shows the seasonal share of each aerosol type estimated by FLARE-GMM. This figure first illustrates that urban aerosols are the main aerosol type present in the Lille atmosphere, and represent more than half cases in each season. This result is expected as the LOA is located next to the city of Lille, the emission of urban aerosols by human activity is therefore the first aerosol source in the observed atmosphere. Regarding smoke aerosols,

Figure 3.22 (b) indicates that they are significantly more frequent during spring and summer compared to fall and winter. This trend can be attributed to higher temperatures in spring and summer, which increase the likelihood of fires, the primary source of smoke aerosols, occurring during these periods. Eventually, regarding dust aerosols, they are the aerosol type the least represented. This is because dust scenarios are rare in Lille, which is not located close to a source of desert dust. The occurrence of dust cases in Lille are due to extreme events such as the advection of Saharan dust by Sirocco winds as mentioned part 3.5.1. Such events generally occur early in spring or in winter, and it is possible to observe that dust cases are more represented at these periods, thus confirming the importance of these phenomena in the observation of dust aerosols in Lille. However, it is important to consider the proportion of dust cases in winter within the context of data availability limitations during this season. Additionally, the colder temperatures in winter increase the likelihood of observing ice or mixed-phase clouds below 6000 m. These clouds can occasionally be misinterpreted as desert dust aerosols with the current classification method, as it has been mentioned. Despite these challenges, these findings are crucial for gaining insights into the composition and distribution of aerosols in the region, showing the benefits from using an automatic aerosol typing process like FLARE-GMM.

### **3.7 Chapter summary**

This chapter details FLARE-GMM, an aerosol typing algorithm designed to identify smoke, urban, and dust aerosols using LILAS data. The chapter covers the evolution of various versions of the algorithm, highlighting challenges such as the negative impact of mixed aerosol cases on cluster identification and the representation of hygroscopic growth by separating the dataset according to humidity levels. The final version of FLARE-GMM has been generalized and tested. First the result of FLARE-GMM has been investigated on three cases, all three correctly classified by the algorithm. Then, FLARE-GMM has been compared to NATALI, another automatic aerosol typing model using lidar data. The comparison has shown that the two algorithms agree in 38% of the cases, which is a low number but still encouraging. Disagreements between FLARE-GMM and NATALI have been analyzed, suggesting that NATALI may be less applicable in this context, as FLARE-GMM has been directly trained on LILAS data, in comparison. Finally, a brief statistical analysis of aerosol types in the Lille atmosphere has been conducted, showing that smoke and urban aerosols are mostly observed at lower altitudes due to local emissions, while dust, originating from distant sources, is found at various altitudes between 1,000 and 6,000 meters above ground level. Seasonal analysis indicates the presence of annual cycles in aerosol type distribution, with more smoke during summer and more urban aerosols in the fall.

The FLARE-GMM algorithm is expected to be integrated into AUSTRAL soon, allowing for automated aerosol typing on each inverted profile. Due to the adaptable nature of this classification method, FLARE-GMM is well-suited to accommodate future technological advancements or algorithmic updates. For example, the upcoming lidar system LIFE (Laser Induced Fluorescence Explorer), set to become operational shortly, will offer enhanced power and the ability to measure fluorescence across different wavelengths. This new capability is crucial for more precise aerosol identification and will significantly deepen our understanding of aerosol types. By applying a protocol similar to the one detailed in this chapter, it would be feasible to develop an updated version of FLARE-GMM that utilizes the LIFE dataset for training, further enhancing its capabilities and accuracy in aerosol typing. Eventually, in order to assess the robustness of this approach to perform aerosol typing, it could be tested on another instrument also measuring aerosols but in a different environment than Lille, to confront its viability in the presence of other aerosol types such as marine, volcanic aerosols or pollen in higher quantity.

## 4. Hygroscopic growth study from Mie-Raman-fluorescence Lidar

This chapter aims to present the potential and advantages of using fluorescence measurements with lidars to study aerosol hygroscopic growth. The chapter begins by the description of hygroscopic growth and the conventional methods used to investigate it with lidars and other instruments. Next, it introduces a novel approach that uses fluorescence to enhance the determination of aerosol hygroscopic properties using lidar data. This method is demonstrated through three distinct cases of hygroscopic growth: two involving urban aerosols and one involving smoke, each case illustrating the added value of fluorescence data. Following this, an automated detection of hygroscopic growth cases is carried out, accompanied by a statistical analysis of the hygroscopic properties of aerosols in the Lille atmosphere. The chapter concludes with a mathematical analysis of the method for estimating hygroscopic properties, highlighting its sensitivity to errors in relative humidity (RH) measurements and offering strategies to model and mitigate these errors.

### 4.1 Aerosol hygroscopic growth

#### 4.1.1 General description

Hygroscopicity is the ability of a material to absorb water vapor. For aerosols, the absorption of water vapor can lead to changes in the physical properties, such as size or optical properties, as well as changes in chemical properties (Hänel, 1976). Aerosol hygroscopic growth corresponds to the change of aerosol size in response to varying ambient relative humidity. This growth depends on the particle affinity for water; some aerosols are highly hygroscopic, such as marine aerosols, while others, like desert dust, are hydrophobic and exhibit little to no hygroscopic growth.

The investigation of aerosol hygroscopic properties is a field of high importance. Indeed, these properties greatly influence aerosol effective radiative forcing through two mechanisms. First, at high humidity, the size of hygroscopic aerosols increases, enhancing their scattering properties which affects  $ER_{\text{Far}}$ . On the other hand, aerosol hygroscopic properties significantly impact the CCN activity of aerosols, thereby affecting  $ER_{\text{Fac}}$  (Dusek et al., 2006; Köhler, 1936; Mochida et al., 2006; Twomey, 1974).

The Köhler theory describes the condensation of water vapor to form liquid cloud droplets, based on equilibrium thermodynamics (Köhler, 1936). This theory considers the Kelvin effect and the Raoult law which both influences the condensation of a water droplet in real conditions, the Köhler curve which illustrates this theory is represented in Figure 4.1.

On the one hand, the Kelvin effect takes into account the droplet curvature and how it increases the saturation pressure necessary to condense water vapor in comparison to a flat surface. On the other hand, the Raoult law describes the decrease of the saturation ratio in the presence of a solute. It can be expressed as:

$$P_i = x_i P_{sat} \quad (4.1)$$

where  $P_i$  is the gaseous partial pressure of a chemical component,  $x_i$  is the mixing ratio of this component in the liquid phase, and  $P_{sat}$  is the equilibrium vapor pressure. In the case of water, in the presence of a solute, the water mixing ratio is then inferior to 1, consequently favoring water condensation. The Köhler equation can take the form:

$$s = a_w \cdot Ke \quad (4.2)$$

Where  $s$  is the saturation ratio, the ratio between the equilibrium water vapor pressure of a realistic water droplet, and equilibrium water vapor pressure of a flat pure water liquid surface. The term depends on both the Raoult term,  $a_w$  which represents the effect of the Raoult law, and the Kelvin term,  $Ke$ , which represents the impact of the Kelvin effect. This equation is fundamental in cloud microphysics as it describes the aerosol CCN activity at a given relative humidity. Indeed, the Raoult term can be expressed using the aerosol hygroscopic parameter  $\kappa$  (Petters and Kreidenweis, 2007):

$$a_w = \left(1 + \kappa \frac{V_s}{V_w}\right)^{-1} \quad (4.3)$$

Where  $V_s$  is the volume of dry soluble, and  $V_w$  is the volume of water. The study of aerosol hygroscopic properties and the determination of  $\kappa$  are therefore essential to characterize the CCN activity of aerosols, and how their presence influence the formation of water droplets.

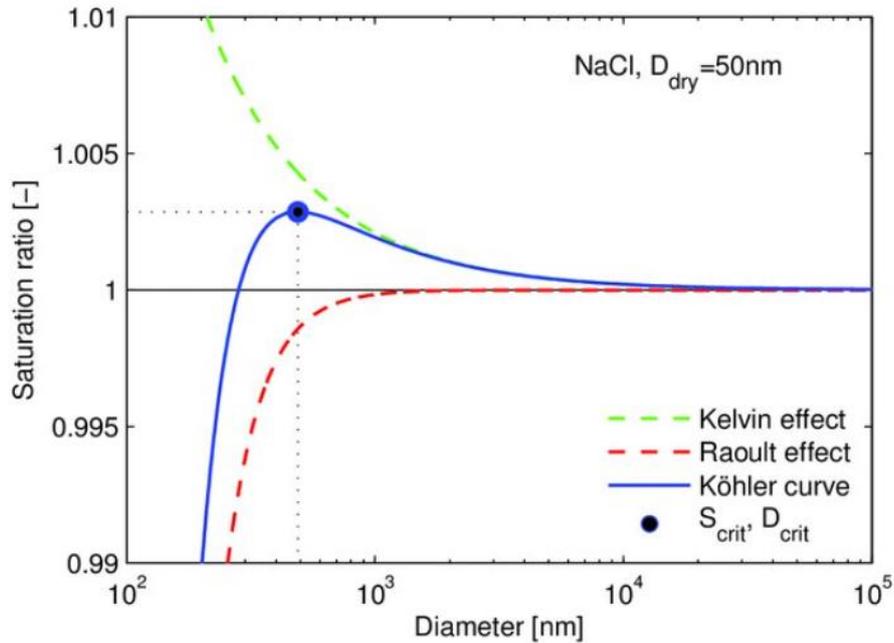


Figure 4.1: Köhler curve (Saturation ratio vs particle diameter) of a sodium chloride (NaCl) particle with a dry size of  $D_{dry} = 50 \text{ nm}$  (at  $T = 298 \text{ K}$ ). The contributions of the Raoult effect (red curve) and the Kelvin effect (green curve) are shown. The critical saturation ratio  $S_{crit}$  and the corresponding critical diameter  $D_{crit}$  are indicated (blue bullet) (From P. C. Zieger (2011)).

#### 4.1.2 Characterization of aerosol hygroscopic growth in the literature

The study of aerosol hygroscopic properties involves a multidisciplinary approach combining experimental techniques, theoretical models, and computational simulations. For a long time, aerosol hygroscopic growth measurements have been performed using mainly *in situ* instruments. The hygroscopicity tandem differential mobility analyzer (HTDMA), directly measures the change in aerosol size as a function of humidity, providing insights into hygroscopic behavior and particle activation. This instrument is composed of an aerosol dryer and an aerosol humidifier, used to control the humidity levels at which the collected aerosols will be examined. Two differential mobility analyzers measure the aerosol electrical mobility, from which it is possible to estimate the aerosol size, in the dry and humid conditions. This method allows us to compute the growth factor, which is the ratio between the dry and the humid aerosol size, and which quantifies the hygroscopic properties of the particle (Liu et al., 1978; Swietlicki et al., 2008; Zardini et al., 2008).

Nephelometers are also largely used to estimate aerosol hygroscopic properties *in situ*. These instruments can measure the impact of hygroscopicity on aerosol backscattering properties. Nephelometers can measure aerosol scattering properties, and when they are used in combination with dryers and humidifiers, it allows us to estimate how these properties evolve at different  $RH$  levels. The aerosol light scattering enhancement factor, noted  $f(RH)$  can then

be estimated. It is a quantity largely used to quantify aerosol hygroscopic growth, and it is defined as the ratio between the aerosol backscatter coefficient at high  $RH$ , and the backscatter coefficient at low  $RH$ . It is different however from the growth factor measured with HTDMA, but is still a valuable information on aerosol properties, which is essential to quantify the effect of aerosols on radiation balance, and can also be used to estimate the aerosol hygroscopic parameter (Burgos et al., 2019; Chen et al., 2014; Covert et al., 1972; Gomez et al., 2018).

However, *in situ* measurements show strong limitations. Apart from the fact that measurements can be performed only at the point where the instrument is present, the use of dryers and humidifiers, and the different measurement techniques themselves, are susceptible to alter aerosol properties in comparison to real atmospheric conditions (Spanu et al., 2020). Therefore, remote sensing techniques, and more precisely lidars, tend to be more and more used to estimate aerosol hygroscopic properties remotely (Feingold 2003; Fernández et al. 2015; Granados-Muñoz et al. 2015; Zieger et al. 2015; Haarig et al. 2017; Navas-Guzmán et al. 2019; Dawson et al. 2020; Düsing et al. 2021; Sicard et al. 2022...). Since lidars measure the aerosol backscatter coefficient, it is also possible to compute  $f(RH)$  from lidar measurements. In order to do so, it is necessary to identify an aerosol layer that is spread spatially, this way the lidar is either able to capture the layers on a large range of altitudes, or during a long time. To estimate  $f(RH)$ , most studies assume that the aerosol layer is homogeneous. Then, any change of backscatter coefficient can therefore be attributed to hygroscopic growth. From these variations, it is possible to estimate  $f(RH)$  in function of  $RH$  in the layer.

The homogeneity of the aerosol layer can be controlled with several criteria. Wind measures and backward trajectory methods, such as NOAA's HYSPLIT trajectory model (Draxler et al., 2023; Stein et al., 2015), are used to verify that different parts of the aerosol layer have the same origin. The water vapor mixing ratio can also be used as a tracer of air masses. In the absence of condensation or evaporation, the water vapor mixing ratio remains constant. Therefore, a constant water vapor mixing ratio means with a good approximation that the air masses have the same origin, and supports the hypothesis that the aerosol layer is homogeneous. Eventually, the potential temperature can be used. It is defined as:

$$\theta = T \left( \frac{P_{std}}{P} \right)^{\frac{R_a}{c_p}} \quad (4.4)$$

Where  $\theta$  is the potential temperature,  $T$  is the temperature,  $P_{std} = 1\,000$  hPa by convention,  $P$  is the atmospheric pressure,  $R_a = 287.05$  J kg<sup>-1</sup>K<sup>-1</sup> is the specific gas constant for dry air, and  $c_p$  is the dry air thermal capacity at constant pressure, approximately equal to 7/2 times  $R_a$ . The potential temperature is essential since it is a tracer of atmospheric stability. Indeed, if it increases with altitude, the atmosphere is stable and stratified, meaning that there is no

exchange of air masses between low and high altitudes. On the contrary, if the potential temperature decreases with altitude, the atmosphere is unstable and upward air motion is likely to happen, therefore favoring good mixing and homogeneity in the layer. Furthermore, due to the low thermal conductivity of air, the potential temperature of an air parcel tends to evolve very slowly, at least on a timescale of a few hours. Consequently, potential temperature can be used to label air parcels in function of their origin, similarly to water vapor mixing ratio. Therefore, decreasing to constant potential temperature with altitude, also suggests that the considered aerosol layer is homogeneous (Andrews, 2012).

Once the homogeneity of the aerosol layer has been confirmed with the methods mentioned above, we can estimate  $f(RH)$ , the aerosol enhancement factor, which quantifies the aerosol hygroscopic properties:

$$f(RH) = \frac{\beta(RH)}{\beta_{dry}} \quad (4.5)$$

Where  $\beta(RH)$  is the aerosol backscattering coefficient at the considered relative humidity, and  $\beta_{dry}$  is the same quantity in dry condition. However, on the contrary to *in situ* measurements where the relative humidity can be fixed by dryers and humidifiers, the remote sensing measurements are performed under ambient humidity, which depends on the atmospheric conditions and cannot be fixed. Therefore, it is not always possible to measure  $\beta_{dry}$  to estimate  $f(RH)$ . Instead, the enhancement factor is often estimated by considering the ratio between the aerosol backscatter coefficients at maximum and minimum  $RH$  in the layer. The main issue is that in this framework, it is complicated to compare the results between different cases and different studies, since the values of  $f(RH)$  depend on the humidity levels of the different aerosol layers. In order to overcome this, we can use the Hänel parameterization (Hänel, 1976). This parameterization estimates  $f(RH)$  as an exponential function of  $RH$  under the form:

$$f(RH) = \frac{\beta(RH)}{\beta(RH_{ref})} = \left( \frac{1 - RH}{1 - RH_{ref}} \right)^{-\gamma} \quad (4.6)$$

Where  $RH_{ref}$  is the reference  $RH$ , generally chosen equal to 0.4, which is a level of relative humidity at which it is estimated that hygroscopic growth does not occur, and is a recommendation of the World Meteorological Organization (World Meteorological Organization and Global Atmosphere Watch, 2023).  $\gamma$  is the Hänel hygroscopic parameter of the aerosol. This parameter is different from  $\kappa$ , the hygroscopic parameter used in the Köhler theory, but can also quantify the hygroscopic properties of the aerosol. The Hänel parameterization allows to compute  $f(RH)$  when  $\gamma$  is known, considering that  $RH$  is not equal or close to 1. We can estimate  $\gamma$  by taking into account the variations of the backscatter coefficients with  $RH$  in a homogeneous aerosol layer:

$$\frac{\beta(RH)}{\beta(RH_{min})} = \left( \frac{1 - RH}{1 - RH_{min}} \right)^{-\gamma} \quad (4.7)$$

Where  $RH_{min}$  is the minimum  $RH$  value in the aerosol layer. The evolution of  $\beta(RH)$  with  $RH$  can then be fitted using (4.7) to estimate  $\gamma$ . This parameter can be either used to compare  $\gamma$  estimates between different cases, or to compute  $f(RH)$  at the same relative humidity for each analyzed case. This way, the values of the enhancement factors  $f(RH)$  can be compared between each other.

## **4.2 Methodology for estimating aerosol hygroscopic properties with LILAS**

The unique characteristics of the LILAS instrument present opportunities to improve the characterization of aerosol hygroscopic properties with lidar technology. First, this instrument simultaneously measures the aerosol backscattering coefficient and humidity throughout the atmospheric column. This aspect enhances the accuracy of hygroscopic properties retrieval compared to methods that use separate instruments for aerosol and humidity measurement. Indeed, humidity can vary rapidly on short distances depending on atmospheric stability, the ability to collocate aerosol and humidity measurements with one instrument is therefore a significant advantage. Furthermore, the fluorescence measurement of LILAS is a valuable tool allowing to gather precious information about aerosols, and improving the study of their hygroscopic properties compared to traditional methods using lidars. As stated previously, the emission of fluorescence by aerosols can be attributed to biogenic material. This characteristic has been used to identify the origin of aerosol layers with FLARE-GMM, described in chapter 3. Indeed, pure water does not produce fluorescence at 466 nm for an excitation at 355 nm. Therefore, in a first approximation, fluorescence signal measured by LILAS is independent from the presence of water vapor in the air, unlike other aerosol optical properties (Veselovskii et al., 2020). Then, considering that the mixing ratio of biogenic particles present in an aerosol layer is constant,  $\beta_{fluor}$  becomes a tracer of aerosol concentration within the layer. It can therefore be used to correct variations in aerosol concentration within the layer. Thus, it allows to study hygroscopic growth even if the layer is not homogeneous, which is a very strong constraint and difficult to properly verify, adding a high degree of uncertainty in studying aerosol hygroscopic properties with previous methods. Instead, the necessary assumption is that the biogenic particles mixing ratio is constant within the aerosol layer, since this assumption allows to use  $\beta_{fluor}$  to follow the aerosol concentration.

In order to assert that the aerosol biogenic particles mixing ratio remains constant within the studied aerosol layer, most methods previously employed to verify the aerosol layer homogeneity can be used. Constant water vapor mixing ratio, which is measured by LILAS,

indicates that air masses have the same origin. It therefore supports that the biogenic particles mixing ratio remains constant in the layer. Similarly, the potential temperature, estimated with ERA-5 and standard atmospheric pressure profiles, indicates strong mixing within the layer if it is constant or if it decreases with altitude, which also suggests that the biogenic particles mixing ratio is likely to be homogeneous. The aerosol source can also be verified using FLARE-GMM result. If this result is consistent across the aerosol layer, it means that the source of the aerosol layer is of the same nature, further supporting that biogenic particles mixing ratio is likely to be constant across the layer. Eventually, in some occasions, backward trajectories can be computed to ensure that the aerosols in the layer have the same origin.

Once the homogeneity of the biogenic particles mixing ratio is confirmed in the aerosol layer, the hygroscopic properties of the aerosol can be studied, given that hygroscopic growth occurs. To ensure its occurrence,  $RH$  can be investigated; the maximum  $RH$  in the layer should be over 0.8. This way, hygroscopic growth is most certainly happening for most aerosol types. Moreover,  $RH$  should vary in the layer over at least a range of 0.1, in order to accurately estimate  $\gamma$ . Furthermore, the LILAS depolarization measurement can also confirm that hygroscopic growth occurs. Indeed, this property depends on the aerosol shape, and in most cases, humidified aerosols have smoother shapes, thus reducing their ability to depolarize light (Cooper et al., 1974). Therefore, decreasing  $PLDR$  supports that hygroscopic growth is occurring within the layer, given that the aerosol type remains constant.

Once these verifications have been performed,  $\beta_{fluo}$  can be used to follow the evolutions of the aerosol concentration within the layer. We can take this evolution into account to improve  $f(RH)$  estimation. To do so, equation (4.7) can be modified as follow:

$$\frac{\beta(RH)}{\beta(RH_{min})} \frac{\beta_{fluo}(RH_{min})}{\beta_{fluo}(RH)} = \left( \frac{1 - RH}{1 - RH_{min}} \right)^{-\gamma} \quad (4.8)$$

$$\frac{\bar{\beta}(RH)}{\bar{\beta}(RH_{min})} = \left( \frac{1 - RH}{1 - RH_{min}} \right)^{-\gamma} \quad (4.9)$$

Where  $\bar{\beta}(RH)$  is the normalized aerosol backscatter coefficient, equal to the aerosol backscatter coefficient divided by the fluorescence backscatter coefficient. Using (3.2), it is possible to estimate  $\gamma$  while considering eventual aerosol concentration variations within the aerosol layer. This method allows to better estimate the Hänel hygroscopic constant, improving the aerosol enhancement factor  $f(RH)$  estimation.

In order to evaluate the efficiency of this method, it has been performed on more than twenty cases, in various conditions covering different aerosol types. The Hänel hygroscopic constant as well as the aerosol enhancement factor for the backscatter coefficient at 532 nm have been

estimated for these different situations, the results are analyzed and compared to the literature in the following part.

## 4.3 Case studies

### 4.3.1 Hygroscopic growth of urban aerosol during the night of 29 July 2021

The first case studied in this work corresponds to the night between 29 and 30 July 2021. Quick-looks of LILAS  $\beta_{532}$  and FLARE-GMM results are displayed Figure 4.2. The quick-looks show the presence of a stable boundary layer, made of urban aerosols, with height located at 2 km above ground level. Small clouds form at the top of the boundary layer around 00:00 UTC and at 4:00 UTC showing high  $RH$  conditions, which are favorable for the occurrence of hygroscopic growth. Liquid clouds are also present higher in altitude in the free troposphere during the night, around 4 km and 6 km. Ice clouds are present above 8 km from 00:00 UTC to the end of the night, as shown by their altitude, large spatial coverage and low optical thickness in comparison to the other clouds. The  $PLDR$  of these clouds is also high (over 30%), further supporting the presence of ice crystals. It is noteworthy that the quick-looks reveal a decrease in boundary layer height during the night. Additionally, the boundary layer appears more homogeneous in the early part of the night, while towards the end, it shows strong backscatter at the top and lower backscatter at lower altitudes. These observations can be explained by the cooling of the ground during the night, which reduces the turbulence in the boundary layer, increasing its stratification and stability, while also reducing its height.

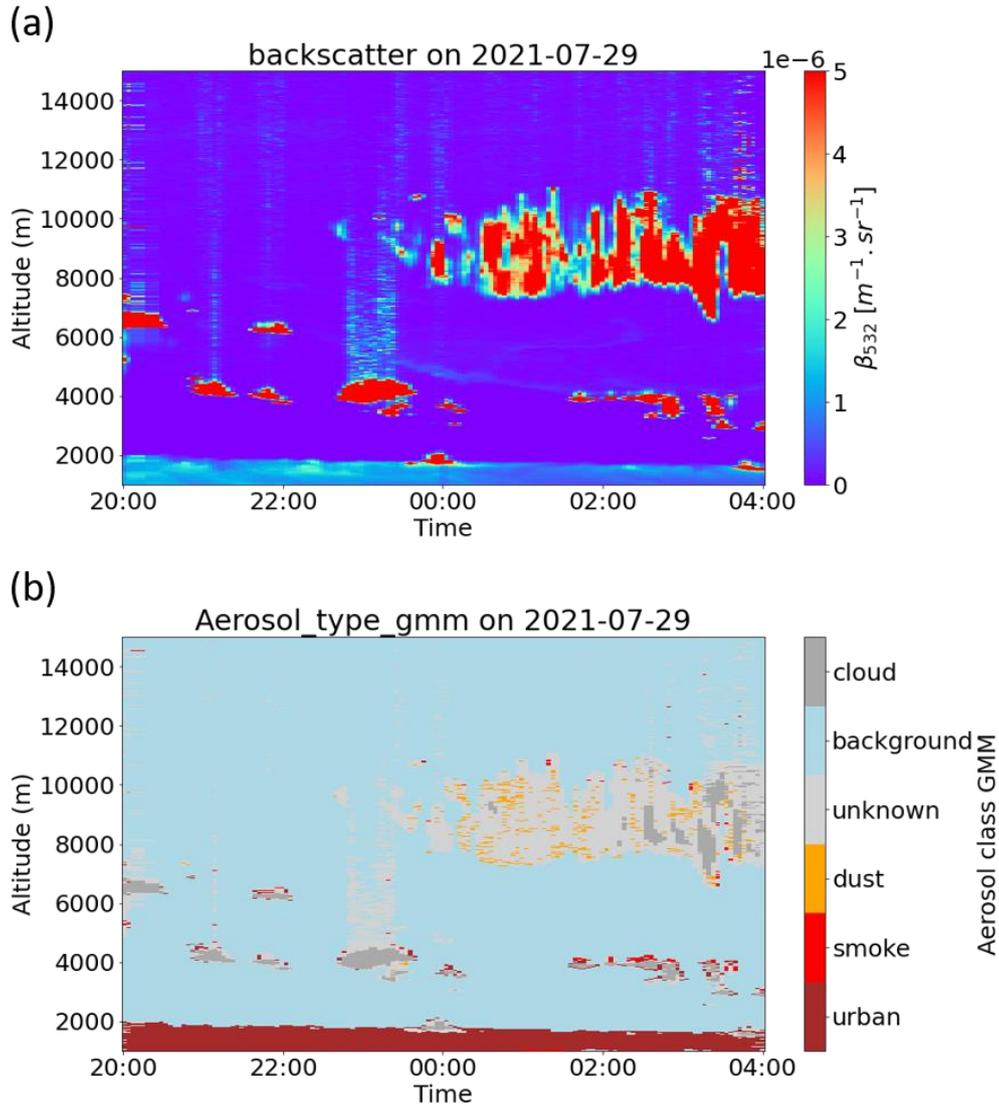


Figure 4.2: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $m^{-1}.sr^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 29 July 2021

In order to determine if this situation is suitable for the study of aerosol hygroscopic properties, the profiles of LILAS inverted products can be investigated. The situation between 22:00 and 23:00 UTC seems to be more appropriate for the study of hygroscopic properties. First, the results of FLARE-GMM are homogeneous on this interval within the boundary layer, and show an absence of clouds. Furthermore, since convective processes are supposed to occur for some time after sunset, the boundary layer is expected to be well mixed in this situation, which is favorable for studying hygroscopic properties of the aerosols. Profiles of LILAS measures averaged on this time period are displayed Figure 4.3.

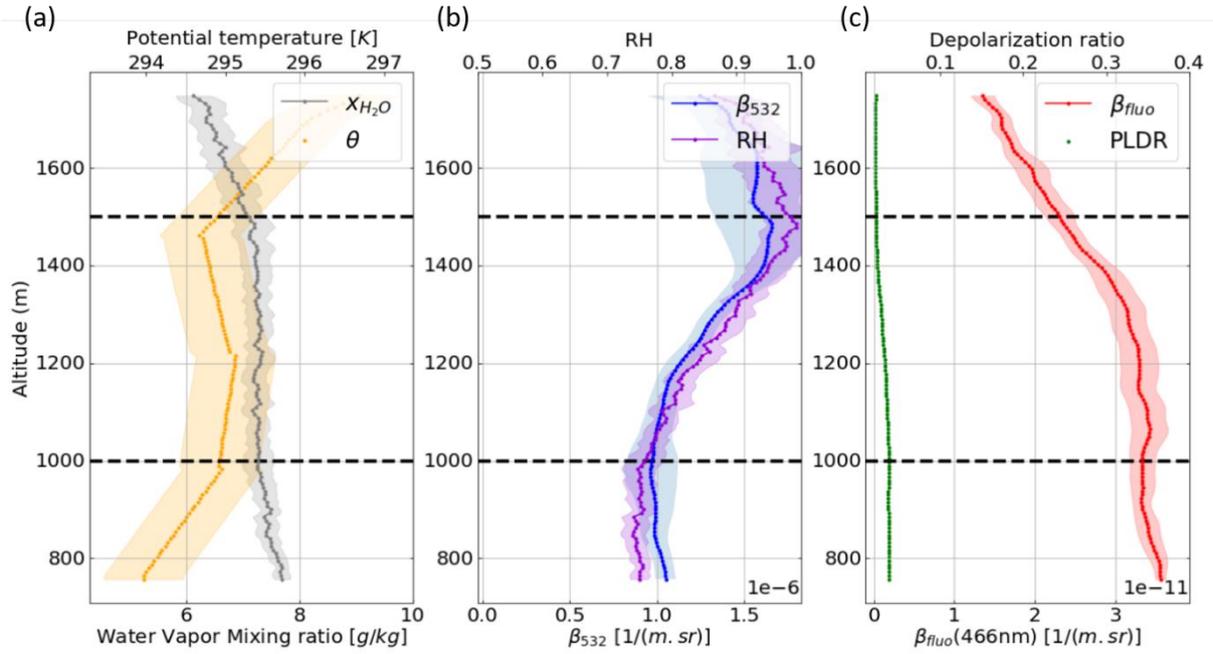


Figure 4.3: Profile of retrieved optical properties in function of altitude above ground level (a) water vapor mixing ratio [g/kg] and potential temperature [K], (b) elastic backscatter coefficient at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and  $RH$ , (c) fluorescence backscatter coefficient at 466 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and particular linear depolarization ratio at 532 nm, on 29 July 2021 from 22:00 to 23:00 UTC, the black dashed lines identify the area where hygroscopic growth is expected to occur

Figure 4.3 displays the atmospheric products used to study hygroscopic growth. Figure 4.3 (a) shows the potential temperature and the water vapor mixing ratio. The potential temperature appears uniform between 1000 m and 1500 m above ground level ( $295 \pm 1$  K), suggesting effective mixing of the boundary layer at these altitudes. This could be due to the fact that the profiles are acquired shortly after sunset, when convective processes might still be active, favoring strong mixing in the boundary layer. The water vapor mixing ratio also appears homogeneous ( $7.2 \pm 0.5$  g.kg $^{-1}$ ), indicating that the air masses between 1000 m and 1500 m are expected to have the same origin. This conclusion is also supported by FLARE-GMM results, which indicate the presence of urban aerosols across the whole boundary layer. These observations are generally used to assess the homogeneity of the layer. Therefore, it allows us to estimate that the biogenic particle mixing ratio remains constant within the considered aerosol layer.

The middle profiles (Figure 4.3 (b)) represent both the aerosol elastic backscatter coefficient at 532 nm, and the relative humidity. A strong increase of  $\beta_{532}$  with altitude is observed between 1000 m and 1500 m, from  $9.7 \pm 0.2$  Mm $^{-1}$ sr $^{-1}$  at 1000 m to  $1.7 \pm 0.3$  Mm $^{-1}$ sr $^{-1}$  at 1500 m.  $RH$  simultaneously increases in the boundary layer, varying from  $76 \pm 3$  % (bottom of the layer) to  $99 \pm 5$  % (top of the layer). The significant variation in  $RH$  (a 23 % difference between the lowest and highest altitudes) and the fact that the maximum  $RH$  approaches 100

% indicate that hygroscopic growth is occurring within this aerosol layer, which can be witnessed by the increase of  $\beta_{532}$ . This is further supported by the slight decrease in aerosol depolarization with altitude, dropping from  $3.7 \pm 0.2$  % to  $2.2 \pm 0.4$  %, shown on the last profiles (Figure 4.3 (c)). This decrease is small but is in fact not relevant in this case. Indeed, given that the depolarization of urban aerosols is already low, a strong reduction in depolarization is not anticipated in the occurrence of hygroscopic growth. The hygroscopic properties of the aerosols can then be investigated, by fitting the evolution of  $\beta_{532}$  with  $RH$  on the Hänel parameterization.

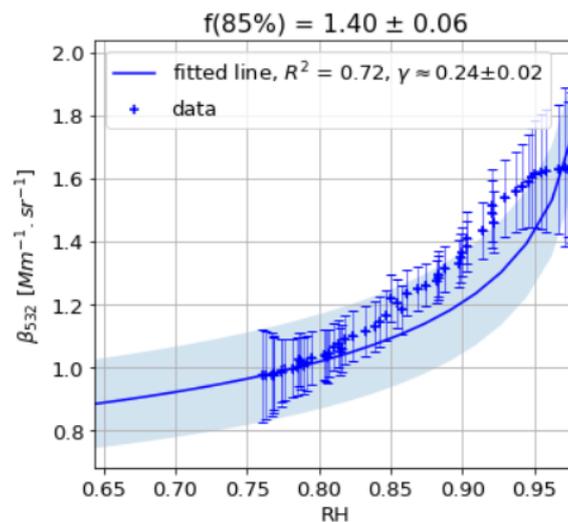


Figure 4.4: Evolution of the elastic backscatter coefficient at 532 nm in function of  $RH$  on 29 July 2021 from 22:00 to 23:00 UTC, between 1000 m and 1500 m above ground level, and results of the fit on the Hänel parameterization

Initially, the hygroscopic properties of the aerosols were studied using the method generally employed in the literature. Figure 4.4 shows the evolution of the elastic backscatter coefficient with  $RH$  and the fit to the Hänel parameterization as expressed equation (4.7), without the correction with the fluorescence backscatter coefficient. This expression supposes that the aerosols are homogeneously distributed in the layer where the hygroscopic growth occurs. This hypothesis is generally confirmed by a constant potential temperature, which suggests that strong mixing processes occur in the layer, and by a constant water vapor mixing ratio, indicating that air masses have the same origin. Both potential temperature and water vapor mixing ratio are constant in this case as stated earlier, the method commonly used to study aerosol hygroscopic properties can then be performed in this case.

Figure 4.4 shows that the fit to the Hänel equation is not ideal in this case. The coefficient of determination, noted  $R^2$ , measures the quality of the fit to the model, and how well the model predicts the data points. It varies generally between 0 and 1, with  $R^2 = 1$  when the model perfectly predicts the data, and  $R^2$  close to or below 0, or above 1 when the model is wrongly

chosen to represent the data. In this case,  $R^2 = 0.72$ , which indicates that the model is not perfect but still represents the data accurately enough. The Hänel hygroscopic parameter is estimated at  $\gamma = 0.24 \pm 0.02$  and the enhancement factor at 85 % is therefore estimated at  $f(85\%) = 1.40 \pm 0.06$  according to the equation (4.6). This value is low compared to previous studies on urban aerosols (Sicard et al., 2022), which estimated that the enhancement factor at 85 % for this type of aerosols, and specifically traffic emissions, is expected to range between 1.5 and 2 at 532 nm. This fact, and the relatively low determination coefficient, both suggest that the situation is not accurately modeled by the Hänel parameterization in this case, and that other processes are likely to occur in the aerosol layer.

Concerning the uncertainties of  $\gamma$  and  $f(RH)$  estimates, only the contribution from the backscatter coefficient uncertainties has been considered. This is because the uncertainties on  $RH$  estimates are difficult to assess, and are mainly discussed part 4.5.

Different hypotheses could explain the poor fit to the Hänel parameterization in this situation. The first explanation is that there could be little or no hygroscopic growth, which would result in an underestimation of  $\gamma$  and thus  $f(RH)$ . However, given the  $RH$  values and the collocated increase in  $RH$  and backscatter with altitude, this hypothesis seems unlikely. A second hypothesis which could explain a poor fit is the occurrence of condensation within the layer. Indeed, the Hänel parameterization is not valid anymore for  $RH$  values close to 100 % (which is the case here). This is because it does not model condensation, which would consequently drastically increase  $\beta_{532}$ . However, this hypothesis is also highly unlikely since it generally leads to an overestimation of the Hänel hygroscopic parameter, as well as an overestimation of  $f(RH)$ , which is the opposite of the results observed in this situation. Eventually, a final explanation of the poor fit to the Hänel parameterization could be that the homogeneity hypothesis, on which it relies, might not hold true in this case. In order to confirm this hypothesis, the fluorescence backscatter can be investigated on the profiles shown in Figure 4.3 (c). Thanks to the arguments previously raised, the organic particles mixing ratio can be assumed as constant within the aerosol layer, therefore,  $\beta_{fluor}$  becomes a tracer of the aerosol concentration evolution. Moreover, we can see that the fluorescence backscatter coefficient decreases in the upper part of the aerosol layer, suggesting a reduction in aerosol concentration in these altitudes. This consequently reduces the elastic backscatter coefficient when  $RH$  is at its peak, compared to a scenario where the aerosol concentration would have stayed uniform. Thus, it explains the underestimation of the Hänel hygroscopic parameter, and of the enhancement factor  $f(RH)$ .

To take this variation into account, the method presented in part 4.2 using equation (3.2) can be applied instead. In this method, the fluorescence backscatter is used to correct the aerosol

concentration variations in the layer. Figure 4.5 shows the evolution of the normalized backscatter coefficient with  $RH$ , as well as the fit of this evolution to the parameterization presented equation (3.2).

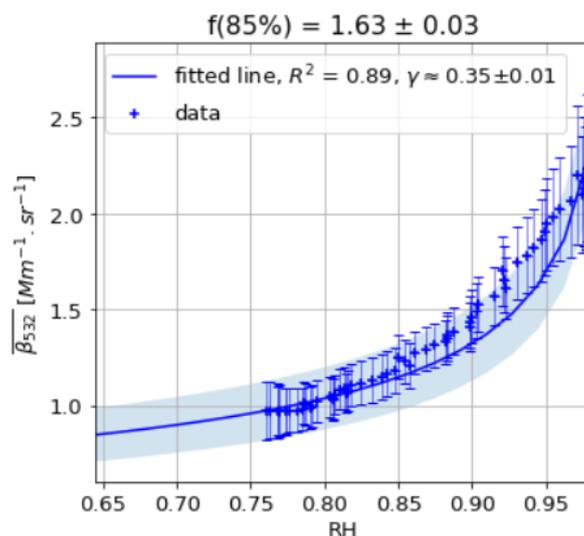


Figure 4.5: Evolution of the normalized elastic backscatter coefficient at 532 nm (normalized with fluorescence backscatter coefficient) in function of  $RH$  on 29 July 2021 from 22:00 to 23:00 UTC, between 1000 m and 1500 m above ground level, and results of the fit to the Hänel parameterization

Figure 4.5 shows that the fit to the Hänel parameterization is improved when the aerosol elastic backscatter coefficient is normalized with the fluorescence. First, concerning the determination coefficient, it increases significantly from  $R^2 = 0.72$  to  $R^2 = 0.89$ . This notable rise indicates that the model now represents the data much more accurately, as the determination coefficient is closer to  $R^2 = 1$ , the value at which the model perfectly fits the data. The Hänel hygroscopic parameter estimation is also impacted by the use of fluorescence, increasing from  $\gamma = 0.24 \pm 0.02$  without fluorescence correction, to  $\gamma = 0.35 \pm 0.01$  with the fluorescence correction. This increase is a direct consequence of the fact that the decreasing aerosol concentration in the upper part of the layer is now balanced by the normalization of the elastic backscatter coefficient with the fluorescence backscatter coefficient. The same behavior is observable on the enhancement factor at 85 % as it increases from  $f(85\%) = 1.40 \pm 0.06$  without fluorescence correction to  $f(85\%) = 1.63 \pm 0.03$  with fluorescence correction, falling precisely in the range of previous estimates for the enhancement factor of urban aerosols.

It is interesting to note that the major concerns that were raised in the characterization of the hygroscopic properties of the aerosol with the previous method are dealt with when  $\beta_{fluo}$  is used to take into account aerosol concentration variations. The determination coefficient increases, witnessing a better fit to the model, and the enhancement factor also increases,

getting in the range between 1.5 to 2, which concurs with previous estimates of  $f$  (85%) for urban aerosols. These facts confirm that the aerosol concentration changes in the aerosol layer between the lowest and the highest altitudes, in spite of the constant potential temperature and constant water vapor mixing ratio. It also confirms that the fluorescence backscatter coefficient can indeed be used to follow these evolutions and correct them to improve the estimation of the aerosol hygroscopic characteristics.

#### 4.3.2 Hygroscopic growth of biomass burning aerosols during the night of 9 March 2021.

The second case analyzed for this study occurred nighttime between the 9 and 10 March 2021. Figure 4.6 shows the quick-looks of the elastic backscatter coefficient at 532 nm and the results of FLARE-GMM during this night.

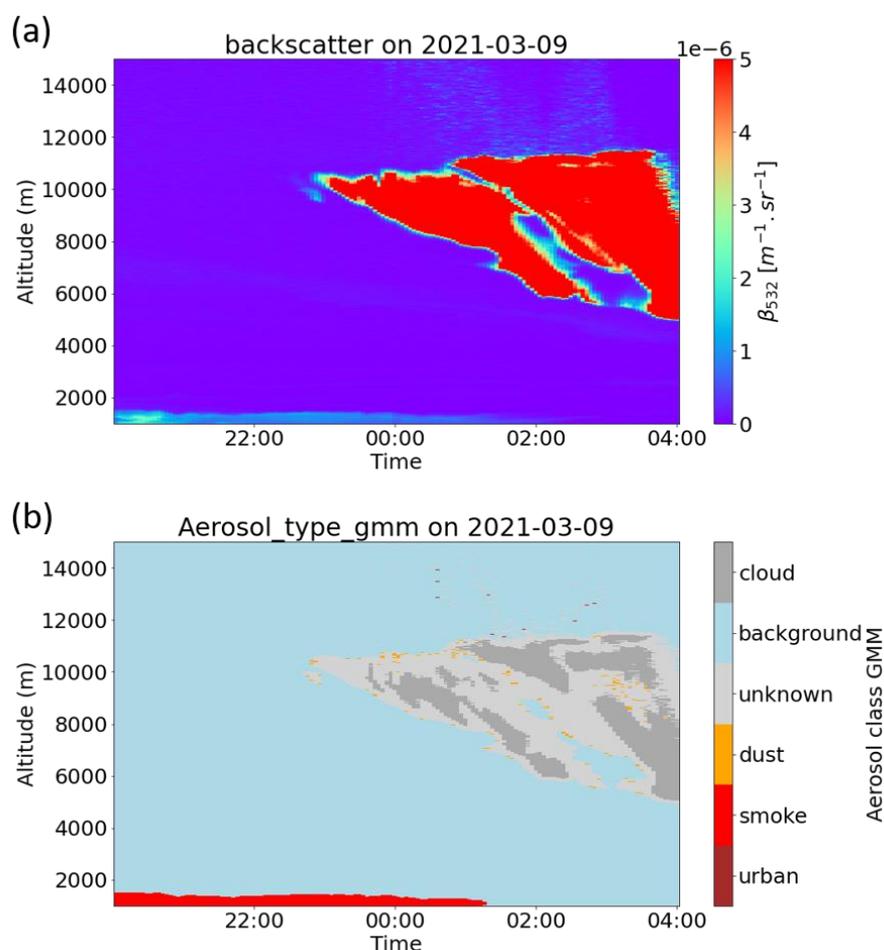


Figure 4.6: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 9 March 2021

From the backscatter quick-look, it is possible to observe the presence of aerosols in the boundary layer, up to 1500 m during the early part of the night. It is interesting to notice that, similarly to the 29 July situation, the boundary layer height decreases with time during the

night, and even disappears from the range of the lidar after 1:00 UTC. This is also a result of the reduction in ground heating, which diminishes after sunset when the sun's radiative power ceases to warm the ground, consequently favoring boundary layer stability, and decreasing its height. An ice cloud is also present during the second part of the night, ranging from 5 km to 12 km, from 23:00 UTC to the end of the night. The presence of ice crystals in the cloud can be confirmed by its high altitude and high depolarization. Its large spatial coverage and low optical thickness also confirm the fact that this cloud is made of ice crystals.

Regarding the results from FLARE-GMM, it estimates that the boundary layer is loaded with biomass burning aerosols, which are present across the whole layer. During winter, heating can be responsible for the emission of biomass burning aerosols, when fireplaces are used for example, supporting the presence of this aerosol type in the boundary layer.

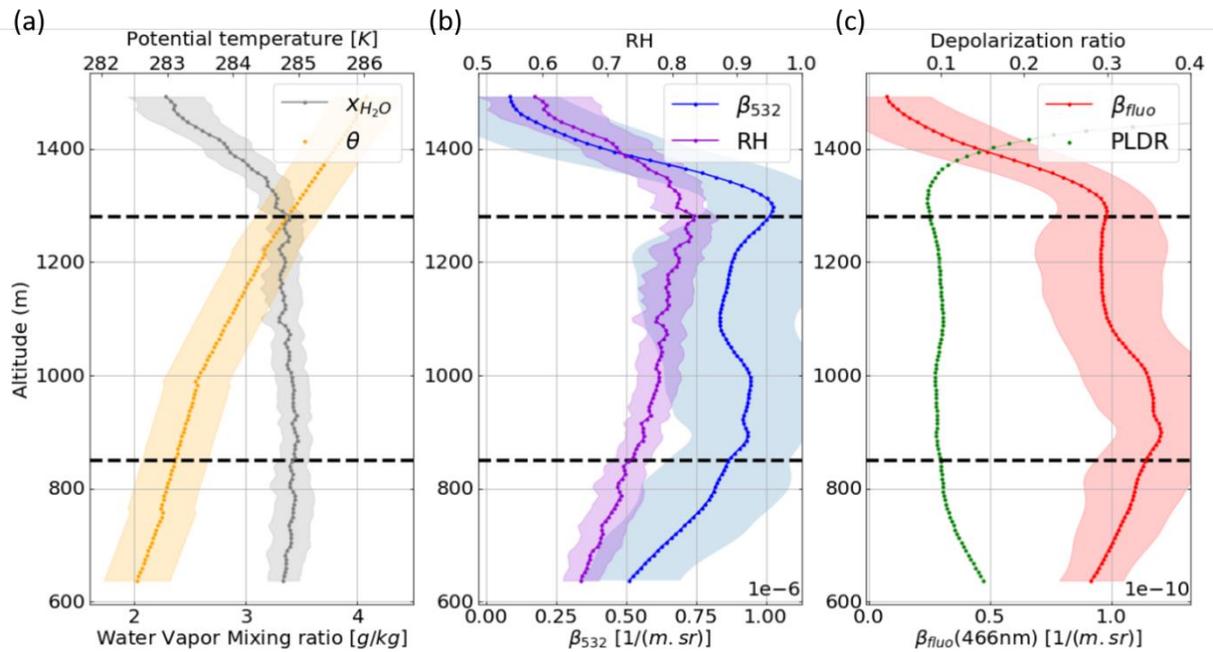


Figure 4.7: Profile of retrieved optical properties in function of altitude above ground level (a) water vapor mixing ratio [g/kg] and potential temperature [K], (b) elastic backscatter coefficient at 532 nm [ $m^{-1} \cdot sr^{-1}$ ] and  $RH$ , (c) fluorescence backscatter coefficient at 466 nm [ $m^{-1} \cdot sr^{-1}$ ] and particular linear depolarization ratio at 532 nm, on 9 March 2021 from 21:00 to 22:00 UTC, the black dashed lines identify the area where hygroscopic growth is expected to occur

Backward trajectories have been computed for this case (in the Appendix), and show that the air masses originate from the Northwest direction, where the coast is located. It suggests that marine aerosols might be present in the atmospheric boundary layer. However, the lack of data of pure marine aerosols makes it difficult to draw definitive conclusions. Moreover, the city of Lille is located 70 km from the coast, making it improbable to find a majority of marine aerosols in the boundary layer. Therefore, the atmospheric boundary layer is assumed as composed of biomass burning aerosols in this case.

In order to determine if this situation is suitable for the study of aerosol hygroscopic properties, the same procedure as for the previous case is used. The profiles of LILAS inverted products between 21:00 and 22:00 UTC can be investigated. This situation seems to be the more appropriate for the study of hygroscopic properties, as the boundary layer appears more stable during this time interval. Furthermore, good mixing could be expected in the layer since it is not long after sunset, therefore convective processes could still occur to favor good mixing of the boundary layer.

Interesting observations can be made from the LILAS profiles displayed on Figure 4.7. Concerning the water vapor mixing ratio present on the first panel (Figure 4.7 (a)), it appears to be constant between 850 m and 1280 m ( $3.4 \pm 0.2$  g/kg), indicating that the air masses located at these altitudes have the same origin, which has already been stated by the analysis of the backward trajectory, but is confirmed here. Regarding the potential temperature, it is increasing with altitude, from  $283.1 \pm 0.5$  K to  $284.8 \pm 0.5$  K. This increase suggests that the boundary layer is likely to be stratified, indicating that there is no mixing between the air masses at the bottom of the layer and the air masses at the top of the layer. This could consequently lead to strong aerosol concentration variations in the layer. The fact that the potential temperature at the top and at the bottom of the layer are different might suggest that the air masses have different origins as the potential temperature is not likely to evolve rapidly in the atmosphere. However, since the difference is less than 2 K, and its evolution is smooth and slow, this conclusion can be disputed, especially when considering the water vapor mixing ratio and backward trajectories. If the air masses have the same origin, the hypothesis that the biogenic particles mixing ratio is homogeneous on the layer can still be formulated.

By looking at the middle panel, (Figure 4.7 (b)), it is possible to observe that  $RH$  increases from  $74 \pm 3$  % to  $83 \pm 4$  % which is almost a 10 % increase on the layer. This is not ideal since a wide variation of  $RH$  helps to better estimate  $\gamma$ . Yet, considering the high values of  $RH$  (being over 80 %), hygroscopic growth is still likely to happen in this situation, and thus, the hygroscopic growth constant can be estimated, at the cost of higher uncertainty. It is possible to observe that the elastic backscatter coefficient increases between the bottom and the top of the layer. However, this increase is not monotonous and does not follow the evolution of  $RH$ . This could be a consequence of the high measurement uncertainty (represented in light blue) that is due to atmospheric instability during the average time and instrumental noise. But the variations of  $\beta_{532}$  can also be explained by a variation of the aerosol concentration in the layer. By looking at the fluorescence backscatter represented on the last panel (Figure 4.7 (c)), it is possible to observe that the fluorescence backscatter coefficient follows the same behavior as the elastic backscatter coefficient, especially in the lowest part of the layer. This correlation indicates that the aerosol concentration, which can be followed by the fluorescence, is likely to be varying in

---

the layer. This variation could be caused by the layer stratification, suggested by the increasing potential temperature. Regarding the depolarization, it is slightly decreasing from  $10.0 \pm 0.2 \%$  at the bottom of the layer to  $8.7 \pm 0.7 \%$  at the top of the layer, which could confirm the presence of hygroscopic growth, however, like for the former case, the value of the *PLDR* being already low, the variations are not relevant. One final comment which can be made on this situation is that LILAS measures are investigated below 1000 m. Below this level, the lidar overlap function is not supposed to be exactly equal to 1, which consequently impacts the estimation of the aerosol backscatter and fluorescence backscatter coefficients. However, due to the low variation range of *RH*, and the high uncertainty of the elastic backscatter coefficient, a high level of uncertainty on the estimations of the aerosol hygroscopic properties is expected in these conditions. Therefore, the overlap function is not expected to have a significant impact on the aerosol hygroscopic property estimates. For this reason, it is approximated as to one for such particular case.

All these elements allow us to conclude that this situation, even if it is not an ideal case, can be used to study the hygroscopic properties of this aerosol layer. However, the low range of variation of *RH* and the high uncertainty on  $\beta_{532}$  indicate that the estimation of these properties is likely to have a high degree of uncertainty. Nevertheless, this case can still be used to evaluate the efficiency of the normalization by the fluorescence backscatter coefficient to study hygroscopic properties, especially in a situation with atmospheric stratification and varying aerosol concentration.

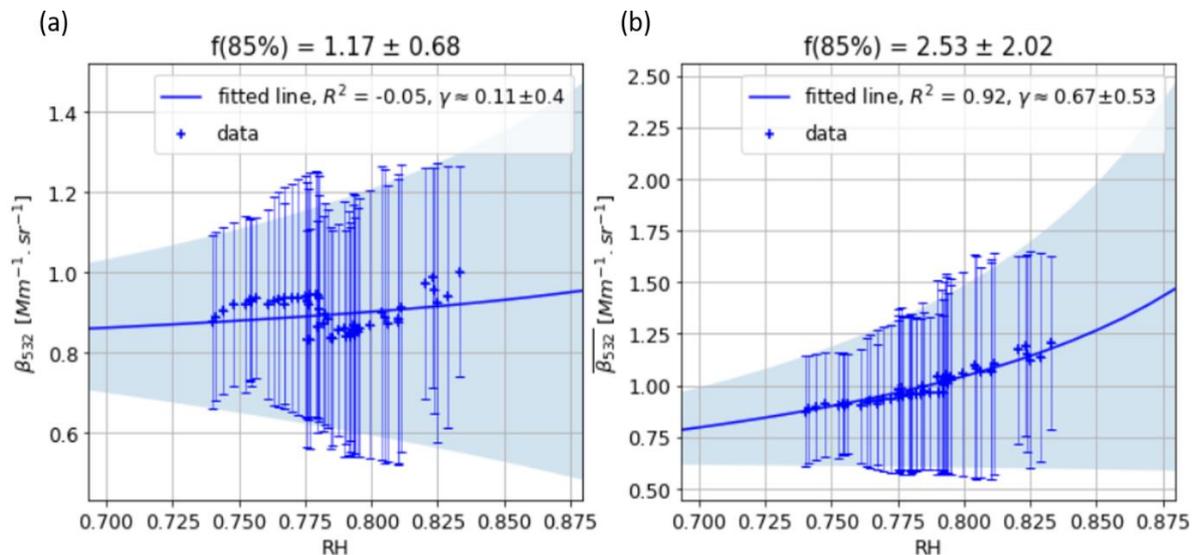


Figure 4.8: Evolution of the elastic backscatter coefficient at 532 nm (a) without normalization and (b) with normalization with the fluorescence backscatter coefficient, in function of *RH* on 9 March 2021 from 21:00 to 22:00 UTC, between 850 m and 1280 m above ground level, and results of the fit to the Hänel parameterization

Figure 4.8 (a) shows that when the fluorescence backscatter coefficient is not used to correct aerosol concentration variations in this case, the determination coefficient of the fit to the Hänel parameterization is negative ( $R^2 = -0.05$ ). A negative determination coefficient indicates that the model is wrongly chosen to represent the data. It means that in this situation, the Hänel parameterization does not represent how the elastic backscatter coefficient evolves with  $RH$ . Therefore, it indicates that other processes are occurring in the layer, like it was the case in the previous situation. In this condition, the evaluation of the hygroscopic properties estimated by this method is not relevant, since the situation is not well represented by this parameterization. This poor fit could be explained by the increase of the potential temperature in the layer, which suggests that the boundary layer is stratified, as previously stated, and aerosol concentration variations are therefore likely to occur. In order to take this variation into account, the fluorescence backscatter coefficient is used to normalize  $\beta_{532}$ .

Figure 4.8 (b) shows the elastic backscatter coefficient normalized by the fluorescence in function of  $RH$ , and the fit to the Hänel parameterization. This figure shows that the model represents the data much more accurately when the normalization is applied. This can be observed with the determination coefficient which largely increases when the elastic backscatter coefficient is normalized ( $R^2 = 0.92$  instead of  $R^2 = -0.05$ ). This increase confirms that the aerosol concentration varies in the aerosol layer, explaining the poor fit to the Hänel parameterization previously observed when this variation is not considered. Moreover, it indicates that these variations are well captured by the normalization of the elastic backscatter coefficient with the fluorescence backscatter coefficient. It therefore supports the fact that aerosol concentration can be followed by the fluorescence backscatter coefficient, considering that the biogenic particle mixing ratio is constant in the aerosol layer, which has been demonstrated here. Furthermore, it suggests that even if the potential temperature within the layer increases, indicating stratification, the hypothesis that the aerosol biogenic particle mixing ratio remains homogeneous still holds true, provided that this temperature increase is minor (less than 2 K) and the water vapor mixing ratio is constant. This condition allows for tracking aerosol concentration using fluorescence. Eventually, it is possible to conclude that even if the potential temperature increase is small, there is still enough stratification in the layer to exhibit strong aerosol concentration variations, which greatly impact the hygroscopic properties estimations with a classic approach.

Concerning the hygroscopic properties, it is difficult to conclude about these estimates since the uncertainties are very important, as anticipated. Nevertheless, the Hänel hygroscopic coefficient is estimated at  $\gamma = 0.7 \pm 0.5$ , and the enhancement factor at 85 % is estimated at  $f(85\%) = 2.5 \pm 2$ . The expected range of enhancement factor for biomass burning aerosols is around 2 at 532 nm (Sicard et al., 2022), the results obtained by the analysis of this case are

---

therefore in this expected range for this aerosol type. However, the high uncertainty on the hygroscopic properties retrieval, due to the uncertainties on  $\beta_{532}$  and  $\beta_{fluor}$ , and the low range of  $RH$  variation, makes it challenging to interpret these results.

Nevertheless, this case is still a nice example of the use of fluorescence data to improve the characterization of aerosol hygroscopic properties. In this type of situation where stratification occurs, the fluorescence helps following the aerosol concentration evolutions in the layer, and therefore mitigate the impact of this variation on the Hänel hygroscopic coefficient, and therefore the enhancement factor estimates, thus improving the characterization of the aerosols hygroscopic properties.

#### 4.3.3 Hygroscopic growth of urban aerosols during the night of 16 April 2021.

The final case presented in this study occurred during the night between 15 to 16 April 2021. The quick-looks of the aerosol elastic backscatter coefficient at 532 nm and FLARE-GMM results are displayed Figure 4.9.

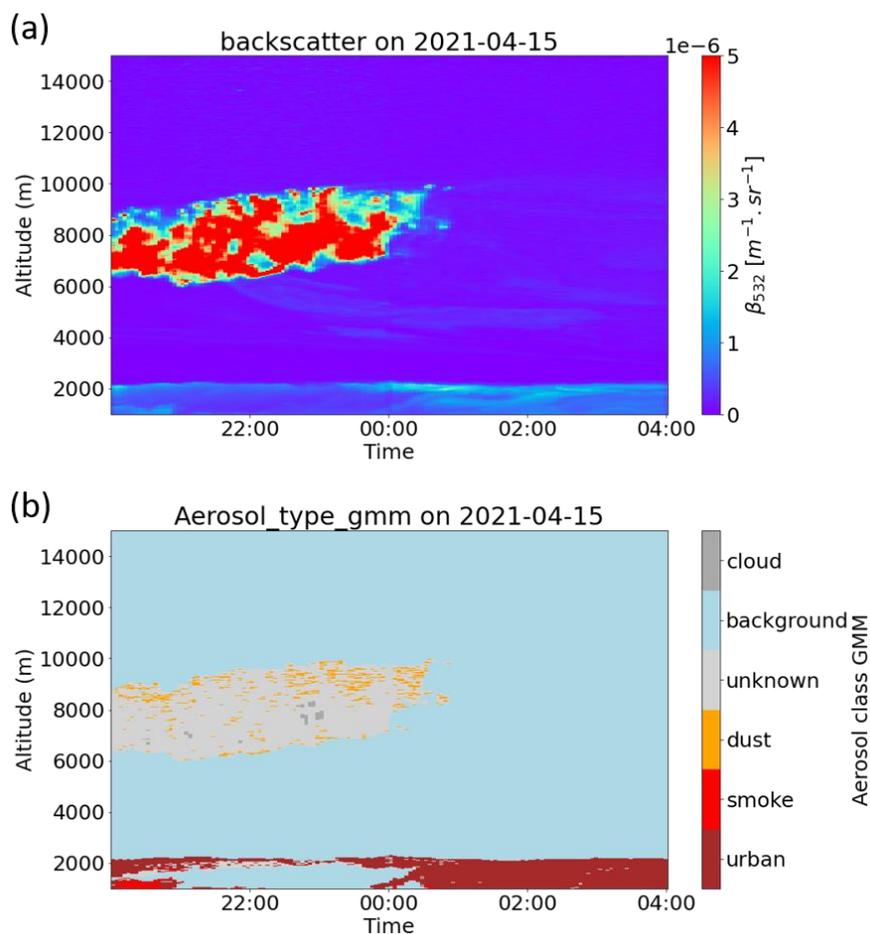


Figure 4.9: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 15 April 2021

The elastic backscatter coefficient quick-look (Figure 4.9 (a)) shows the presence of an ice cloud during the early part of the night between 6 km and 10 km until 00:00 UTC. It can be identified by its high depolarization, large vertical coverage and low optical thickness. Figure 4.9 (a) also shows the presence of a boundary layer, with height located around 2000 m above ground level. It is interesting to observe that the boundary layer height does not decrease during the night, as it is generally expected due to ground cooling, and can be observed in the two first cases presented in this study. One more interesting observation concerns the variation of aerosol elastic backscatter coefficient in the boundary layer during the night. Between 20:00 and 00:00 UTC, the aerosol elastic backscatter coefficient is highly heterogeneous in the boundary layer, it is stronger in the top of the boundary layer, and quickly decreases at lower altitudes. However, after 00:00,  $\beta_{532}$  increases in the lower part of the boundary layer, and the boundary layer appears more homogeneous. This could be the consequence of convective processes lifting air masses from lower altitudes, which are more loaded in aerosols. However, these processes are generally driven by radiative forces from the ground, when it is heated by the sunlight, which cannot be the case here as it occurs in the middle of the night. Another possible explanation for this phenomenon is that aerosols have been emitted at this time. The results of FLARE-GMM (Figure 4.9 (b)) show that before 00:00 UTC, the aerosol type is not obvious; there might be a mixture between urban aerosols and biomass burning aerosols. However, after 00:00 UTC, the type of the aerosols present in the boundary layer is estimated as being mainly urban. This could indicate that urban aerosols have been emitted from the ground around 00:00 UTC, and have loaded the boundary layer. This supports the fact that the increase of the elastic backscatter in the boundary layer is due to the emission of urban aerosols during the night.

To study hygroscopic growth in this case, the profiles averaged between 01:00 UTC and 02:00 UTC have been investigated. This interval has been chosen because the boundary layer appears to be the more homogeneous at this time. Profiles of LILAS measures averaged on this time period are displayed Figure 4.10.

The potential temperature profile displayed Figure 4.10 (a) shows that similarly to the former case, the potential temperature increases in the boundary layer, indicating the presence of stratification between the low altitudes and the high altitudes. However, the increase of potential temperature is inferior than 2 k (from 280.9 K to 281.4 K), which suggests that the stratification is weak and that it does not contradict the fact that the biogenic particles mixing ratio remains constant in the layer. On the other hand, the water vapor mixing ratio is constant (equal to  $2.4 \pm 0.2$  g/kg), indicating that the air masses have the same origin. This fact is confirmed by the backward trajectory (in the Appendix), suggesting that the air masses present in the boundary layer all come from the North direction. These observations, combined with

---

FLARE-GMM result being homogeneous on the layer, allow us to conclude that the biogenic particles mixing ratio is expected to be constant in the considered layer. The fluorescence backscatter coefficient can therefore be used to follow the evolution of the aerosol concentration.

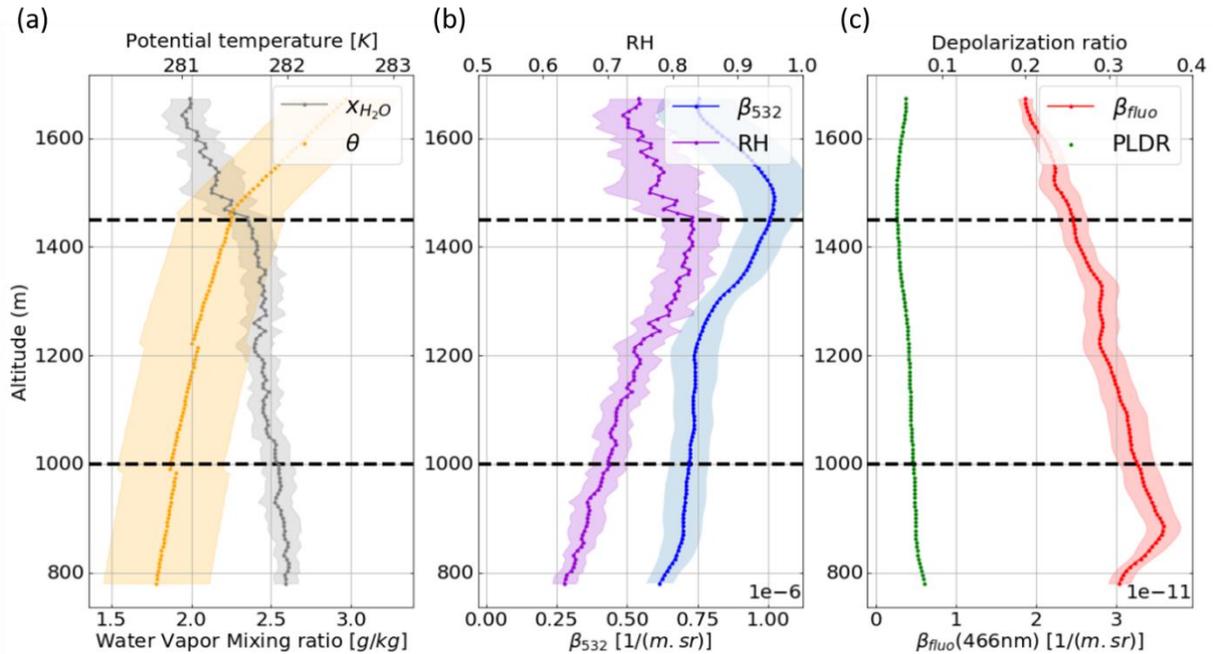


Figure 4.10: Profile of retrieved optical properties in function of altitude above ground level (a) water vapor mixing ratio [g/kg] and potential temperature [K], (b) elastic backscatter coefficient at 532 nm [ $m^{-1}.sr^{-1}$ ] and  $RH$ , (c) fluorescence backscatter coefficient at 466 nm [ $m^{-1}.sr^{-1}$ ] and particular linear depolarization ratio at 532 nm, on 16 April 2021 from 01:00 to 02:00 UTC, the black dashed lines identify the area where hygroscopic growth is expected to occur

Figure 4.10 (b) shows that both the elastic backscatter coefficient and  $RH$  increase in the layer. The elastic backscatter increases from  $0.7 \pm 0.1 Mm^{-1}sr^{-1}$  to  $1.0 \pm 0.1 Mm^{-1}sr^{-1}$  which is a small increase but which can still be characterized. On the other hand,  $RH$  increases from  $70 \pm 3 \%$  to  $83 \pm 4 \%$ . It indicates that hygroscopic growth is expected to occur since the maximum  $RH$  reached in the layer is over 80 %. Furthermore, the span of the  $RH$  variations in the layer is over 10 %, allowing to accurately estimate the hygroscopic properties of the aerosols. The  $PLDR$  profile represented Figure 4.10 (c) shows that the depolarization decreases from  $6.5 \pm 0.2 \%$  at the bottom of the layer, to  $4.6 \pm 0.2 \%$  at the top of the layer. Once again, this decrease is not relevant considering the low value of  $PLDR$  for this type of aerosols, but still comfort the fact that hygroscopic growth is occurring. These elements indicate that this situation is ideal to study the hygroscopic properties of the aerosols present in the boundary layer during that night.

In order to quantify the hygroscopic properties of the aerosols, the evolution of the elastic backscatter coefficient with  $RH$ , both not normalized (Figure 4.11 (a)) and normalized (Figure 4.11 (b)) by the fluorescence are fitted on the Hänel parameterization. Interesting observations can be made from the results of these fits. First, concerning the case which is not normalized by the fluorescence, we can see that the fit to the Hänel parameterization is relatively good, with a determination coefficient equal to  $R^2 = 0.87$ . The Hänel hygroscopic parameter is estimated at  $\gamma = 0.46 \pm 0.02$  which corresponds to an enhancement factor of  $f(85\%) = 1.9 \pm 0.1$  which is exactly in the expected range for urban aerosols, between 1.5 and 2. These results could indicate that the situation is well characterized by the approach usually used in the literature, as the fit to the Hänel parameterization is good, and the result is in the range of what is expected. However, this result lies on the hypothesis that the aerosol layer is homogeneous, which can be refuted by the increasing potential temperature shown Figure 4.10 (a), indicating a stratification of the aerosol layer. Moreover, the homogeneity of the layer can also be contradicted by the fluorescence backscatter coefficient shown Figure 4.10 (c), which decreases with altitude. In the case of a homogeneous layer, the fluorescence backscatter coefficient should remain constant, which means that in this case, the aerosol concentration is expected to evolve in the layer.

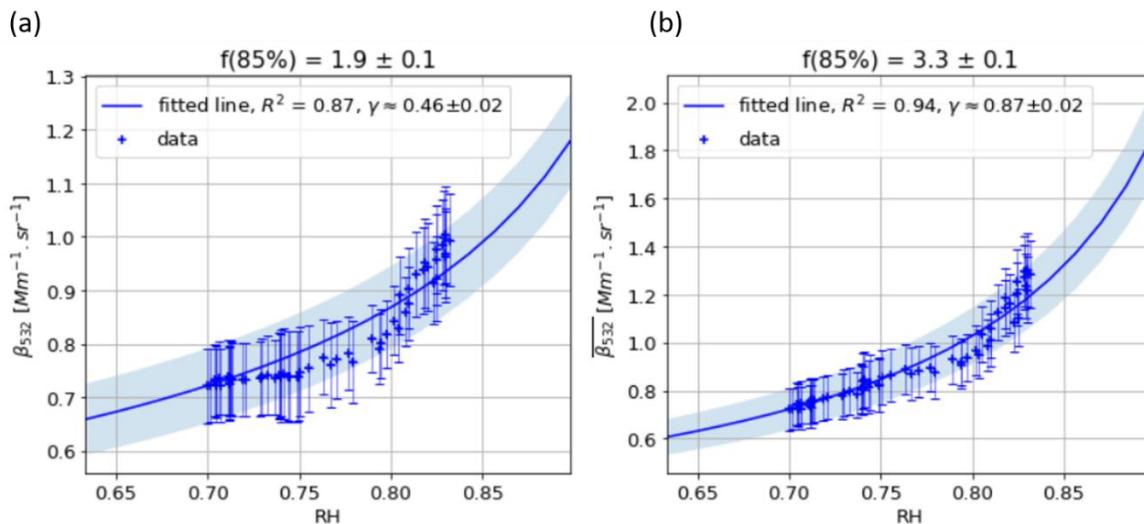


Figure 4.11: Evolution of the elastic backscatter coefficient at 532 nm (a) without normalization and (b) with normalization with the fluorescence backscatter coefficient, in function of  $RH$  on 9 March 2021 from 21:00 to 22:00 UTC, between 1000 m and 1450 m above ground level, and results of the fit to the Hänel parameterization

On the other hand, the fit of the normalized elastic backscatter coefficient to the Hänel parameterization can be analyzed. The determination coefficient is equal to  $R^2 = 0.94$  which is an increase compared to the case where the backscatter coefficient is not normalized ( $R^2 = 0.87$ ). This increase supports the fact that the aerosol concentration varies in the layer, and that this variation is accurately corrected by the normalization of  $\beta_{532}$  with the fluorescence

backscatter coefficient. The Hänel hygroscopic parameter is estimated at  $\gamma = 0.87 \pm 0.02$  which corresponds to an enhancement factor of  $f(85\%) = 3.3 \pm 0.1$ . This value of enhancement factor is much more important than what is expected for urban aerosols. These values of enhancement factor at 532 nm are characteristic of situations where marine aerosols are mixed with urban aerosols (Sicard et al., 2022). This could be the case here as the backward trajectories indicate that the air masses present in the Lille boundary layer during this night originate from the North, *i.e.* from the sea. However, considering the distance between Lille and coast, it is highly doubtful that the marine aerosol concentration could be high enough that it would impact this much the hygroscopic growth properties of the aerosol. These values are generally attained in the littoral, while Lille is 70 km inland, and pure marine aerosols layers are almost never observed. Another factor which could explain this result is the uncertainty on  $RH$ . Indeed, an error on the humidity calibration coefficient, or on the temperature can have important impacts on  $RH$  estimate, and therefore on the estimation of aerosol hygroscopic properties. This aspect will be explained deeply in part 4.5.

## 4.4 Automatic study of hygroscopic cases

### 4.4.1 Method

The different cases presented in the previous part show the efficiency of the method developed to study aerosol hygroscopic growth from LILAS, which enables a better characterization of the aerosol hygroscopic properties compared to other common methods employed with lidar data. The objective of this section is now to establish a routine that allows for the automatic identification of hygroscopic growth cases. This way, it would be possible to automatically estimate the aerosol hygroscopic properties of a wide variety of situations. This would enable the study of correlations between the hygroscopic properties of aerosols and other properties, such as aerosol type, fluorescence capacity or depolarization.

The method starts with the determination of intervals of altitudes showing monotonous increase of  $RH$ . In order to mitigate the impact of measurement noise, moving average on 70.5 m (10 altitude ranges) is applied to LILAS profiles of  $RH$ . Then, the growth rate can be computed to determine altitudes where  $RH$  increases. Once these altitudes are identified, a list of criteria can be used to estimate if each situation can be studied to determine the hygroscopic properties of the aerosol.

- \_ The average elastic backscatter coefficient on the layer is over  $0.5 \text{ Mm}^{-1}\text{sr}^{-1}$
- \_ The maximum value of  $RH$  in the layer is over 80 %
- \_ The variation of  $RH$  in the layer is over 10 %
- \_ The variation of potential temperature is below 2 K

- \_ The variation of water vapor mixing ratio is below 10 % of the average on the layer
- \_ FLARE-GMM aerosol type remains constant on the layer

Once the cases are identified, the aerosol hygroscopic properties can be estimated. Fluorescence can be utilized to correct potential changes in aerosol concentration within the layer, as described in equation (3.2). Additionally, fitting the Hänel parameterization to the normalized elastic backscatter coefficient enables to estimate the Hänel hygroscopic parameter  $\gamma$ . From this parameter, it is possible to compute the enhancement factor,  $f(RH)$ , which quantifies the aerosol hygroscopic properties.

In order to consider only relevant cases, if the determination coefficient of the fit to the Hänel parameterization is below  $R^2 = 0.8$ , the situation is not treated. This generally corresponds to cases where a problem occurred with the inversion or in the humidity calibration. In order to identify situations where the atmosphere is highly unstable during the average over a 1-hour period, cases which exhibit high uncertainty on the enhancement factor estimates are not considered. If the uncertainty is over 30 % of the  $f(85\%)$  value, the case is removed. Eventually, to further eliminate irrelevant cases, it is important to identify situations where condensation occurs, as it is not modeled well by the Hänel parameterization. Condensation of water droplets consequently introduces overestimations of  $\gamma$  and  $f(85\%)$ . In order to take this aspect into account, the cases for which the Hänel hygroscopic parameter estimate is over  $\gamma = 1$  are not considered. This value of  $\gamma$  corresponds to highly hydrophilic aerosols, and is reached only in the case of pure sea salts (Haarig et al., 2017). These conditions are not expected to be met in Lille due to its distance to the sea as previously stated. Therefore, these values of Hänel hygroscopic parameter correspond to situations where condensation occurs, which leads to overestimations of  $\gamma$ , and can be removed in order to analyze hygroscopic growth cases only.

#### **4.4.2 Results**

The method previously described to automatically identify and analyze hygroscopic growth cases has been applied to all available profiles between 2021 and 2023. It allowed us to identify 36 cases in 2021 and 2022. Unfortunately, due to the absence of humidity measurements for several months in the spring of 2023, no hygroscopic cases have been identified for this year.

Figure 4.12 shows the histogram of the hygroscopic growth occurrences in function of the time (Figure 4.12 (a)) and the histogram of the mean altitudes at which hygroscopic growth cases are identified (Figure 4.12 (b)). The first observation that can be made is the absence of dust cases. This can be explained by the fact that dust events generally occur in Lille during specific meteorological conditions, such as the manifestation of Sirocco winds bringing desert dust from the Sahara. For that reason, desert dust aerosol layers are generally dry in Lille, and no hygroscopic growth can be witnessed. Moreover, the hygroscopicity of desert dust aerosols is

---

still debated in the scientific community (Herich et al., 2009; Chen et al., 2020). For this reason, rare instances of desert dust with high humidity are not expected to exhibit the same hygroscopic behavior as other aerosol types. Therefore, their fit to the Hänel parameterization is likely to be poor, and these cases would be filtered out because their determination coefficient would not meet the necessary threshold.

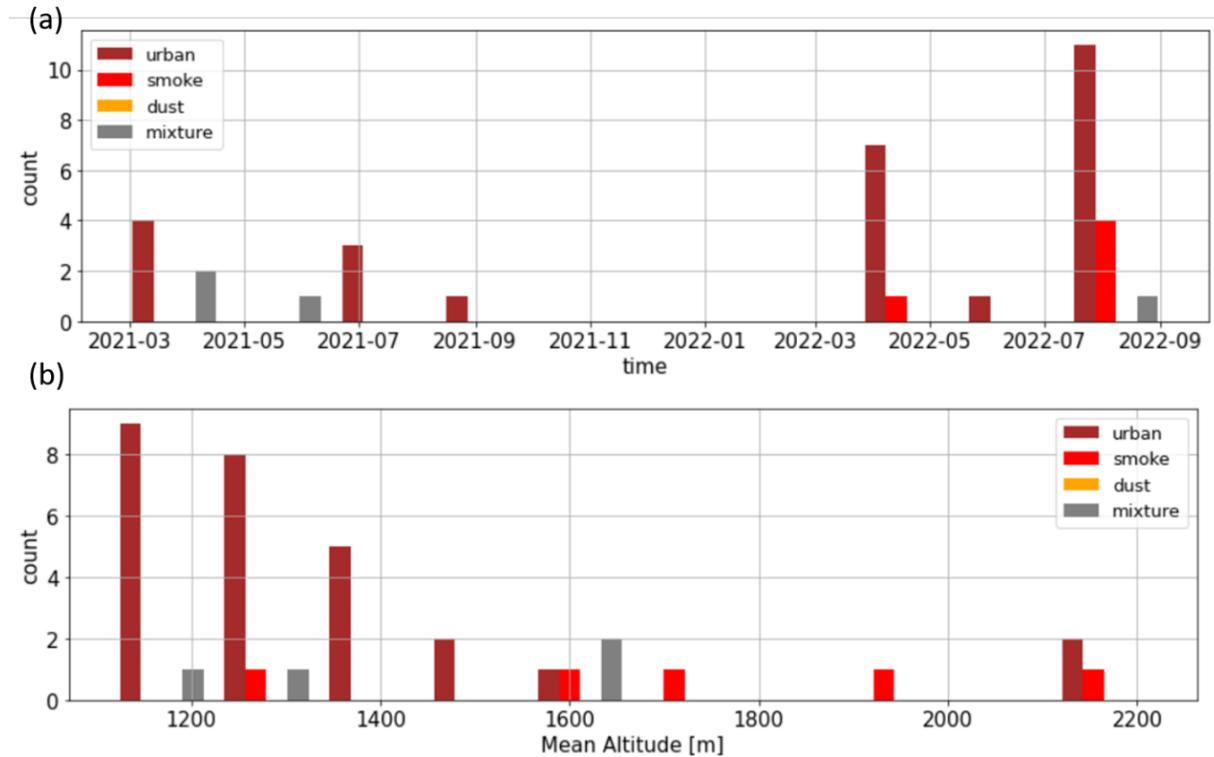


Figure 4.12: Histograms of hygroscopic growth cases (a) time of occurrences (b) mean altitude above ground level (in meter) between February 2021 and December 2023

Figure 4.12 (a) also shows that most cases occur during spring and summer, and that no hygroscopic growth cases seem to be identified between October to March. This can be explained by the fact that in Lille during the winter, the precipitation rate is generally high, reducing the opportunities for lidar measurements under humid atmospheric conditions. Additionally, during clear sky conditions, the low temperature typically results in low relative humidity levels, preventing hygroscopic growth situations. Eventually, the lower boundary layer height in winter implies that cases of hygroscopic growth may fall below the detection range of the lidar, as they are typically found within the boundary layer, as explained below.

Figure 4.12 (b) shows that the average altitude of hygroscopic growth events in Lille is generally below 2 km. This result is expected as the boundary layer shows high humidity and high aerosol concentration, gathering therefore favorable conditions to witness hygroscopic growth. On the contrary, as the altitude increases, the humidity tends to quickly decrease in the free troposphere during clear-sky conditions, and the aerosol load is generally lower than in the

boundary layer, which has been evidenced in chapter 3. As most hygroscopic growth cases occur in the boundary layer, it mainly concerns urban aerosols which tend to be the most commonly present at low altitudes in Lille atmosphere. Therefore, over the 36 identified cases, 27 are for urban aerosols, only 5 are for biomass burning aerosols, and 4 concern unknown aerosols that can be mixtures of different types of aerosols.

Figure 4.13 shows the results of the enhancement factor estimates for the different kinds of aerosols across all identified hygroscopic growth cases. Due to the limited number of events involving biomass burning aerosols and mixtures, the boxplots are unfortunately not relevant for these aerosol types, and it is difficult to analyze the results. Moreover, for the same reason, the results for all cases combined are very similar to those for the urban plot. Nevertheless, it is possible to observe that except for mixtures, all distributions exhibit high standard deviation.

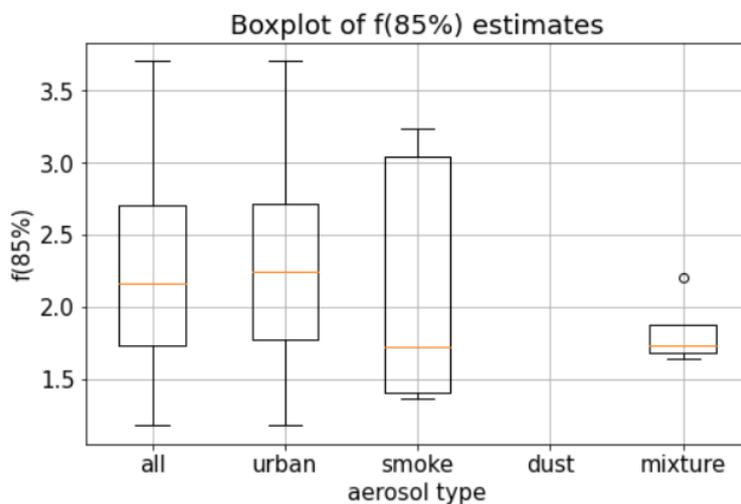


Figure 4.13: boxplots of the different estimates of the enhancement factor at 85 % for  $\beta_{532}$ ,  $f(85\%)$  for all aerosols and each aerosol type estimated from FLARE-GMM

Concerning urban aerosols, the enhancement factor estimates are averaged at  $f(85\%) = 2.3 \pm 0.2$  with a standard deviation of 0.6. This average value is higher than previous estimates for urban aerosols, which are expected to range between 1.5 and 2 (Sicard et al., 2022). Moreover, the enhancement factors estimate for urban aerosols vary from 1.2 to 3.7, representing a significant range for a single aerosol type. The same observation can be made for biomass burning aerosols. The enhancement factor is averaged at  $f(85\%) = 2.1 \pm 0.3$ , which aligns with the expected value for biomass burning aerosols, typically ranging around 2 (Sicard et al., 2022). However, the standard deviation is equal to 0.9 and the estimates range from 1.4 to 3.2, which also represent much more important variations than what is anticipated for a single aerosols type.

Several hypotheses could explain why the results differ from the expectations. The first one is that aerosols in Lille exhibit a large variety of compositions, explaining why their hygroscopic

properties can vary on such large ranges, and why the average estimate of urban aerosol enhancement factor differs from what has been determined in previous work. However, values of enhancement factors at 85 % above 3 correspond to aerosols containing marine particles, as previously stated, which is not expected to be the case in the Lille atmosphere. Therefore, these extreme values could not be explained only by the aerosol composition. On the other hand, potential errors on  $RH$  estimate could also explain why the estimations of aerosol hygroscopic properties are highly variable. Indeed, errors on  $RH$ , due to calibration issues or uncertainty in the temperature estimate can significantly impact the assessment of the Hänel hygroscopic constant, and can therefore potentially strongly affect the enhancement factor estimate. Unfortunately, the uncertainties in  $RH$ , which are expected to range from 1% to 10% with our current setup, significantly impact  $\gamma$  estimates. This makes it challenging to draw definitive conclusions about the aerosol hygroscopic properties and their potential correlations with other aerosol properties, such as type, fluorescence capacity, or depolarization.

The next part of the chapter focuses on evaluating the influence of  $RH$  uncertainties on the estimation of aerosol hygroscopic properties.

## **4.5 Influence of $RH$ uncertainty on hygroscopic growth properties estimation**

### **4.5.1 Impact of a flat adjustment in $RH$**

The objective of this part is to evaluate how the uncertainties on the determination of  $RH$  can influence the estimation of the Hänel hygroscopic constant, and therefore the enhancement factor. In order to do so, the  $RH$  profile has been adjusted by a fixed value  $\Delta_{RH}$  ( $RH' = RH + \Delta_{RH}$ ), ranging from plus 0.1 to minus 0.1. The fit to the Hänel parameterization of  $\beta_{532}$  normalized by  $\beta_{flu0}$  is then performed with the new  $RH$  value. The modification of the Hänel hygroscopic growth constant and of the enhancement factor can then be analyzed in function of  $\Delta_{RH}$ .

The case presented in part 4.3.3 which occurred during the night between 15 and 16 April 2021 has been chosen to study how the Hänel hygroscopic constant is changed by the adjustment on  $RH$ . This case is interesting as the estimation of the aerosol hygroscopic properties is not in the expected range for urban aerosols, and because this difference can be explained by an error on  $RH$ . It could be either because of an error on the LILAS humidity calibration or on the assessment of the temperature, which would both impact the estimation of  $RH$ .

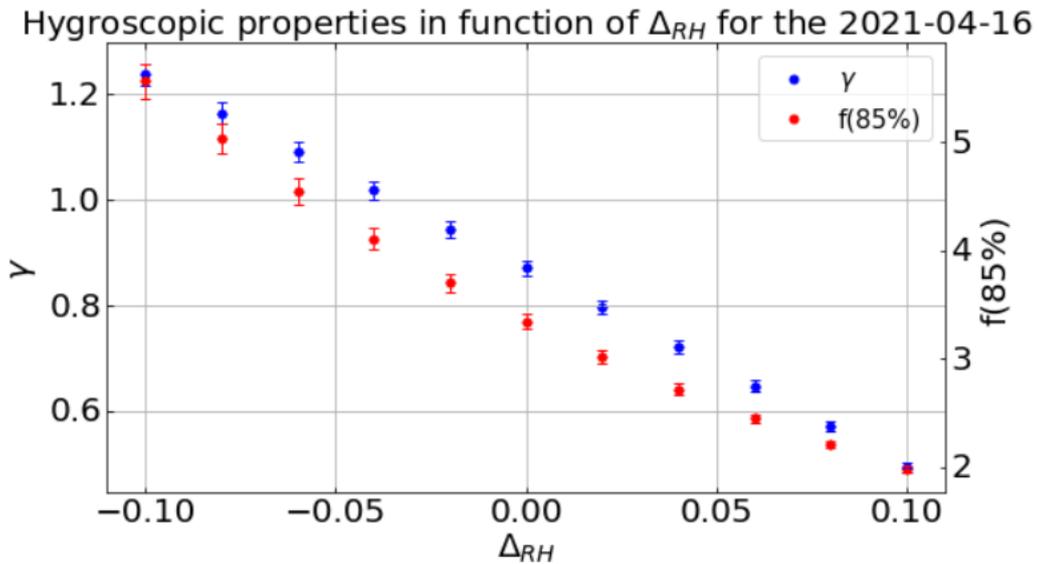


Figure 4.14: Variation of the Hänel hygroscopic constant  $\gamma$  (blue) and the enhancement factor at 85 % (red) at 532 nm, in function of the adjustment on the humidity ( $RH' = RH + \Delta_{RH}$ ) for the 16 April 2021 between 01:00 and 02:00 UTC between 1000 m and 1450 m

Figure 4.14 shows how the estimations of the Hänel hygroscopic constant and the enhancement factor at 85 % evolve with the shift in  $RH$ . The first observation that can be made is that  $\gamma$  varies from  $\gamma = 1.24 \pm 0.02$  for  $\Delta_{RH} = -0.1$ , to  $\gamma = 0.49 \pm 0.01$  for  $\Delta_{RH} = +0.1$ , which correspond to variations of enhancement factor from  $f(85\%) = 5.6 \pm 0.2$  for  $\Delta_{RH} = -0.1$ , to  $f(85\%) = 1.98 \pm 0.03$  for  $\Delta_{RH} = +0.1$ . This wide variation confirms that the results presented part 4.3.3 and 4.4.2 can indeed be explained by errors on the estimation of  $RH$ . Moreover, as the enhancement factor falls in the expected values for urban aerosols when  $\Delta_{RH}$  is around plus 0.1, it suggests that  $RH$  is certainly underestimated during the night between 15 and 16 April 2021, either because of an error on LILAS calibration, or on the estimation of the temperature.

Another interesting observation concerns the evolution of the Hänel hygroscopic growth constant with  $\Delta_{RH}$ , which is smooth and appears almost linear. The next part is dedicated to the approximation of the slope of this variation, in order to quantify the dependency of  $\gamma$  to  $RH$ .

Eventually, it is also interesting to investigate how the determination coefficient of the fit to the Hänel parameterization varies with errors on  $RH$ . Figure 4.15 shows how  $R^2$  evolves in function of  $\Delta_{RH}$ . The first observation which can be made is that, contrary to  $\gamma$ ,  $R^2$  is much less impacted by errors on  $RH$ . This can be witnessed with the range on which the determination coefficient varies with  $\Delta_{RH}$ , from  $R^2 = 0.93$  at  $\Delta_{RH} = -0.1$ , to  $R^2 = 0.95$  at  $\Delta_{RH} = 0.1$ . This means that the main conclusions drawn from the analyses of  $R^2$  in the former parts are robust to potential errors on  $RH$ . The benefit of fluorescence to study aerosol hygroscopic properties is therefore supported by this observation. Moreover, it is interesting to note that in this case,  $R^2$  increases with  $RH$ , which means that when  $RH$  increases, the fit to the Hänel

parameterization of the situation is improved. This fact is difficult to interpret but could also support that  $RH$  is currently underestimated in this case. Eventually, it is interesting to note that the value of  $\gamma$  estimated for the maximum  $R^2$  value is equal to  $\gamma = 0.49 \pm 0.01$ , which corresponds to an enhancement factor of  $f(85\%) = 1.98 \pm 0.03$ , falling in the range of the expected value for urban aerosols. This aspect also suggests that  $RH$  is probably underestimated in this case.

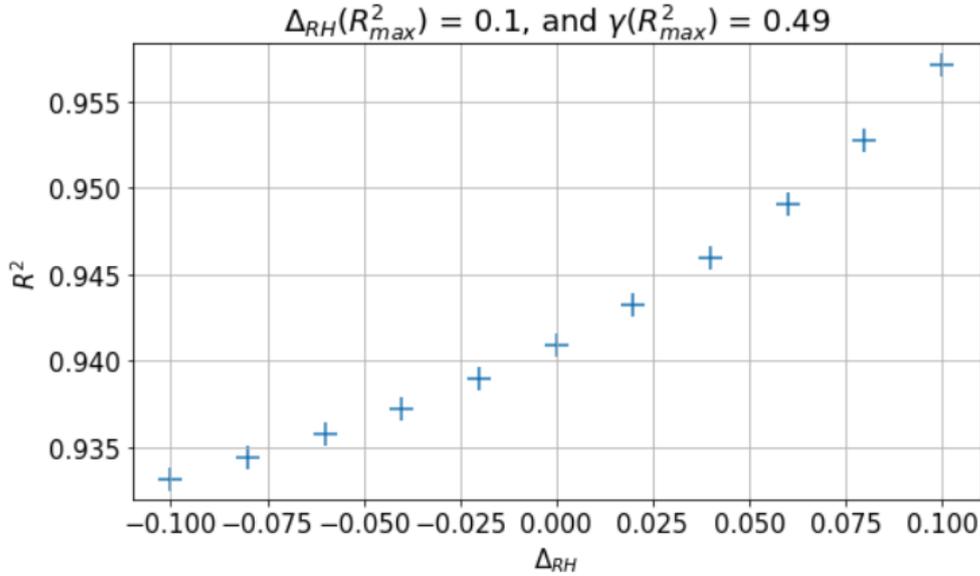


Figure 4.15: Variation of the determination coefficient of the fit to the Hänel parameterization in function of the adjustment on the humidity ( $RH' = RH + \Delta_{RH}$ ) for the 16 April 2021 between 01:00 and 02:00 UTC between 1000 m and 1450 m

#### 4.5.2 Modeling $\gamma$ variations with $RH$ and approximation of the slope

The objective of this part is to estimate the slope of the Hänel hygroscopic constant variation with the adjustment of  $RH$  highlighted in the previous part. The approximation of this slope would allow to quantify the dependency of  $\gamma$  to the uncertainty on  $RH$ .

The Hänel hygroscopic growth constant is estimated using equation (4.7). When the adjustment on  $RH$  is considered, this equation changes from:

$$\frac{\beta(RH)}{\beta(RH_{min})} = \left( \frac{1 - RH}{1 - RH_{min}} \right)^{-\gamma_0} \quad (4.10)$$

Where  $\gamma_0$  is the Hänel hygroscopic constant estimated when  $\Delta_{RH} = 0$ , to:

$$\frac{\beta'(RH + \Delta_{RH})}{\beta'(RH_{min} + \Delta_{RH})} = \left( \frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})} \right)^{-\gamma(\Delta_{RH})} \quad (4.11)$$

Where  $\beta'(RH + \Delta_{RH}) = \beta(RH)$  as the values of the elastic backscatter coefficient are not changed with the adjustment of  $RH$ , and  $\gamma(\Delta_{RH})$  is the Hänel hygroscopic constant, as a function of  $\Delta_{RH}$ .

$$\frac{\beta(RH)}{\beta(RH_{min})} = \left( \frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})} \right)^{-\gamma(\Delta_{RH})} \quad (4.12)$$

Therefore, from equation (4.10), we have:

$$\left( \frac{1 - RH}{1 - RH_{min}} \right)^{-\gamma_0} = \left( \frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})} \right)^{-\gamma(\Delta_{RH})} \quad (4.13)$$

$$\gamma(\Delta_{RH}) = \gamma_0 \frac{\ln\left(\frac{1 - RH}{1 - RH_{min}}\right)}{\ln\left(\frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})}\right)} \quad (4.14)$$

Now from the equation (4.14), in order to express  $\gamma(\Delta_{RH})$  in function of  $\Delta_{RH}$ , it is necessary to simplify the ratio between the two logarithm functions, which is not straightforward.

Figure 4.16 shows how the two logarithm functions correlate depending on  $\Delta_{RH}$ . It can be observed that the relationship between the two functions appears mostly linear, and that the slope of this correlation varies with the value of  $\Delta_{RH}$ . It is also possible to note that the fit appears to be the best when  $RH$  is close to the  $RH_{min}$  value, and that the functions tend to deviate from the linear fit as  $RH$  increases.

In order to approximate the value of the slope in function of  $\Delta_{RH}$ , it is therefore possible to study the limit of the logarithm functions when  $RH$  is close to  $RH_{min}$ . To do so, it is possible to express  $RH = RH_{min} + \epsilon$ , with  $\epsilon \ll RH_{min}$ . Therefore:

$$\lim_{RH \rightarrow RH_{min}} \ln\left(\frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})}\right) = \lim_{\epsilon \rightarrow 0} \ln\left(1 - \frac{\epsilon}{1 - (RH_{min} + \Delta_{RH})}\right) \quad (4.15)$$

By expressing the first-order Taylor expansion of the logarithm function, when considering that  $\epsilon \ll 1 - (RH_{min} + \Delta_{RH})$ , it is then possible to write:

$$\lim_{RH \rightarrow RH_{min}} \ln\left(\frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})}\right) \underset{\epsilon \rightarrow 0}{=} \frac{-\epsilon}{1 - (RH_{min} + \Delta_{RH})} + o(\epsilon) \quad (4.16)$$

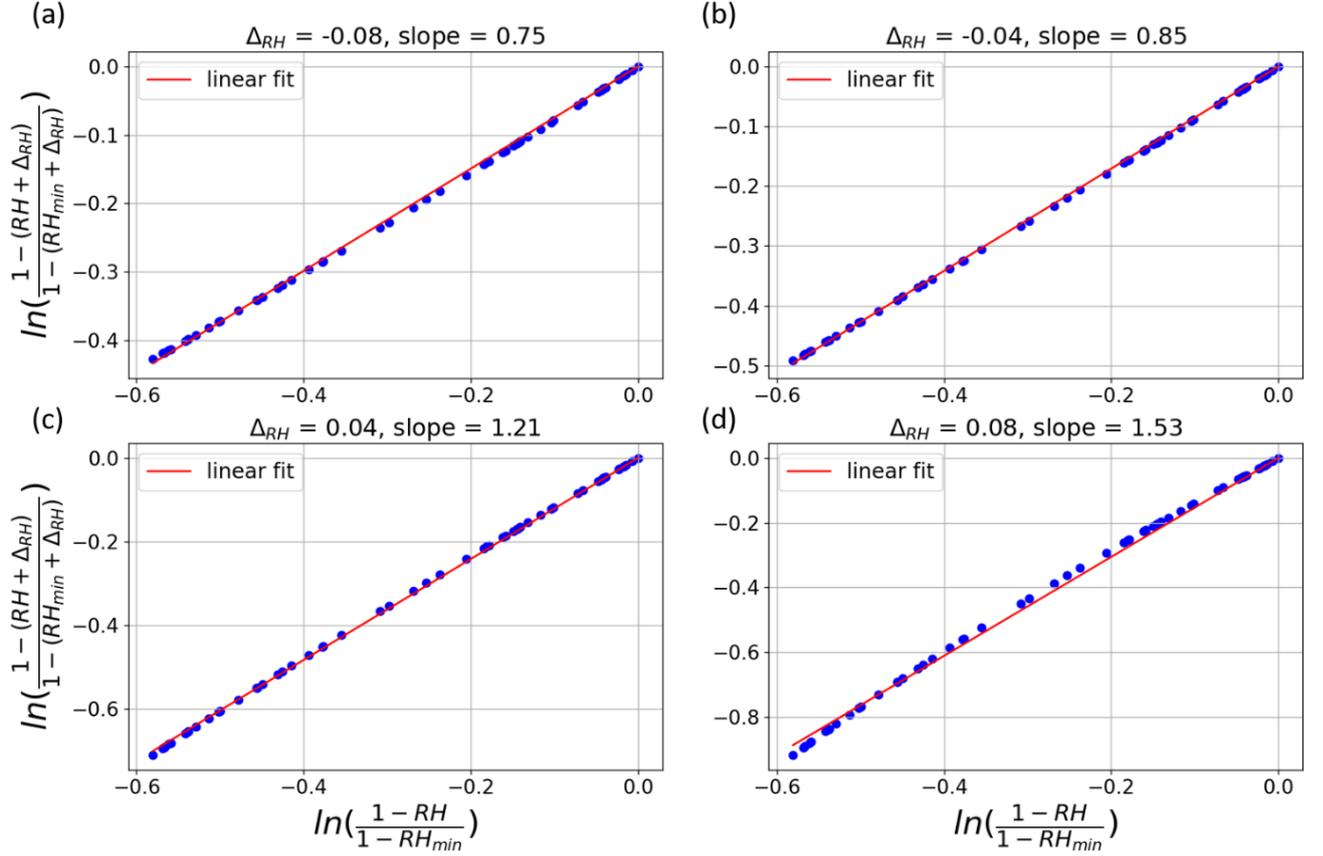


Figure 4.16:  $\ln\left(\frac{1-(RH+\Delta_{RH})}{1-(RH_{min}+\Delta_{RH})}\right)$  in function of  $\ln\left(\frac{1-RH}{1-RH_{min}}\right)$  for different values of  $\Delta_{RH}$  (a)  $\Delta_{RH} = -0.08$ , (b)  $\Delta_{RH} = -0.04$ , (c)  $\Delta_{RH} = 0.04$ , (d)  $\Delta_{RH} = 0.08$ , the linear fit and the values of the slope

On the other hand, by expressing the second-order Taylor expansion of the logarithm function, we can express the limit of the other logarithm function:

$$\lim_{RH \rightarrow RH_{min}} \ln\left(\frac{1-RH}{1-RH_{min}}\right) \underset{\epsilon \rightarrow 0}{=} \frac{-\epsilon}{1-RH_{min}} - \frac{1}{2} \left(\frac{\epsilon}{1-RH_{min}}\right)^2 + o(\epsilon^2) \quad (4.17)$$

By taking the inverse of equation (4.17) and using the first-order Taylor expansion of the inverse function it is possible to write:

$$\lim_{RH \rightarrow RH_{min}} \left(\ln\left(\frac{1-RH}{1-RH_{min}}\right)\right)^{-1} \underset{\epsilon \rightarrow 0}{=} \frac{1-RH_{min}}{-\epsilon} \frac{1}{1 + \frac{\epsilon}{2(1-RH_{min})} + o(\epsilon)} \quad (4.18)$$

Leading to:

$$\lim_{RH \rightarrow RH_{min}} \left(\ln\left(\frac{1-RH}{1-RH_{min}}\right)\right)^{-1} \underset{\epsilon \rightarrow 0}{=} \frac{1-RH_{min}}{-\epsilon} \left(1 - \frac{\epsilon}{2(1-RH_{min})} + o(\epsilon)\right) \quad (4.19)$$

$$\lim_{RH \rightarrow RH_{min}} \left( \ln \left( \frac{1 - RH}{1 - RH_{min}} \right) \right)^{-1} \underset{\epsilon \rightarrow 0}{=} \frac{1 - RH_{min}}{-\epsilon} + \frac{1}{2} + o(1) \quad (4.20)$$

Therefore, when multiplying equation (4.16) with equation (4.20), it is possible to approximate the limits as:

$$\lim_{RH \rightarrow RH_{min}} \ln \left( \frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})} \right) \left( \ln \left( \frac{1 - RH}{1 - RH_{min}} \right) \right)^{-1} \approx \frac{1 - RH_{min}}{1 - (RH_{min} + \Delta_{RH})} \quad (4.21)$$

Equation (4.21) allows us to approximate the slope of the linear fit between the two linear functions presented in Figure 4.16. Figure 4.17 shows how this approximation compares to the real fit between the two functions. It is possible to observe that the approximation of the slope is accurate, especially for smaller values of  $\Delta_{RH}$ , and for  $RH$  values close to  $RH_{min}$ . It can therefore be used in equation (4.14) to express how  $\gamma(\Delta_{RH})$  evolves with  $\Delta_{RH}$ .

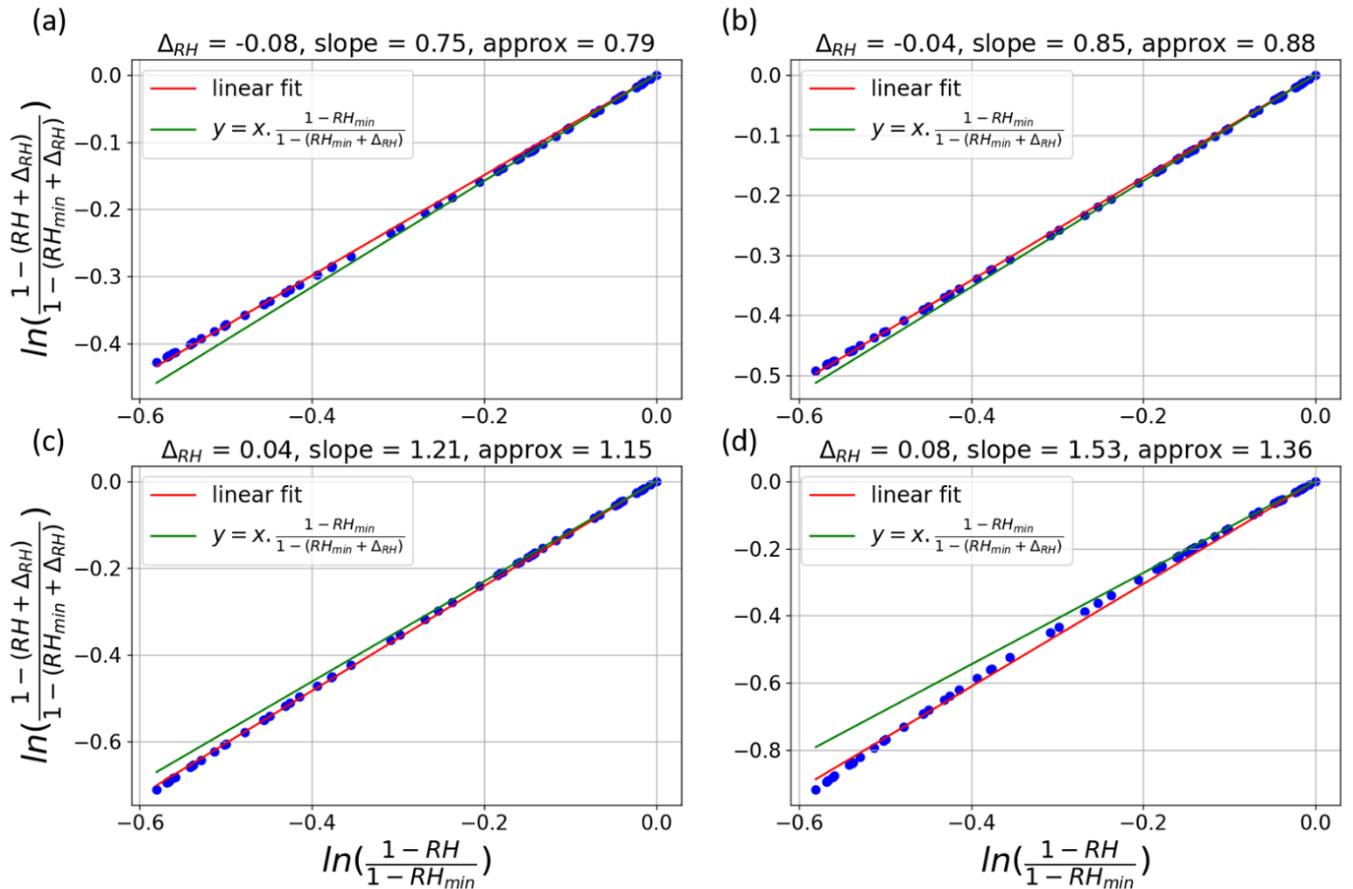


Figure 4.17:  $\ln\left(\frac{1 - (RH + \Delta_{RH})}{1 - (RH_{min} + \Delta_{RH})}\right)$  in function of  $\ln\left(\frac{1 - RH}{1 - RH_{min}}\right)$  for different values of  $\Delta_{RH}$  (a)  $\Delta_{RH} = -0.08$ , (b)  $\Delta_{RH} = -0.04$ , (c)  $\Delta_{RH} = 0.04$ , (d)  $\Delta_{RH} = 0.08$ , with the linear fit, and the approximation of the linear relationship expressed equation (4.21)

By combining equation (4.14) and (4.21) it is possible to write:

$$\gamma(\Delta_{RH}) \approx \gamma_0 \left( 1 - \frac{\Delta_{RH}}{1 - RH_{min}} \right) \quad (4.22)$$

Therefore, it is possible to approximate the slope of the variation of  $\gamma(\Delta_{RH})$  with  $\Delta_{RH}$  as  $\gamma_0/(1 - RH_{min})$ . Figure 4.18 shows how the linear fit of this variation compares to this approximation for the case of 16 April 2021. We can see that the comparison is not perfect, with a difference of 0.81 between the slope and the approximation. This difference can be explained by the fact that the Taylor expansion has been performed at  $RH \rightarrow RH_{min}$ , and also the fact that only the first order has been expressed. Nevertheless, this estimation remains an accurate approximation of the slope in the first order. This approximation allows us to identify that a relevant criterion to quantify the dependency of  $\gamma$  to errors on  $RH$  is the value of  $RH_{min}$ . Indeed, the slope, in first order approximation, varies with the inverse of  $1 - RH_{min}$ , meaning that as  $RH_{min}$  gets closer to 1, the estimation of  $\gamma$  becomes more sensitive to errors on  $RH$ .

To conclude, this study reveals that in order to reduce the dependency of the Hänel hygroscopic constant estimation to errors in  $RH$ , it is important to consider cases where  $RH_{min}$  is as low as possible. Otherwise, the estimates of  $\gamma$  are likely to be more sensitive to errors on  $RH$  and determining accurately the aerosol hygroscopic properties is then expected to be more difficult.

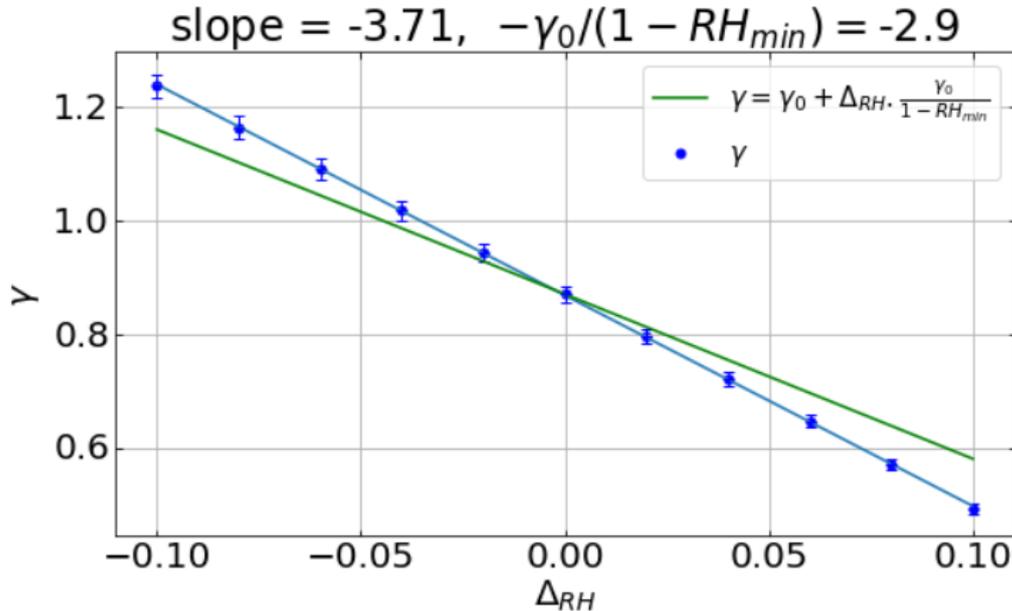


Figure 4.18: Variation of the Hänel hygroscopic constant  $\gamma$  at 532 nm, in function of the adjustment on the humidity ( $RH' = RH + \Delta_{RH}$ ) for the 16 April 2021 between 01:00 and 02:00 UTC between 1000 m and 1450 m, and the approximation of the linear fit to this variation

## 4.6 Chapter summary

This chapter explores aerosol hygroscopic growth using LILAS data. It presents a method for estimating aerosol hygroscopic properties, which relies on fluorescence measurements to account for aerosol concentration changes within a layer, thereby improving the accuracy of these estimates. The method has been applied to three case studies: one involving smoke and two involving urban aerosols. These cases demonstrate the advantages of using fluorescence to correct for aerosol concentration variations in the layer. The method has then been applied to automatically identified cases using LILAS data and FLARE-GMM aerosol typing. The enhancement factor for urban aerosols in Lille was found to be  $f(85\%) = 2.3 \pm 0.2$ , while for biomass burning aerosols, it was  $f(85\%) = 2.1 \pm 0.3$ . However, the high standard deviation in these estimates indicates significant uncertainty, likely due to errors in  $RH$  estimation. To address this, a theoretical analysis has been conducted to examine how uncertainty in  $RH$  affects the enhancement factor estimate. A model has been developed to show that the uncertainty in the Hanel hygroscopic growth constant  $\gamma$  depends on the minimum  $RH$  value  $RH_{min}$  in the layer, suggesting that cases with lower  $RH_{min}$  values should be prioritized to improve the accuracy of  $\gamma$  estimates.

One of the primary limitations of this work is the estimation of  $RH$ , which is based on comparisons between lidar and radiometer data. The uncertainty in  $RH$  is likely the main source of errors in the  $\gamma$  estimates. To improve  $RH$  accuracy, it would be beneficial to apply this method at a site equipped with both fluorescence measurements and radiosoundings. With more precise  $\gamma$  estimates and a large dataset, it would be possible to explore correlations between hygroscopic properties and other aerosol characteristics, such as  $PLDR$  or fluorescence capacity. These relationships could offer valuable insights for modeling interactions between aerosols and water vapor, laying the groundwork for studying aerosol-cloud interactions (Dusek et al., 2006; Petters and Kreidenweis, 2007).

Finally, the study of hygroscopic growth is expected to be adapted and refined for the upcoming LIFE lidar (Laser Induced Fluorescence Explorer). This new lidar system will feature increased power and additional fluorescence channels, providing more comprehensive data and significantly enhancing the accuracy and performance of retrievals.

# 5. Fluorescence measurements in cloud layers

This chapter explores the potential applications of lidar fluorescence measurements within cloud layers. We first give an overview of how lidar systems have been traditionally used in cloud studies. Following this, the chapter introduces fluorescence measurements specifically conducted within cloud layers, demonstrating LILAS capability to detect fluorescence in thin liquid cloud layers. The chapter then provides interpretations of fluorescence behavior within these clouds, offering various hypotheses that explore phenomena likely to impact fluorescence signals in cloudy conditions. Finally, the chapter introduces a method for determining the elastic backscatter coefficient of aerosols within clouds, presenting a valuable approach to study aerosols in cloud layers without the influence of cloud droplets.

## 5.1 Generalities on cloud and aerosol-cloud interaction studies with lidars

Clouds can be studied remotely with a variety of instruments, including lidars. Lidars offer multiple opportunities as they are able to measure cloud properties such as their base and top altitudes, shape, or extinction coefficient and optical depth (Ansmann et al., 1992; Cadet et al., 2005; Wu et al., 2009; Young, 1995). Lidars are therefore a valuable instrument to characterize clouds and study their properties as well as radiative impact, however it suffers from strong limitations. Indeed, many important properties, such as microphysical parameters, or temperature, are difficult to estimate with classic Mie-Raman lidars (Su and McCormick, 2019). For this reason, synergies between lidars and radars are widely used for cloud studies, as radars are able to penetrate the cloud and estimate cloud microphysical properties, such as particle size and mass content, thus completing lidar measurements (Intrieri et al., 1993, 2002; Tinel et al., 2005; Bühl et al., 2013). However, at the LOA, only the lidar is available as no radar is currently operating on the ATOLL platform. This study therefore uses only lidar data to study clouds, aerosols, and their interactions.

The main challenges regarding studying clouds with lidars concerns their optical thickness. In most situations, especially for low liquid clouds, cloud layers are too thick optically for the laser to penetrate through it, making it therefore impossible for the instrument to get any information. For this reason, cloud studies with lidars are mostly dedicated to cirrus, which

are ice clouds optically thinner than low liquid clouds, allowing the laser to penetrate and study their properties. Another challenge regarding studying clouds with lidars concerns their identification. Indeed, accurately distinguishing between clouds and aerosols requires sophisticated algorithms and often supplementary data from other instruments to ensure precise classification, this is especially the case for ice clouds as illustrated in chapter 3, as their backscatter coefficients can reach levels similar to aerosol layers (Wang et al., 2022).

Eventually, the study of aerosol-cloud interactions with lidars is extremely difficult. Indeed, while Mie-Raman lidars are valuable instruments for the observation of both aerosols and clouds, the optical thickness of the latter makes it impossible to observe aerosol-cloud co-existence, the aerosol backscatter being masked by the presence of the cloud. For this reason, most aerosol-cloud interaction studies with remote sensing techniques rely on observations of aerosols in the vicinity of the cloud (J. Schmidt et al. 2014,2015; Su and Patrick McCormick 2019; Wang et al. 2022). While this method allows us to study aerosol-cloud interactions, the lack of remote sensing collocated observations of aerosols and clouds is a high source of uncertainty, and a strong limitation to our understanding of these interactions.

## **5.2 Fluorescence signals in clouds**

### **5.2.1 Observations with LILAS**

As previously mentioned in this work, the main particularity of the LILAS instrument is its ability to measure fluorescence signals. According to our observations, these signals are independent from the presence of water in most subsaturated situations, allowing us to gather information on aerosols in humid conditions. Given this feature, possibility arises of measuring aerosol fluorescence signals within cloud layers to gather valuable information about the particles present there. This raises intriguing questions about the instrument potential applications and the exploitation of fluorescence signals in cloud studies. Fluorescence signals have been measured in cloudy conditions, and three cases are going to be presented, one concerning a cloud located in a smoke aerosol layer, and two showing a cloud present in an urban aerosol layer. These cases have been chosen as they exhibit situations in which an optically thin cloud has been observed inside of an aerosol layer, which are ideal conditions for the study of fluorescence signals in cloud layers with LILAS.

#### **a) Smoke particles in clouds: 17 April 2021 case**

The first presented situation corresponds to the night between 17 and 18 April 2021. During this night, a smoke aerosol layer was present in the Lille atmosphere above the boundary layer, spanning from 2 km to 6 km, as determined by FLARE-GMM, shown Figure 5.1 (b). The boundary layer is mainly composed of urban aerosols up to 1.5 km during the night, with the

presence of smoke aerosols around 1:00 UTC, suggesting that urban aerosols in this layer might be mixed with smoke. Figure 5.1 also shows the presence of ice clouds after 1:00 UTC between 6 km and 8 km. They can be identified with their high depolarization, low optical thickness and large spatial coverage. Eventually, a low liquid cloud can be observed on this figure around 2.2 km between 0:30 and 2:00 UTC, inside of the smoke layer.

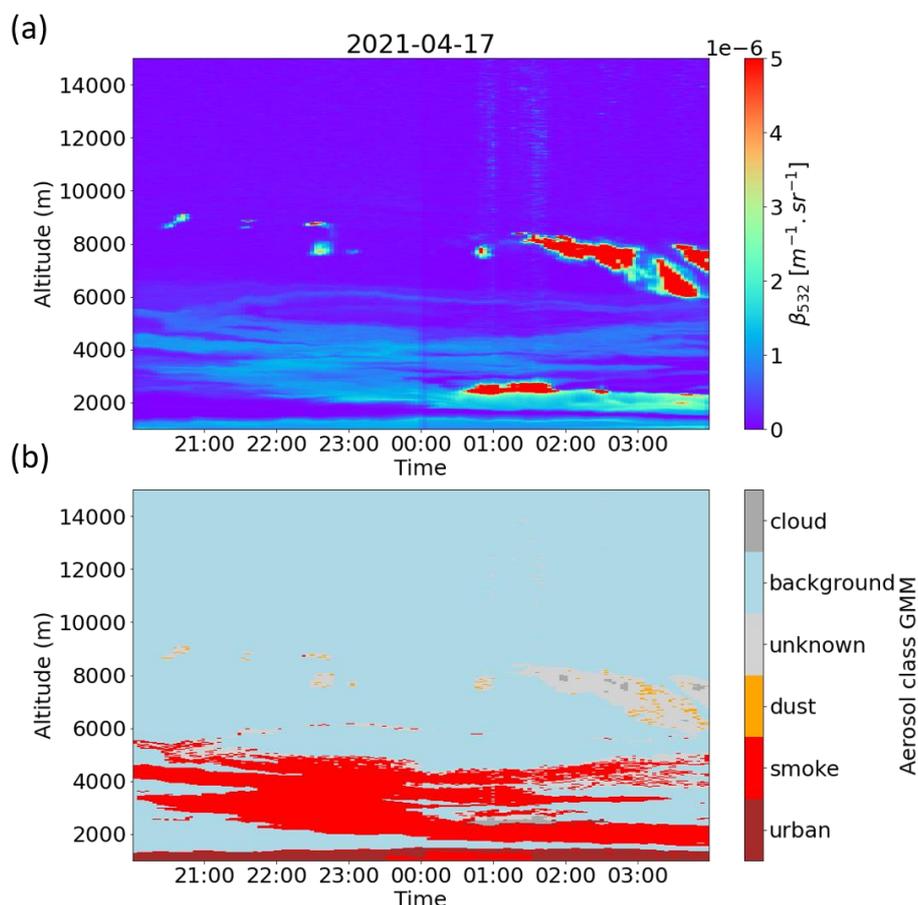


Figure 5.1: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $m^{-1}.sr^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 17 and 18 April 2021

The particularity of the low liquid cloud present in this situation is its low optical thickness, which can be observed on Figure 5.1 (a) as no shadow effect is visible, meaning that the lidar signal above the cloud is not impacted by its presence. Furthermore, this situation is interesting as the cloud layer is located inside an aerosol layer. This suggests that aerosols are expected to be present in the cloud as well, making it an ideal situation for the study of fluorescence signals in cloud layers.

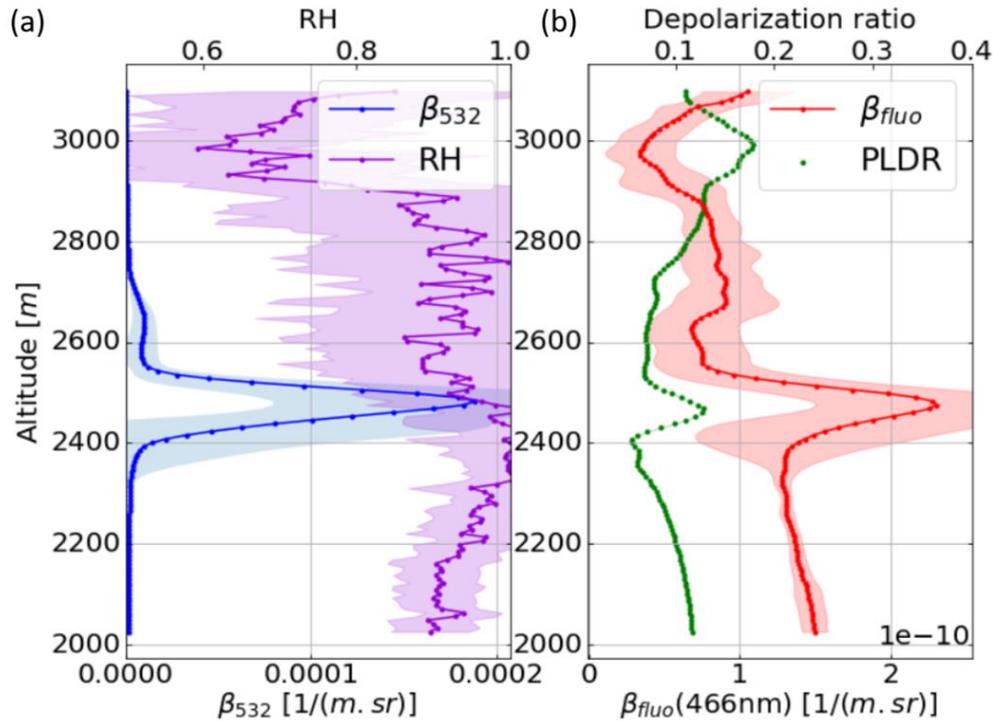


Figure 5.2: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [ $m^{-1} \cdot sr^{-1}$ ] and *RH*, (b) fluorescence backscatter coefficient at 466 nm [ $m^{-1} \cdot sr^{-1}$ ] and particular linear depolarization ratio at 532 nm on 18 April 2021 from 1:00 to 2:00 UTC

Figure 5.2 shows the profiles of the different retrieved products from LILAS between 2 km and 3 km, averaged from 1:00 to 2:00 UTC on 18 April 2021. Figure 5.2 (a) shows that the cloud is located between 2300 m and 2750 m above ground level approximately. Figure 5.2 (b) shows the fluorescence backscatter coefficient, as well as the depolarization ratio at 532 nm measured by LILAS. The measured fluorescence signal indicates the presence of fluorescing aerosols in the cloud, which is what has been expected since the cloud is located within a smoke aerosol layer. Interestingly, there is an increase in the fluorescence backscatter coefficient within the cloud region, and we can observe that both maximum of the cloud elastic backscatter coefficient and of the aerosol fluorescence backscatter coefficient coincide in altitude. The interpretation of this increase and its implications will be discussed in the next section, providing a deeper understanding of the observed phenomenon. We can also observe an increase of the depolarization, with similar characteristics to the fluorescence increase, in the cloud layer. This depolarization increase could suggest the presence of ice in the cloud. However, considering its low altitude, the presence of ice crystals is highly doubtful. Instead, another explanation of the depolarization increase could be a detector saturation. Since the backscattered light is much more intense in cloud regions, the detection system, designed primarily for cloud-free conditions, might become saturated. Given that liquid cloud depolarization is typically low due to the spherical shape of water droplets, most of the

backscattered light has a polarization parallel to that of the emitted light. Consequently, the detection channel measuring light polarized parallel to the laser is more prone to saturation than the channel for cross-polarized light, leading to an artificial increase in depolarization. Eventually, the increase of depolarization in the cloud could also be a consequence of multiple scattering. Even though the impact of multiple scattering is difficult to assess, it can be responsible for the increase of depolarization measured by the lidar, due to scattering at other scattering angles (Sassen and Petrilla, 1986; Hu et al., 2001, 2006; Reichardt and Reichardt, 2003).

### **b) Urban particles in clouds: 2 July 2021 case**

The second presented case occurred during the night between 2 and 3 July 2021. Figure 5.3 shows the quick-looks of elastic backscatter coefficient and FLARE-GMM aerosol type estimate for this case. This figure shows the presence of the boundary layer going up to 3 km, composed mainly of urban aerosols, with occasional traces of smoke. Ice clouds can be observed over 8 km above ground level between 21:00 and 23:00 UTC. During the first part of the night, liquid clouds can be observed between 5 and 7 km around 22:00 UTC. These clouds are not collocated with an aerosol layer and are optically thick as suggested by the alteration of the lidar signal above them illustrated Figure 5.3 (a). These clouds are therefore not ideal to study fluorescence signals. During the late part of the night, after 01:00 UTC, low liquid clouds can also be observed in the boundary layer, however, such as the former one, these clouds are too optically thick, which can be observed by the lidar signal above them which is highly deteriorated, making it difficult to use them to study fluorescence signals accurately.

Eventually, between 23:00 and 00:00 UTC, we can observe the presence of much smaller clouds in the boundary layer between 1.5 and 2.5 km. These clouds are much smaller than the previous one, and more importantly much optically thinner, and will therefore be used to study fluorescence signals.

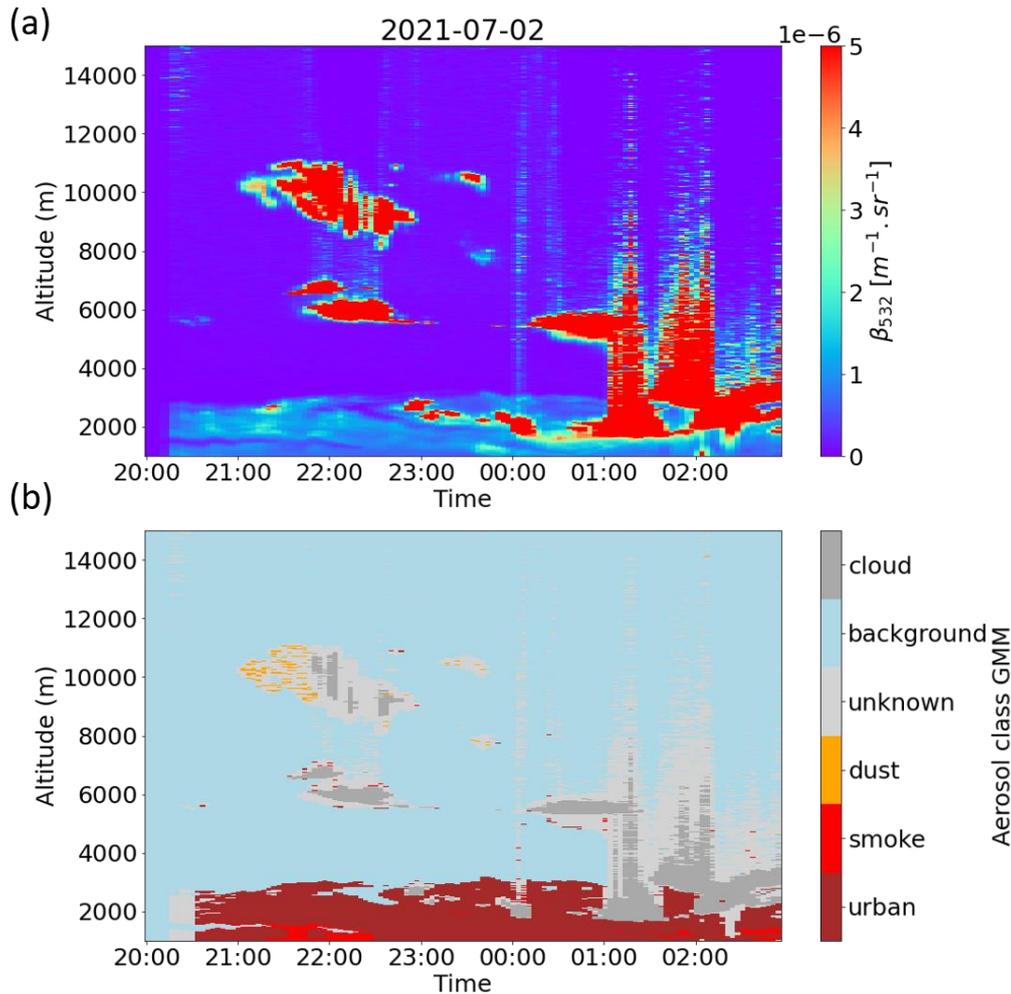


Figure 5.3: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 3:00 UTC, on 2 and 3 July 2021

Unfortunately, unlike the former situation presented on Figure 5.1 and Figure 5.2, these clouds are not located at a constant altitude and do not align with an averaging window of one hour. Therefore, the profiles averaged between 22:00 and 23:00 UTC are not ideal to characterize the situation, but will be used nevertheless as they are still relevant for the fluorescence signal behavior. Therefore, these lidar profiles can be observed on Figure 5.4. This figure shows that the clouds included in this average are located between 2550 m and 2700 m approximately. Figure 5.4 (b) shows that in these altitudes, fluorescence signals are measured, suggesting once again the presence of fluorescing aerosols in the layer. Moreover, the same fluorescence signal behavior as previously can be observed. Even though it is certainly attenuated by the averaging process, an increase of the retrieved fluorescence backscatter coefficient can still be observed Figure 5.4 (b). Moreover, similarly to the situation illustrated on Figure 5.2, the maximum of the cloud elastic backscatter coefficient coincides in altitude with the maximum of the fluorescence backscatter coefficient. It is interesting to note that the same fluorescence backscatter behavior can be observed in this situation as well, even though the aerosol type is

different. This could suggest that this behavior is not due to the nature of the aerosols, but to other effects that are discussed in part 5.2.2.

Increases of depolarization in the cloud regions can also be observed, potentially due to multiple scattering, or indicating a saturation of the detection system.

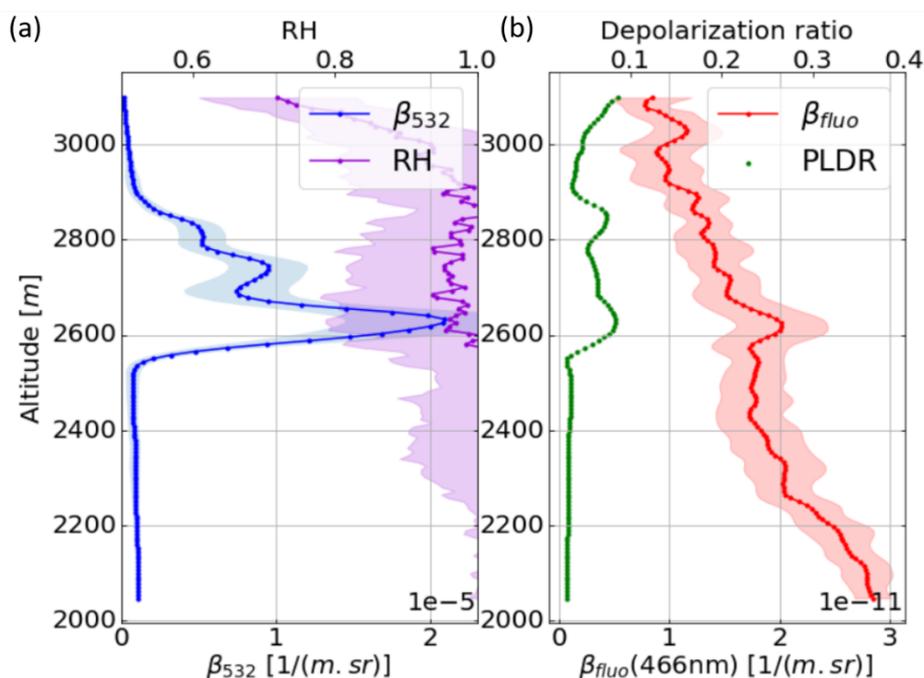


Figure 5.4: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [ $m^{-1} \cdot sr^{-1}$ ] and *RH*, (b) fluorescence backscatter coefficient at 466 nm [ $m^{-1} \cdot sr^{-1}$ ] and particular linear depolarization ratio at 532 nm on 2 July 2021 from 22:00 to 23:00 UTC

### c) Urban particles in clouds: 29 July 2021 case

Eventually, the last situation studied in this section occurred during the night between 29 and 30 July 2021. Figure 5.5 shows the quick-looks of the elastic backscatter coefficient and FLARE-GMM aerosol type estimate for this case. This figure shows the presence of a boundary layer composed of urban aerosols going up to 2 km during the whole night. Ice clouds can be observed over 7 km, from 00:00 to the end of the night. They can be identified with their high depolarization, low optical thickness and large spatial coverage. Liquid clouds can also be observed in the free troposphere during the whole night. Most of them are located around 4 km, however, as these clouds are located in areas with low aerosol loads, they are not ideal to study fluorescence signals. However, we can see that liquid clouds form at the top of the boundary layer during this night around 00:00 UTC and 04:00 UTC. The cloud case at 00:00 UTC is optically thin and located in an urban layer, making it an ideal case for the observation of fluorescence signals in clouds.

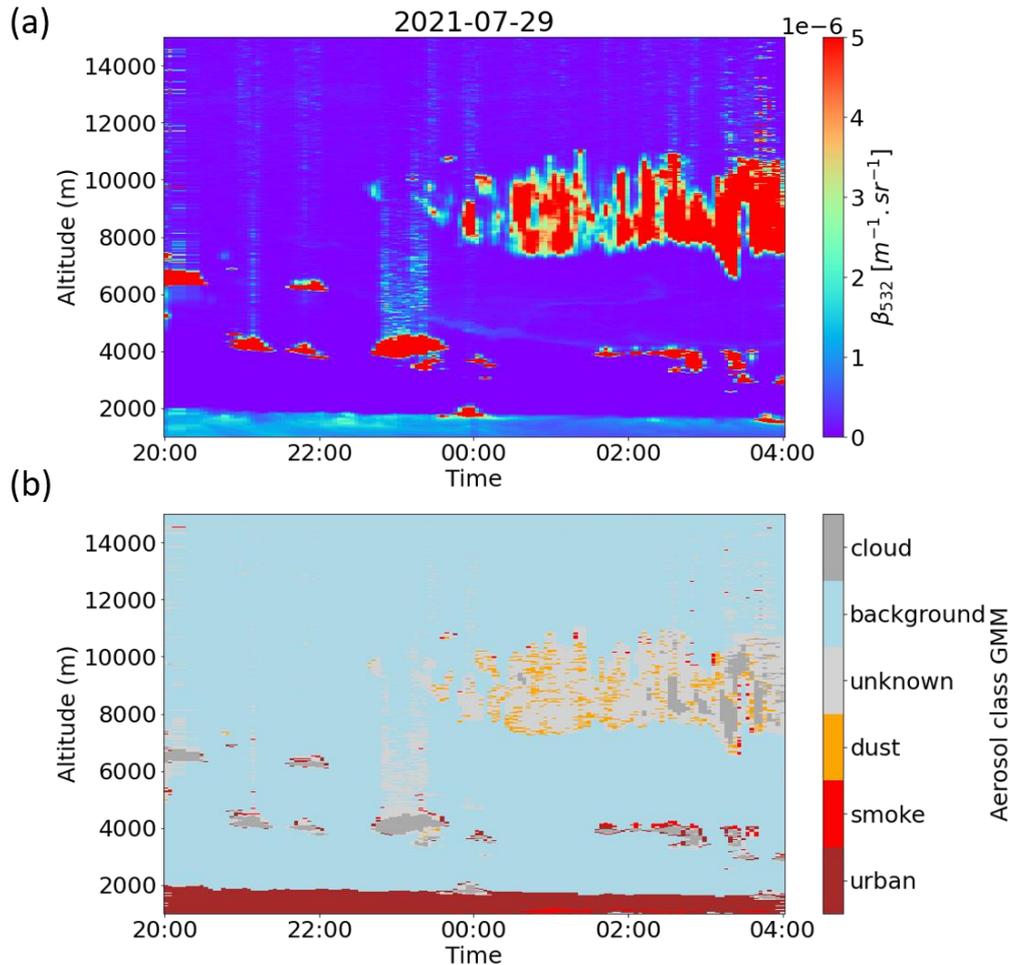


Figure 5.5: (a) aerosol elastic backscatter coefficient quick-look at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and (b) results of the FLARE-GMM typing, between 20:00 UTC and 4:00 UTC, on 29 and 30 July 2021

Similarly to the second presented case, in this situation, the cloud does not match a lidar profile averaging window. However, it is still possible to observe the fluorescence signal and the impact of the cloud on its profile. Figure 5.6 shows the lidar profiles averaged from 23:00 to 00:00 UTC. We can see from Figure 5.6 (a) that the cloud spans from 1700 m to 1850 m approximately, and we can observe that similarly to the previous cases, fluorescence backscatter coefficient is measured and increases inside the cloud. It is also possible to note that the maximum of the cloud elastic backscatter coefficient is at the same altitude as the maximum of fluorescence backscatter coefficient, such as it has been the case for the previous analyzed situations. Similarly to the two other cases, an increase of the depolarization, similar to the fluorescence increase, can be observed. This increase can be potentially attributed to multiple scattering or to detector saturation, such as stated previously.

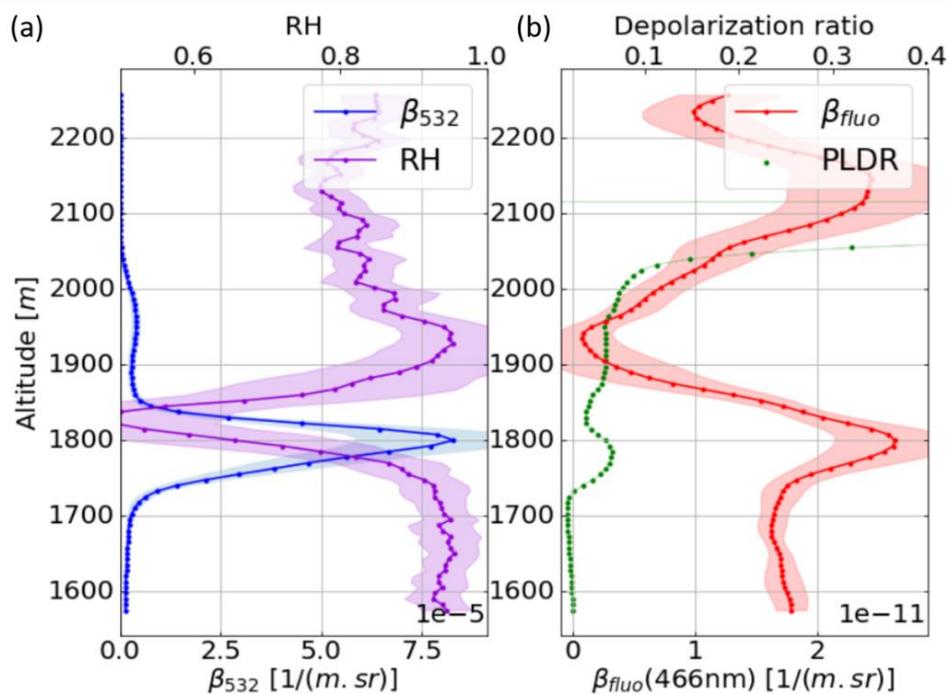


Figure 5.6: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and *RH*, (b) fluorescence backscatter coefficient at 466 nm [ $\text{m}^{-1} \cdot \text{sr}^{-1}$ ] and particular linear depolarization ratio at 532 nm on 29 and 30 July 2021 from 23:00 to 00:00 UTC

### 5.2.2 Fluorescence increase in cloud layers

The three cases analyzed in the former section show that in the presence of clouds inside an aerosol layer, the fluorescence backscatter coefficient measured by LILAS increases. It has also been observed that the maximum of the cloud elastic backscatter coefficient has consistently been located at the same altitude as the fluorescence backscatter coefficient maximum. This suggests a direct connection between this increase and the presence of the cloud. The aim of this section is to present mechanisms susceptible to impact fluorescence that should be considered for the exploitation of these signals. However, determining the most important processes and evaluating the magnitude of their impact remains an extremely complex task that has not been addressed in this study.

The first hypothesis that could explain an increase of the fluorescence signal is an increase of the aerosol concentration, which would be collocated with the presence of the cloud. This concentration increase could be the result of dynamic effects, and would directly be linked to an increase of fluorescence signal as the amount of fluorophore molecules would mechanically increase inside of the cloud. However, this hypothesis seems unlikely to explain the main increase of the fluorescence signal in the cloud layer. Indeed, the fluorescence backscatter coefficient increase is always aligned with the cloud location, and the peaks of both the elastic

and fluorescence backscatter consistently coincide, which would be highly coincidental if this increase were solely due to a rise in aerosol concentration.

Another hypothesis for the alteration of the fluorescence backscatter coefficient within the cloud is the occurrence of chemical reactions that might change the nature of the fluorophores present in the aerosols. Depending on the aerosol type, it may be soluble in water and dissolve within cloud droplets, potentially initiating chemical reactions. The fluorophore chemical systems are extremely complex, and their behavior in aqueous solvents is not well understood. Estimating the products of these reactions, as well as their kinetic and thermodynamic properties, is challenging. Therefore, it is difficult to determine whether chemical reactions involving fluorophores are likely to occur within the cloud and, if so, their impact on the observed fluorescence. However, it can be anticipated that a chemical change in fluorophores would potentially result in a decrease in fluorescence, as the emitting molecule is transformed, while an increase is observed in our case. For this reason, we can assume that even if such reactions are possible, their impact on the fluorescence signal measured by the lidar is probably not the main explanation for the increase of the fluorescence signals in cloud layers (Agostiano et al., 2002; Fredj and Ruiz-López, 2010; Miyatake and Tamiaki, 2010; Wu et al., 2021).

Multiple scattering could also be an explanation for the fluorescence signal increase in the cloud. Indeed, in clouds, the single scattering approximation, on which the lidar equations depend, no longer applies due to the close proximity of scattering particles and their strong scattering power. Therefore, depending on the cloud droplet concentration, and the incident light properties, multiple scattering can strongly impact the cloud elastic backscatter coefficient estimation (Bissonnette, 1996; Bissonnette et al., 1995; Shcherbakov et al., 2022). Therefore, one might argue the fact that multiple scattering also affects fluorescence measurements with lidar. However, former studies have shown that Raman scattering, which is several orders of magnitude less efficient than elastic scattering, is not impacted by multiple scattering in clouds (Wandinger, 1998; Rizi et al., 2004). Considering that fluorescence is 4 to 5 orders of magnitude less efficient than elastic backscatter, the impact of multiple scattering could be considered as negligible. However, since this process depends on the light wavelength and is extremely complex, assessing the impact of multiple scattering on fluorescence measurements remains difficult.

Other phenomena that could influence fluorescence signals include optical effects caused by the presence of water droplets surrounding aerosols, especially if the aerosols are not dissolved. One of them is the change of the aerosol phase function due to the presence of water. This phenomenon has been studied by Griaznov et al. (2002) and Veselovskii et al. (2002), who have shown that the phase function in the backward direction can be impacted by the presence of the water droplet, resulting in an increase of the Raman signal by a factor of 2. As

this increase is due to an optical effect, it can be estimated that fluorescence signals should be impacted in the same order of magnitude as the Raman signals. However, Figure 5.2 (a), Figure 5.4 (a) and Figure 5.6 (a) show that the estimated  $RH$  in the cloud layer is not overestimated. Except for the last case in which  $RH$  strongly decreases inside of the cloud, the other cases show  $RH$  close to 100%, which is expected inside clouds. In the case of an increase of the Raman signal by optical effects due to the water droplet, the retrieved  $RH$  should then be over 100% inside of the cloud, which is not the case here. Compensation between different phenomenon could occur, but this observation supports the fact that in these cases, these optical processes could be neglected.

Eventually, the last aspect presented in this work which could explain the increase of the fluorescence signal inside of cloud layers is possible leakage of elastic signal inside of the fluorescence channel. In spite of the presence of dichroic mirrors and interference filters in the reception section of the instrument, some photons from elastic backscatter might still reach the detector of the fluorescence channel. In clear sky conditions, this leakage is negligible, however in clouds, as the intensity of the elastic backscatter light increases, so does the contamination by elastic photons, therefore impacting the fluorescence signal. The different presented cases have shown that the maximum of the cloud backscatter coefficient always coincides with the maximum of fluorescence backscatter coefficient inside of the cloud. Furthermore, the different profiles of elastic backscatter and fluorescence backscatter coefficients appear to be correlated, potentially supporting this hypothesis. However, it is important to remember that this correlation does not necessarily imply that the increase in fluorescence is due to elastic contamination. In fact, in case of multiple scattering, as the cloud thickens, both the elastic backscatter coefficient and the impact of multiple scattering to the fluorescence increase would rise. The observation of a correlation between fluorescence and elastic backscatter coefficients could then be a consequence of this aspect, and not necessarily a confirmation of elastic leakage.

As illustrated in this section, many factors can influence fluorescence signals in clouds, resulting in numerous uncertainties and unknown variables. These complexities make correcting the contribution of the cloud to the fluorescence increase too ambitious in our current situation. Despite these challenges, the next section presents a method that attempts at exploiting the fluorescence signal in clouds. This method offers a glimpse of what could be achieved with fluorescence measurements to retrieve information about aerosols in cloud layers, assuming we can develop a correction technique to mitigate the fluorescence increase caused by the cloud presence. By exploring these methods, we aim to advance our understanding and capabilities in aerosol characterization within cloudy environments.

### 5.3 Use of fluorescence to retrieve aerosol properties in cloud layers

As explained in the previous section, fluorescence signals are impacted by the presence of clouds through numerous interactions. Unfortunately, it is currently impossible to correct these effects, meaning that the possible exploitation of the fluorescence signals in cloud layers remains limited. Nevertheless, this chapter presents an approach that utilizes fluorescence signals within clouds to retrieve information about aerosols present in the cloud layer. This method shows perspectives for future applications, assuming that effective corrections for the cloud-induced effects on fluorescence can be developed. The objective of this approach is to use the information given by the fluorescence signal measured by the lidar in clouds to obtain information on the aerosols present in the cloud layer.

To do so, we can rely on the fact that the fluorescence signal is primarily issued from fluorophore matter contained in the aerosols. In specific occasions, when a cloud located inside of an aerosol layer, such as it is the case for the situation of 17 April 2021, it is expected that aerosols should be present inside of the cloud. Furthermore, it can be assumed that these aerosols are of the same nature as the ones outside the cloud. This assumption allows us to use the fluorescence signal to invert aerosol properties inside of the cloud, such as their backscatter coefficient, or potentially concentration and size distribution. Precisely, we can use the fluorescence capacity of the aerosols in the vicinity of the cloud. As this quantity is intensive and expresses the ratio between fluorescent matter and non-fluorescent matter in the aerosol, we can suppose that in the absence of chemical reactions changing the nature of the fluorophores, the fluorescence capacity of the aerosols outside and inside of the cloud should be the same. Fluorescence capacity being the ratio between the fluorescence backscatter coefficient and the elastic backscatter coefficient, it becomes then possible to retrieve the elastic backscatter coefficient of the aerosols that are present inside of the cloud.

In order to apply this process, it is first necessary to estimate the fluorescence capacity of the aerosols in the vicinity of the cloud. However, as shown previously in this work, in chapter 3 and chapter 4, the presence of humidity strongly impacts the aerosol fluorescence capacity through hygroscopic growth. In order to estimate precisely the aerosol fluorescence capacity, it is therefore necessary to take this aspect into account. To do so, it is possible to use equation (4.7)(4.6) to estimate  $\beta(RH_{ref})$  from  $\gamma$  and  $RH$ . The fluorescence capacity of the aerosols in dry conditions can then be estimated, without the impact of the presence of water vapor. This estimation can then be used to approximate the fluorescence capacity of the dry particles present in the cloud more accurately.

---

This method has been tested on the case occurring during the night between 17 and 18 April 2021. In order to estimate the dry fluorescence capacity of the smoke aerosols present in the vicinity of the cloud, an approximation of the Hänel hygroscopic growth constant needs to be chosen. In this case, it has been approximated to 0.7, which is the expected value for most smoke aerosols at 532 nm (Sicard et al., 2022). Using the  $RH$  measured with LILAS,  $\beta(RH_{ref})$  can then be estimated with equation (4.6) and used to determine the dry fluorescence capacity of the considered aerosols. The value of the fluorescence capacity has been obtained by averaging the profiles between 1850 m and 2200 m above ground level in order to remain below the cloud while being in the smoke layer. The estimated dry fluorescence capacity of the aerosols below the cloud is  $(3.6 \pm 0.6)10^{-4}$ , which is what is expected from dry smoke aerosols. This value has then been used as the approximated one for the aerosols present in the cloud layer. Then, this value, as well as the fluorescence backscatter coefficient, can be used to estimate the elastic backscatter of the dry aerosols present in the cloud.

Figure 5.7 (b) shows the estimated dry aerosol elastic backscatter coefficient using the a priori smoke Hänel hygroscopic growth constant, the  $RH$ , the fluorescence backscatter coefficient, and the fluorescence capacity of the aerosols in the vicinity of the cloud. The method described in this section has been used to estimate the backscatter coefficient of dry aerosols between 2300 m and 2800 m, where the cloud is approximately localized. This case provides an example of the application of this method, however, the results presented should be interpreted with caution. As highlighted previously, a significant increase in fluorescence can be observed where the cloud is located. As discussed in the former section, it is unlikely that this increase can be solely explained by a higher concentration of aerosols within the cloud. Consequently, the value of the elastic backscatter coefficient derived from this approach is subject to numerous biases related to the cloud impact on fluorescence signals. Therefore, the reliability of this estimate is currently very low. Moreover, the important level of uncertainty of this retrieval also mitigates its significance. Nevertheless, this result demonstrates the potential of this approach, which could eventually enable the estimation of aerosol properties such as concentration or size distribution, in cloudy conditions, allowing us to observe aerosol-cloud coexistence. This kind of information could greatly improve our understanding of aerosol-clouds interactions, and further work should be dedicated to the correction of fluorescence increase due to the cloud presence, allowing for the improvement of this method.

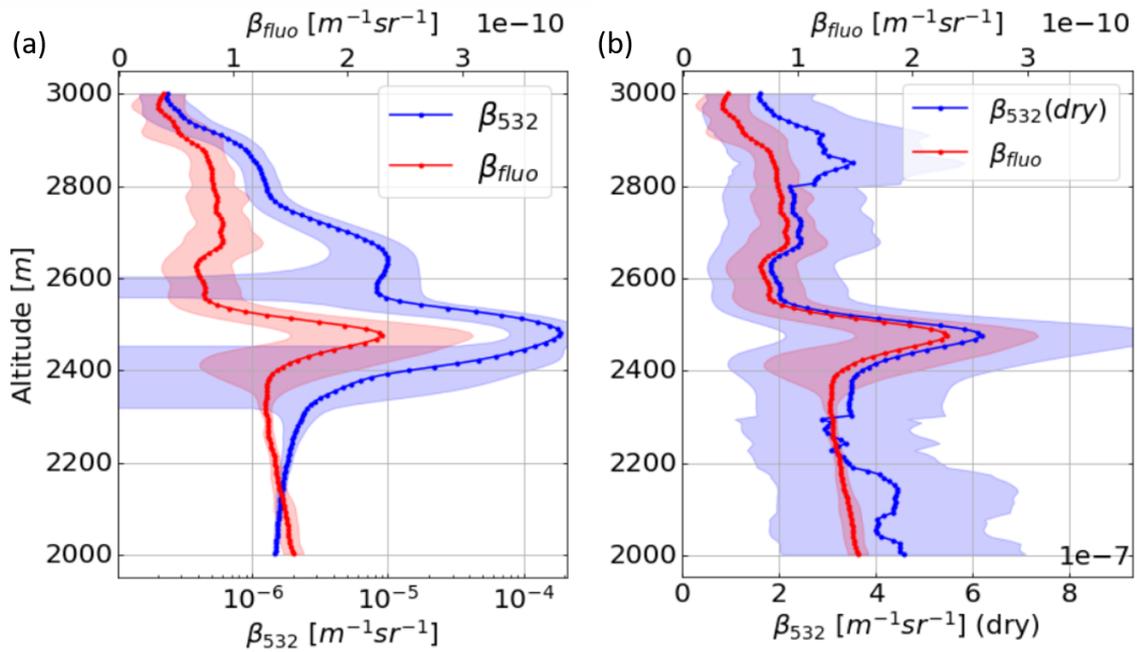


Figure 5.7: Profile of retrieved optical properties in function of altitude above ground level (a) elastic backscatter coefficient at 532 nm [ $m^{-1}.sr^{-1}$ ] and fluorescence backscatter coefficient at 466 nm [ $m^{-1}.sr^{-1}$ ] (b) dry elastic backscatter coefficient of dry aerosol particles in the cloud layer at 532 nm [ $m^{-1}.sr^{-1}$ ] and fluorescence backscatter coefficient at 466 nm [ $m^{-1}.sr^{-1}$ ] on 18 April 2021 from 1:00 to 2:00 UTC

## 5.4 Chapter summary

In this chapter, generalities on cloud studies with lidars have been presented, highlighting the strong limits of this instrument to determine cloud properties, while also pointing out the difficulty of observing collocated aerosols and clouds simultaneously. Following this, examples of fluorescence measurements in thin low liquid clouds have been presented, revealing an increase in the fluorescence backscatter coefficient within cloud layers, which corresponds with a similar increase in the elastic backscatter coefficient. Various interpretations for this increase have been explored, including potential chemical reactions, multiple scattering, or contamination by elastic light. Unfortunately, the lack of knowledge regarding these processes makes it extremely difficult to evaluate their impact, and to correct them. Finally, an approach using the fluorescence backscatter to estimate the elastic backscatter coefficient of aerosols within cloud layers has been presented. This approach also relies on the hygroscopic properties and the fluorescence capacity of the aerosols in the vicinity of the cloud to retrieve information on dry aerosols present in the cloud layer. Unfortunately, the cases presented exhibit high levels of uncertainty, and in the absence of correction for the fluorescence increase due to cloud effects, the exploitation of this method remains extremely limited. Moreover, unfortunately no reference is available to proceed to a validation of this approach.

Further development of this method is crucial, as it enables the simultaneous observation of aerosols and clouds. Accurately estimating the elastic backscatter coefficient within clouds could be used to determine aerosol concentration or size distribution, key aerosol properties that are essential for advancing studies on aerosol-cloud interactions.

In the future, this method could be enhanced with data from the LIFE lidar, which is anticipated to have more power than LILAS, enabling it to penetrate through thicker clouds and to have more case studies. Moreover, the ability of LIFE to measure fluorescence at different wavelengths is also expected to provide much more detailed information about aerosols within clouds, further enriching our understanding of how fluorescence signals are impacted by the presence of the cloud, and increase the amount of information on aerosols in cloud layers that could be retrieved.

## 6. Conclusion and perspectives

### 6.1 Conclusion

The study has been dedicated to lidar fluorescence measurement and its exploitation in aerosol characterization. The key instrument, the Mie-Raman-fluorescence lidar LILAS, is maintained and operated with a high automation level at the LOA in Lille and operated at ATOLL. The acquired lidar data has been inverted using AUSTRAL, a processing server developed internally, to obtain aerosol information, and the RPG-HATPRO G5 radiometer has been used as reference to calibrate the lidar water vapor measurement.

The fluorescence measurements and high level of automation in LILAS enable automatic aerosol typing using machine learning methods. FLARE-GMM, an aerosol typing model based on a clustering approach developed in this work, used LILAS data to automatically distinguish between dust, smoke, and urban aerosols. To train this model, different datasets have been tested. The final version of the algorithm has been trained on a hand selected dataset, to mitigate the impact of mixtures on the cluster identification. This automatic aerosol typing method has been validated by comparing its results with reference data, either through the use of backward trajectories for isolated cases or by comparison with NATALI, another automatic aerosol typing model, on a larger dataset. These validation results are promising, and FLARE-GMM has been applied to examine the types of aerosols present in the Lille atmosphere from 2021 to 2023. The analysis revealed that urban aerosols are predominant, with smoke also present, both being observed mainly at low altitudes, while dust is observed on rarer occasions. Further work dedicated to the improvement of this model could improve its performance, with a wider training set or more features.

Fluorescence measurements have then been used to characterize aerosol hygroscopic properties. The method benefits from the fact that fluorescence emission is independent from the presence of water vapor, enabling the tracking and the compensation of aerosol concentration changes within layers where humidity varies. This approach allows for more accurate estimates of aerosol hygroscopic properties, provided that the aerosol type remains consistent throughout the layer, which can be confirmed using FLARE-GMM. The method has been tested and compared with the standard approach for characterizing aerosol hygroscopic properties in various case studies, demonstrating the advantages of using fluorescence for more precise estimates. A statistical analysis of hygroscopic properties derived from automatically detected cases has been conducted, revealing that urban aerosols in Lille have

an average enhancement factor of  $f(85\%) = 2.3 \pm 0.2$ , while smoke aerosols exhibit an enhancement factor of  $f(85\%) = 2.1 \pm 0.3$ . While these results are consistent with the literature, they highlight the significance of uncertainty in this method, primarily due to noise in humidity estimation. A theoretical analysis and modeling indicate that the estimate of aerosol hygroscopic properties is influenced by the minimum value of  $RH$  within the aerosol layer. This finding suggests that cases with lower minimum  $RH$  values should be prioritized to achieve more accurate estimates.

Eventually, fluorescence signals have been observed in clouds. LILAS fluorescence signals in thin low liquid cloud layers have been investigated in three situations, demonstrating the ability of the instrument to measure such signals, but also showing that the measured fluorescence signals increase in cloud layers. Interpretations of this increase have been proposed, such as chemical reactions, multiple scattering, or leakage of elastic backscattered light. However, the evaluation of each of these possible contributions to fluorescence increase is extremely difficult. However, accurately quantifying the contribution of each of these interactions is extremely challenging. Therefore, attempting to correct the increase in fluorescence due to the presence of the cloud seems too ambitious at this stage. Nevertheless, we can use the fluorescence backscatter coefficient inside of the cloud to estimate the elastic backscatter coefficient of the aerosols present in it. To do so, the fluorescence capacity and hygroscopic properties of the aerosols in the vicinity of the cloud need to be considered. This approach has been tested on a smoke case from 17 April 2021. This case shows the feasibility of this method to approximate the elastic backscatter coefficient of the dry aerosols present in the cloud. Unfortunately, in the absence of correction for the fluorescence increase due to the cloud presence, the relevance of this inversion remains limited. Moreover, the large uncertainties involved and the strong hypothesis used in this approach also limit its significance. However, this method is a first step to invert aerosol properties in liquid clouds such as concentration or size distribution of the aerosols, which would significantly contribute to the enhancement of aerosol-cloud interactions understanding.

## 6.2 Perspectives

This work is an innovative approach dedicated to the exploitation of lidar fluorescence measurements for aerosol characterization. To this end, it introduces a variety of tools and methods that improve aerosol studies. Further work could be dedicated to develop these methods, with a more complete database, or with instruments with different specificities.

First, regarding the database used in this study, it is based on the modified Raman technique, that allows us to invert profiles in conditions, that are otherwise difficult to invert with the classic Raman inversion method. However, the inversions performed in this study are limited

to only a small number of parameters. Consequently, other aerosol properties such as the elastic backscatter coefficients and *PLDR* at 355 nm and 1064 nm, or extinction, are not accessible in the dataset.

Aerosol characterization could greatly benefit from additional information on aerosols such as the *LR* or the Ångström exponent, that respectively require extinction coefficients and elastic backscatter coefficients at different wavelengths. FLARE-GMM has the advantage of being a clustering model that can be easily updated to incorporate more features. The implementation of features such as *LR* and the Ångström exponent are expected to directly impact the aerosol typing performance, since these are intensive quantities that vary with the aerosol type, and are used in other automatic aerosol typing models such as NATALI.

With the use of more features, deconvolution of aerosol mixture lidar measurements becomes possible. This means that the algorithm could be able to identify the share of each aerosol type composing a mixture, which can be performed automatically with data driven approaches (Thiébaud, 2006; Karl et al., 2023), but which cannot be performed in the current set up, given the low number of features.

Hygroscopic growth study would also benefit from more information on aerosols. In the current work, only the impact of hygroscopic growth on the elastic backscatter coefficient at 532 nm is investigated. In order to complete this study, a similar approach could be used to evaluate the impact of hygroscopic growth on the elastic backscatter coefficients at 355 nm and 1064 nm, or on extinction coefficients.

Finally, the estimation of the aerosol elastic backscatter coefficient in clouds could be extended to estimate it at 355 nm and 1064 nm, providing better insights into aerosols in clouds, and increasing the amount of retrievable information from these layers.

Then, another limitation of this study concerns the fact that it has been acquired only in Lille, which limits the application of the methods presented in this work. More specifically, FLARE-GMM has been trained to distinguish between urban, smoke and dust because of the nature of the aerosols present in the Lille atmosphere. Other aerosol types present in other environments, such as volcanic, or marine aerosols, might be difficult to identify and differentiate from dust or other aerosol types without the use of other features, potentially limiting the application of FLARE-GMM. Investigating the performances of FLARE-GMM in other contexts would therefore be highly beneficial to evaluate the relevance of this approach in a more general and global framework.

Concerning the study of hygroscopic properties, the identification of more cases is essential. With the use of the automatic identification of aerosol cases, and supposing the noise issue from *RH* would be mitigated, it could be interesting to investigate potential correlations

between hygroscopic properties and other optical parameters such as the fluorescence capacity, the *PLDR* or the *LR*. Such complete study would potentially allow us to estimate the hygroscopic properties of an aerosol based on its optical properties, measured by the lidar. This correlation would allow us to estimate the aerosol hygroscopic properties in any ambient situation much more accurately, compared to the current approach in which the estimation of the aerosol hygroscopic properties in ambient situations can be performed only based on the aerosol type.

Eventually, regarding the study of fluorescence signals in clouds, optimal situations for the investigation of these signals are extremely rare, especially for other aerosols than urban. In the current dataset, only one smoke case and one dust case have been identified. The presented approach would benefit from being applied to more cases. Furthermore, evaluating the amplitude of the cloud impact on fluorescence, such as the impact of multiple scattering, would require information on cloud properties, such as droplet concentration or droplet size distribution. In the current situation, these properties are extremely difficult to obtain. In order to perform corrections of fluorescence increase due to the cloud presence, information on the cloud properties is therefore essential.

Eventually, LIFE (Laser Induced Fluorescence Explorer), another Mie-Raman-fluorescence lidar developed at the LOA in the frame of OBS4CLIM, proposed by ACTRIS, is expected to start operating in 2024. This lidar, based on the design of LILAS, will exhibit more power, and more importantly, will be able to measure fluorescence at different wavelengths. Different fluorophores having different fluorescence spectra, measuring fluorescence at different wavelengths will therefore allow us to identify much more precisely the origin of the aerosols observed by the lidar. This information will also help deconvolute mixture cases, allowing us to identify much more accurately the mixing ratio of each aerosol type in a given layer. Eventually, the investigation of fluorescence signals in clouds would be easier with LIFE. The main benefit being its increased laser power, allowing it to penetrate in thicker liquid clouds, permitting the investigation of more diverse situations. The use of different fluorescence wavelengths will also be greatly beneficial, allowing us to have much more complex information about the aerosols in clouds. For example, it could be possible to investigate the occurrence of chemical reactions involving the fluorophores as such reactions are expected to impact the fluorescence spectra. This instrument is therefore expected to give many opportunities regarding the study of fluorescence in clouds.

---

# References

Aalto, P., Hämeri, K., Paatero, P., Kulmala, M., Bellander, T., Berglind, N., Bouso, L., Castaño-Vinyals, G., Sunyer, J., Cattani, G., Marconi, A., Cyrus, J., Klot, S. von, Peters, A., Zetzsche, K., Lanki, T., Pekkanen, J., Nyberg, F., Sjövall, B., and Forastiere, F.: Aerosol Particle Number Concentration Measurements in Five European Cities Using TSI-3022 Condensation Particle Counter over a Three-Year Period during Health Effects of Air Pollution on Susceptible Subpopulations, *J. Air Waste Manag. Assoc.*, 55, 1064–1076, <https://doi.org/10.1080/10473289.2005.10464702>, 2005.

Agostiano, A., Cosma, P., Trotta, M., Monsù-Scolaro, L., and Micali, N.: Chlorophyll *a* Behavior in Aqueous Solvents: Formation of Nanoscale Self-Assembled Complexes, *J. Phys. Chem. B*, 106, 12820–12829, <https://doi.org/10.1021/jp026385k>, 2002.

Albrecht, B. A.: Aerosols, Cloud Microphysics, and Fractional Cloudiness, *Science*, 245, 1227–1230, <https://doi.org/10.1126/science.245.4923.1227>, 1989.

Anderson, T. L. and Ogren, J. A.: Determining Aerosol Radiative Properties Using the TSI 3563 Integrating Nephelometer, *Aerosol Sci. Technol.*, 29, 57–69, <https://doi.org/10.1080/02786829808965551>, 1998.

Andrews, D. G.: *An Introduction to Atmospheric Physics*, Second Edition, 2nd ed., Cambridge University Press, 237 pp., 2012.

World Meteorological Organization and Global Atmosphere Watch: WMO/GAW aerosol measurement procedures: guidelines and recommendations, WMO/TD- No. 1178, GAW Report-No. 153, [https://library.wmo.int/viewer/41221?medianame=wmo-td\\_1178\\_#page=1&viewer=picture&o=bookmark&n=0&q=](https://library.wmo.int/viewer/41221?medianame=wmo-td_1178_#page=1&viewer=picture&o=bookmark&n=0&q=), last access: 16 May 2024.

Ansmann, A., Riebesell, M., Wandinger, U., Weitkamp, C., Voss, E., Lahmann, W., and Michaelis, W.: Combined raman elastic-backscatter LIDAR for vertical profiling of moisture, aerosol extinction, backscatter, and LIDAR ratio, *Appl. Phys. B Photophysics Laser Chem.*, 55, 18–28, <https://doi.org/10.1007/BF00348608>, 1992a.

Ansmann, A., Wandinger, U., Riebesell, M., Weitkamp, C., and Michaelis, W.: Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar, *Appl. Opt.*, 31, 7113, <https://doi.org/10.1364/AO.31.007113>, 1992b.

Ansmann, A., Petzold, A., Kandler, K., Tegen, I., Wendisch, M., Müller, D., Weinzierl, B., Müller, T., and Heintzenberg, J.: Saharan Mineral Dust Experiments SAMUM–1 and SAMUM–2: what have we learned?, *Tellus B Chem. Phys. Meteorol.*, 63, 403, <https://doi.org/10.1111/j.1600-0889.2011.00555.x>, 2011.

Arnott, W. P., Hamasha, K., Moosmüller, H., Sheridan, P. J., and Ogren, J. A.: Towards Aerosol Light-Absorption Measurements with a 7-Wavelength Aethalometer: Evaluation with a Photoacoustic Instrument and 3-Wavelength Nephelometer, *Aerosol Sci. Technol.*, 39, 17–29, <https://doi.org/10.1080/027868290901972>, 2005.

Baars, H., Kanitz, T., Engelmann, R., Althausen, D., Heese, B., Komppula, M., Preißler, J., Tesche, M., Ansmann, A., Wandinger, U., Lim, J.-H., Ahn, J. Y., Stachlewska, I. S., Amiridis,

- 
- V., Marinou, E., Seifert, P., Hofer, J., Skupin, A., Schneider, F., Bohlmann, S., Foth, A., Bley, S., Pfüller, A., Giannakaki, E., Lihavainen, H., Viisanen, Y., Hooda, R. K., Pereira, S. N., Bortoli, D., Wagner, F., Mattis, I., Janicka, L., Markowicz, K. M., Achtert, P., Artaxo, P., Pauliquevis, T., Souza, R. A. F., Sharma, V. P., van Zyl, P. G., Beukes, J. P., Sun, J., Rohwer, E. G., Deng, R., Mamouri, R.-E., and Zamorano, F.: An overview of the first decade of Polly&sup&NET&/sup&: an emerging network of automated Raman-polarization lidars for continuous aerosol profiling, *Atmospheric Chem. Phys.*, 16, 5111–5137, <https://doi.org/10.5194/acp-16-5111-2016>, 2016.
- Baars, H., Ansmann, A., Ohneiser, K., Haarig, M., Engelmann, R., Althausen, D., Hanssen, I., Gausa, M., Pietruczuk, A., Szkop, A., Stachlewska, I. S., Wang, D., Reichardt, J., Skupin, A., Mattis, I., Trickl, T., Vogelmann, H., Navas-Guzmán, F., Haefele, A., Acheson, K., Ruth, A. A., Tatarov, B., Müller, D., Hu, Q., Podvin, T., Goloub, P., Veselovskii, I., Pietras, C., Haeffelin, M., Fréville, P., Sicard, M., Comerón, A., Fernández García, A. J., Molero Menéndez, F., Córdoba-Jabonero, C., Guerrero-Rascado, J. L., Alados-Arboledas, L., Bortoli, D., Costa, M. J., Dionisi, D., Liberti, G. L., Wang, X., Sannino, A., Papagiannopoulos, N., Boselli, A., Mona, L., D'Amico, G., Romano, S., Perrone, M. R., Belegante, L., Nicolae, D., Grigorov, I., Gialitaki, A., Amiridis, V., Soupiona, O., Papayannis, A., Mamouri, R.-E., Nisantzi, A., Heese, B., Hofer, J., Schechner, Y. Y., Wandinger, U., and Pappalardo, G.: The unprecedented 2017–2018 stratospheric smoke event: decay phase and aerosol properties observed with the EARLINET, *Atmospheric Chem. Phys.*, 19, 15183–15198, <https://doi.org/10.5194/acp-19-15183-2019>, 2019.
- Bishop, C. M.: *Pattern recognition and machine learning*, Springer, New York, 738 pp., 2006.
- Bissonnette, L. R.: Multiple-scattering lidar equation, *Appl. Opt.*, 35, 6449, <https://doi.org/10.1364/AO.35.006449>, 1996.
- Bissonnette, L. R., Bruscaioni, P., Ismaelli, A., Zaccanti, G., Cohen, A., Benayahu, Y., Kleiman, M., Egert, S., Flesia, C., Schwendimann, P., Starkov, A. V., Noormohammadian, M., Oppel, U. G., Winker, D. M., Zege, E. P., Katsev, I. L., and Polonsky, I. N.: LIDAR multiple scattering from clouds, *Appl. Phys. B Laser Opt.*, 60, 355–362, <https://doi.org/10.1007/BF01082271>, 1995.
- Bohlmann, S., Shang, X., Vakkari, V., Giannakaki, E., Leskinen, A., Lehtinen, K. E. J., Pätsi, S., and Komppula, M.: Lidar depolarization ratio of atmospheric pollen at multiple wavelengths, *Atmospheric Chem. Phys.*, 21, 7083–7097, <https://doi.org/10.5194/acp-21-7083-2021>, 2021.
- Bohren, C. F. and Clothiaux, E. E.: *Fundamentals of atmospheric radiation: an introduction with 400 problems*, Wiley-VCH-Verl, Weinheim, 472 pp., 2006.
- Bouteiller, J.: Pourquoi le ciel pourrait subitement devenir jaune à Lille et dans le Nord - Pas-de-Calais, ACTU.fr, 15th March, 2022.
- Bromberg, J. L.: The Birth of the Laser, *Phys. Today*, 41, 26–33, <https://doi.org/10.1063/1.881155>, 1988.
- Budisulistiorini, S. H., Canagaratna, M. R., Croteau, P. L., Baumann, K., Edgerton, E. S., Kollman, M. S., Ng, N. L., Verma, V., Shaw, S. L., Knipping, E. M., Worsnop, D. R., Jayne, J. T., Weber, R. J., and Surratt, J. D.: Intercomparison of an Aerosol Chemical Speciation Monitor (ACSM) with ambient fine aerosol measurements in downtown Atlanta, Georgia, *Atmospheric Meas. Tech.*, 7, 1929–1941, <https://doi.org/10.5194/amt-7-1929-2014>, 2014.
- Bühl, J., Ansmann, A., Seifert, P., Baars, H., and Engelmann, R.: Toward a quantitative characterization of heterogeneous ice formation with lidar/radar: Comparison of CALIPSO/CloudSat with ground-based observations, *Geophys. Res. Lett.*, 40, 4404–4408, <https://doi.org/10.1002/grl.50792>, 2013.
-

- 
- Burgos, M. A., Andrews, E., Titos, G., Alados-Arboledas, L., Baltensperger, U., Day, D., Jefferson, A., Kalivitis, N., Mihalopoulos, N., Sherman, J., Sun, J., Weingartner, E., and Zieger, P.: A global view on the effect of water uptake on aerosol particle light scattering, *Sci. Data*, 6, 157, <https://doi.org/10.1038/s41597-019-0158-7>, 2019.
- Burton, S. P., Ferrare, R. A., Hostetler, C. A., Hair, J. W., Rogers, R. R., Obland, M. D., Butler, C. F., Cook, A. L., Harper, D. B., and Froyd, K. D.: Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples, *Atmospheric Meas. Tech.*, 5, 73–98, <https://doi.org/10.5194/amt-5-73-2012>, 2012.
- Cadet, B., Giraud, V., Haeffelin, M., Keckhut, P., Rechou, A., and Baldy, S.: Improved retrievals of the optical properties of cirrus clouds by a combination of lidar methods, *Appl. Opt.*, 44, 1726, <https://doi.org/10.1364/AO.44.001726>, 2005.
- Calvo, A. I., Alves, C., Castro, A., Pont, V., Vicente, A. M., and Fraile, R.: Research on aerosol sources and chemical composition: Past, current and emerging issues, *Atmospheric Res.*, 120–121, 1–28, <https://doi.org/10.1016/j.atmosres.2012.09.021>, 2013.
- Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q., Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L. R., Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer, *Mass Spectrom. Rev.*, 26, 185–222, <https://doi.org/10.1002/mas.20115>, 2007.
- Charlson, R. J., Ahlquist, N. C., Selvidge, H., and MacCready, P. B.: Monitoring of Atmospheric Aerosol Parameters with the Integrating Nephelometer, *J. Air Pollut. Control Assoc.*, 19, 937–942, <https://doi.org/10.1080/00022470.1969.10469360>, 1969.
- Chen, J., Zhao, C. S., Ma, N., and Yan, P.: Aerosol hygroscopicity parameter derived from the light scattering enhancement factor measurements in the North China Plain, *Atmospheric Chem. Phys.*, 14, 8105–8118, <https://doi.org/10.5194/acp-14-8105-2014>, 2014.
- Chen, L., Peng, C., Nenes, A., Gu, W., Fu, H., Jian, X., Zhang, H., Zhang, G., Zhu, J., Wang, X., and Tang, M.: On mineral dust aerosol hygroscopicity, *Aerosols/Laboratory Studies/Troposphere/Chemistry (chemical composition and reactions)*, <https://doi.org/10.5194/acp-2020-442>, 2020.
- Chepfer, H., Brogniez, G., and Fouquart, Y.: Cirrus clouds' microphysical properties deduced from POLDER observations, *J. Quant. Spectrosc. Radiat. Transf.*, 60, 375–390, [https://doi.org/10.1016/S0022-4073\(98\)00013-2](https://doi.org/10.1016/S0022-4073(98)00013-2), 1998.
- Choi, W., Lee, H., and Park, J.: A First Approach to Aerosol Classification Using Space-Borne Measurement Data: Machine Learning-Based Algorithm and Evaluation, *Remote Sens.*, 13, 609, <https://doi.org/10.3390/rs13040609>, 2021.
- Cooper, D. W., Davis, J. W., and Byers, R. L.: Measurements of depolarization by dry and humidified salt aerosols using a lidar analogue, *J. Aerosol Sci.*, 5, 117–123, [https://doi.org/10.1016/0021-8502\(74\)90043-3](https://doi.org/10.1016/0021-8502(74)90043-3), 1974.
- Cornet, C., C-Labonnote, L., and Szczap, F.: Three-dimensional polarized Monte Carlo atmospheric radiative transfer model (3DMCPOL): 3D effects on polarized visible reflectances of a cirrus cloud, *J. Quant. Spectrosc. Radiat. Transf.*, 111, 174–186, <https://doi.org/10.1016/j.jqsrt.2009.06.013>, 2010.
-

- 
- Covert, D. S., Charlson, R. J., and Ahlquist, N. C.: A study of the Relationship of Chemical Composition and Humidity to Light Scattering by Aerosols, *J. Appl. Meteorol.*, 11, 968–976, [https://doi.org/10.1175/1520-0450\(1972\)011<0968:ASOTRO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<0968:ASOTRO>2.0.CO;2), 1972.
- Crumeyrolle, S., Kontkanen, J. S. S., Rose, C., Velazquez Garcia, A., Bourrienne, E., Catalfamo, M., Riffault, V., Tison, E., Ferreira de Brito, J., Visez, N., Ferlay, N., Auriol, F., and Chiapello, I.: Measurement report: Atmospheric new particle formation at a peri-urban site in Lille, northern France, *Atmospheric Chem. Phys.*, 23, 183–201, <https://doi.org/10.5194/acp-23-183-2023>, 2023.
- Dawson, K. W., Ferrare, R. A., Moore, R. H., Clayton, M. B., Thorsen, T. J., and Eloranta, E. W.: Ambient Aerosol Hygroscopic Growth From Combined Raman Lidar and HSRL, *J. Geophys. Res. Atmospheres*, 125, e2019JD031708, <https://doi.org/10.1029/2019JD031708>, 2020.
- Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data Via the *EM* Algorithm, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 39, 1–22, <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>, 1977.
- Dicke, R. H., Beringer, R., Kyhl, R. L., and Vane, A. B.: Atmospheric Absorption Measurements with a Microwave Radiometer, *Phys. Rev.*, 70, 340–348, <https://doi.org/10.1103/PhysRev.70.340>, 1946.
- Dinh, D.-T., Fujinami, T., and Huynh, V.-N.: Estimating the Optimal Number of Clusters in Categorical Data Clustering by Silhouette Coefficient, in: *Knowledge and Systems Sciences*, vol. 1103, edited by: Chen, J., Huynh, V. N., Nguyen, G.-N., and Tang, X., Springer Singapore, Singapore, 1–17, [https://doi.org/10.1007/978-981-15-1209-4\\_1](https://doi.org/10.1007/978-981-15-1209-4_1), 2019.
- Draxler, R., Stunder, B., Rolph, G., Stein, A., Taylor, A., Zinn, S., Loughner, C., and Crawford, A.: *HYSPLIT User's Guide*, 2023.
- Dubuisson, P., Cornet, C., Goloub, P., Chiapello, I., and Pujol, O.: Le Laboratoire d'optique atmosphérique (LOA), *La Météorologie*, 047, <https://doi.org/10.37053/lameteorologie-2021-0068>, 2021.
- Ducos, F., Hu, Q., and Popovici, I.: *AUSTRAL User's Guide - AUtomated Server for the Treatment of Atmospheric Lidars*, 2022.
- Dusek, U., Frank, G. P., Hildebrandt, L., Curtius, J., Schneider, J., Walter, S., Chand, D., Drewnick, F., Hings, S., Jung, D., Borrmann, S., and Andreae, M. O.: Size Matters More Than Chemistry for Cloud-Nucleating Ability of Aerosol Particles, *Science*, 312, 1375–1378, <https://doi.org/10.1126/science.1125261>, 2006.
- Düsing, S., Ansmann, A., Baars, H., Corbin, J. C., Denjean, C., Gysel-Beer, M., Müller, T., Poulain, L., Siebert, H., Spindler, G., Tuch, T., Wehner, B., and Wiedensohler, A.: Measurement report: Comparison of airborne, in situ measured, lidar-based, and modeled aerosol optical properties in the central European background – identifying sources of deviations, *Atmospheric Chem. Phys.*, 21, 16745–16773, <https://doi.org/10.5194/acp-21-16745-2021>, 2021.
- Edgerton, E. S., Hartsell, B. E., Saylor, R. D., Jansen, J. J., Hansen, D. A., and Hidy, G. M.: The Southeastern Aerosol Research and Characterization Study: Part II. Filter-Based Measurements of Fine and Coarse Particulate Matter Mass and Composition, *J. Air Waste Manag. Assoc.*, 55, 1527–1542, <https://doi.org/10.1080/10473289.2005.10464744>, 2005.
-

- 
- Feingold, G.: Aerosol hygroscopic properties as measured by lidar and comparison with in situ measurements, *J. Geophys. Res.*, 108, 4327, <https://doi.org/10.1029/2002JD002842>, 2003.
- Felidj, N.: Introduction à... La spectroscopie Raman et la diffusion exaltée de surface, *Photoniques*, 46–49, <https://doi.org/10.1051/photon/20168146>, 2016.
- Fernández, A. J., Apituley, A., Veselovskii, I., Suvorina, A., Henzing, J., Pujadas, M., and Artíñano, B.: Study of aerosol hygroscopic events over the Cabauw experimental site for atmospheric research (CESAR) using the multi-wavelength Raman lidar Caeli, *Atmos. Environ.*, 120, 484–498, <https://doi.org/10.1016/j.atmosenv.2015.08.079>, 2015.
- Ferrare, R., Hair, J., Hostetler, C., Shingler, T., Burton, S. P., Fenn, M., Clayton, M., Scarino, A. J., Harper, D., Seaman, S., Cook, A., Crosbie, E., Winstead, E., Ziemba, L., Thornhill, L., Robinson, C., Moore, R., Vaughan, M., Sorooshian, A., Schlosser, J. S., Liu, H., Zhang, B., Diskin, G., DiGangi, J., Nowak, J., Choi, Y., Zuidema, P., and Chellappan, S.: Airborne HSRL-2 measurements of elevated aerosol depolarization associated with non-spherical sea salt, *Front. Remote Sens.*, 4, 1143944, <https://doi.org/10.3389/frsen.2023.1143944>, 2023.
- Fiocco, G. and Smullin, L. D.: Detection of Scattering Layers in the Upper Atmosphere (60–140 km) by Optical Radar, *Nature*, 199, 1275–1276, <https://doi.org/10.1038/1991275a0>, 1963.
- Foth, A., Baars, H., Di Girolamo, P., and Pospichal, B.: Water vapour profiles from Raman lidar automatically calibrated by microwave radiometer data during HOPE, *Atmospheric Chem. Phys.*, 15, 7753–7763, <https://doi.org/10.5194/acp-15-7753-2015>, 2015.
- Fredj, A. B. and Ruiz-López, M. F.: Theoretical Study of Chlorophyll *a* Hydrates Formation in Aqueous Organic Solvents, *J. Phys. Chem. B*, 114, 681–687, <https://doi.org/10.1021/jp909380t>, 2010.
- Freudenthaler, V., Esselborn, M., Wiegner, M., Heese, B., Tesche, M., Ansmann, A., Müller, D., Althausen, D., Wirth, M., Fix, A., Ehret, G., Knippertz, P., Toledano, C., Gasteiger, J., Garhammer, M., and Seefeldner, M.: Depolarization ratio profiling at several wavelengths in pure Saharan dust during SAMUM 2006, *Tellus B Chem. Phys. Meteorol.*, 61, 165, <https://doi.org/10.1111/j.1600-0889.2008.00396.x>, 2009.
- Freudenthaler, V., Linné, H., Chaikovski, A., Rabus, D., and Groß, S.: EARLINET lidar quality assurance tools, <https://doi.org/10.5194/amt-2017-395>, 3 January 2018.
- Gast, B., Jimenez, C., Ansmann, A., Haarig, M., Engelmann, R., Fritzsche, F., Floutsi, A. A., Griesche, H., Ohneiser, K., Hofer, J., Radenz, M., Baars, H., Seifert, P., and Wandinger, U.: Invisible aerosol layers: improved lidar detection capabilities by means of laser-induced aerosol fluorescence, <https://doi.org/10.5194/egusphere-2024-2586>, 27 August 2024.
- Gentle, J. E., Härdle, W. K., and Mori, Y. (Eds.): *Handbook of Computational Statistics: Concepts and Methods*, Springer Berlin Heidelberg, Berlin, Heidelberg, <https://doi.org/10.1007/978-3-642-21551-3>, 2012.
- Gislason, S. R., Hassenkam, T., Nedel, S., Bovet, N., Eiríksdóttir, E. S., Alfredsson, H. A., Hem, C. P., Balogh, Z. I., Dideriksen, K., Oskarsson, N., Sigfusson, B., Larsen, G., and Stipp, S. L. S.: Characterization of Eyjafjallajökull volcanic ash particles and a protocol for rapid risk assessment, *Proc. Natl. Acad. Sci.*, 108, 7307–7312, <https://doi.org/10.1073/pnas.1015053108>, 2011.
- Glickman, T. S., Glickman, T. S., and American Meteorological Society (Eds.): *Glossary of meteorology*, 2. ed., American Meteorological Society, Boston, Mass, 855 pp., 2000.
-

- 
- Gobbi, G. P.: Polarization lidar returns from aerosols and thin clouds: a framework for the analysis, *Appl. Opt.*, **37**, 5505, <https://doi.org/10.1364/AO.37.005505>, 1998.
- Gomez, S. L., Carrico, C. M., Allen, C., Lam, J., Dabli, S., Sullivan, A. P., Aiken, A. C., Rahn, T., Romonosky, D., Chylek, P., Sevanto, S., and Dubey, M. K.: Southwestern U.S. Biomass Burning Smoke Hygroscopicity: The Role of Plant Phenology, Chemical Composition, and Combustion Properties, *J. Geophys. Res. Atmospheres*, **123**, 5416–5432, <https://doi.org/10.1029/2017JD028162>, 2018.
- Granados-Muñoz, M. J., Navas-Guzmán, F., Bravo-Aranda, J. A., Guerrero-Rascado, J. L., Lyamani, H., Valenzuela, A., Titos, G., Fernández-Gálvez, J., and Alados-Arboledas, L.: Hygroscopic growth of atmospheric aerosol particles based on active remote sensing and radiosounding measurements: selected cases in southeastern Spain, *Atmospheric Meas. Tech.*, **8**, 705–718, <https://doi.org/10.5194/amt-8-705-2015>, 2015.
- Griaznov, V., Veselovskii, I., Kolgotin, A., and Whiteman, D. N.: Angle- and size-dependent characteristics of incoherent Raman and fluorescent scattering by microspheres 1 General expressions, *Appl. Opt.*, **41**, 5773, <https://doi.org/10.1364/AO.41.005773>, 2002.
- Haarig, M., Ansmann, A., Gasteiger, J., Kandler, K., Althausen, D., Baars, H., Radenz, M., and Farrell, D. A.: Dry versus wet marine particle optical properties: RH dependence of depolarization ratio, backscatter, and extinction from multiwavelength lidar measurements during SALTRACE, *Atmospheric Chem. Phys.*, **17**, 14199–14217, <https://doi.org/10.5194/acp-17-14199-2017>, 2017.
- Haarig, M., Ansmann, A., Engelmann, R., Baars, H., Toledano, C., Torres, B., Althausen, D., Radenz, M., and Wandinger, U.: First triple-wavelength lidar observations of depolarization and extinction-to-backscatter ratios of Saharan dust, *Atmospheric Chem. Phys.*, **22**, 355–369, <https://doi.org/10.5194/acp-22-355-2022>, 2022.
- Hagan, D. H. and Kroll, J. H.: Assessing the accuracy of low-cost optical particle sensors using a physics-based approach, *Atmospheric Meas. Tech.*, **13**, 6343–6355, <https://doi.org/10.5194/amt-13-6343-2020>, 2020.
- Hänel, G.: The Properties of Atmospheric Aerosol Particles as Functions of the Relative Humidity at Thermodynamic Equilibrium with the Surrounding Moist Air, in: *Advances in Geophysics*, vol. 19, Elsevier, 73–188, [https://doi.org/10.1016/S0065-2687\(08\)60142-9](https://doi.org/10.1016/S0065-2687(08)60142-9), 1976.
- Harris, D. C. and Bertolucci, M. D.: *Symmetry and spectroscopy: an introduction to vibrational and electronic spectroscopy*, Unabridged, corr. republ. of the work first publ. by Oxford Univ. Press, New York, 1978., Dover, New York, NY, 550 pp., 1989.
- Harrison, R. M., Beddows, D. C. S., Alam, M. S., Singh, A., Brean, J., Xu, R., Kotthaus, S., and Grimmond, S.: Interpretation of particle number size distributions measured across an urban area during the FASTER campaign, *Atmospheric Chem. Phys.*, **19**, 39–55, <https://doi.org/10.5194/acp-19-39-2019>, 2019.
- Herich, H., Tritscher, T., Wiacek, A., Gysel, M., Weingartner, E., Lohmann, U., Baltensperger, U., and Cziczo, D. J.: Water uptake of clay and desert dust aerosol particles at sub- and supersaturated water vapor conditions, *Phys. Chem. Chem. Phys.*, **11**, 7804, <https://doi.org/10.1039/b901585j>, 2009.
- Hill, S. C., Pinnick, R. G., Niles, S., Pan, Y.-L., Holler, S., Chang, R. K., Bottiger, J., Chen, B. T., Orr, C.-S., and Feather, G.: Real-time measurement of fluorescence spectra from single
-

- 
- airborne biological particles, *Field Anal. Chem. Technol.*, 3, 221–239, [https://doi.org/10.1002/\(SICI\)1520-6521\(1999\)3:4/5<221::AID-FACT2>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1520-6521(1999)3:4/5<221::AID-FACT2>3.0.CO;2-7), 1999.
- Hoekstra, A., Maltsev, V., and Videen, G. (Eds.): *Optics of biological particles*, Springer, Dordrecht, 284 pp., 2007.
- Hofer, J., Ansmann, A., Althausen, D., Engelmann, R., Baars, H., Fomba, K. W., Wandinger, U., Abdullaev, S. F., and Makhmudov, A. N.: Optical properties of Central Asian aerosol relevant for spaceborne lidar applications and aerosol typing at 355 and 532 nm, *Atmospheric Chem. Phys.*, 20, 9265–9280, <https://doi.org/10.5194/acp-20-9265-2020>, 2020.
- Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, *Remote Sens. Environ.*, 66, 1–16, [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5), 1998.
- Hu, Q.: *Advanced aerosol characterization using sun/sky photometer and multi-wavelength Mie-Raman lidar measurements*, 2018.
- Hu, Q., Goloub, P., Veselovskii, I., and Podvin, T.: The characterization of long-range transported North American biomass burning plumes: what can a multi-wavelength Mie–Raman-polarization-fluorescence lidar provide?, *Atmospheric Chem. Phys.*, 22, 5399–5414, <https://doi.org/10.5194/acp-22-5399-2022>, 2022.
- Hu, Y., Liu, Z., Winker, D., Vaughan, M., Noel, V., Bissonnette, L., Roy, G., and McGill, M.: Simple relation between lidar multiple scattering and depolarization for water clouds, *Opt. Lett.*, 31, 1809, <https://doi.org/10.1364/OL.31.001809>, 2006.
- Hu, Y., Winker, D., Vaughan, M., Lin, B., Omar, A., Trepte, C., Flittner, D., Yang, P., Nasiri, S. L., Baum, B., Holz, R., Sun, W., Liu, Z., Wang, Z., Young, S., Stamnes, K., Huang, J., and Kuehn, R.: CALIPSO/CALIOP Cloud Phase Discrimination Algorithm, *J. Atmospheric Ocean. Technol.*, 26, 2293–2309, <https://doi.org/10.1175/2009JTECHA1280.1>, 2009.
- Hu, Y.-X., Winker, D., Yang, P., Baum, B., Poole, L., and Vann, L.: Identification of cloud phase from PICASSO-CENA lidar depolarization: a multiple scattering sensitivity study, *J. Quant. Spectrosc. Radiat. Transf.*, 70, 569–579, [https://doi.org/10.1016/S0022-4073\(01\)00030-9](https://doi.org/10.1016/S0022-4073(01)00030-9), 2001.
- Husar, R. B.: Intercontinental Transport of Dust: Historical and Recent Observational Evidence, 4, 277–294, <https://doi.org/10.1007/b94531>, 2004.
- Immler, F., Engelbart, D., and Schrems, O.: Fluorescence from atmospheric aerosol detected by a lidar indicates biogenic particles in the lowermost stratosphere, *Atmospheric Chem. Phys.*, 5, 345–355, <https://doi.org/10.5194/acp-5-345-2005>, 2005.
- Intergovernmental Panel on Climate Change: *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1st ed., Cambridge University Press, <https://doi.org/10.1017/9781009157896>, 2023.
- Intrieri, J. M., Stephens, G. L., Eberhard, W. L., and Uttal, T.: A Method for Determining Cirrus Cloud Particle Sizes Using Lidar and Radar Backscatter Technique, *J. Appl. Meteorol.*, 32, 1074–1082, [https://doi.org/10.1175/1520-0450\(1993\)032<1074:AMFDCC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1993)032<1074:AMFDCC>2.0.CO;2), 1993.
-

- 
- Intrieri, J. M., Shupe, M. D., Uttal, T., and McCarty, B. J.: An annual cycle of Arctic cloud characteristics observed by radar and lidar at SHEBA, *J. Geophys. Res. Oceans*, 107, <https://doi.org/10.1029/2000JC000423>, 2002.
- Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P., and Morisette, J. T.: An overview of MODIS Land data processing and product status, *Remote Sens. Environ.*, 83, 3–15, [https://doi.org/10.1016/S0034-4257\(02\)00084-6](https://doi.org/10.1016/S0034-4257(02)00084-6), 2002.
- Karl, W. C., Fowler, J. E., Bouman, C. A., Çetin, M., Wohlberg, B., and Ye, J. C.: The Foundations of Computational Imaging: A signal processing perspective, *IEEE Signal Process. Mag.*, 40, 40–53, <https://doi.org/10.1109/MSP.2023.3274328>, 2023.
- Kim, S.-W., Berthier, S., Raut, J.-C., Chazette, P., Dulac, F., and Yoon, S.-C.: Validation of aerosol and cloud layer structures from the space-borne lidar CALIOP using a ground-based lidar in Seoul, Korea, *Atmospheric Chem. Phys.*, 8, 3705–3720, <https://doi.org/10.5194/acp-8-3705-2008>, 2008.
- Köhler, H.: The nucleus in and the growth of hygroscopic droplets, *Trans Faraday Soc*, 32, 1152–1161, <https://doi.org/10.1039/TF9363201152>, 1936.
- Li, Z., Hou, W., Hong, J., Zheng, F., Luo, D., Wang, J., Gu, X., and Qiao, Y.: Directional Polarimetric Camera (DPC): Monitoring aerosol spectral optical properties over land from satellite observation, *J. Quant. Spectrosc. Radiat. Transf.*, 218, 21–37, <https://doi.org/10.1016/j.jqsrt.2018.07.003>, 2018.
- Liou, K. N.: *An Introduction to Atmospheric Radiation*, 2nd ed., Academic Press, California, USA, 2002.
- Liu, B. Y. H., Pui, D. Y. H., Whitby, K. T., and Kittelson, D. B.: DETECTOR FOR SULFURIC ACID AEROSOLS, *Atmos. Environ.*, 12, 99–104, [https://doi.org/10.1016/0004-6981\(78\)90192-0](https://doi.org/10.1016/0004-6981(78)90192-0), 1978.
- Mallet, P.-É., Pujol, O., Brioude, J., Evan, S., and Jensen, A.: Marine aerosol distribution and variability over the pristine Southern Indian Ocean, *Atmos. Environ.*, 182, 17–30, <https://doi.org/10.1016/j.atmosenv.2018.03.016>, 2018.
- Mascaut, F., Pujol, O., Verreyken, B., Peroni, R., Metzger, J. M., Blarel, L., Podvin, T., Goloub, P., Sellegri, K., Thornberry, T., Dufлот, V., Tulet, P., and Brioude, J.: Aerosol characterization in an oceanic context around Reunion Island (AEROMARINE field campaign), *Atmos. Environ.*, 268, 118770, <https://doi.org/10.1016/j.atmosenv.2021.118770>, 2022.
- Mascaut, F., Pujol, O., Brioude, J., Jensen, A., Lefranc, M., Evan, S., and Crumeyrolle, S.: A competition–species model for water vapour-aerosol-cloud-rain interactions, *Atmospheric Res.*, 284, 106588, <https://doi.org/10.1016/j.atmosres.2022.106588>, 2023.
- Mattis, I., Ansmann, A., Althausen, D., Jaenisch, V., Wandinger, U., Müller, D., Arshinov, Y. F., Bobrovnikov, S. M., and Serikov, I. B.: Relative-humidity profiling in the troposphere with a Raman lidar, *Appl. Opt.*, 41, 6451, <https://doi.org/10.1364/AO.41.006451>, 2002.
- McFarquhar, G. M., Baumgardner, D., Bansemer, A., Abel, S. J., Crosier, J., French, J., Rosenberg, P., Korolev, A., Schwarzenboeck, A., Leroy, D., Um, J., Wu, W., Heymsfield, A. J., Twohy, C., Detwiler, A., Field, P., Neumann, A., Cotton, R., Axisa, D., and Dong, J.: Processing of Ice Cloud In Situ Data Collected by Bulk Water, Scattering, and Imaging Probes: Fundamentals, Uncertainties, and Efforts toward Consistency, *Meteorol. Monogr.*, 58, 11.1–11.33, <https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0007.1>, 2017.
-

- 
- McNeil, W. R. and Carsweil, A. I.: Lidar polarization studies of the troposphere, *Appl. Opt.*, 14, 2158, <https://doi.org/10.1364/AO.14.002158>, 1975.
- Mie, G.: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen, *Ann. Phys.*, 330, 377–445, <https://doi.org/10.1002/andp.19083300302>, 1908.
- Miyatake, T. and Tamiaki, H.: Self-aggregates of natural chlorophylls and their synthetic analogues in aqueous media for making light-harvesting systems, *Coord. Chem. Rev.*, 254, 2593–2602, <https://doi.org/10.1016/j.ccr.2009.12.027>, 2010.
- Mochida, M., Kuwata, M., Miyakawa, T., Takegawa, N., Kawamura, K., and Kondo, Y.: Relationship between hygroscopicity and cloud condensation nuclei activity for urban aerosols in Tokyo, *J. Geophys. Res. Atmospheres*, 111, 2005JD006980, <https://doi.org/10.1029/2005JD006980>, 2006.
- Müller, D., Ansmann, A., Mattis, I., Tesche, M., Wandinger, U., Althausen, D., and Pisani, G.: Aerosol-type-dependent lidar ratios observed with Raman lidar, *J. Geophys. Res. Atmospheres*, 112, 2006JD008292, <https://doi.org/10.1029/2006JD008292>, 2007.
- Murayama, T., Furushima, M., Oda, A., Iwasaka, N., and Kai, K.: Depolarization Ratio Measurements in the Atmospheric Boundary Layer by Lidar in Tokyo, *J. Meteorol. Soc. Jpn. Ser II*, 74, 571–578, [https://doi.org/10.2151/jmsj1965.74.4\\_571](https://doi.org/10.2151/jmsj1965.74.4_571), 1996.
- Navas-Guzmán, F., Martucci, G., Collaud Coen, M., Granados-Muñoz, M. J., Hervo, M., Sicard, M., and Haeferle, A.: Characterization of aerosol hygroscopicity using Raman lidar measurements at the EARLINET station of Payerne, *Atmospheric Chem. Phys.*, 19, 11651–11668, <https://doi.org/10.5194/acp-19-11651-2019>, 2019.
- Ng, N. L., Herndon, S. C., Trimborn, A., Canagaratna, M. R., Croteau, P. L., Onasch, T. B., Sueper, D., Worsnop, D. R., Zhang, Q., Sun, Y. L., and Jayne, J. T.: An Aerosol Chemical Speciation Monitor (ACSM) for Routine Monitoring of the Composition and Mass Concentrations of Ambient Aerosol, *Aerosol Sci. Technol.*, 45, 780–794, <https://doi.org/10.1080/02786826.2011.560211>, 2011.
- Nicolae, D., Vasilescu, J., Talianu, C., Biniotoglou, I., Nicolae, V., Andrei, S., and Antonescu, B.: A neural network aerosol-typing algorithm based on lidar data, *Atmospheric Chem. Phys.*, 18, 14511–14537, <https://doi.org/10.5194/acp-18-14511-2018>, 2018a.
- Nicolae, D., Talianu, C., Vasilescu, J., Nicolae, V., and Stachlewska, I. S.: Strengths and limitations of the NATALI code for aerosol typing from multiwavelength Raman lidar observations, *EPJ Web Conf.*, 176, 05005, <https://doi.org/10.1051/epjconf/201817605005>, 2018b.
- Paatero, P. and Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, 5, 111–126, <https://doi.org/10.1002/env.3170050203>, 1994.
- Pan, Y., Holler, S., Chang, R. K., Hill, S. C., Pinnick, R. G., Niles, S., and Bottiger, J. R.: Single-shot fluorescence spectra of individual micrometer-sized bioaerosols illuminated by a 351- or a 266-nm ultraviolet laser, *Opt. Lett.*, 24, 116, <https://doi.org/10.1364/OL.24.000116>, 1999.
- Pan, Y.-L., Huang, H., and Chang, R. K.: Clustered and integrated fluorescence spectra from single atmospheric aerosol particles excited by a 263- and 351-nm laser at New Haven, CT, and Adelphi, MD, *J. Quant. Spectrosc. Radiat. Transf.*, 113, 2213–2221, <https://doi.org/10.1016/j.jqsrt.2012.07.028>, 2012.
-

- 
- Papagiannopoulos, N., Mona, L., Amodeo, A., D'Amico, G., Gumà Claramunt, P., Pappalardo, G., Alados-Arboledas, L., Guerrero-Rascado, J. L., Amiridis, V., Kokkalis, P., Apituley, A., Baars, H., Schwarz, A., Wandinger, U., Biniotoglou, I., Nicolae, D., Bortoli, D., Comerón, A., Rodríguez-Gómez, A., Sicard, M., Papayannis, A., and Wiegner, M.: An automatic observation-based aerosol typing method for EARLINET, *Atmospheric Chem. Phys.*, 18, 15879–15901, <https://doi.org/10.5194/acp-18-15879-2018>, 2018.
- Parworth, C., Fast, J., Mei, F., Shippert, T., Sivaraman, C., Tilp, A., Watson, T., and Zhang, Q.: Long-term measurements of submicrometer aerosol chemistry at the Southern Great Plains (SGP) using an Aerosol Chemical Speciation Monitor (ACSM), *Atmos. Environ.*, 106, 43–55, <https://doi.org/10.1016/j.atmosenv.2015.01.060>, 2015.
- Patel, E. and Kushwaha, D. S.: Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model, *Procedia Comput. Sci.*, 171, 158–167, <https://doi.org/10.1016/j.procs.2020.04.017>, 2020.
- Petit, J.-E., Favez, O., Sciare, J., Canonaco, F., Croteau, P., Močnik, G., Jayne, J., Worsnop, D., and Leoz-Garziandia, E.: Submicron aerosol source apportionment of wintertime pollution in Paris, France by double positive matrix factorization (PMF<sup>2</sup>) using an aerosol chemical speciation monitor (ACSM) and a multi-wavelength Aethalometer, *Atmospheric Chem. Phys.*, 14, 13773–13787, <https://doi.org/10.5194/acp-14-13773-2014>, 2014.
- Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmos Chem Phys*, 7, 1961–1971, <https://doi.org/10.5194/acp-7-1961-2007>, 2007.
- Polivka, B. J.: The Great London Smog of 1952, *AJN Am. J. Nurs.*, 118, 57–61, <https://doi.org/10.1097/01.NAJ.0000532078.72372.c3>, 2018.
- Raman, C. V.: A new radiation, 1928.
- Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., and Barkstrom, B. R.: Cloud-Radiative Forcing and Climate: Results from the Earth Radiation Budget Experiment, 243, 1989.
- Rao, Z., He, T., Hua, D., Wang, Y., Wang, X., Chen, Y., and Le, J.: Preliminary measurements of fluorescent aerosol number concentrations using a laser-induced fluorescence lidar, *Appl. Opt.*, 57, 7211, <https://doi.org/10.1364/AO.57.007211>, 2018.
- Reff, A., Eberly, S. I., and Bhawe, P. V.: Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods, *J. Air Waste Manag. Assoc.*, 57, 146–154, <https://doi.org/10.1080/10473289.2007.10465319>, 2007.
- Reichardt, S. and Reichardt, J.: Effect of multiple scattering on depolarization measurements with spaceborne lidars, *Appl. Opt.*, 42, 3620, <https://doi.org/10.1364/AO.42.003620>, 2003.
- Rizi, V., Iarlori, M., Rocci, G., and Visconti, G.: Raman lidar observations of cloud liquid water, *Appl. Opt.*, 43, 6440, <https://doi.org/10.1364/AO.43.006440>, 2004.
- Rousseuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, 20, 53–65, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7), 1987.
- Rundell, K. W., Hoffman, J. R., Caviston, R., Bulbulian, R., and Hollenbach, A. M.: Inhalation of Ultrafine and Fine Particulate Matter Disrupts Systemic Vascular Function, *Inhal. Toxicol.*, 19, 133–140, <https://doi.org/10.1080/08958370601051727>, 2007.
-

- Sassen, K.: Lidar Backscatter Depolarization Technique for Cloud and Aerosol Research, in: *Light scattering by nonspherical particles: theory, measurements, and applications*, San Diego, 2000.
- Sassen, K. and Petrilla, R. L.: Lidar depolarization from multiple scattering in marine stratus clouds, *Appl. Opt.*, 25, 1450, <https://doi.org/10.1364/AO.25.001450>, 1986.
- Schmidt, J., Ansmann, A., Bühl, J., Baars, H., Wandinger, U., Müller, D., and Malinka, A. V.: Dual-FOV Raman and Doppler lidar studies of aerosol-cloud interactions: Simultaneous profiling of aerosols, warm-cloud properties, and vertical wind, *J. Geophys. Res. Atmospheres*, 119, 5512–5527, <https://doi.org/10.1002/2013JD020424>, 2014.
- Schmidt, J., Ansmann, A., Bühl, J., and Wandinger, U.: Strong aerosol–cloud interaction in altocumulus during updraft periods: lidar observations over central Europe, *Atmospheric Chem. Phys.*, 15, 10687–10700, <https://doi.org/10.5194/acp-15-10687-2015>, 2015.
- Schoeberl, M. R.: *The Afternoon Constellation: A Formation of Earth Observing Systems for the Atmosphere and Hydrosphere*, IEEE Int. Geosci. Remote Sens. Symp., <https://doi.org/10.1109/IGARSS.2002.1025038>, 2002.
- Shaw, G. E.: Sun Photometry, *Bull. Am. Meteorol. Soc.*, 64, 4–10, [https://doi.org/10.1175/1520-0477\(1983\)064<0004:SP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1983)064<0004:SP>2.0.CO;2), 1983.
- Shcherbakov, V., Szczap, F., Alkasem, A., Mioche, G., and Cornet, C.: Empirical model of multiple-scattering effect on single-wavelength lidar data of aerosols and clouds, *Atmospheric Meas. Tech.*, 15, 1729–1754, <https://doi.org/10.5194/amt-15-1729-2022>, 2022.
- Shiraiwa, M., Ueda, K., Pozzer, A., Lammel, G., Kampf, C. J., Fushimi, A., Enami, S., Arangio, A. M., Fröhlich-Nowoisky, J., Fujitani, Y., Furuyama, A., Lakey, P. S. J., Lelieveld, J., Lucas, K., Morino, Y., Pöschl, U., Takahama, S., Takami, A., Tong, H., Weber, B., Yoshino, A., and Sato, K.: Aerosol Health Effects from Molecular to Global Scales, *Environ. Sci. Technol.*, 51, 13545–13567, <https://doi.org/10.1021/acs.est.7b04417>, 2017.
- Sicard, M., Fortunato dos Santos Oliveira, D. C., Muñoz-Porcar, C., Gil-Díaz, C., Comerón, A., Rodríguez-Gómez, A., and Dios Otín, F.: Measurement Report: Spectral and statistical analysis of aerosol hygroscopic growth from multi-wavelength lidar measurements in Barcelona, Spain, *Aerosols/Field Measurements/Troposphere/Physics (physical properties and processes)*, <https://doi.org/10.5194/acp-2021-990>, 2022.
- Skou, N. and Le Vine, D.: *Microwave radiometer systems: design and analysis*, 2. ed., Artech House, Boston, 222 pp., 2006.
- Smekal, A.: *Zur Quantentheorie der Dispersion*, *Naturwissenschaften*, <https://doi.org/10.1007/BF01576902>, 1923.
- Somsen, G. A., van Rijn, C. J. M., Kooij, S., Bem, R. A., and Bonn, D.: Measurement of small droplet aerosol concentrations in public spaces using handheld particle counters, *Phys. Fluids*, 32, 121707, <https://doi.org/10.1063/5.0035701>, 2020.
- Spanu, A., Dollner, M., Gasteiger, J., Bui, T. P., and Weinzierl, B.: Flow-induced errors in airborne in situ measurements of aerosols and clouds, *Atmospheric Meas. Tech.*, 13, 1963–1987, <https://doi.org/10.5194/amt-13-1963-2020>, 2020.
- Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System, *Bull. Am. Meteorol. Soc.*, 96, 2059–2077, <https://doi.org/10.1175/BAMS-D-14-00110.1>, 2015.
-

- 
- Stephens, G. L., Vane, D. G., Tanelli, S., Im, E., Durden, S., Rokey, M., Reinke, D., Partain, P., Mace, G. G., Austin, R., L'Ecuyer, T., Haynes, J., Lebsock, M., Suzuki, K., Waliser, D., Wu, D., Kay, J., Gettelman, A., Wang, Z., and Marchand, R.: CloudSat mission: Performance and early science after the first year of operation, *J. Geophys. Res. Atmospheres*, 113, 2008JD009982, <https://doi.org/10.1029/2008JD009982>, 2008.
- Stohl, A., Akimoto, H., and Hutzinger, O. (Eds.): *Intercontinental transport of air pollution*, Springer, Berlin, 325 pp., 2004.
- Su, J. and Patrick McCormick, M.: Using multi-wavelength Mie–Raman lidar to measure low-level cloud properties, *J. Quant. Spectrosc. Radiat. Transf.*, 237, 106610, <https://doi.org/10.1016/j.jqsrt.2019.106610>, 2019.
- Sugimoto, N., Huang, Z., Nishizawa, T., Matsui, I., and Tatarov, B.: Fluorescence from atmospheric aerosols observed with a multi-channel lidar spectrometer, *Opt. Express*, 20, 20800, <https://doi.org/10.1364/OE.20.020800>, 2012.
- Swietlicki, E., Hansson, H.-C., Hämeri, K., Svenningsson, B., Massling, A., McFiggans, G., McMurry, P. H., Petäjä, T., Tunved, P., Gysel, M., Topping, D., Weingartner, E., Baltensperger, U., Rissler, J., Wiedensohler, A., and Kulmala, M.: Hygroscopic properties of submicrometer atmospheric aerosol particles measured with H-TDMA instruments in various environments—a review, *Tellus B Chem. Phys. Meteorol.*, 60, 432, <https://doi.org/10.1111/j.1600-0889.2008.00350.x>, 2008.
- Thiébaud, E.: INTRODUCTION TO IMAGE RECONSTRUCTION AND INVERSE PROBLEMS, in: *Optics in Astrophysics*, vol. 198, edited by: Foy, R. and Foy, F. C., Springer Netherlands, Dordrecht, 397–422, [https://doi.org/10.1007/1-4020-3437-7\\_25](https://doi.org/10.1007/1-4020-3437-7_25), 2006.
- Tie, X. and Cao, J.: Aerosol pollution in China: Present and future impact on environment, *Particuology*, 7, 426–431, <https://doi.org/10.1016/j.partic.2009.09.003>, 2009.
- Tinel, C., Testud, J., Pelon, J., Hogan, R. J., Protat, A., Delanoë, J., and Bouniol, D.: The Retrieval of Ice-Cloud Properties from Cloud Radar and Lidar Synergy, *J. Appl. Meteorol.*, 44, 860–875, <https://doi.org/10.1175/JAM2229.1>, 2005.
- Twomey, S.: The nuclei of natural cloud formation part II: The supersaturation in natural clouds and the variation of cloud droplet concentration, *Geophys. Pura E Appl.*, 43, 243–249, <https://doi.org/10.1007/BF01993560>, 1959.
- Twomey, S.: POLLUTION AND THE PLANETARY ALBEDO, *Atmos. Environ.*, 8, 1251–1256, 1974.
- Vaughan, M. A., Young, S. A., Winker, D. M., Powell, K. A., Omar, A. H., Liu, Z., Hu, Y., and Hostetler, C. A.: Fully automated analysis of space-based lidar data: an overview of the CALIPSO retrieval algorithms and data products, *Remote Sensing*, Maspalomas, Canary Islands, Spain, 16, <https://doi.org/10.1117/12.572024>, 2004.
- Veselovskii, I., Griaznov, V., Kolgotin, A., and Whiteman, D. N.: Angle- and size-dependent characteristics of incoherent Raman and fluorescent scattering by microspheres 2 Numerical Simulation, *Appl. Opt.*, 41, 5783, <https://doi.org/10.1364/AO.41.005783>, 2002.
- Veselovskii, I., Goloub, P., Podvin, T., Bovchaliuk, V., Derimian, Y., Augustin, P., Fourmentin, M., Tanre, D., Korenskiy, M., Whiteman, D. N., Diallo, A., Ndiaye, T., Kolgotin, A., and Dubovik, O.: Retrieval of optical and physical properties of African dust from multiwavelength Raman lidar measurements during the SHADOW campaign in Senegal, *Atmospheric Chem. Phys.*, 16, 7013–7028, <https://doi.org/10.5194/acp-16-7013-2016>, 2016.
-

- 
- Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Korenskiy, M., Pujol, O., Dubovik, O., and Lopatin, A.: Combined use of Mie–Raman and fluorescence lidar observations for improving aerosol characterization: feasibility experiment, *Atmospheric Meas. Tech.*, 13, 6691–6701, <https://doi.org/10.5194/amt-13-6691-2020>, 2020.
- Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Barchunov, B., and Korenskiy, M.: Combining Mie–Raman and fluorescence observations: a step forward in aerosol classification with lidar technology, *Atmospheric Meas. Tech.*, 15, 4881–4900, <https://doi.org/10.5194/amt-15-4881-2022>, 2022a.
- Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Barchunov, B., and Korenskiy, M.: Combining of Mie-Raman and fluorescence observations: a step forward in aerosol classification with lidar technology, *Aerosols/Remote Sensing/Data Processing and Information Retrieval*, <https://doi.org/10.5194/amt-2022-81>, 2022b.
- Voudouri, K. A., Siomos, N., Michailidis, K., Papagiannopoulos, N., Mona, L., Cornacchia, C., Nicolae, D., and Balis, D.: Comparison of two automated aerosol typing methods and their application to an EARLINET station, *Atmospheric Chem. Phys.*, 19, 10961–10980, <https://doi.org/10.5194/acp-19-10961-2019>, 2019.
- Wandinger, U.: Multiple-scattering influence on extinction- and backscatter-coefficient measurements with Raman and high-spectral-resolution lidars, *Appl. Opt.*, 37, 417, <https://doi.org/10.1364/AO.37.000417>, 1998.
- Wandinger, U., Freudenthaler, V., Baars, H., Amodeo, A., Engelmann, R., Mattis, I., Groß, S., Pappalardo, G., Giunta, A., D’Amico, G., Chaikovskiy, A., Osipenko, F., Slesar, A., Nicolae, D., Belegante, L., Talianu, C., Serikov, I., Linné, H., Jansen, F., Apituley, A., Wilson, K. M., de Graaf, M., Trickl, T., Giehl, H., Adam, M., Comerón, A., Muñoz, C., Rocadenbosch, F., Sicard, M., Tomás, S., Lange, D., Kumar, D., Pujadas, M., Molero, F., Fernández, A. J., Alados-Arboledas, L., Bravo-Aranda, J. A., Navas-Guzmán, F., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Preißler, J., Wagner, F., Gausa, M., Grigorov, I., Stoyanov, D., Iarlori, M., Rizi, V., Spinelli, N., Boselli, A., Wang, X., Lo Feudo, T., Perrone, M. R., De Tomasi, F., and Burlizzi, P.: EARLINET instrument intercomparison campaigns: overview on strategy and results, <https://doi.org/10.5194/amtd-8-10473-2015>, 9 October 2015.
- Wang, C., Soden, B., Yang, W., and Vecchi, G.: Compensation Between Cloud Feedback and AerosolCloud Interaction in CMIP6 Models, *Geophys. Res. Lett.*, 48, <https://doi.org/10.1029/2020GL091024>, 2021.
- Wang, N., Zhang, K., Shen, X., Wang, Y., Li, J., Li, C., Mao, J., Malinka, A., Zhao, C., Russell, L. M., Guo, J., Gross, S., Liu, C., Yang, J., Chen, F., Wu, L., Chen, S., Ke, J., Xiao, D., Zhou, Y., Fang, J., and Liu, D.: Dual-field-of-view high-spectral-resolution lidar: Simultaneous profiling of aerosol and water cloud to study aerosol–cloud interaction, *Proc. Natl. Acad. Sci.*, 119, e2110756119, <https://doi.org/10.1073/pnas.2110756119>, 2022.
- Waquet, F., Cornet, C., Deuzé, J.-L., Dubovik, O., Ducos, F., Goloub, P., Herman, M., Lapyonok, T., Labonnote, L. C., Riedi, J., Tanré, D., Thieuleux, F., and Vanbauce, C.: Retrieval of aerosol microphysical and optical properties above liquid clouds from POLDER/PARASOL polarization measurements, *Atmospheric Meas. Tech.*, 6, 991–1016, <https://doi.org/10.5194/amt-6-991-2013>, 2013.
- Weitkamp, C. (Ed.): *Lidar: range-resolved optical remote sensing of the atmosphere*, Springer, New York, 455 pp., 2005.
- Weitkamp, C.: *Lidar Range-Resolved Optical Remote Sensing of the Atmosphere*, Springer, 2006.
-

- 
- Wu, G., Fu, P., Ram, K., Song, J., Chen, Q., Kawamura, K., Wan, X., Kang, S., Wang, X., Laskin, A., and Cong, Z.: Fluorescence characteristics of water-soluble organic carbon in atmospheric aerosol☆, *Environ. Pollut.*, 268, 115906, <https://doi.org/10.1016/j.envpol.2020.115906>, 2021.
- Wu, Y., Chaw, S., Gross, B., Moshary, F., and Ahmed, S.: Low and optically thin cloud measurements using a Raman-Mie lidar, *Appl. Opt.*, 48, 1218, <https://doi.org/10.1364/AO.48.001218>, 2009.
- Yli-Tuomi, T., Venditte, L., Hopke, P. K., Basunia, M. S., Landsberger, S., Viisanen, Y., and Paatero, J.: Composition of the Finnish Arctic aerosol: collection and analysis of historic filter samples, *Atmos. Environ.*, 37, 2355–2364, [https://doi.org/10.1016/S1352-2310\(03\)00164-X](https://doi.org/10.1016/S1352-2310(03)00164-X), 2003.
- Young, S. A.: Analysis of lidar backscatter profiles in optically thin clouds, *Appl. Opt.*, 34, 7019, <https://doi.org/10.1364/AO.34.007019>, 1995.
- Yttri, K. E., Schnelle-Kreis, J., Maenhaut, W., Abbaszade, G., Alves, C., Bjerke, A., Bonnier, N., Bossi, R., Claeys, M., Dye, C., Evtyugina, M., García-Gacio, D., Hillamo, R., Hoffer, A., Hyder, M., Iinuma, Y., Jaffrezou, J.-L., Kasper-Giebl, A., Kiss, G., López-Mahia, P. L., Pio, C., Piot, C., Ramirez-Santa-Cruz, C., Sciare, J., Teinilä, K., Vermeylen, R., Vicente, A., and Zimmermann, R.: An intercomparison study of analytical methods used for quantification of levoglucosan in ambient aerosol filter samples, *Atmospheric Meas. Tech.*, 8, 125–147, <https://doi.org/10.5194/amt-8-125-2015>, 2015.
- Zardini, A. A., Sjogren, S., Marcolli, C., Krieger, U. K., Gysel, M., Weingartner, E., Baltensperger, U., and Peter, T.: A combined particle trap/HTDMA hygroscopicity study of mixed inorganic/organic aerosol particles, *Atmos Chem Phys*, 2008.
- Zhou, H. B. and Gao, J. T.: Automatic Method for Determining Cluster Number Based on Silhouette Coefficient, *Adv. Mater. Res.*, 951, 227–230, <https://doi.org/10.4028/www.scientific.net/AMR.951.227>, 2014.
- Zieger, P., Aalto, P. P., Aaltonen, V., Äijälä, M., Backman, J., Hong, J., Komppula, M., Krejci, R., Laborde, M., Lampilahti, J., de Leeuw, G., Pfüller, A., Rosati, B., Tesche, M., Tunved, P., Väänänen, R., and Petäjä, T.: Low hygroscopic scattering enhancement of boreal aerosol and the implications for a columnar optical closure study, *Atmospheric Chem. Phys.*, 15, 7247–7267, <https://doi.org/10.5194/acp-15-7247-2015>, 2015.
- Zieger, P. C.: Effects of relative humidity on aerosol light scattering, ETH Zurich, <https://doi.org/10.3929/ETHZ-A-006668068>, 2011.
-

# Appendix A: supplementary material



Figure Appendix 3.1: Silhouette coefficient from K-Means partitions on the training set containing all available data from 2021 and 2022, for a number of clusters ranging from 3 to 8

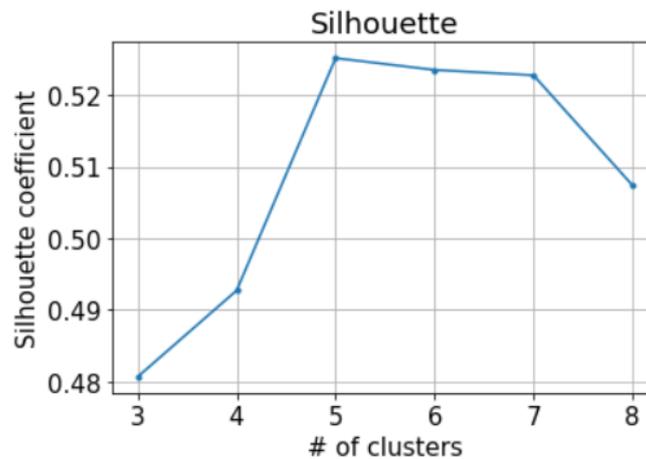


Figure Appendix 3.2: Silhouette coefficient from K-Means partitions on the training set containing data without the boundary layer from 2021, 2022 and 2023, for a number of clusters ranging from 3 to 8

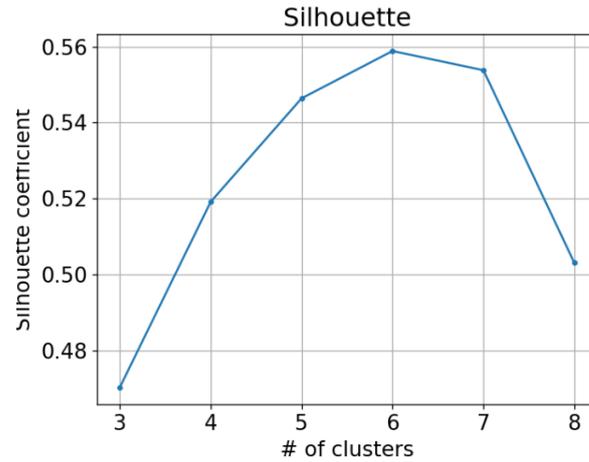


Figure Appendix 3.3: Silhouette coefficient from K-Means partitions on the training set containing data without the boundary layer from 2022 and 2023, for a number of clusters ranging from 3 to 8



Figure Appendix 3.4: Silhouette coefficient from K-Means partitions on the training set containing hand-selected data of pure cases, for a number of clusters ranging from 3 to 8

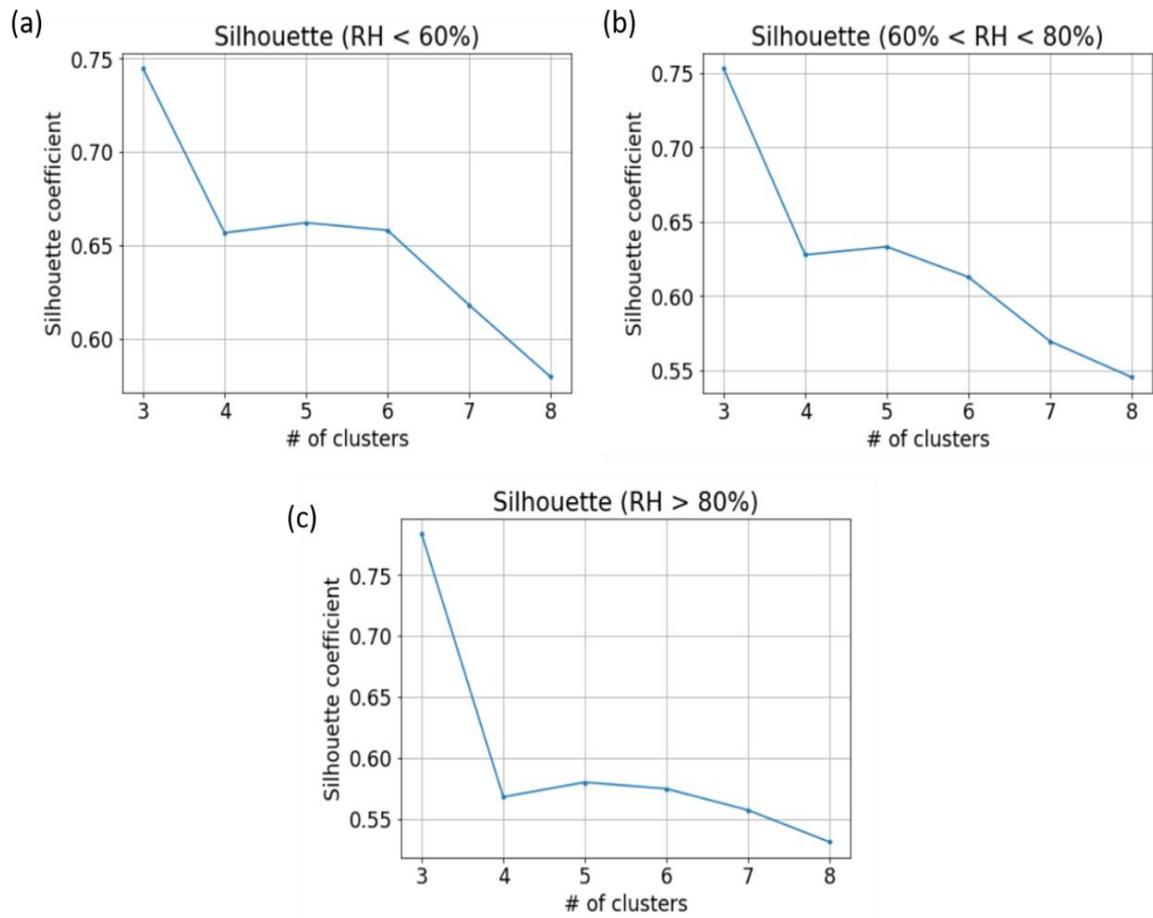


Figure Appendix 3.5: Silhouette coefficient from K-Means partitions for a number of clusters ranging from 3 to 8 on the sections of the training set (a) cases with RH < 60 %, (b) 60 % < RH < 80 %, (c) RH > 80 %

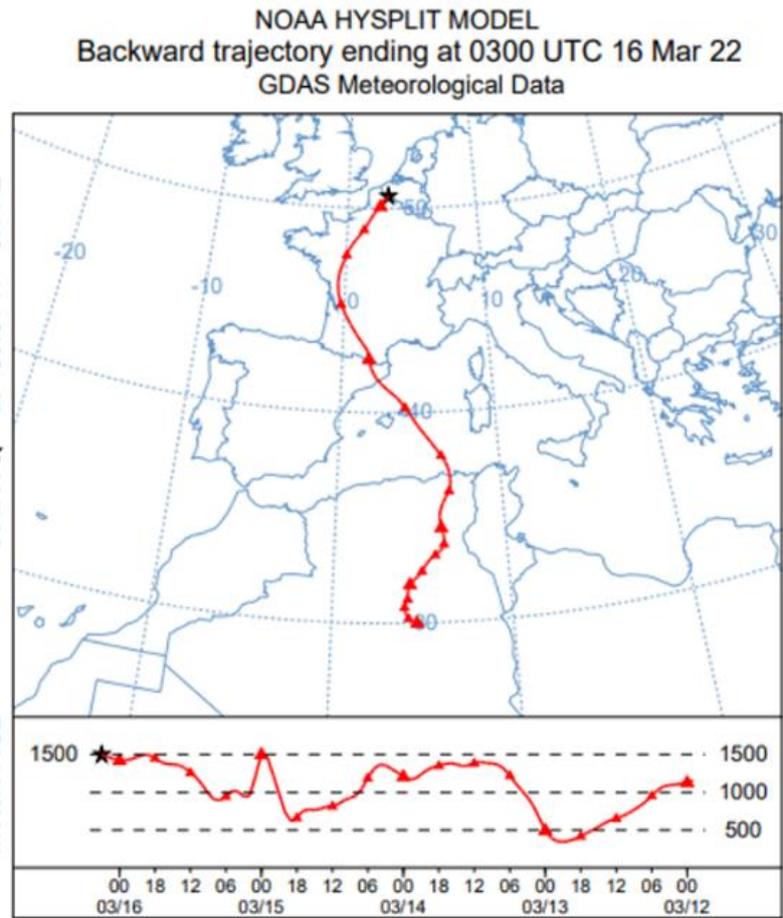


Figure Appendix 3.6: 100 hours backward trajectory at 1500 m above ground level at 03:00 UTC on 16 March 2022

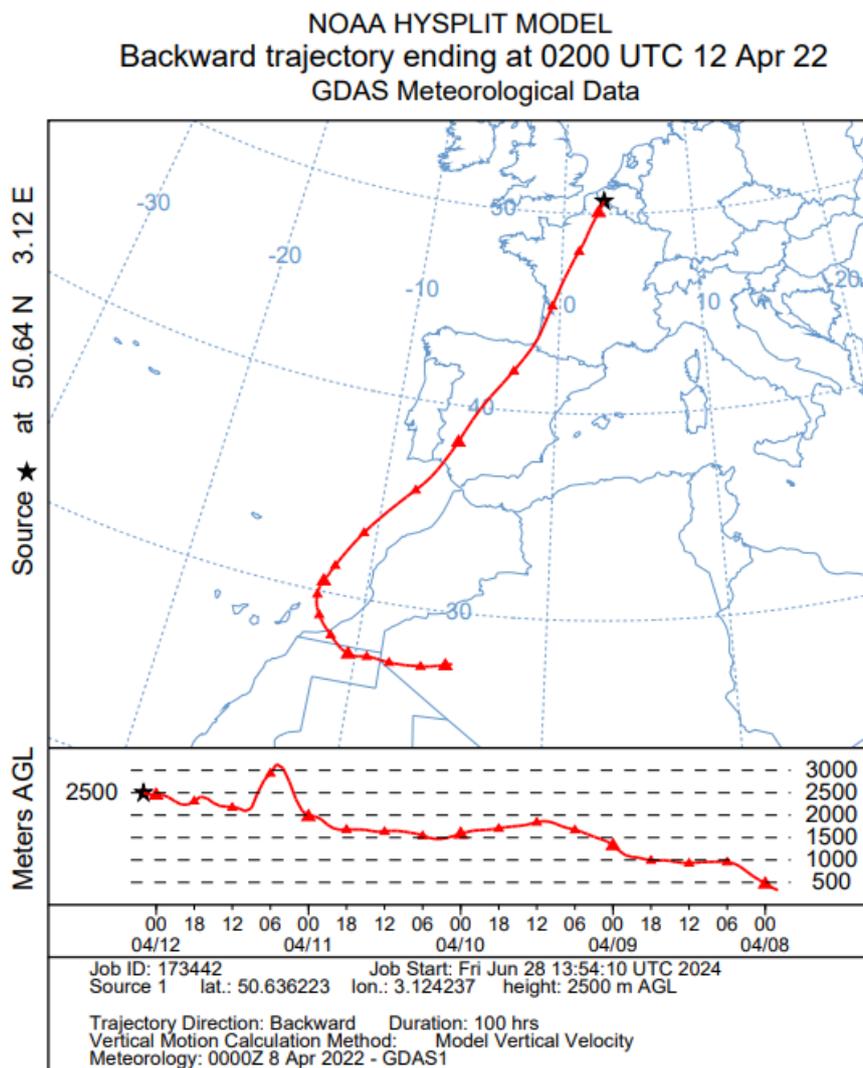


Figure Appendix 3.7: 100 hours backward trajectory at 2500 m above ground level at 02:00 UTC on 12 April 2022

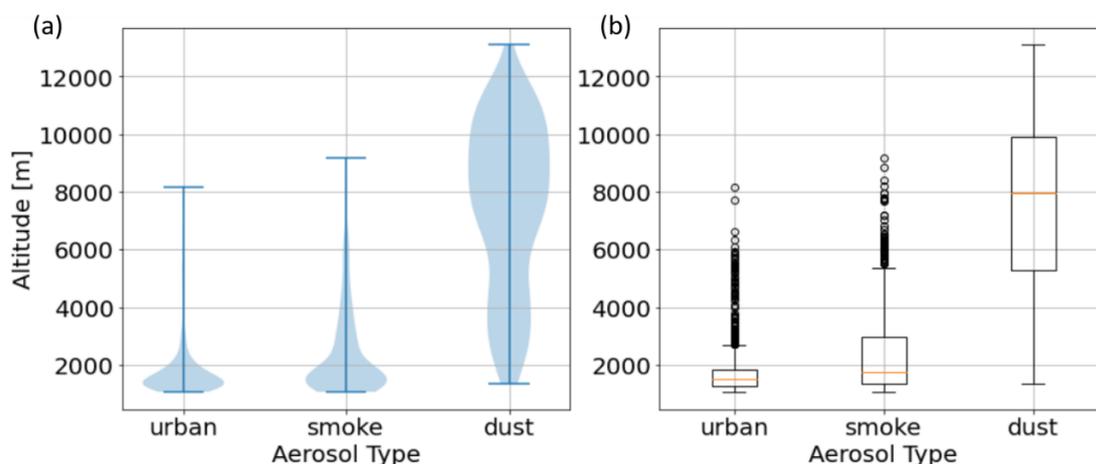


Figure Appendix 3.8: (a) Violin plots and (b) Box plots of averaged altitudes above the ground in function of the aerosol type for all available data from 2021 to 2023 between ground and 15 km above ground level

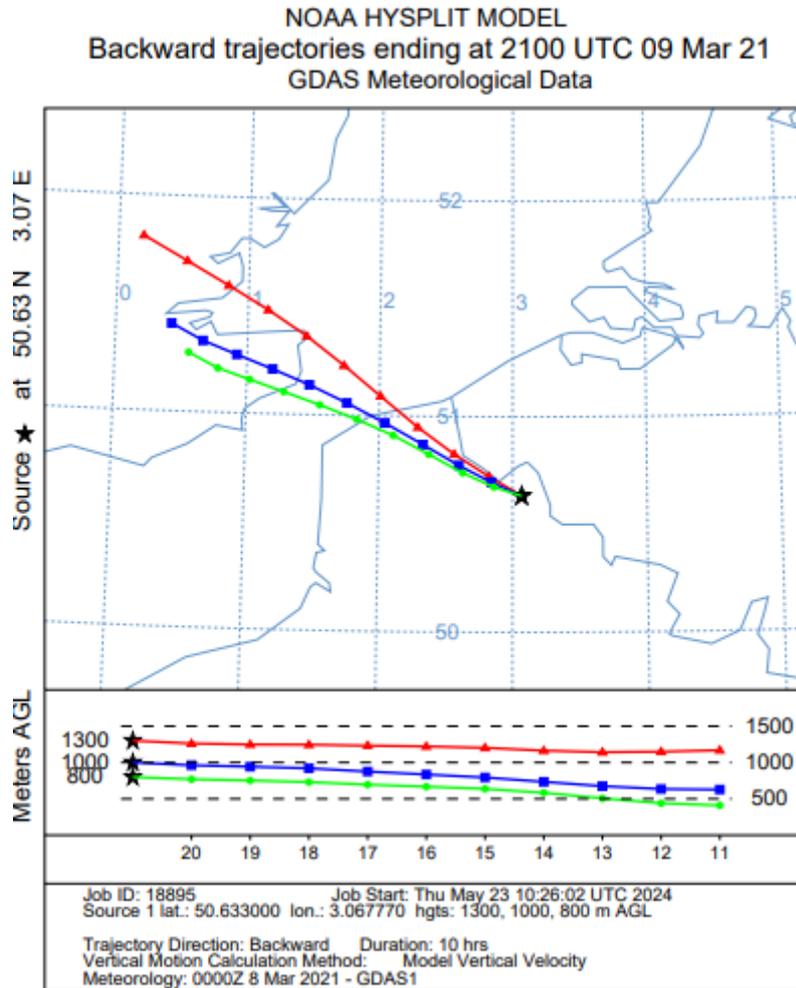


Figure Appendix 4.1: 10 hours backward trajectories computed for the boundary layer during the night of 9 to 10 March 2021 at 800 m, 1000 m, and 1300 m, at 22:00 UTC

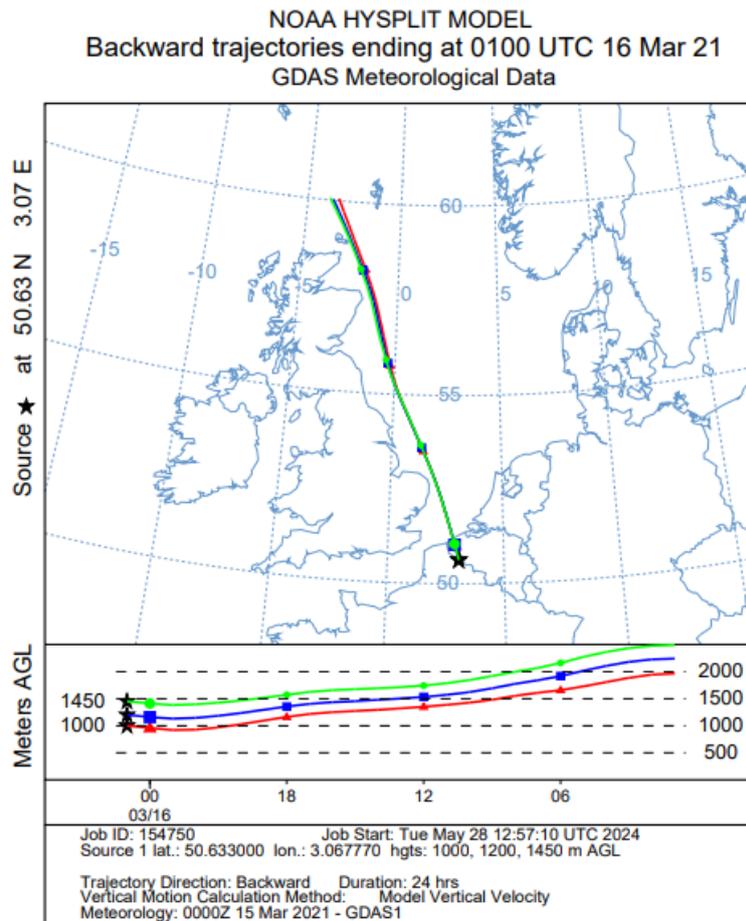


Figure Appendix 4.2: 10 hours backward trajectories computed for the boundary layer during the night of 15 to 16 April 2021 at 1000 m, 1200 m, and 1450 m, at 01:00 UTC

# **Appendix B: Article published in AMT**



# Innovative aerosol hygroscopic growth study from Mie–Raman–fluorescence lidar and microwave radiometer synergy

Robin Miri<sup>1</sup>, Olivier Pujol<sup>1</sup>, Qiaoyun Hu<sup>1</sup>, Philippe Goloub<sup>1</sup>, Igor Veselovskii<sup>2,3</sup>, Thierry Podvin<sup>1</sup>, and Fabrice Ducos<sup>1</sup>

<sup>1</sup>Univ. Lille, CNRS, UMR 8518 – LOA – Laboratoire d’Optique Atmosphérique, 59650 Villeneuve d’Ascq, France

<sup>2</sup>Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow 119991, Russia

<sup>3</sup>Cimel Electronique, 172 rue de Charonne, 75011 Paris, France

**Correspondence:** Robin Miri (robin.miri@univ-lille.fr)

Received: 10 January 2024 – Discussion started: 15 January 2024

Revised: 5 April 2024 – Accepted: 15 April 2024 – Published: 3 June 2024

**Abstract.** This study focuses on the characterization of aerosol hygroscopicity using remote sensing techniques. We employ a Mie–Raman–fluorescence lidar (Lille Lidar for Atmospheric Study, LILAS), developed at the ATOLL platform, Laboratoire d’Optique Atmosphérique, Lille, France, in combination with the RPG-HATPRO-G5 microwave radiometer to enable continuous aerosol and water vapor monitoring. We identify hygroscopic growth cases when an aerosol layer exhibits an increase in both aerosol backscattering coefficient and relative humidity. By examining the fluorescence backscattering coefficient, which remains unaffected by the presence of water vapor, the potential temperature, and the absolute humidity, we verify the homogeneity of the aerosol layer. Consequently, the change in the backscattering coefficient is solely attributed to water uptake. The Hänel theory is employed to describe the evolution of the backscattering coefficient with relative humidity and introduces a hygroscopic coefficient,  $\gamma$ , which depends on the aerosol type. The particularity of this method revolves around the use of the fluorescence which is employed to take into account and correct the aerosol concentration variations in the layer. Case studies conducted on 29 July and 9 March 2021 examine, respectively, an urban and a smoke aerosol layer. For the urban case,  $\gamma$  is estimated as  $0.47 \pm 0.03$  at 532 nm; as for the smoke case, the estimation of  $\gamma$  is  $0.5 \pm 0.3$ . These values align with those reported in the literature for urban and smoke particles. Our findings highlight the efficiency of the Mie–Raman–fluorescence lidar and microwave radiometer synergy in characterizing

aerosol hygroscopicity. The results contribute to advance our understanding of atmospheric processes, aerosol–cloud interactions, and climate modeling.

## 1 Introduction

Aerosols play a crucial role in our understanding of climate dynamics. Their impact on the radiation budget is classified into direct and semi-direct effects (Hansen et al., 1997; Thorsen et al., 2020), with additional contributions arising from aerosol–cloud interactions, commonly known as indirect effects. Certain aerosols can act as cloud condensation nuclei (CCN) or ice nucleating particle (INP), altering cloud properties including albedo and lifetime (Twomey and Warner, 1967). These complex processes remain significant challenges in the interpretation of the Earth energy balance. To advance our comprehension of aerosol–cloud interactions is crucial for improving climate models and accurately accounting for their influence on the energy balance of our planet. A key process in the understanding of these interactions is hygroscopic growth, which consists in aerosol uptake of water vapor in high relative humidity (RH) conditions, resulting in changes in size and, in some cases, chemical composition (Hänel, 1976). Hygroscopic growth efficiency varies depending on the aerosol type, with hydrophobic aerosols like dust and hydrophilic aerosols like marine particles (Chen et al., 2019, 2020). This variability is linked to their potential as CCN and INP, highlighting the importance of under-

standing the hygroscopic properties of aerosols (Dusek et al., 2006).

Hygroscopic growth properties of aerosols can be effectively investigated using a range of instruments. Traditionally, humidified nephelometers and spectrometers have been widely used to study aerosol hygroscopicity (Covert et al., 1972; Burgos et al., 2019). However, active remote sensing systems have tended to appear more advantageous since the last decade, as they allow us to measure with high vertical and temporal resolution without interfering with the observed system. Lidars, in particular, have gained prominence in remotely studying these properties (Feingold and Morley, 2003; Fernández et al., 2015; Granados-Muñoz et al., 2015; Zieger et al., 2015; Navas-Guzmán et al., 2019; Dawson et al., 2020; Düsing et al., 2021; Sicard et al., 2022, etc.) and offer several advantages compared to other methods. In particular, lidars provide high vertical and temporal resolution, allowing for detailed analysis of aerosol characteristics. Moreover, lidars offer the unique capability of simultaneously measuring aerosol properties and water vapor mixing ratio using a single instrument. At the Laboratoire d'Optique Atmosphérique (LOA) in Lille, France, the ATOLL platform (ATmospheric Observations in LiLLe) features a Mie–Raman–fluorescence lidar (Lille Lidar for Atmospheric Study, LILAS) employed in the frame of EARLINET/ACTRIS-FR (European Aerosol Research Lidar Network/Aerosols, Clouds, and Trace Gases Research Infrastructure – France). This multiwavelength lidar system measures elastic, depolarized, and Raman signals, providing comprehensive information on aerosol properties and water vapor. Additionally, LILAS captures aerosol fluorescence signal at 466 nm, which is triggered by the lidar UV wavelength at 355 nm. The fluorescence signal possesses distinctive characteristics that contribute to its utility in aerosol studies. Its intensity correlates with aerosol concentration and type, with biological aerosols like pollen or biomass burning smoke exhibiting higher fluorescence, while pure dust or urban aerosols demonstrate lower fluorescence. Furthermore, the fluorescence signal at 466 nm does not arise from pure water, enabling the extraction of aerosol-specific information without the influence of water vapor, which proves to be essential in studying aerosol hygroscopic growth (Veselovskii et al., 2020). In combination with an RPG-HATPRO-G5 microwave radiometer, also part of the ATOLL platform, it is possible to monitor both aerosol characteristics and water vapor, allowing us to study aerosol hygroscopicity.

The first part of this paper introduces the instruments and outlines a novel method for the study of aerosol hygroscopic growth using LILAS measurements. Following the instrument and method description, case studies are presented to demonstrate the efficiency and potential of the proposed approach. These case studies illustrate the practical implementation and feasibility of this innovative methodology, highlighting the added value brought by aerosol fluorescence

measurement in offering valuable insights into the hygroscopic growth characteristics of these aerosols. Finally, the paper concludes with a summary of the findings and offers comments on the obtained results. The conclusions will also discuss the potential further advancements and applications of the developed method, emphasizing its importance in enhancing our understanding of hygroscopic growth phenomena and its broader implications for atmospheric research.

## 2 Instrumentation and methodology

### 2.1 Experimental setup and data treatment

All the measurements presented in this paper were performed at the ATOLL platform in Lille (50.611° N, 3.138° E). The first instrument used in this study is the lidar LILAS. Its emission component consists of a tripled Nd:YAG laser operating at a repetition rate of 20 Hz, with a pulse energy of 70 mJ at 355 nm. The lidar system is configured in the  $3\beta + 2\alpha + 3\delta$  arrangement, meaning that it measures the elastic backscatter coefficient at three wavelengths (355, 532, and 1064 nm); it also measures the extinction at 355 and 532 nm, as well as the volume depolarization ratios for these wavelengths. This instrument also includes an additional channel for aerosol fluorescence detection, featuring a dedicated interference filter centered at 466 nm with a width of 44 nm. For this study, the aerosol elastic backscatter coefficients ( $\beta$ ) and the particulate linear depolarization ratio (PLDR) were computed at 532 nm from a Mie–Raman observation (Ansmann et al., 1992), due to the high signal-to-noise ratio at this wavelength in comparison with the two others. Furthermore, the detection part of the lidar includes a channel specifically designed to measure the vibrational–rotational Raman scattering of water at 408 nm, allowing for the retrieval of water vapor mixing ratio profiles (Ansmann et al., 1992; Whiteman et al., 1992). The obtained profiles were acquired during nighttime only and averaged over a period of 60 min. General details about the system can be found in Hu et al. (2018) and Veselovskii et al. (2020).

The proximity of the ATOLL platform to the airport prohibits the use of radiosounding. This poses a challenge for the inversion of water vapor using the LILAS lidar, as the computation of the instrumental constant requires a reference. Moreover, radiosoundings traditionally provide temperature profiles which are crucial for calculating RH but are difficult to obtain otherwise.

The second instrument used in this study is the RPG-HATPRO-G5 microwave radiometer, developed by RPG Radiometer Physics GmbH and present at the ATOLL platform, which provides integrated information like integrated water vapor content (IWV) or liquid water content (LWC) but also uses an integrated neural network model to retrieve atmospheric profiles of temperature, humidity, and liquid water. In situ sensors allow for ground level measurement of temper-

ature, humidity, and pressure. Finally, an infrared radiometer extension allows us to detect cloud base height and ice clouds. Data are acquired at high temporal resolution (every 10 min), and profiles range from the ground to 10 km (Louf et al., 2015).

This instrument has been considered to compensate for the lack of radiosounding measurement at the ATOLL platform for the calibration of the lidar water vapor measurement. Unfortunately, after considering using the radiometer humidity and temperature profiles for the lidar calibration, these ones turned out to be insufficiently accurate.

Consequently, temperature profiles from the ERA5 reanalysis database were also collected, and the IWV measurement of the radiometer has been used to calibrate the lidar.

In order to compute the instrumental constant necessary for the lidar water vapor calibration, the radiometer IWV has been compared to the integral of absolute humidity (AH) between the ground and 6 km height (above which humidity is negligible), as derived from the lidar-measured water vapor mixing ratio and the ERA5 temperature. Following the calibration procedure described in Foth et al. (2015), the calibration constant of the instrument is determined as the ratio between IWV and the integral value. The calibrated water vapor mixing ratio can be computed with

$$x_{\text{H}_2\text{O}}(h) = x'_{\text{H}_2\text{O}}(h) \text{IWV} \left[ \int_0^{z_{\text{max}}} x'_{\text{H}_2\text{O}}(h) \frac{P(h)}{R_a T(h)} dh \right]^{-1}, \quad (1)$$

where  $h$  is the height,  $x_{\text{H}_2\text{O}}(h)$  and  $x'_{\text{H}_2\text{O}}(h)$  are the calibrated and not-calibrated water vapor mixing ratios, respectively,  $z_{\text{max}}$  is equal to 6 km,  $P$  is the atmospheric pressure estimated with the hydrostatic approximation,  $R_a$  is the air perfect gas constant, and  $T$  is the temperature (all given in the International System of units). The calibration has been exclusively conducted under clear-sky conditions and taking into account the signal-to-noise ratio of the lidar's water vapor mixing ratio; if the signal-to-noise ratio on the profile is lower than 3, then the calibration constant is not computed. This threshold has been determined to ensure both data quality and a sufficient number of calibration constant computations. An interpolation has then been performed to estimate the calibration constants of cloudy and noisy situations.

## 2.2 Hygroscopic growth identification and study

In order to identify and analyze hygroscopic growth cases, a widely used method consists of searching for a homogeneous aerosol layer that spans either in time or altitude. When RH and elastic backscatter coefficient both increase, or decrease, it serves as a key indicator of hygroscopic growth. In the case of a homogeneous aerosol layer, the elastic backscatter coefficient evolution can then be attributed only to hygroscopic growth. This approach enables us to relate the elastic backscatter coefficient and RH, characterizing the hygroscopic properties of the considered aerosol particles.

The verification of the homogeneous nature of the considered aerosol layer is generally performed by investigating two key variables, absolute humidity, and potential temperature. Absolute humidity is investigated in order to identify any changes in the air mass that would lead to a change in the absolute humidity. A constant or decreasing potential temperature means that strong mixing is occurring within the layer, thus supporting the idea that the layer is homogeneous (Granados-Muñoz et al., 2015; Navas-Guzmán et al., 2019; Sicard et al., 2022).

The focus of this paper deals with the valuable insights provided by  $\beta_{\text{fluo}}$ . As stated in Veselovskii et al. (2020), a fluorescence signal emitted by aerosols at around 466 nm is not expected to be impacted by the presence of water, as pure water does not fluoresce. Therefore, by assuming that hygroscopic growth does not impact aerosol fluorescence and that the aerosol mixing state remains the same in the considered layer,  $\beta_{\text{fluo}}$  becomes a reliable proxy for monitoring the concentration of dry material within the aerosol layer. Under the hypothesis of a constant aerosol mixture and chemical composition in the layer, normalizing  $\beta_{532}$  by  $\beta_{\text{fluo}}$  enables the study of hygroscopic growth properties while also accounting for any possible changes in aerosol concentration within the layer.

Once the hygroscopic growth case has been identified, it becomes possible to examine the correlations between aerosol optical properties and RH. In this paper, particular attention has been given in the investigation of  $\beta_{532}$ . In order to explore this correlation efficiently, the Hänel parameterization has been used to express the changes in  $\beta_{532}$  as a function of RH. It introduces a parameter  $\gamma$ , known as the hygroscopic growth factor, which depends on the wavelength and the type of aerosol (Hänel, 1976). The Hänel parameterization is represented by

$$\frac{\beta_{532}(\text{RH})}{\beta_{532}(\text{RH}_{\text{ref}})} = \left( \frac{100 - \text{RH}}{100 - \text{RH}_{\text{ref}}} \right)^{-\gamma}, \quad (2)$$

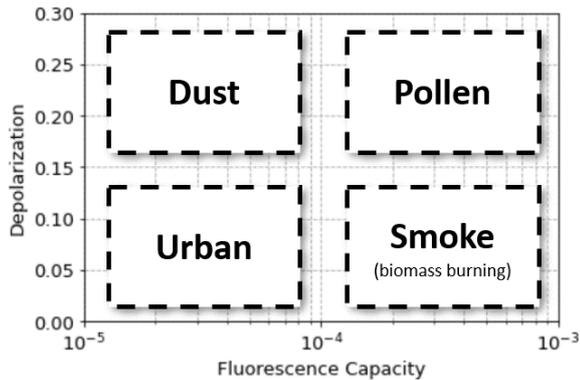
where  $\text{RH}_{\text{ref}}$  is the reference relative humidity. From this parameterization, it is possible to use  $\beta_{\text{fluo}}$  to account for aerosol concentration changes within the layer by normalizing  $\beta_{532}$ , so that

$$\frac{\beta_{532}(\text{RH})}{\beta_{532}(\text{RH}_{\text{ref}})} \frac{\beta_{\text{fluo}}(\text{RH}_{\text{ref}})}{\beta_{\text{fluo}}(\text{RH})} = \left( \frac{100 - \text{RH}}{100 - \text{RH}_{\text{ref}}} \right)^{-\gamma}. \quad (3)$$

We obtain an accurate estimation of the hygroscopic parameter  $\gamma$  by fitting the variation of  $\beta_{532}$  with respect to RH to the function as follows:

$$\beta_{532}(\text{RH}) = \beta_{532}(\text{RH}_{\text{min}}) \frac{\beta_{\text{fluo}}(\text{RH})}{\beta_{\text{fluo}}(\text{RH}_{\text{min}})} \left( \frac{100 - \text{RH}}{100 - \text{RH}_{\text{min}}} \right)^{-\gamma}, \quad (4)$$

where  $\text{RH}_{\text{min}}$  represents the minimum relative humidity observed within the analyzed aerosol layer. Subsequently, the estimations of  $\gamma$  values can be compared to hygroscopic growth parameter estimations from previous studies, the type



**Figure 1.** Schematic empirical aerosol typing repartition in function of the depolarization (PLDR) and fluorescence capacity (adapted from Veselovskii et al., 2022).

of the aerosol can be estimated from the optical properties, mainly its fluorescence capacity (the ratio between the fluorescence and elastic backscatter coefficients), and PLDR (Veselovskii et al., 2022) as shown Fig. 1. This comparative analysis offers valuable insights into how the hygroscopic growth of aerosols relates to their specific compositions and sources, contributing to a deeper understanding of aerosol behavior in changing environmental conditions.

### 3 Results and discussions

In order to experiment the potential of the hygroscopic growth study approach, this method has been tested on two potential hygroscopic growth cases, with the first occurring during 29 July 2021, averaged from 22:00 to 23:00 UTC, and the second occurring during 9 March 2021, averaged from 21:00 to 22:00 UTC.

#### 3.1 Hygroscopic growth study during the event on 29 July 2021 from 22:00 to 23:00 UTC

Figure 2 shows the profiles of the different measurements for the case study of 29 July 2021. In this particular scenario, both the water vapor mixing ratio and potential temperature exhibit relative stability, which are the two criteria commonly used to assess that the considered aerosol layer, are homogeneous (Granados-Muñoz et al., 2015; Navas-Guzmán et al., 2019; Sicard et al., 2022). Even though the potential temperature is derived from ERA5 reanalysis temperature profiles rather than direct measurements, this still provides a strong indication of the aerosol layer homogeneity. Moreover,  $\beta_{\text{fluo}}$  does not show strong variations within the defined region (the standard deviation of  $\beta_{\text{fluo}}$  in the considered layer is about 10 % of the average), further supporting this conclusion.

Conversely, there is an increase in both  $\beta_{532}$  and RH, suggesting a potential case of hygroscopic growth. RH rises

from 74 % to 96 %, which is a significant growth and strongly supports the hypothesis that hygroscopic growth occurs. Last, depolarization is not expected to decrease as it is already low, even if hygroscopic growth occurs.

Finally, it is possible to estimate the aerosol type of the considered layer by looking at its optical properties. Figure 3 shows the scatterplot of the PLDR at 532 nm and the fluorescence capacity, which is here the ratio between  $\beta_{\text{fluo}}$  and  $\beta_{532}$ . The considered aerosol layer exhibits low PLDR, as well as low fluorescence capacity characteristics, that allow it to be identified as an urban aerosol layer.

Figure 4 presents the outcomes of the fitting process for the relationship between  $\beta_{532}$  and RH using the Hänel parameterization. These results indicate a good fit to the Hänel parameterization in this particular case, as evidenced by the determination coefficient being close to 1 ( $R^2 = 0.91$ ). However, the estimated value of  $\gamma$ , which is expected to fall between 0.3 and 0.5 for a case of urban particles (Navas-Guzmán et al., 2019), is equal to 0.31 in this instance, which is very close to the lower limit for this type of aerosol. It also comes along with a slight divergence between the fit and the data, especially at high RH.

Several factors may contribute to the deterioration of the results and explain the low value of  $\gamma$ . First, it is possible that, in this case, there is merely no significant hygroscopic growth occurring for this particular type of aerosol within the observed range of RH. However, given the substantial RH variation, starting at 74 % and reaching up to 96 %, this hypothesis becomes less plausible.

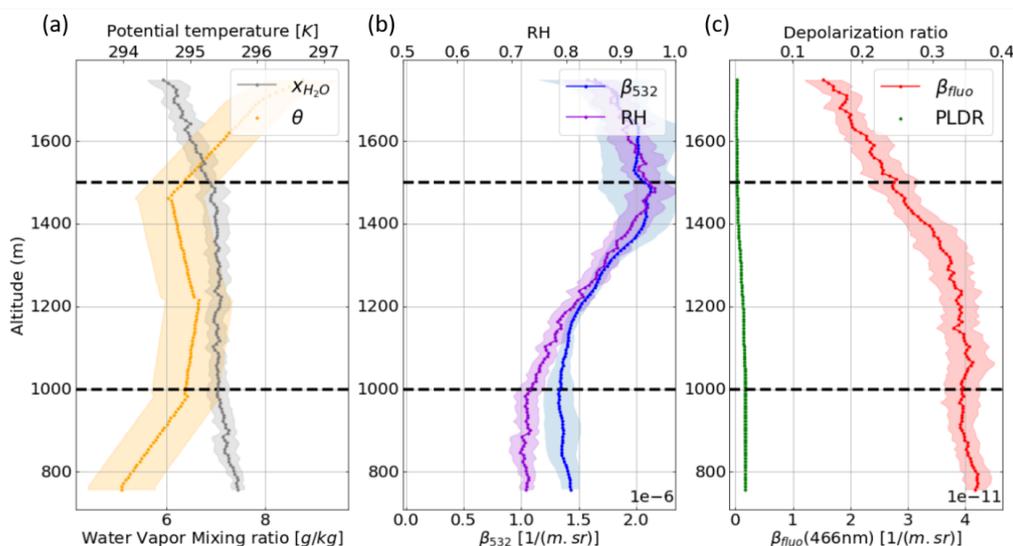
Second, it is possible that the main assumption on which this parameterization lies, i.e., constant aerosol concentration within the observed layer, may not hold true. Even with stable potential temperature and water vapor mixing ratio, there is a possibility that the aerosol concentration varies within the designated area, especially considering that potential temperature is derived from ERA5 reanalysis profiles and not directly measured. This potential aerosol concentration variation could potentially account for the low estimation of the hygroscopic growth factor.

In order to investigate this, we can assume that aerosol mixing remains constant within the study area and that  $\beta_{\text{fluo}}$  varies solely with changes in aerosol concentration. In doing so, it becomes possible to normalize  $\beta_{532}$  based on variations in aerosol concentration according to

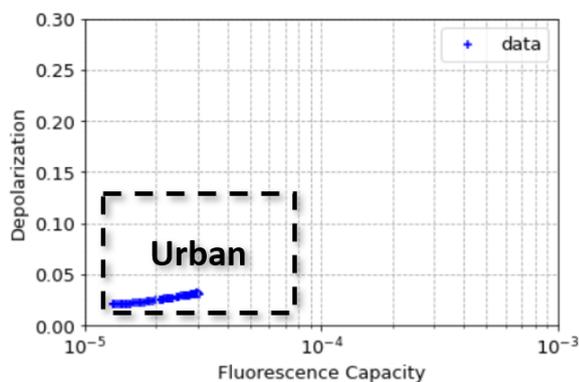
$$\overline{\beta_{532}}(\text{RH}) = \beta_{532}(\text{RH}) \frac{\beta_{\text{fluo}}(\text{RH}_{\text{min}})}{\beta_{\text{fluo}}(\text{RH})}. \quad (5)$$

Here,  $\overline{\beta_{532}}(\text{RH})$  is the normalized elastic backscatter coefficient, and  $\beta_{\text{fluo}}(\text{RH}_{\text{min}})$  is  $\beta_{\text{fluo}}$  at the minimum value of RH in the studied area.

It is now possible to apply the Hänel parameterization to  $\overline{\beta_{532}}$  instead of  $\beta_{532}$  in order to take into account the aerosol concentration variations within the layer and assess whether this normalization yields improved results.

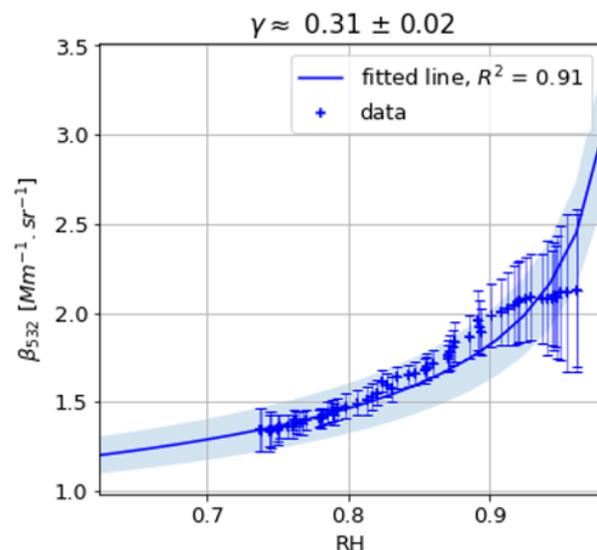


**Figure 2.** Profile of retrieved optical properties as a function of altitude above ground level. **(a)** Water vapor mixing ratio [ $\text{g kg}^{-1}$ ] and potential temperature [K], **(b)** elastic backscatter coefficient at 532 nm [ $\text{m}^{-1} \text{sr}^{-1}$ ] and RH, and **(c)** fluorescence backscatter coefficient at 466 nm [ $\text{m}^{-1} \text{sr}^{-1}$ ] and PLDR at 532 nm on 29 July 2021 from 22:00 to 23:00 UTC. The dashed black lines identify the area where hygroscopic growth is expected to occur.



**Figure 3.** Fluorescence capacity ( $\beta_{flu0}/\beta_{532}$ ) and PLDR at 532 nm between 1000 and 1500 m above ground level on 29 July 2021 from 22:00 to 23:00 UTC, characteristic of an urban aerosol layer.

The relationship between  $\overline{\beta_{532}}$  and RH, along with its fit to the Hänel parameterization presented in Fig. 5, demonstrates a significantly improved fit to the Hänel parameterization, with a notably higher determination coefficient ( $R^2 = 0.98$  instead of 0.91). Furthermore, the estimation of  $\gamma$  is found to be equal to  $0.47 \pm 0.03$ , falling precisely within the range of estimations conducted at 532 nm by previous studies (Navas-Guzmán et al., 2019; Sicard et al., 2022). These findings support the hypothesis that aerosol concentration varies within the aerosol layer and that such fluctuations are traceable through  $\beta_{flu0}$ , corroborating the efficiency of the presented approach for investigating hygroscopic growth phenomena.

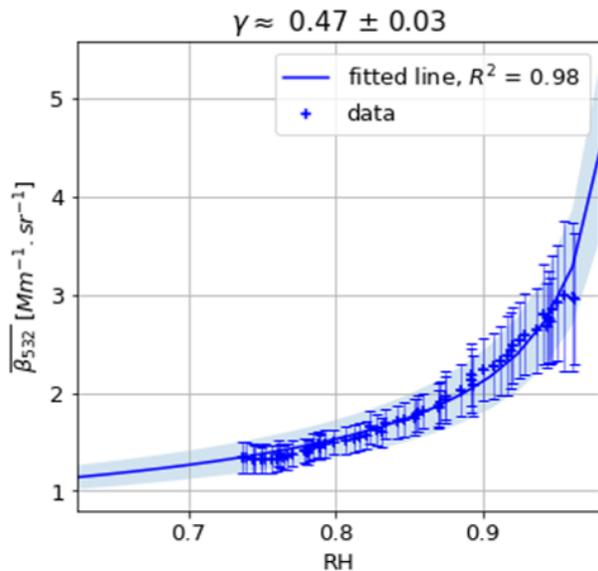


**Figure 4.** Evolution of the elastic backscatter coefficient at 532 nm as a function of RH on 29 July 2021 from 22:00 to 23:00 UTC between 1000 and 1500 m above ground level and the results of the fit to the Hänel parameterization.

### 3.2 Hygroscopic growth study during the event on 9 March from 21:00 to 22:00 UTC

Another case study can be presented to further support the validity of this approach. It is the case occurring on 9 March 2021 and averaged between 21:00 and 22:00 UTC.

Figure 6 shows the profiles of the different measurements for the case study on 9 March 2021. In this situation, both



**Figure 5.** Evolution of the normalized elastic backscatter coefficient at 532 nm (normalized with fluorescence backscatter coefficient) as a function of RH on 29 July 2021 from 22:00 to 23:00 UTC between 1000 and 1500 m above ground level and the results of the fit to the Hänel parameterization.

the water vapor mixing ratio and the potential temperature are relatively stable in the layer (potential temperature variations remain under 2 K, and the water vapor mixing ratio variations remain under  $1 \text{ g kg}^{-1}$ ), indicating a good mixing in the considered layer. An increase in both RH and  $\beta_{532}$  can also be noticed. On the other hand, there are fluctuations of  $\beta_{\text{fluo}}$ , mostly in the lower part of the layer, and the PLDR remains stable, but once again, given its already low value, it is not expected to decrease with hygroscopic growth. These elements together indicate that a hygroscopic growth scenario is most likely to occur in this layer.

The aerosol type can be investigated once again by looking at the fluorescence capacity and the PLDR. Both are represented Fig. 7 and show characteristics indicating that the aerosol layer comes from biomass burning smoke with low PLDR and strong fluorescence. Something worth noticing, however, is the low value of the fluorescence capacity. Indeed, the fluorescence capacity is the ratio between fluorescence backscatter coefficient and elastic backscatter coefficient. While the first is expected to remain stable with hygroscopic growth, the second increases in a high-humidity condition, consequently decreasing the fluorescence capacity and potentially leading to misclassification as indicated in Veselovskii et al. (2020). In this case, however, it is still possible to identify the biomass burning smoke aerosol layer.

The results of the fit to the Hänel parameterization on both  $\beta_{532}$  and  $\overline{\beta}_{532}$  (shown in Fig. 8) indicate a significant improvement brought by the normalization with the fluorescence. Without this process, the fit to the Hänel parameter-

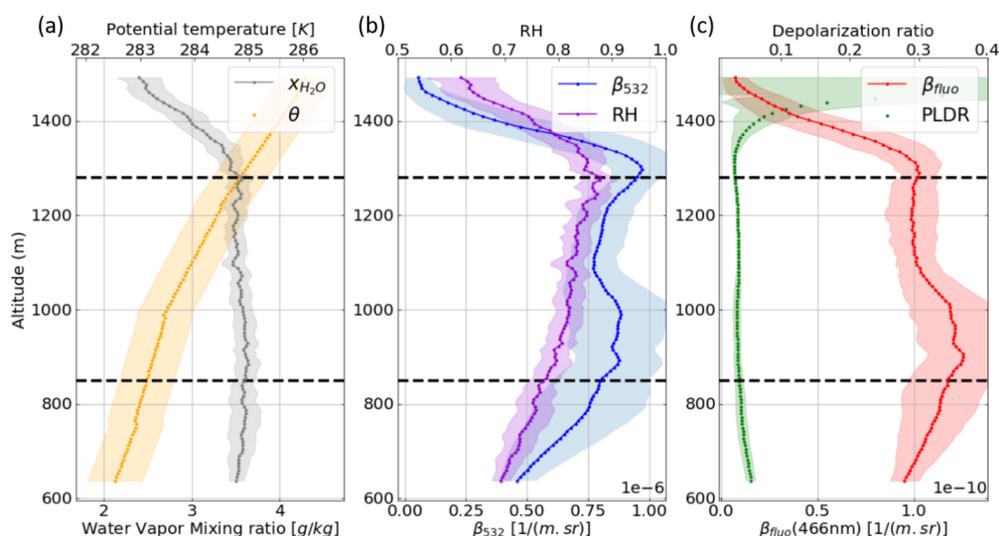
ization is extremely poor, with  $R^2 = -0.06$ . Furthermore, the estimation of the hygroscopic growth parameter is much lower than expected ( $\gamma = 0.1 \pm 0.3$ ), while the value is expected to fall around 0.5 at 532 nm for smoke aerosols according to Gomez et al. (2018). On the other hand, using the information given by the fluorescence to normalize the elastic backscatter coefficient, it is possible to obtain a much better fit to the Hänel parameterization, with  $R^2 = 0.93$  and a better estimation of the hygroscopic growth parameter, with  $\gamma = 0.5 \pm 0.3$  falling in the expected range for smoke aerosols. These findings suggest that it is indeed possible to use  $\beta_{\text{fluo}}$  to correct the variation in the aerosol concentration within the aerosol layer to study hygroscopic growth. The only drawback of this case lies in its high uncertainty.

An explanation for this high uncertainty could be instrumental noise, the span of RH covered being narrower than the first presented case (RH varying from 77.8 % to 87.6 %), or even the atmospheric variability being more important in this situation. Nevertheless, the point demonstrated in this analysis relies in the utility of the fluorescence correction for the hygroscopic factor estimation which is well emphasized here.

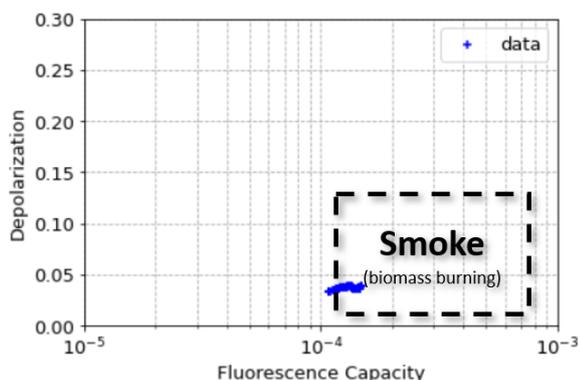
The influence of a shift in RH on  $\gamma$  has also been examined. For the case on 9 March 2021, when a bias of minus 0.1 is manually introduced to the RH, the corresponding  $\gamma$  estimation becomes 0.82, while a positive bias of 0.1 in RH results in a  $\gamma$  value of 0.23. The estimation of RH is based on both measurement from LILAS and the radiometer but also on ERA5 reanalysis data, which heavily rely on computational models. While this estimation provides valuable insights, it inherently introduces a level of uncertainty to the results. It is anticipated that the uncertainty associated with this estimation falls within the range of 10 %. The estimation of  $\gamma$  and the conclusions drawn from this estimation should then be considered with caution. Future studies might focus on refining the methods used for RH estimation, aiming at minimizing this inherent uncertainty and enhancing the accuracy of these findings. However, even if a shift in RH introduces high variability in  $\gamma$ , the determination coefficient  $R^2$  remains almost unchanged ( $R^2 = 0.92$  when RH is decreased by 0.1 and  $R^2 = 0.91$  for a 0.1 increase), meaning that the conclusions drawn on the use of the fluorescence correction are still valid in spite of the uncertainty in RH.

## 4 Conclusion

In this article, we have examined the possibility of using LILAS data for aerosol hygroscopic growth studies. The calibration of LILAS's water vapor channel has been addressed using thermodynamic data from the RPG-HATPRO-G5 microwave radiometer and temperature data from ERA5 reanalysis. A new approach to analyze aerosol hygroscopicity, relying on the fluorescence profiles measured by LILAS, has been developed and tested on two situations.



**Figure 6.** Profile of retrieved optical properties as a function of altitude above ground level. (a) Water vapor mixing ratio [ $\text{g kg}^{-1}$ ] and potential temperature [K], (b) elastic backscatter coefficient at 532 nm [ $\text{m}^{-1} \text{sr}^{-1}$ ] and RH, (c) fluorescence backscatter coefficient at 466 nm [ $\text{m}^{-1} \text{sr}^{-1}$ ] and PLDR at 532 nm on 9 March 2021 from 21:00 to 22:00 UTC. The dashed black lines identify the area where hygroscopic growth is expected to occur.



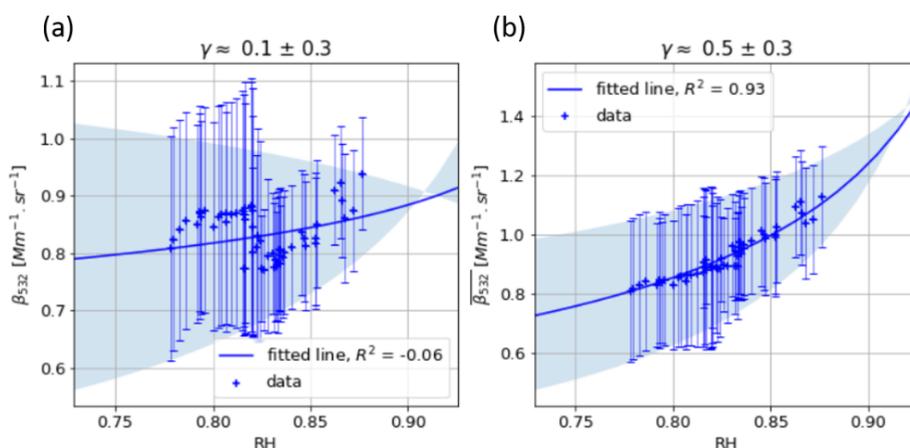
**Figure 7.** Fluorescence capacity ( $\beta_{fluo}/\beta_{532}$ ) and PLDR at 532 nm between 850 and 1280 m above ground level on 9 March 2021 from 21:00 to 22:00 UTC, characteristic of a biomass burning smoke aerosol layer.

The unique feature of the method presented in this article hinges on its use of the fluorescence backscatter coefficient. This coefficient serves as a weighting factor in tracking the evolution of aerosol concentration within the aerosol layer. Consequently, it leads to a significantly improved representation of the hygroscopic state evolution of the aerosols, thereby enhancing the characterization of the Hanel hygroscopic coefficient,  $\gamma$ . To validate this approach, evaluations were performed on two cases from July and March 2021, yielding promising results and highlighting the value brought by the fluorescence backscatter coefficient measurement with the lidar. In the first case, an estimation of  $\gamma$  of  $0.47 \pm 0.03$  with the fluorescence correction fell in the expected range

of the hygroscopic growth parameter for urban aerosols at 532 nm. In the second case, the estimation of  $\gamma$  is  $0.5 \pm 0.3$ , which, despite the higher uncertainty, is in the expected value for smoke particles at 532 nm and, most importantly, is a great improvement compared to the estimation carried on without the fluorescence correction.

In order to further increase the accuracy of our results, this method could be applied to a site featuring both fluorescence lidar measurement, as well as radiosoundings, in order to better estimate RH, a variable that significantly influences  $\gamma$  estimation and which is really complicated to estimate accurately otherwise. Based on the presented approach, values of  $\gamma$  can be calculated for various types of aerosols, and the assessment of the relationship between  $\gamma$  and aerosol optical properties like PLDR or fluorescence capacity can be considered. These relationships are expected to provide valuable insights for modeling interactions between aerosols and water vapor, serving as an initial step in studying aerosol–cloud interactions (Dusek et al., 2006; Petters and Kreideweis, 2007).

However, a current limitation of the present work arises in the identification of hygroscopic growth cases (which is done manually). Future efforts could focus on automatically identifying hygroscopic growth cases using lidar measurements, simplifying the study of  $\gamma$  dependency with aerosol parameters in a large number of situations (Gysel et al., 2007). From this perspective, an automatic classification method is also currently being developed, using a clustering approach, in order to automatically classify aerosol layers based on their optical properties, as well as thermodynamic parameters, accounting for humidity impact on the fluorescence capacity, as illustrated in the analysis of Fig. 7.



**Figure 8.** Evolution of the elastic backscatter coefficient at 532 nm (a) without normalization and (b) with normalization with the fluorescence backscatter coefficient and as a function of RH on 9 March 2021 from 21:00 to 22:00 UTC between 850 and 1280 m above ground level and the results of the fit to the Hanel parameterization.

Furthermore, the hygroscopic growth study will be adapted and improved for the LIFE lidar (laser-induced fluorescence explorer), which is anticipated to be operational by 2024. This upcoming lidar system is set to have more power and include additional fluorescence channels, thereby increasing the amount of information available, which will significantly enhance the retrieval performance.

*Code and data availability.* Data and code will be made available upon the request to the corresponding author.

*Author contributions.* RM analyzed the LILAS data, prepared the figures, and wrote the manuscript. PG, OP, IV, and QH supervised the work and contributed to the writing of the manuscript. PG, QH, IV, and TP designed, conceived, and ran the LILAS instrument. FD and QH developed and supported LILAS algorithms. OP is in charge of the RPG-HATPRO-G5 microwave radiometer.

*Competing interests.* The contact author has declared that none of the authors has any competing interests.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

*Acknowledgements.* We acknowledge the CaPPA project funded by the ANR through the PIA (under contract no. ANR-11-LABX-0005-01). The authors thank the Région Hauts-de-France, the Ministère de l'Enseignement Supérieur et de la Recherche, and the

European Fund for Regional Economic Development for their financial support to the CPER CLIMIBIO and ECRIN programs. The contribution from Qiaoyun Hu has been supported by Agence Nationale de Recherche ANR (grant no. ANR-21-ESRE-0013) through the OBS4CLIM project.

ChatGPT has been employed for drafting purposes in this document.

*Financial support.* This research has been supported by the Agence Nationale de la Recherche (grant nos. ANR-11-LABX-0005-01 and ANR-21-ESRE-0013).

*Review statement.* This paper was edited by Ulla Wandinger and reviewed by two anonymous referees.

## References

- Ansmann, A., Riebesell, M., Wandinger, U., Weitkamp, C., Voss, E., Lahmann, W., and Michaelis, W.: Combined raman elastic-backscatter LIDAR for vertical profiling of moisture, aerosol extinction, backscatter, and LIDAR ratio, *Appl. Phys. B-Lasers O.*, 55, 18–28, <https://doi.org/10.1007/BF00348608>, 1992.
- Burgos, M. A., Andrews, E., Titos, G., Alados-Arboledas, L., Baltensperger, U., Day, D., Jefferson, A., Kalivitis, N., Mihalopoulos, N., Sherman, J., Sun, J., Weingartner, E., and Zieger, P.: A global view on the effect of water uptake on aerosol particle light scattering, *Sci. Data*, 6, 157, <https://doi.org/10.1038/s41597-019-0158-7>, 2019.
- Chen, J., Li, Z., Lv, M., Wang, Y., Wang, W., Zhang, Y., Wang, H., Yan, X., Sun, Y., and Cribb, M.: Aerosol hygroscopic growth, contributing factors, and impact on haze events in a severely polluted region in northern China, *Atmos. Chem. Phys.*, 19, 1327–1342, <https://doi.org/10.5194/acp-19-1327-2019>, 2019.
- Chen, L., Peng, C., Gu, W., Fu, H., Jian, X., Zhang, H., Zhang, G., Zhu, J., Wang, X., and Tang, M.: On mineral dust

- aerosol hygroscopicity, *Atmos. Chem. Phys.*, 20, 13611–13626, <https://doi.org/10.5194/acp-20-13611-2020>, 2020.
- Covert, D. S., Charlson, R. J., and Ahlquist, N. C.: A study of the Relationship of Chemical Composition and Humidity to Light Scattering by Aerosols, *J. Appl. Meteorol.*, 11, 968–976, [https://doi.org/10.1175/1520-0450\(1972\)011<0968:ASOTRO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<0968:ASOTRO>2.0.CO;2), 1972.
- Dawson, K. W., Ferrare, R. A., Moore, R. H., Clayton, M. B., Thorsen, T. J., and Eloranta, E. W.: Ambient Aerosol Hygroscopic Growth From Combined Raman Lidar and HSRL, *J. Geophys. Res.-Atmos.*, 125, e2019JD031708, <https://doi.org/10.1029/2019JD031708>, 2020.
- Dusek, U., Frank, G. P., Hildebrandt, L., Curtius, J., Schneider, J., Walter, S., Chand, D., Drewnick, F., Hings, S., Jung, D., Borrmann, S., and Andreae, M. O.: Size Matters More Than Chemistry for Cloud-Nucleating Ability of Aerosol Particles, *Science*, 312, 1375–1378, <https://doi.org/10.1126/science.1125261>, 2006.
- Düsing, S., Ansmann, A., Baars, H., Corbin, J. C., Denjean, C., Gysel-Beer, M., Müller, T., Poulain, L., Siebert, H., Spindler, G., Tuch, T., Wehner, B., and Wiedensohler, A.: Measurement report: Comparison of airborne, in situ measured, lidar-based, and modeled aerosol optical properties in the central European background – identifying sources of deviations, *Atmos. Chem. Phys.*, 21, 16745–16773, <https://doi.org/10.5194/acp-21-16745-2021>, 2021.
- Feingold, G. and Morley, B.: Aerosol hygroscopic properties as measured by lidar and comparison with in situ measurements, *J. Geophys. Res.*, 108, 4327, <https://doi.org/10.1029/2002JD002842>, 2003.
- Fernández, A. J., Apituley, A., Veselovskii, I., Suvorina, A., Henzing, J., Pujadas, M., and Artíñano, B.: Study of aerosol hygroscopic events over the Cabauw experimental site for atmospheric research (CESAR) using the multi-wavelength Raman lidar Caeli, *Atmos. Environ.*, 120, 484–498, <https://doi.org/10.1016/j.atmosenv.2015.08.079>, 2015.
- Foth, A., Baars, H., Di Girolamo, P., and Pospichal, B.: Water vapour profiles from Raman lidar automatically calibrated by microwave radiometer data during HOPE, *Atmos. Chem. Phys.*, 15, 7753–7763, <https://doi.org/10.5194/acp-15-7753-2015>, 2015.
- Gomez, S. L., Carrico, C. M., Allen, C., Lam, J., Dabli, S., Sullivan, A. P., Aiken, A. C., Rahn, T., Romonosky, D., Chylek, P., Sevanto, S., and Dubey, M. K.: Southwestern U.S. Biomass Burning Smoke Hygroscopicity: The Role of Plant Phenology, Chemical Composition, and Combustion Properties, *J. Geophys. Res.-Atmos.*, 123, 5416–5432, <https://doi.org/10.1029/2017JD028162>, 2018.
- Granados-Muñoz, M. J., Navas-Guzmán, F., Bravo-Aranda, J. A., Guerrero-Rascado, J. L., Lyamani, H., Valenzuela, A., Titos, G., Fernández-Gálvez, J., and Alados-Arboledas, L.: Hygroscopic growth of atmospheric aerosol particles based on active remote sensing and radiosounding measurements: selected cases in southeastern Spain, *Atmos. Meas. Tech.*, 8, 705–718, <https://doi.org/10.5194/amt-8-705-2015>, 2015.
- Gysel, M., Crosier, J., Topping, D. O., Whitehead, J. D., Bower, K. N., Cubison, M. J., Williams, P. I., Flynn, M. J., McFiggans, G. B., and Coe, H.: Closure study between chemical composition and hygroscopic growth of aerosol particles during TORCH2, *Atmos. Chem. Phys.*, 7, 6131–6144, <https://doi.org/10.5194/acp-7-6131-2007>, 2007.
- Hänel, G.: The Properties of Atmospheric Aerosol Particles as Functions of the Relative Humidity at Thermodynamic Equilibrium with the Surrounding Moist Air, *Adv. Geophys.*, 19, 73–188, [https://doi.org/10.1016/S0065-2687\(08\)60142-9](https://doi.org/10.1016/S0065-2687(08)60142-9), 1976.
- Hansen, J., Sato, M., and Ruedy, R.: Radiative forcing and climate response, *J. Geophys. Res.*, 102, 6831–6864, <https://doi.org/10.1029/96JD03436>, 1997.
- Hu, Q., Goloub, P., and Dubovik, O.: École doctorale Sciences de la matière, N., du rayonnement et de l’environnement (Villeneuve d’Ascq, and atmosphérique (LOA), L. d’optique: Advanced Aerosol Characterization Using Sun/sky Photometer and Multi-wavelength Mie-Raman Lidar Measurements, PhD thesis, 2018.
- Louf, V., Pujol, O., Sauvageot, H., and Riédi, J.: Seasonal and diurnal water vapour distribution in the Sahelian area from microwave radiometric profiling observations, *Q. J. Roy. Meteor. Soc.*, 141, 2643–2653, <https://doi.org/10.1002/qj.2550>, 2015.
- Navas-Guzmán, F., Martucci, G., Collaud Coen, M., Granados-Muñoz, M. J., Hervo, M., Sicard, M., and Haeferle, A.: Characterization of aerosol hygroscopicity using Raman lidar measurements at the EARLINET station of Payerne, *Atmos. Chem. Phys.*, 19, 11651–11668, <https://doi.org/10.5194/acp-19-11651-2019>, 2019.
- Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmos. Chem. Phys.*, 7, 1961–1971, <https://doi.org/10.5194/acp-7-1961-2007>, 2007.
- Sicard, M., Fortunato dos Santos Oliveira, D. C., Muñoz-Porcar, C., Gil-Díaz, C., Comerón, A., Rodríguez-Gómez, A., and Dios Otín, F.: Measurement report: Spectral and statistical analysis of aerosol hygroscopic growth from multi-wavelength lidar measurements in Barcelona, Spain, *Atmos. Chem. Phys.*, 22, 7681–7697, <https://doi.org/10.5194/acp-22-7681-2022>, 2022.
- Thorsen, T. J., Ferrare, R. A., Kato, S., and Winker, D. M.: Aerosol Direct Radiative Effect Sensitivity Analysis, *J. Climate*, 33, 6119–6139, <https://doi.org/10.1175/JCLI-D-19-0669.1>, 2020.
- Twomey, S. and Warner, J.: Comparison of Measurements of Cloud Droplets and Cloud Nuclei, *J. Atmos. Sci.*, 24, 702–703, 1967.
- Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Korenskiy, M., Pujol, O., Dubovik, O., and Lopatin, A.: Combined use of Mie–Raman and fluorescence lidar observations for improving aerosol characterization: feasibility experiment, *Atmos. Meas. Tech.*, 13, 6691–6701, <https://doi.org/10.5194/amt-13-6691-2020>, 2020.
- Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Barchunov, B., and Korenskiy, M.: Combining Mie–Raman and fluorescence observations: a step forward in aerosol classification with lidar technology, *Atmos. Meas. Tech.*, 15, 4881–4900, <https://doi.org/10.5194/amt-15-4881-2022>, 2022.
- Whiteman, D. N., Melfi, S. H., and Ferrare, R. A.: Raman lidar system for the measurement of water vapor and aerosols in the Earth’s atmosphere, *Appl. Optics*, 31, 3068–3082, <https://doi.org/10.1364/AO.31.003068>, 1992.
- Zieger, P., Aalto, P. P., Aaltonen, V., Äijälä, M., Backman, J., Hong, J., Komppula, M., Krejci, R., Laborde, M., Lampilahti, J., de Leeuw, G., Pfüller, A., Rosati, B., Tesche, M., Tunved, P., Väänänen, R., and Petäjä, T.: Low hygroscopic scattering enhancement of boreal aerosol and the implications for a columnar optical closure study, *Atmos. Chem. Phys.*, 15, 7247–7267, <https://doi.org/10.5194/acp-15-7247-2015>, 2015.

## *Résumé de la thèse*

Cette thèse explore l'utilisation des signaux de fluorescence mesurés par le lidar LILAS pour étudier les aérosols et leurs interactions avec la vapeur d'eau et les nuages. D'abord, un algorithme de classification automatique des aérosols, basé sur un modèle d'apprentissage automatique, a été développé. Ensuite, la fluorescence a été exploitée pour améliorer l'estimation des propriétés hygroscopiques des aérosols. Indépendante de la présence d'eau, la fluorescence permet de suivre la concentration d'un aérosol, permettant d'en compenser les variations pour une meilleure détermination des propriétés hygroscopiques. Enfin, le sujet des signaux de fluorescence dans les nuages a été abordé. Une méthode est proposée pour extraire des informations sur les aérosols présents dans la couche nuageuse à partir de leurs signaux de fluorescence. Cette inversion comporte de grandes incertitudes et repose sur des hypothèses fortes, mais elle offre une perspective prometteuse pour l'étude des aérosols en conditions nuageuses.

**Mots-clés :** Télédétection, Aérosols, Nuages, Lidar

## *Summary of the thesis*

This thesis explores the use of fluorescence signals measured by the LILAS lidar to study aerosols and their interactions with water vapor and clouds. First, an automatic aerosol classification algorithm based on a machine learning model was developed. Next, fluorescence was utilized to improve the estimation of aerosol hygroscopic properties. Independent of the presence of water, fluorescence allows for tracking aerosol concentration, which helps compensate for variations and thus better determine hygroscopic properties. Finally, the topic of fluorescence signals in clouds was addressed. A method is proposed to extract information about aerosols within the cloud layer based on their fluorescence signals. While this inversion involves significant uncertainties and relies on strong assumptions, it offers a promising perspective for studying aerosols in cloudy conditions.

**Key words :** Remote sensing, Aerosols, Clouds, Lidar