

Atmospheric Imaging with Differential Absorption Lidar Off-beam Returns



THE UNIVERSITY
of EDINBURGH

Robert Lung

School of Engineering

Institute for Imaging, Data and Communications

This dissertation is submitted for the degree

Doctor of Philosophy

December 2023

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text, references and acknowledgments.

December 2023,

Robert Lung

Acknowledgements

I would like to acknowledge the advice and support of my advisor Dr. Nicholas Polydorides, who suggested the topic of this research and has been very supportive in the pursuit and development of my own ideas. The work presented here has without a doubt benefited from our discussions.

I would also like to acknowledge the financial support of the James Clerk Maxwell Foundation and the School of Engineering who were very accommodating in the difficult times during the lockdown caused by the global pandemic.

Finally, I would like to thank my examiners, Prof. Simon Arridge and Prof. Konstantinos Zygalakis, who provided thoughtful suggestions and comments that have significantly contributed to the quality and clarity of the work in this thesis.

Abstract

Environmental monitoring has become increasingly important in recent years. The quantification and analysis of particle concentrations in the atmosphere are not only fundamental to our understanding of global climate but also have applications in hazard assessment after chemical accidents where they can be critical in mitigating the risk for first responders and those in the vicinity of the event. The underpinning three-dimensional imaging problem in which one seeks to determine the concentration of gas at every point within a given spatial domain is especially challenging because the typically large scales of these gas plumes make individual concentration measurements at each location virtually impossible. Although remote sensing modalities can overcome this challenge by being able to yield information about a particular region without having to make physical contact with it, the recorded data is typically corrupted by noise, caused either by the instrument itself or the inability to control the dynamic nature of the atmosphere, and are thus accompanied by their own set of difficulties. The primary subject of this work is the inverse problem of fitting parameters related to atmospheric dispersion along with the associated image, based on time-resolved back-scattered differential absorption Lidar (DIAL), an optical remote sensing modality suited for measurements of small molecules with narrow absorption features. In contrast to other state-of-the-art DIAL methods and motivated by the fact that atmospheric dispersion behaves in a way such that long-term averages inevitably have diminishing sensitivity to certain gas features, the problem is approached without making a single scattering assumption but rather a new modality that includes the collection of multiply scattered photons from wider/multiple fields-of-view (FOV) is

proposed. It is argued that this data when paired with a time-dependent radiative transfer equation (RTE) model, can be beneficial for the reconstruction of certain key aspects of the image. The resulting inverse problem is formulated as a semi-parametric statistical model in which the quantity of interest is reduced to a small number of dispersion-related parameters accompanied by a high-dimensional but computationally convenient nuisance component. This not only effectively avoids a high-dimensional inverse problem but simultaneously provides a natural regularisation mechanism along with parameters that are directly related to the dispersion model. These can be associated with meaningful physical units while spatial concentration profiles can be obtained employing forward evaluation of the dispersion process. The typically high cost associated with RTE-based inverse problems is overcome with the development of a randomised solver which makes use of concentration properties for smooth low-dimensional functions. This facilitates fast estimation of the gas concentration and at the same time paves the way for quantifying the uncertainties within those estimates.

Lay summary

It is a well-established fact that many gases absorb light at certain wavelengths. If the gas of interest is known, then this observation can be exploited to determine its concentration by measuring the intensity of light after it passes through a region of interest and comparing it to a reference that corresponds to the absence of gas. This quantitative relationship, also known as Beer-Lambert's law, is at the core of many modern instruments for the quantification of gases not just in a controlled laboratory setting but also in the atmosphere. One such technique is Differential Absorption Lidar (DIAL) which uses a time-resolving detector that can determine how far a photon has travelled within the atmosphere before it was scattered into the detector. It is paired with a pulsed laser located nearby operating at two wavelengths, one acting as a reference while the other is being absorbed.

It should not come as much of a surprise that when the measurements are taken out in the open, not only is the region of interest typically very large, which can result in relatively weak signals recorded by the detector, but many other variables will influence the measurement, yet cannot be controlled through the experimental setup. Modern DIAL instruments are typically designed in a way to avoid many such effects by using very precise lasers and sophisticated detectors which can "remove" several potential sources of uncertainty, in particular light that has undergone multiple scattering. As the trajectory of a photon that is scattered only once must be straight, this results in data that is in theory very precise and yields high resolution by avoiding the blurriness typically caused by multiple scattering. In practice, this means that less light is collected and therefore

comes at the cost of having to average the data over prolonged periods of time.

This work, which explores the possibility of using light that has been scattered more than once in the reconstruction process rather than discarding it, is motivated by the observation that DIAL typically uses measurements taken over relatively long periods. Virtually all plumes in the atmosphere will, due to turbulence, change throughout the measurement period which will result in a qualitatively similar blurring as multiple scattering would cause. Studies of turbulence have revealed that, if observed over long periods of time, the effects average out and may result in simpler structures that preserve only certain low-resolution aspects of the instantaneous concentration profiles.

To formalise this motivation this thesis introduces a framework that can describe gas plumes through such features and how the differential absorption measurements are impacted by the atmospheric and optical quantities that affect the measurement. It is essentially shown that it isn't necessary to remove light that scattered more often unless a very high resolution image is sought which in turn requires very strong lasers. On the flip side, this implies a potential reduction in the error compared to a state-of-the-art like-for-like measurement in instances where the concentration is sufficiently low to necessitate extensive averaging. A major challenge that is often encountered when image reconstruction is performed in conjunction with scattering is the actual computation of the images, or in this case gas concentration profiles. In this work the problem is mitigated based on the same ideas that were used to motivate it, i.e. it is possible to use fast approximations based on random sampling as long as only high-level features of the image need to be reconstructed.

Symbols & abbreviations

Reoccurring symbols

X	spatial domain
Θ	space of feasible values for $\theta \in \Theta \subseteq \mathbb{R}^{N_\theta}$
θ	finite-dimensional parameter of interest, $N_\theta := \dim(\theta)$ (Eq. (2.11))
θ_0	θ -component, controls differential absorption weight (Eqs. (3.28) and (5.44))
θ_s	θ -component, controls scattering weight (Eq. (5.44))
θ_{source}	θ -component, contains source coordinates (Eq. (5.44))
θ_{up}	θ -component, controls upwards drift (Eqs. (5.44) and (5.45))
θ_h	θ -component, controls plume width (Eqs. (5.44) and (5.46))
θ_α	sub-component $\theta_\alpha = (\theta_0, \theta_{\text{source}}, \theta_{\text{up}}, \theta_h)$, determines α (Eq. (5.44))
x_D	light source/detector location (Fig. 3.1)
ϕ	kernel function (Definition 2.1.1)
$K(\phi \cdot)$	kernel space (Definition 2.1.1)
u_δ	instantaneous puff (Eqs. (2.11) and (2.13))
u	plume obtained from prolonged release (Eq. (2.18))
κ	diffusion coefficient (Eqs. (2.11) and (2.13))
κ_0	isotropic diffusion coefficient (Eqs. (2.15) and (2.16))
η	drift coefficient (Eq. (2.11))
$\vec{\eta}_{\text{amb}}$	measured/known wind vector (Eq. (2.11))
ρ	point source of plume located at b^0 of strength ρ_0 (Eq. (2.12))
ρ_0	strength of point source ρ (Eq. (2.12))
b^0	location of point source ρ (Eq. (2.12))
b	plume/kernel centre (Eqs. (2.14), (2.16), (2.18), and (2.19))
h	plume/kernel width parameter (Eqs. (2.15), (2.16), (2.18), and (2.19))
w	plume/kernel weights (Eqs. (2.16), (2.18), and (2.19))
\mathbf{T}	turbulence process (Appendix A.2.2 contains an example)
\mathbf{R}	turbulence ratio (Eq. (4.64))

σ_a	non-differential absorption coefficient (Eq. (2.22))
σ_s	non-differential scattering coefficient (Eq. (2.22))
f_p	phase function (Eq. (2.22))
α	differential absorption coefficient (Eq. (3.3))
ξ	complete set of optical parameters $\xi = (\alpha, \sigma_a, \sigma_s, f_p)$ (Eqs. (3.8) and (3.10))
Ξ	feasible space for optical parameters $\xi \in \Xi$ (Eqs. (3.8) and (3.10))
ξ_{off}	off parameter $\xi_{\text{off}} = (\sigma_a, \sigma_s, f_p)$ (Eq. (3.9))
$\Xi_{\text{off}}(\alpha)$	feasible space $\xi_{\text{off}} \in \Xi_{\text{off}}(\alpha) \iff (\alpha, \xi_{\text{off}}) \in \Xi$ (Eq. (3.9))
ψ	nuisance parameter (Eqs. (3.12a) to (3.12c))
\mathcal{G}	dispersion forward operator $\Theta \rightarrow \Xi$ (Eq. (3.10))
\mathcal{M}	optical forward operator $\Xi \rightarrow C(Z_{(-)}(x_D)) \times C(Z_{(-)}(x_D))$ (Eq. (3.8))
\mathcal{Q}	differential forward operator $\Xi \rightarrow C(Z_{(-)}(x_D))$ (Eq. (4.1) and Lemma 4.2.3)
$Z_{(-)}(x_D)$	observable domain from x_D (Eq. (3.7))
$Z_{[-]}(x_D)$	sensing domain, compact subset of $Z_{(-)}(x_D)$ (Eqs. (4.9) and (4.12))
$\mathbf{m}_{\text{on/off}}$	measurement intensity (Lemma 3.1.1 and Eq. (3.8))
$\bar{q}[\theta]$	detection sensitivity (Eq. (3.31))
$\bar{L}[\theta]$	detection precision (Eq. (3.30))
$\mathbf{m}^{\text{on/off}}$	discretised noisy on/off observations (Eq. (3.15))
\mathbf{y}	discretised noisy DIAL data (Eq. (3.25))
\mathbf{z}	discretised noisy residual (Eq. (3.26))
$\mathbf{M}^{\text{on/off}}$	continuous noisy on/off observations (Definition 4.1.2 and Eq. (4.8))
$\Lambda_{\text{on/off}}$	on/off intensity measure (Eq. (4.8))
β	non-parametric differential deviation (Definition 4.2.1)
$\partial_u^i D$	detection sets for $i \in \{\pm, (0, \pm), \emptyset\}$ (Definition 4.2.2)
$B_{\infty,0}(\alpha)$	unobscured β constraint (Eq. (4.35))
$B_\infty(\alpha)$	obscured β constraint (Eq. (4.36))

Abbreviations

Lidar	light detection and ranging
ToF	time of flight
DOAS	differential optical absorption spectroscopy
DIAL	differential absorption Lidar
on/off	online (absorbing) and offline (non-absorbing) wavelengths in DIAL
FOV	field of view
nFOV	narrow field of view (collects single scattering)
wFOV	wide field of view (collects single and multiple scattering in aggregated form)
mFOV	multiple fields of view (collect single and multiple scattering individually)
PDE	partial differential equation
RTE	radiative transfer equation
HG	Henyey-Greenstein
MLE	maximum likelihood estimator
MC	Monte-Carlo
UQ	uncertainty quantification
SDE	stochastic differential equation
PPP	Poisson point process

Contents

Symbols & abbreviations	vii
1 Introduction	1
1.1 Optical inverse problems	1
1.1.1 Absorption spectroscopy, remote sensing & DIAL	1
1.1.2 Multiple scattering & ill-posedness	5
1.2 Particle transport in the atmosphere	7
1.2.1 Plume models	7
1.2.2 Model reduction & regularity	9
1.2.3 Probabilities, entropy & concentration of measure	9
1.3 Problem statement & contributions	11
1.4 Outline	14
2 Notation & mathematical fundamentals	15
2.1 Generic problem description	15
2.2 Advection-diffusion based dispersion models	20
2.3 Radiative transfer	27
3 Point estimation	34
3.1 DIAL type measurements with wide FOVs	34
3.1.1 Wide FOVs & angularly averaged measurements	35
3.1.2 Semi-parametric form of the RTE-based inverse problem	39
3.2 Signal quality and noise considerations	47

3.2.1	Poisson noise model	48
3.2.2	Trivial edge cases	51
3.2.3	Weak absorption errors	54
4	Consistency & asymptotic behaviour	67
4.1	Continuity & convergence of the estimators	68
4.2	Convergence Rates	83
4.3	Modelling discrepancies & estimation bias	98
4.4	Uncertainty quantification for estimates in Θ	102
5	Algorithms & experiments	111
5.1	Computational modelling and reconstruction	111
5.1.1	Variational form for regularised DIAL	111
5.1.2	Computational complexity	120
5.1.3	Randomisation induced errors	132
5.1.4	Image approximation errors	138
5.2	A numerical example	143
5.2.1	Bias & convergence rates	143
5.2.2	Steady state plume & turbulence	147
5.2.3	Uncertainty estimation	153
6	Conclusion & outlook	161
A	Additional mathematical details	165
A.1	Supplementary proofs and remarks	165
A.1.1	Continuity of RTE solutions	165
A.1.2	Proof of Lemma 3.1.2 & Lemma 3.1.3	171
A.1.3	Semi-parametric relaxation and proof of Theorem 3.1.4	176
A.2	Implementation & simulation details	179
A.2.1	Regularisation terms and hyper-parameter selection	179
A.2.2	Turbulence process	181

Chapter 1

Introduction

We start by giving a mostly informal introduction to the scientific areas and methods related to the problem under consideration along with an introduction to the relevant physical processes that will be formalised and used in later developments.

1.1 Optical inverse problems

For the purpose of this thesis, we may think of light as a collection of particles, i.e. photons, which can interact with the environment by means of elastic scattering, i.e. a change in direction but not energy, or by absorption in which case the photon “vanishes”, i.e. will not be observed in a measurement. Broadly speaking, the goal of optical inverse problems is the retrieval of optical and/or material properties of a physical system/scene based on the intensity of light, which can be thought of as a photon count, measured at an array of detectors and generated by an active or passive source of illumination.

1.1.1 Absorption spectroscopy, remote sensing & DIAL

Arguably one of the most fundamental and ubiquitous optical inverse problems is that of (in-situ) absorption spectroscopy for determining the concentration of a known material based on the relationship described by Beer-Lambert’s law which states that the absorbance is proportional to the concentration of the absorber in the sample [136, sec-

tion 6.5]. The resulting inverse problem is, assuming the relevant proportionality constants are known (which can be found in databases such as HITRAN [53]), of course rather trivial. During the past several decades several measurement modalities have been developed based on this fundamental technique and today it comes in many different forms, each tailored to different scenarios and requirements, and has played an important role towards a better understanding of the Earth’s atmosphere. For an overview, we refer to [116, 136] and references therein.

In many situations, especially in the context of investigations related to Earth’s atmosphere, it is non-trivial to make contact with all locations about which information is sought. Although the principles behind absorption spectroscopy are in theory not reliant on a localised sample of the material of interest, it certainly requires a source/detector configuration which, one way or another, allows for a measurement of light *after* it has passed the region of interest. In such situations one is left with either of two choices:

- (a) Place a sensor in the region of interest and analyse the sample in-situ, e.g. through the use of unmanned aerial vehicles [14, 52].
- (b) Collect data from a source/detector configuration such that it can be guaranteed that the measurement is related to the object of interest.

The latter idea, i.e. inference about an object without physical contact, is commonly referred to as remote sensing and often relies on some form of electromagnetic radiation [121]. Perhaps the simplest way to ensure that condition (b) will be met, is by putting a passive reflector in a remote location which leads to an open path system where the in-situ measurement cell is replaced by the atmosphere which has led to compact and commercially viable monitoring systems [43]. Approaches using such reflectors are rather limited in the sense that they will only contain information about the path between source and reflector. A more generally applicable alternative is to measure light that was scattered from an arbitrary surface [144].

If scattered photons from a collimated beam are collected by a time-of-flight (ToF) resolving detector, such as found in a Lidar [100, 146] (short for LIght Detection And Ranging),

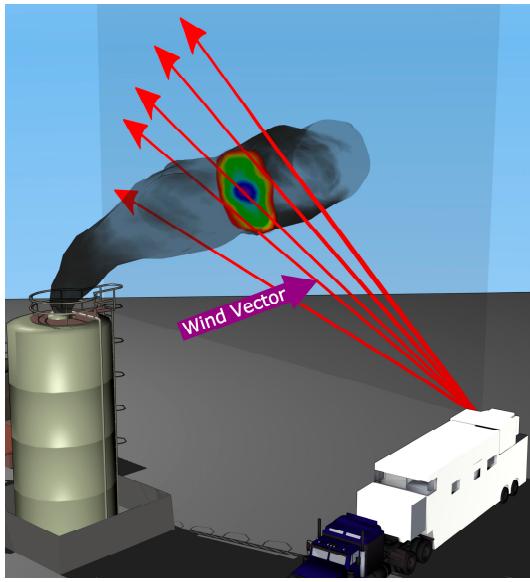


Figure 1.1: Scanning of a plume cross-section (taken from [70])

then it is possible to assign a path length to each measured photon. If it is additionally assumed that the photons only scattered once, then this trajectory is uniquely determined, and, under suitable assumptions on the particle distributions, these return signals can be used to determine the concentration of atmospheric constituents such as aerosols using simple analytic inversion formulas [39, 40].

In addition to the difficulty related to the acquisition of localised samples, when performing remote sensing measurements in the open atmosphere it is often not known what constituents are present and losses in the measured intensity can be caused by unknown materials. Differential methods in spectroscopy, such as Differential Optical Absorption Spectroscopy (DOAS) or Differential Absorption Lidar (DIAL), exploit the fact that some materials have narrow-band absorption structures (see e.g. Fig. 1.2) that can be “extracted” from a suitable collection of measurements and/or filtering procedures [116]. In the case of DIAL, which will be the primary method of interest in this work, this is accomplished by measurement of two wavelengths, typically referred to as online and offline wavelengths, one of which is absorbed by the gas of interest while the other serves as a reference for the expected return of the signal in the absence of the absorber. If we assume single scattering and the optical properties of the medium associated with the online and offline wavelengths differ only with respect to absorption of a target gas,

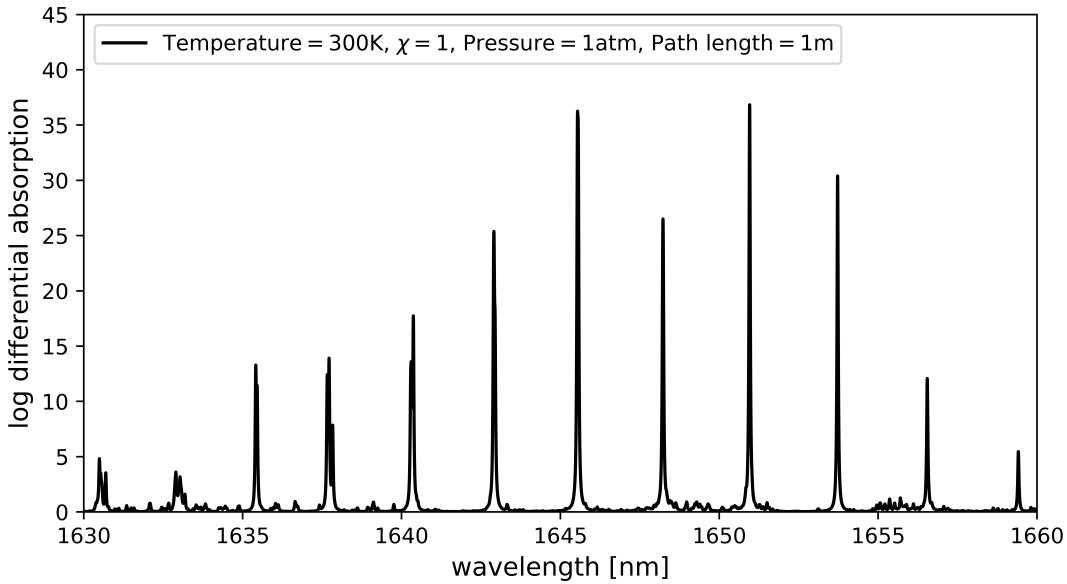


Figure 1.2: Absorption of methane based on Beer-Lambert's law for wavelengths between 1630nm and 1660nm at temperature 300K and pressure 1atm corresponding to a mole fraction $\chi = 1$ and a path length of 1m. It shows the narrow absorption features that can be exploited by differential spectroscopy methods, e.g. a DIAL system operating in the $1.6\mu\text{m}$ range [122]. Data taken from <https://www.spectraplot.com> [51].

then the measurements can in principle be inverted by means of simple analytic formulas and used for determining the spatial concentration profile of the targeted gas. The large distances that light must travel in order to be useful in atmospheric remote sensing applications will inevitably result in weaker optical signals (photon counts) and, combined with potentially low concentrations/absorptions, lead to significant measurement noise. Gas concentrations are obtained by computing what is essentially a derivative of the data which is a process that will typically amplify any noise that is present in the measurement. A low signal-to-noise ratio (SNR) thus presents an obstacle that can be overcome either through prolonged (temporal) averaging of measurements, i.e. combining the response of multiple Lidar pulses sent in the same direction in order to obtain a more accurate one, or computational smoothing/regularisation techniques [70, 124]. Although the details will be discussed later on, Fig. 1.3 should give a sense of the problem at hand and the potential difficulties that one may encounter when dealing with the data.

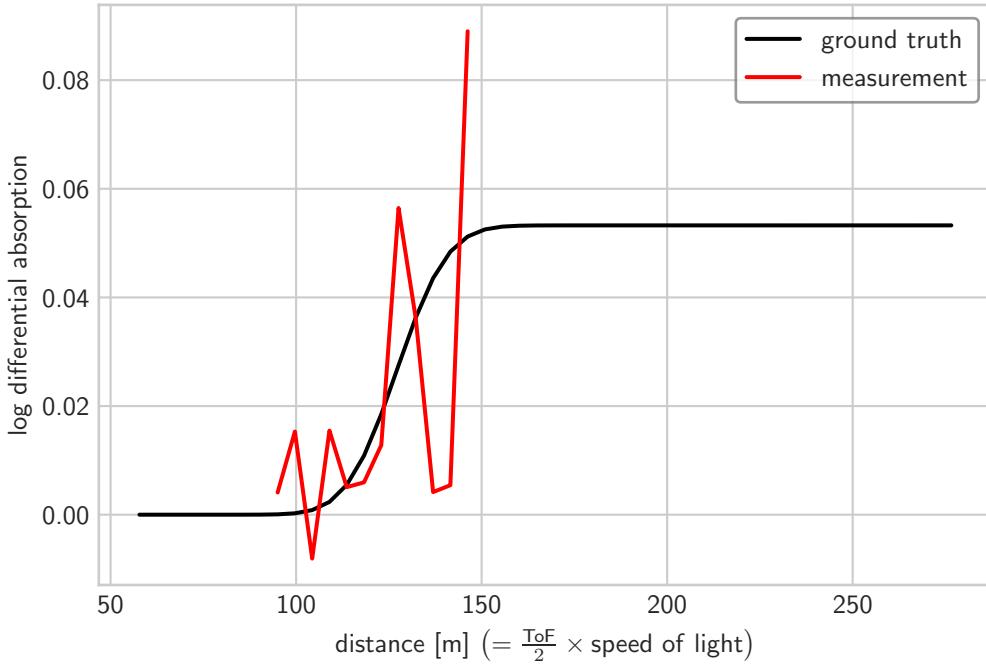


Figure 1.3: The graphic shows an example of simulated data for a plume located roughly 100m away from the detector and measurements subject to Poisson noise with an assumed SNR relevant to the problem at hand. The black line is the expected absorption according to Beer-Lambert's law while the red line is the analogous quantity computed from noisy data. The spatial concentration is obtained by computing derivatives which, in the case of noisy observations, would result in negative values without significant prior smoothing.

1.1.2 Multiple scattering & ill-posedness

With the notable exception of optical diffusion tomography [3] (which in effect does the exact opposite), there is a general tendency in the field of optical inverse problems to exploit/rely on the simple nature of single scattering models. Indeed, considerable effort in the design of experimental setups, by either manipulation of the acquisition modality or material itself, to extract/separate single scattering effects [60, 103, 106]. Classical Lidar methods, be it aerosol remote sensing or DIAL, typically also make a single scattering assumption for the recorded photons which is fundamental to the Lidar equation and, in case the measurement of multiple scattering cannot be avoided, the remainder is usually dealt with by multiple scattering approximations/corrections based on simplifying assumptions [65, 66]. Multiply scattered photons, although recognised as something that

does not just influence the recorded photon count but also spectroscopic measurements [113], are usually also regarded as a nuisance for concentration measurements based on differential methods [1, 97] and either assumed negligible [130] or removed through various filtering procedures [2, 37].

Even though the equation of radiative transfer (RTE) [25], a model for global illumination and thus also known as the rendering equation [5, 75], which can be used to model multiple scattering-related phenomena was well known for decades, the reliance on simplified models can arguably be explained, at least in part, with historically much lower computing resources [18] and the limited availability of analytical solutions which have only been found under rather restrictive conditions on the optical parameters and domain geometries [90, 91]. Arguably the most popular way of solving the RTE is by means of Monte Carlo methods, i.e. through stochastic approximations (see [5, 75] and references therein) but deterministic alternatives, usually classified by what type of discretisation is used, have been developed [7, 77, 117]. As the RTE models the radiance in space, time and direction, it is inherently high-dimensional and due to the presence of non-local integral operators, there is no method, stochastic or deterministic, which can easily deliver high-fidelity solutions quickly. This makes the resulting inverse problems computationally challenging even on modern hardware components.

Aside from the computational burden that comes with the RTE, it is well understood that forward operators involving (multiple) scattering are essentially smoothing their inputs which makes the corresponding inverse problems more ill-posed [8]. This is of course not desirable and, all things being equal, less scattering is consequently preferable. Given that all things are of course never equal the reality is somewhat more nuanced, especially in remote sensing applications that involve a Lidar and thus rely on scattering as unscattered light will not reach the detector. Indeed, wider fields of view (FOVs) have been shown to contain information beyond what is contained in single scattering [20, 69] but are only useful if the optical forward model is chosen appropriately so that this data can be considered as an information signal as opposed to a nuisance. In the context of differential spectroscopy, it is also not hard to see that if the concentration of the trace gas is homogeneous in the atmosphere then any photon registered by a Lidar at a given time

is equally useful in determining the concentration and any sort of filtering is wasteful. A more general form of this observation, i.e. the increase in utility of multiply scattered light with decreased resolution requirements, will be a common theme throughout this work.

1.2 Particle transport in the atmosphere

For many practical purposes, it is prohibitively expensive to acquire accurate measurements of a 3-dimensional region which means that some regularisation of the inverse problem is required in order to be able to deal with data such as shown in Fig. 1.3 which often results in penalisation of solutions that are highly variable and thus non-smooth in a certain sense [12, 26]. In the Bayesian context this is equivalent to the specification of prior distribution [134] which should ideally be based on information available to us beforehand. To address that issue, this work proposes the use of a regularisation of the optical inverse problem based on a dispersion process that enforces smoothness while at the same time reducing the concentration profile to a low-dimensional set of parameters which benefits the computationally challenging reconstruction process when complex optical forward models are used.

1.2.1 Plume models

The starting point for most atmospheric dispersion models is the advection-diffusion equation [132] which, under suitable assumptions, allows for simplified treatment of the underlying transport processes by means of analytical solutions leading to the family of so-called Gaussian plume models (see Fig. 1.4). Due to their simplicity, and ease of use, these models have been adapted to different applications, validated with field data and are now arguably the de-facto standard in environmental monitoring [21, 29, 48, 86, 89, 111, 114, 115, 119, 152]. Of course, it would make little sense to assume that such a model including all unknown input parameters is known *a priori* as this would uniquely determine the concentration profile without any (optical) measurements whatsoever. However,

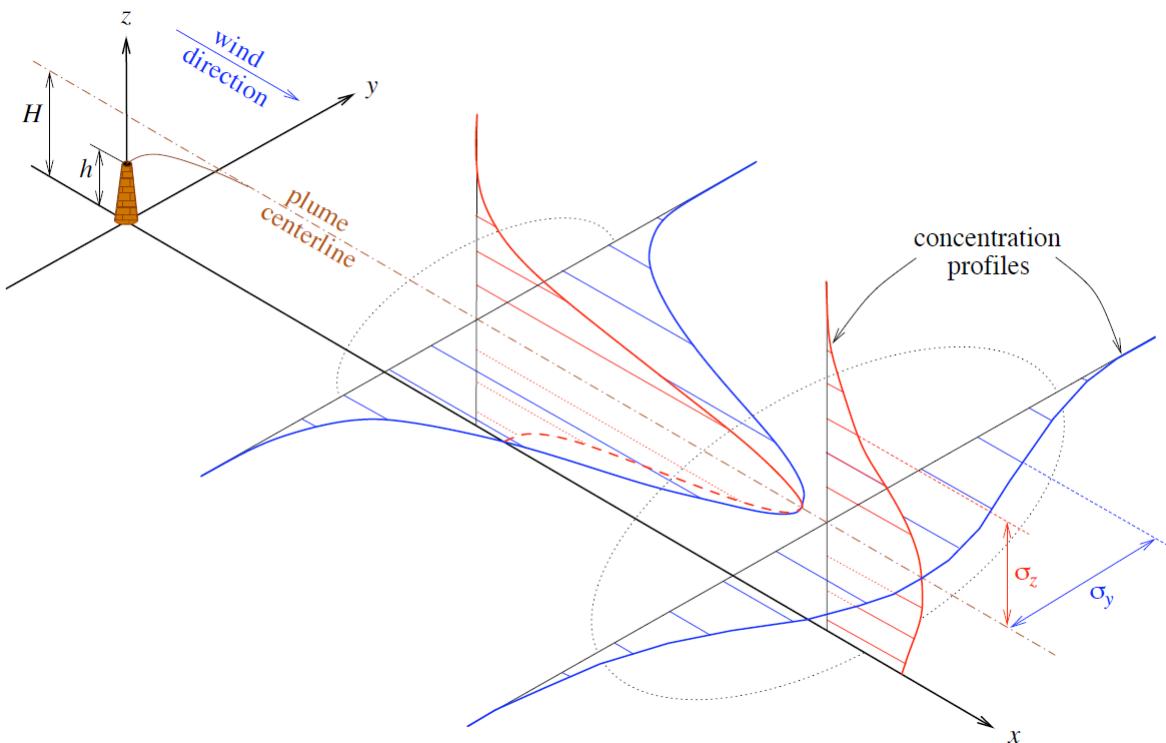


Figure 1.4: Schematic of a Gaussian plume: The main parameters are a 1-dimensional curve describing the centre of the plume as well as width parameters that capture the spread of the contaminant perpendicular to the direction of the wind. The cross-sectional profiles are assumed to be Gaussian (taken from [132]).

the existence and empirical success of such models is certainly a good motivation for the use of model reduction techniques parameterised by quantities similar to those present in a Gaussian plume model. In other words, simplifications like the Gaussian plume model can be used to get an idea regarding which degrees of freedom are the most important if what we seek is to maximise the variance in the spatial concentration that is captured by each fitted parameter.

1.2.2 Model reduction & regularity

Intuitively it seems plausible that a reduction in the number of parameters should make the model easier to fit, i.e. should have a similar effect as classical regularisation methods. This intuition can indeed be formalised by the introduction of effective degrees of freedom, see [58, sec. 7.6] for a formal definition. Considering that Gaussian plumes are derived from a differential equation, it is easily seen how such models could be incorporated as regularisation terms into variational reconstruction methods by replacing the gradient penalty terms, found e.g. as part of Sobolev-type norms, with informed differential operators.

However, a large reduction in model complexity is particularly (and only) suitable for instances where the observations are heavily corrupted by noise as the inability of the resulting estimator to fit complex shapes can result in considerable bias. This is a well-known model selection dilemma also known as the bias-variance trade-off [46, 57].

1.2.3 Probabilities, entropy & concentration of measure

Given the highly restrictive nature of the simple plume models and the ability of sophisticated measurement modalities such as DIAL to, at least in theory (i.e. ignoring instrument noise), fully resolve 3-D concentration profiles in space as well as time, one should ask whether those regularisation methods are perhaps not too much of a restriction. Answering this question is non-trivial as it would arguably require a better understanding of turbulent flow, yet we don't even have an affirmative answer regarding the existence, uniqueness and regularity of solutions to the Navier-Stokes equations which are believed

to describe the related phenomena [38].

At this point it should be mentioned that, although the equations that were used to derive the aforementioned models describe atmospheric transport mimic those of diffusion, it is recognised that the spread of the plume is predominantly caused by turbulent fluctuations rather than molecular diffusion (which is typically negligible). That diffusion is not even close to a perfect model for the effects that govern turbulent flow should come as no surprise, after all virtually no plume truly looks smooth and Gaussian. The discrepancy between what is commonly observed and the used models can to some extent be explained by the presence of averages, an approach that extends well beyond simple Gaussian plume models, see [85] for an overview. Loosely speaking, it is assumed that turbulent flow can be written as the sum of a steady “average” as well as a mean zero “random” component and model only the steady component along with essential assumptions that allow for a closure of the resulting system of equations [135].

The aforementioned averaging can either be done over (purely hypothetical) independent realisations of the same process or over time. Even spatial averages can, at least to some extent, be used in the above line of reasoning by assuming that components of the plume are dragged over a sensor which makes spatial averages behave very similarly to those over time [138]. Given the deterministic nature of the proposed governing equations, such an ergodicity assumption may seem somewhat odd but can be explained by the probabilistic nature of deterministic chaos [84] which has a long history in the field of turbulence modelling [42].

We conclude this section with a neat, primarily information-theoretic, observation which should at least partly justify the approach taken later on in this work. If we accept that our measurement is going to be corrupted by noise, be it from the measurement or atmospheric turbulence, and at the same time we only care, by choice or necessity, for high-level features akin to those in simple plume models then one may ask what the “correct” way of parameterising the inverse problem would be. If all we are able to explain/constrain is insufficient to evaluate the concentration profile at each point in space, then the principle of maximum entropy provides a way to select distributions that are in some sense unbiased [71]. It is not hard to see that the parameters in the plume

parameterisation are not strictly local and, under suitable assumptions on the acquisition modality, their reconstruction is not impacted too much by any given measurement point. In such situations, we can justify the preservation of these high-level features by the concentration of measure phenomenon [137]. It is worth noticing that if we model each plume cross section by its first two moments, i.e. width and position, then this derivation which assumes no functional form on the governing equations is not only consistent with the observation that entropy increases in irreversible dynamical systems [95] and averages over independent copies of the same random element [4] but yields the Gaussian plume model [128].

1.3 Problem statement & contributions

Before proceeding with the more technical parts of the work we give a succinct statement of the problem in consideration and a list of the main contributions followed by a brief summary of relevant literature to highlight the primary differences and similarities between this and previous work.

Problem statement: The primary goal of this thesis is to examine, using radiative transfer theory, the utility that may (or may not) be provided by collecting off-beam photons, i.e. light that reaches the detector from directions different from that of the initially released collimated beam, in reconstructing gas concentration profiles in a DIAL-like setting. Thereafter, to develop a computationally efficient algorithm that exploits the structure of atmospheric dispersion processes to yield noise-robust 3-dimensional estimates of gas concentration from temporally resolved measurements of differential absorption.

Contributions: The main contributions of this research regarding the aforementioned problem can be summarised as:

- A new physics-constrained regularised formulation of the inverse problem, exploiting the intrinsic features of the targeted dispersion process.

- A new, semi-parametric model to generalise the DIAL measurement to the RTE setting allowing for multiple scattering and measurements with wider FOVs.
- A uniqueness result for noise-free data in the above model established under regularity assumptions motivated by the diffusive nature of atmospheric dispersion.
- A theoretical investigation of the inverse problem under relaxed conditions showcasing the limitations of the proposed measurement.
- A quasi-Newton based estimation procedure that uses a mesh-free Monte-Carlo RTE solver which efficiently incorporates the problem's regularity with the ability to (additionally) compute uncertainty estimates.

Related work: The investigation conducted in this thesis is motivated and makes use of previous work regarding radiative transfer theory and inverse problems in general.

- Atmospheric imaging problems with RTE forward models have been investigated in the context of clouds and aerosols in the classical [67, 87, 88] as well as the Bayesian framework [11, 82, 83] where the model was parameterised by quantities similar to those appearing in plume models. A notable departure from previous approaches is that in this work we only consider differential absorption-based measurements, i.e. we seek to utilise only information that can be extracted from the change in absorption at two wavelengths rather than the returned signals itself. To that end, we introduce a semi-parametric model that generalises the DIAL equation in a consistent way to an RTE-based forward model and we give an interpretation of the roles that each optical parameter plays in the context of such a parameterisation. We argue that “lost” information is traded for more flexibility in certain optical parameters and greater resilience against errors caused by turbulence-induced perturbations.
- The rigorous assessment of inverse problems revolving around the RTE are typically based on the singular decomposition of the forward operator [16, 17, 27, 28]. By

adapting these ideas to the differential framework we derive rigorous results regarding the injectivity of the forward operator for data that involves multiple scattering. The results can be used to find semi-explicit inversion formulas for the dispersion parameters.

- By analysing the unregularised problem we show that the dispersion-based constraints affect the reconstruction in a non-trivial way and are effectively a necessity as exact recovery is in general only possible around a small region close to an appropriately defined boundary. A somewhat similar, although *much* stronger and precise, non-uniqueness result was obtained for the time-independent RTE in the case of angularly dependent absorption where it was shown that the optical parameters can be stably recovered up to a gauge transformation which “preserves” the optical parameters at the boundary [99, 131].
- We analyse the transition from the unregularised to the fully regularised formulation by looking at convergence rates for the parameter of interest and show that in situations when optical noise is dominant, in particular when the concentration of the trace gas tends to 0, the rates match those of single scattering based models with potentially better constants. A similar comparison has previously been carried out for X-ray tomography [47].
- The computational challenge associated with multiple scattering forward operators is addressed by adapting the stochastic optimisation methods [49, 50] to our differential measurements and subsequently exploiting that our low-dimensional problem benefits from the same concentration properties as standard approximations in numerical linear algebra [96, 98, 149].
- The feasibility of our proposal and good scalability on GPUs, allowing us to solve the inverse problems within a few minutes on a state-of-the-art personal gaming PC, is demonstrated with simulations that are implemented in Python/Numpy using the Numba package [80] for GPU access. The performance is reported for various optical noise levels as well as perturbations similar to those considered in [118] which ought

to mimic instantaneous snapshots of a turbulent dispersion process. The codes are currently not public but can be made available on request.

1.4 Outline

The thesis is organised as follows. In Chapter 2 we introduce the necessary notation and mathematical basics and we give a formal description of the dispersion and radiative transfer forward problems.

Starting with a coupled formulation of dispersion and optical models, Chapter 3 investigates conditions under which the inverse problem is well behaved. In Section 3.1 we show injectivity of the forward operator, i.e. uniqueness of the inverse problems when noise is absent, under suitable regularity conditions. Contrasting that, Section 3.2 deals with a simpler detection related problem in the presence of noise which highlights conditions in which the use of wider FOVs can be beneficial.

In Chapter 4 we analyse the behaviour of the reconstructed parameters for situations in which the conditions of our primary uniqueness result are violated. To that end we show in Section 4.1 that the unregularised DIAL problem can become more ill-posed and effectively not unique when multiple scattering, as opposed to only single scattering, is present. Such relaxations may, in less extreme cases, also manifest themselves in the form of different, typically slower, convergence rates which are investigated in Section 4.2 under a Poisson noise assumption. Although it is not a primary focus of this work, the chapter concludes with a section regarding the effect of model mis-specification due to turbulent perturbations.

In Chapter 5 we give a description of our algorithm which is capable of efficiently utilising prior knowledge in the form of a suitable dispersion process. Based on the analysis and methodology derived in Section 5.1 we provide some numerical results to validate our findings in Section 5.2.

Chapter 2

Notation & mathematical fundamentals

2.1 Generic problem description

Throughout this work, we will almost exclusively work with continuous or even (infinitely) differentiable functions on a set $X \subseteq \mathbb{R}^d$, where typically $d = 3$, and X may be regarded as topological space whose open sets are induced by a norm (for basics on topology see e.g. Hatcher's notes [59]). We will write for the spaces of k -times, where $k \in \mathbb{N}_0 \cup \{\infty\}$, continuously differentiable functions

$$C^k(X) := \{f : X \rightarrow \mathbb{R} : u \text{ is } k \text{ times continuously differentiable}\}$$

equipped with the topology induced by the canonical family of semi-norms $s_{n,l}$ for multi-indexes $n \in \mathbb{N}_0^d$ and $l \in \mathbb{N}$

$$s_{n,l}(f) := \sup_{x \in X_l} \left| \frac{\partial^{n_1 + \dots + n_d}}{\partial x_1^{n_1} \dots x_d^{n_d}} u(x) \right| \quad (2.1)$$

where $X_1 \subseteq X_2 \subseteq X_3 \subseteq \dots$ is an arbitrary exhaustion of X by compact sets, i.e. each set is compact and their union is X . We will also define the subset $C_c^k(X)$ of compactly

supported functions in $C^k(X)$ as

$$C_c^k(X) := \{u \in C^k(X) : \text{supp } u \text{ is compact in } X\}.$$

which, much like $C^k(X)$, is a Fréchet space (see also [31, ch. IV] and [125, ch. 1]). Here $\text{supp } u$ is the closure of $\{x : u(x) \neq 0\}$. Note that if X is compact then for all k we have $C_c^k(X) = C^k(X)$. We also want to introduce a shorthand for integration which is inspired by how the indexing of arrays in scientific computing software is performed. For $X \subseteq \mathbb{R}^d$ and a finite set $J = \{j_1, \dots, j_{|J|}\}$ consider the set

$$X^J := \{(x_{j_1}, \dots, x_{j_{|J|}}) \in X^{|J|}\},$$

i.e. the collection of $|J|$ -dimensional arrays indexed by distinct elements of J or (equivalently) functions $J \rightarrow \mathbb{R}$. For some $I \subseteq J$ and $x \in X^J$ we set

$$x_I := (x_{i_1}, \dots, x_{i_{|I|}}) \in X^{|I|}, \quad (2.2)$$

i.e. the $|I|$ -dimensional array whose components are identical to those of the parent array x in the components $I \cap J = I$. As a case of special importance, we want to consider for $n_1, n_2 \in \mathbb{Z}$

$$n_1:n_2 := \{k \in \mathbb{Z} : n_1 \leq k \leq n_2\} \quad (2.3)$$

so that for all intents and purposes $X^{1:n}$ and X^n can be seen as identical for $n \in \mathbb{N}$. The operations above may also be performed in reverse, i.e. given $x_{n_1}, \dots, x_{n_2} \in X$ we will write $x_{n_1:n_2}$ for the array containing those vectors. We will also consider integration w.r.t. dx_I of a function $f : X^J \rightarrow \mathbb{R}$ which is defined via

$$\int_{X^{|I|}} f(x) dx_I := \int_X \cdots \int_X f(x) dx_{i_1} \cdots dx_{i_{|I|}}, \quad (2.4)$$

i.e. the integration is performed with respect to the canonical product measure on the $|I|$ -dimensional subspace spanned by the unit vectors corresponding to elements in I . This may not necessarily be the product of Lebesgue measures on \mathbb{R}^d , e.g. X is a lower dimensional sub-manifold such as the sphere $\mathbb{S}^{d-1} \subseteq \mathbb{R}^d$. This makes Eq. (2.4) effectively a function of $x_{J \setminus I}$.

In order to have explicit control of the “size” of the function spaces, i.e. the complexity/regularity of functions within, we will often work with the following discrete space of smooth and well-behaved functions.

Definition 2.1.1. For any function $\phi \in C_c^\infty(\mathbb{R})$, i.e. ϕ is a smooth function with compact support, we define the set of functions $K(\phi) \subseteq C^\infty(X)$ as

$$K(\phi) := \left\{ \sum_{j=1}^N w_j \phi \left(\frac{\|\cdot - b_j\|}{h_j} \right) : N \in \mathbb{N}, b_j \in X, h_j > 0 \text{ and } w_j \in \mathbb{R} \right\}. \quad (2.5)$$

We may, without loss of generality, assume that $\text{supp } \phi \subseteq [-1, +1]$ and also define for any point $x \in \partial X$ the set $K(\phi | x)$ of functions that don't extend beyond x for such kernels ϕ . Formally,

$$K(\phi | x) := \{w_0 + u \in K(\phi) : w_0 \geq 0, x \notin \text{supp } u\} \quad (2.6)$$

where w_0 is constant (as a function on X). A kernel $\phi \in C_c^\infty(\mathbb{R}) \setminus \{0\}$ is called a good kernel, if its non-negative and decreases on $[0, \infty)$, i.e. for all $z \geq 0$ we require $\phi(z) \geq 0$ as well as $\phi'(z) \leq 0$.

Note that the additional condition in Eq. (2.6) is inspired by the setting encountered in remote sensing. Indeed, if x_D is the location of our sensor then we can express $K(\phi | x_D)$ equivalently as

$$K(\phi | x_D) = \left\{ w_0 + \sum_{j=1}^N w_j \phi \left(\frac{\|\cdot - b_j\|}{h_j} \right) \in K(\phi) : h_j < \|x_D - b_j\| \forall j = 1, \dots, N \right\}$$

which shows explicitly that the constraint is chosen precisely in a way so that the non-constant part of $u \in K(\phi | x_D)$ mustn't make contact with the sensor.

(Non)-parametric estimation problems Before making things more explicit we start with a general form of the imaging problem at hand. Assume that the plume, or *image*, which we are trying to determine is given by a smooth function, i.e. $u \in \mathcal{V} \subseteq C^\infty(X)$ for $X \subseteq \mathbb{R}^d$ where the dimension d depends on, for example, whether there is a temporal component or the plume is assumed to be stationary. Regardless of what type of measurements are being used, we can be all but certain that they are going to be corrupted by a significant amount of noise.

If we assume that this noise is *random* and that its distribution is known, then we can assume to have access to data in the form of realisations $\mathbf{y}_1, \dots, \mathbf{y}_n$ of a random element \mathbf{Y} . If we additionally assume that the distribution $\mathbb{P}(\mathbf{Y} \in \cdot \mid u)$ of \mathbf{Y} given the image u is known (or at least its expectation), then we can evaluate to the forward operator

$$\mathcal{F} : u \mapsto \mathbb{E}(\mathbf{Y} \mid u)$$

and construct a reasonable loss functional $u \mapsto \text{Loss}(u \mid \mathbf{y}_1, \dots, \mathbf{y}_n)$ and find an estimator \hat{u} for the image by solving the optimisation problem

$$\hat{\mathbf{u}} \in \arg \min_{u \in \mathcal{V}} \text{Loss}(u \mid \mathbf{y}_1, \dots, \mathbf{y}_n). \quad (2.7)$$

where the loss can be a least squares penalty, negative log-likelihood, or indeed any function that reasonably captures the discrepancy between the observed data $\mathbf{y}_1, \dots, \mathbf{y}_n$ and the forward evaluation $\mathcal{F}(u)$ given image u .

There are two competing objectives when dealing with estimators as the one given in Equation (2.7). On one hand, we would like the set \mathcal{V} to be large, i.e. cover a broad range of images $u \in C^\infty(X)$ in order to accommodate as much flexibility in the image reconstruction as possible, after all, the bigger our feasible set \mathcal{V} is the better our chance of capturing image that generated the data inside of it.

On the other hand, it is well known that for ill-posed problems, which is, unfortunately, something that one encounters with most forward operators of practical interest, one is better off solving a regularised problem resulting in a modified version of the optimisation

problem Equation (2.7) given by

$$\hat{\mathbf{u}} \in \arg \min_{u \in \mathcal{V}} \text{Loss}(u \mid \mathbf{y}_1, \dots, \mathbf{y}_n) + \gamma \text{Reg}(u) \quad (2.8)$$

with $\gamma > 0$ and where the regularisation penalty $\text{Reg}(u)$ will generally be based on the structure of the feasible set as well as prior knowledge which in the present case could mean a penalty term based on the semi-norms from Eq. (2.1), for example, the norm of a differential operator such as $\|\Delta u\|^2$. The smoothness enforcing regularisation term remedies some of the ill-posedness but the problem in Equation (2.8) is often even more computationally demanding than that of Equation (2.7). Furthermore, when the data are extremely noisy it may become necessary to increase γ to an extent where too much structure gets lost for it to be of actual value for the intended application.

Conceptually, our approach can be derived from Equation (2.8) by first replacing a generic regularisation term such as $\|\Delta u\|^2$ with $\|D_\theta(u)\|$ where the operator D_θ is a parameter dependent differential operator and $\theta \in \Theta \subseteq \mathbb{R}^{N_\theta}$ is a finite dimensional quantity of interest. We may write

$$(\hat{\boldsymbol{\theta}}, \hat{\mathbf{u}}) \in \arg \min_{(u, \theta) \in \mathcal{V} \times \Theta} \text{Loss}(u \mid \mathbf{y}_1, \dots, \mathbf{y}_n) + \gamma \|D_\theta(u)\| \quad (2.9)$$

which would result in an estimator for both θ and u at the same time. If we assume that for all $\theta \in \Theta$ the solution $D_\theta(u) = 0$ is unique in \mathcal{V} , then this induces an implicit mapping $\Theta \rightarrow \mathcal{V}$, $\theta \mapsto u[\theta]$ which is then determined by the relation $D_\theta(u[\theta]) = 0$. Then, letting the regularisation parameter $\gamma \rightarrow \infty$ in, Eq. (2.9) becomes equivalent to solving

$$\hat{\boldsymbol{\theta}} \in \arg \min_{\theta \in \Theta} \text{Loss}(u[\theta] \mid \mathbf{y}_1, \dots, \mathbf{y}_n) \quad (2.10)$$

and setting $\hat{\mathbf{u}} = u[\hat{\boldsymbol{\theta}}]$. In contrast to more generic regularisation terms (like a Laplace operator Δ), the use of D_θ results in an informed regularisation penalty which has the potential to preserve the structure of a plume even for large regularisation parameters γ . The map $\theta \mapsto u[\theta]$ will, unlike parameterisations based on local (voxel-based) functions, be non-linear but more importantly it is often possible to select a parameter space Θ of

a rather low dimension and still obtain meaningful images which makes Equation (2.10) a low-dimensional or *parametric* problem. The use of parametric models is, of course, nothing new [11, 56, 82, 83] or in itself particularly exciting but as will be shown later, it is the transition from high- to low-dimensional problems where the comparison of wide and narrow FOV based DIAL approaches makes most sense. Note that the behaviour of such estimators can be quite different from the situation when a \mathcal{V} is parameterised via local basis functions, which technically makes it finite-dimensional at a given resolution, since the dimension of Θ is fixed. If all we do is discretise an infinite-dimensional function space then the behaviour of the estimator is much better described in the non-parametric setting where convergence rates are typically slower (see e.g. [141]).

As such one may hope for good robustness towards noise, at least for well-behaving operators D_θ . If the latter is not the case and D_θ is related to some atmospheric transport process, essentially the same issues that arise in Equation (2.10) will be present in any method for plume tracking and localisation based on that dispersion model regardless of how concentration estimates are obtained. Note that we essentially require that if θ is of interest then its recovery from $u[\theta]$, i.e. the image itself, should be well-posed. The primary danger with an approach such as Equation (2.10) is placing constraints that are too restrictive and result in unrealistic images with large biases.

2.2 Advection-diffusion based dispersion models

In this section, we introduce a family of widely used gas dispersion models, which under suitable boundary conditions and domains have a closed-form solution. Following the developments in [132] we consider the advection-diffusion operator given by

$$D_\theta(u_\delta) = \frac{\partial}{\partial t} u_\delta + \nabla_x \cdot (\eta u_\delta) - \frac{1}{2} \nabla_x \cdot (\kappa \nabla_x u_\delta) - \rho \quad (2.11)$$

with $\theta = (\eta, \rho, \kappa)$, or more precisely it is assumed that these functions are parameterised by an N_θ dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^{N_\theta}$ in which case $\eta = \eta[\theta]$, $\rho = \rho[\theta]$ and $\kappa = \kappa[\theta]$ are θ -dependent inputs of Eq. (2.11). We use η to model the drift, ρ is a source

term, κ is a diagonal matrix with diffusion coefficients and u_δ is the gas concentration as a function of space and time. If ρ is a point source located at $b^0 = (b_1^0, b_2^0, b_3^0) \in \mathbb{R}^3$, then

$$\rho(x, t) = \rho_0 \cdot \delta(x_1 - b_1^0) \delta(x_2 - b_2^0) \delta(x_3 - b_3^0) \delta(t) \quad (2.12)$$

where $\rho_0 > 0$ models the amount of released gas and δ are single variable delta distributions such that ρ ultimately depends on 4 scalar parameters. Typically plume models assume also that κ is a function of downwind distance [132] and the primary challenge is parameterisation of this quantity using measurable (atmospheric) quantities [41, 44, 142]. In the time-dependent setting, this is essentially the same as assuming dependence on time since particles move downwind as they are advected by the wind modelled through η . Equally, we shall also assume position-independent wind although we may want to allow time-dependence, i.e. $\eta = \eta(t)$, which can thus be used to express plume rise and effective stack height. As both κ and η can reasonably be assumed continuous and consist of single variable functions, they can easily be approximated by piece-wise polynomials (see e.g. [19, ch. 3.5]).

Note that we do not assume that diffusion in the wind direction is much smaller than advection and can thus be ignored, which is assumed in the derivation of steady-state Gaussian plume models [129, 132], as this would prevent us from using the dispersion model in situations with negligible wind. We chose a somewhat more general, Lagrangian, approach to model steady-state plumes via super-positions of instantaneous releases, which is also a commonly used approach for modelling dispersion in practice [85]. Given the above assumption on η and κ and assuming an infinite domain (without spatial boundary) we can solve Eq. (2.11) in order to obtain

$$u_\delta[\theta](x, t) = \rho_0 \prod_{j=1}^3 \frac{1}{\sqrt{2\pi \int_0^t \kappa_{jj}(s) ds}} \exp \left(-\frac{(x_j - [b_j^0 + \int_0^t \eta_j(s) ds])^2}{2 \int_0^t \kappa_{jj}(s) ds} \right) \quad (2.13)$$

In the situation where the diffusion is isotropic, i.e. κ is a scalar function κ_0 multiplied

by the identity matrix, we define

$$b(t) = b^0 + \int_0^t \eta(s) ds \quad (2.14)$$

$$h(t) = \sqrt{\int_0^t \kappa_0(s) ds} \quad (2.15)$$

so that we can simplify Equation (2.13) and obtain

$$u_\delta[\theta](x, t) = \frac{\rho_0}{(\sqrt{2\pi}h(t))^3} \exp\left(-\frac{1}{2}\left[\frac{\|x - b(t)\|_2}{h(t)}\right]^2\right). \quad (2.16)$$

The above developments are equivalent to the assumption that the plume is made up of (infinitely many) particles whose trajectory in the j -th coordinate $\mathbf{U}_j(\cdot)$ obeys the SDE

$$\begin{aligned} \mathbf{U}_j(0) &= b_j^0 \\ d\mathbf{U}_j(t) &= \eta_j(t)dt + \sqrt{\kappa_{jj}(t)}d\mathbf{B}(t) \end{aligned} \quad (2.17)$$

for a Brownian motion $\mathbf{B}(\cdot)$. In this case, the Fokker-Planck equation related to its solution is given by Eq. (2.11) (see [109, sec 3] for details). In situations where the boundary is not negligible, e.g. when the plume is close to a reflecting or absorbing flat ground at $x_3 = 0$, similar relations such as those between Eq. (2.13) and Eq. (2.17) can be obtained [92, ch. 3.5].

Steady-state models and super-positions The Gaussian puff model from the previous section considers an instantaneous release and as such it isn't immediately suitable for modelling longer or ongoing releases. Instead of using the solution presented in the previous paragraph, we may express a continuous release as the integrated superposition of instantaneous releases at different times. In fact, by allowing a sufficient amount of independent spatially distributed sources we could essentially approximate any smooth function with puffs such as in Eq. (2.16), regardless of boundary conditions or even dispersion dynamics. This means that the complexity of the dispersion controls the variability of the images and thus the amount of regularisation in Eq. (2.10).

Although much of the theory presented can be viewed independently, we will typically assume that $h(t)$ as well as $b(t)$ are piece-wise linear and that we can use a single source to model the structure of the dispersion. The primary reason for this assumption is that point estimation (and any quantification of uncertainty based solely on expansions of those estimations) becomes increasingly unreliable for complicated dispersion structures in the presence of noise. Such situations are therefore better dealt with by using a Bayesian approach coupled with a more sophisticated uncertainty quantification approach, something that is beyond the scope of this work.

In case of a steady/prolonged release, this means we have $u \approx \sum_{j=1}^K u_{\delta,i}$ where each $u_{\delta,i}$ has the form of Equation (2.16). For the sake of consistency with our later developments, we shall also include a constant, i.e. (x,t) -independent, term $u_0 = u_0[\theta]$ to model a homogeneous concentration field. If we assume that the source releases gas continuously during an interval $[0, T]$ for some $T \in (0, \infty]$ and use multiple puffs to approximate the continuous release then, for $t \leq T$, we end up with a dispersion model of the form

$$u[\theta](x, t) := u_0[\theta] + \int_0^t u_\delta[\theta](x, s) ds \approx w_0 + \sum_{j=1}^{N_\phi} w_j \phi \left(\frac{\|x - b_j\|_2}{h_j} \right) \quad (2.18)$$

for some suitable, typically t -dependent, values for w_j, h_j and b_j . These can be associated with a discretisation $t_j = \frac{t}{2N_\phi} + j \frac{t}{N_\phi}, j = 0, \dots, N_\phi - 1$ by putting

$$(w_j = w[\theta](t_j) t N_\phi^{-1}, b_j = b[\theta](t_j), h_j = h[\theta](t_j))_{j \in 1:N_\phi} \quad (2.19)$$

for $w[\theta], b[\theta]$ and $h[\theta]$ in accordance with Eq. (2.16) and the squared exponential kernel $\phi(x) = \exp(-x^2/2)$. A more explicit formulation is not needed at this stage and will be given in Section 5.1.4. Although $\exp(-x^2/2)$ is strictly speaking not a good kernel we may think of it as one as it meets all requirements aside from being compactly supported and has rapidly decaying tails. Indeed, the only reason for the introduction of good kernels as in Definition 2.1.1 is to have a way of isolating perturbations in the optical parameters and being able to assign meaningful boundaries to them. When the plume is in steady-state we have $t = \infty$ in Eq. (2.18) and there is no dependence on time.

Note that if we ignore the difference between a Gaussian kernel and a voxel indicator, then Equation (2.18) differs from a voxel-based parameterisation only in that the basis functions are not fixed in place or size and instead are allowed to vary in these quantities and any potential non-uniqueness caused by allowing this additional variability is dealt with through explicit parameterisations of w_j , h_j and b_j . One compromise that we have made is the use of isotropic covariances in κ which, unlike the diagonal structure (see also [129, sec. 18.4.1]), is in general not an entirely realistic assumption in Gaussian plume models [85].

Recovery of source parameters from concentration In some situations it might be of interest to not only have an estimate of the gas concentration but one may also seek information regarding its source. For this work we will focus on ambiguities regarding the release rate of a steady, i.e. time-independent, source.

It is not hard to see that in general we cannot extract the sought information uniquely from steady-state concentration measurements alone. The core issue is that by adapting timescales and velocities at which particles are transported through space we can obtain identical concentrations despite having different source terms. To make this more concrete we may consider the advection-diffusion equation (2.11) for drift $\eta(t)$, diffusion $\kappa(t)$ and source parameter ρ and put $\tilde{\eta}(t) = 2\eta(2t)$, $\tilde{\kappa}(t) = 2\kappa(2t)$ and $\tilde{\rho} = 2\rho$. Denote by u_δ and \tilde{u}_δ the solution of Eq. (2.11) for (η, κ, ρ) and $(\tilde{\eta}, \tilde{\kappa}, \tilde{\rho})$ respectively. By construction it must be true that for all $(t, x) \in (0, \infty) \times \mathbb{R}^3$

$$\frac{\partial}{\partial t} u_\delta(2t, x) = -\nabla_x \cdot (\tilde{\eta}(t) u_\delta(2t, x)) + \frac{1}{2} \nabla_x \cdot (\tilde{\kappa}(t) \nabla_x u_\delta(2t, x)) \quad (2.20)$$

which means that $u_\delta(2t, x)$ solves Eq. (2.11) for $(\tilde{\eta}, \tilde{\kappa}, \tilde{\rho})$, i.e. $u_\delta(2t, x) = \frac{1}{2} \tilde{u}_\delta(t, x)$, and it immediately follows that for all $(t, x) \in (0, \infty) \times \mathbb{R}^3$ we have

$$2u_\delta(2t, x) = \tilde{u}_\delta(t, x) \implies \int_0^\infty \tilde{u}_\delta(t, x) dt = \int_0^\infty u_\delta(t, x) dt. \quad (2.21)$$

If, for example, $\inf_{t \in (0, \infty)} \|\eta(t)\|_\infty > 0$ and $\sup_{t \in (0, \infty)} \|\kappa(t)\|_\infty < \infty$ then the above integrals converge for any $x \in \mathbb{R}^3$ and can be seen as steady-state concentrations, i.e.

obtained after gas has been continuously released for an infinite period of time. In other words, the function $u(\cdot, \infty)$ as in Eq. (2.18) does not determine the dispersion parameters uniquely. Eq. (2.21) has an intuitive interpretation. The parameters η and κ describe the rate of change in terms of linear and mean-square displacement respectively and thereby describe the rate at which particles are moving (see also Eq. (2.17)). Thus Eq. (2.21) states that if particles are “removed” at twice the rate, then this can be compensated by “creating” them at twice the rate so that steady-state concentration remains unchanged. The underlying concept can of course be generalised to scalars other than 2 and it does not necessarily rely on a particular atmospheric transport model either, i.e. it is not an artifact of Eq. (2.11) as we may mimic the developments after replacing the right-hand side of Eq. (2.20) with a different operator. The resulting ambiguity, at least for the model and assumptions under consideration, can be overcome by assuming that the drift term η is (partially) known. This requires a separate measurement of wind, e.g. at the location of the DIAL instrument, and has been implemented as part of operational DIAL instruments [70, 124]. Note that one does not have to assume full knowledge of $\eta(t) = (\eta_1(t), \eta_2(t), \eta_3(t))$ and for gases released at different pressure/temperature we can instead consider the case where only the projection $(\eta_1(t), \eta_2(t))$ onto the x_1x_2 -plane is known while the unknown plume rise is modelled by a variable x_3 -component (see also [129, sec. 18.11] and the schematic Fig. 1.4 taken from [132]). We will get back to this assumption in Chapter 5 for our simulation studies.

Entropy of atmospheric dispersion It is worth noticing that the above model is most useful and accurate in scenarios where the plume is observed over longer periods of time which obviously isn’t possible when the release is instant or inhomogeneous. The above model will therefore inevitably have some error and not necessarily represent the true dispersion process accurately. Instead of viewing the above dispersion model as an approximate truth to the real-world particle transport one might also consider reconstructing certain aspects $F_0(u), \dots, F_k(u)$ of the image u , where F_j can in principle evaluate any real-valued feature of the image u , regardless of any knowledge about underlying phenomena responsible for the motion of airborne particles. If F_j evaluates localised integrals

over equal patches of the domain, i.e. voxels, then $F_j(u)$ corresponds to local averages. Most standard methods seek to reconstruct these aspects of the image and assume that u is constant within each voxel which coincides with the maximum entropy distribution given the constraints $F_j(u)$ while the number of constraints used corresponds to the image resolution in the usual sense.

The concept of filling in the missing information by means of finding the distribution with maximal entropy that satisfies a set of constraints can be generalised in a straightforward way to functionals other than voxel averages and represents in a fairly strong sense the least biased and optimal way of choosing distributions given partial or incomplete knowledge [71]. Similarly, we can interpret the number of real-value constraints as a more general form of image resolution (informally one may think of this as coefficients in a different basis even though the F_j can be non-linear) where the number of parameters needed to represent the function to a certain degree of accuracy heavily depends on the selected features. As an alternative to local averages over voxels, we can consider F_j which corresponds to the amount of gas, position and width of the plume which can be expressed through moments of the function u . In domains that are mostly unconstrained, i.e. open space or uncluttered environments, the maximum entropy solution given the first three moment constraints will be (approximately) Gaussian and take a form similar to Equation (2.16). As such this can be considered as our best guess for the gas distribution given only the amount, the location and a (homogeneous) dispersion rate independently of any knowledge about the underlying atmospheric transport.

The entropy-based derivation of Equations Eq. (2.13) or Eq. (2.16) is arguably more general in that it doesn't assume the existence of a hypothetical average, i.e. ergodicity. The empirical observations that support such assumptions merely suggest that the chosen parameterisation is sensible and turbulence-induced deviations behave much like irreversible, entropy-increasing operations resulting in errors that in a way resemble random noise. Although it may seem unnecessary at this point to derive the dispersion model in two different ways, the above arguments will help us understand the reasoning behind our treatment of missing information for the optical transport problem where averaging becomes essentially meaningless (much like for instantaneous gas releases). In particu-

lar, the mechanism responsible for errors due to regularisation of turbulence is essentially identical to what will be used to handle missing information regarding optical parameters due to our inability to measure light to a degree sufficient for the reconstruction of all involved parameters.

2.3 Radiative transfer

As of now, we haven't discussed how the optical measurements are influenced by a plume. In principle, the forward model is indifferent to any other optical imaging or tomography setup where it's given by the time-dependent Radiative Transfer Equation (RTE) [25]. The time-dependent RTE describes the intensity of light at location $x \in X$ travelling in direction $v \in \mathbb{S}^2$ at any time $t \geq 0$ and is given by the integro-differential equation

$$\left(\frac{\partial}{\partial t} + v \cdot \nabla_x + \sigma_{a+s}(x) \right) H(x, v, t) = \sigma_s(x) \int_{\mathbb{S}^2} H(x, v', t) f_p(x, v' \rightarrow v) dv' \quad (2.22)$$

for $(x, v, t) \in X \times \mathbb{S}^2 \times [0, \infty)$, where the spatial domain of interest $X \subseteq \mathbb{R}^3$ is a bounded open set with sufficiently regular boundary ∂X . In Eq. (2.22) we have normalised the time component such that the speed of light is equal to 1, which is no loss of generality. The functions $\sigma_a, \sigma_s : X \rightarrow [0, \infty)$ are the absorption and scattering coefficients of the material, $\sigma_{a+s} := \sigma_a + \sigma_s$ and $f_p : X \times \mathbb{S}^2 \times \mathbb{S}^2 \rightarrow [0, \infty)$ is called phase function and describes the distribution of angles for scattered light. Since f_p is a probability density on the sphere it satisfies

$$\int_{\mathbb{S}^2} f_p(x, v' \rightarrow v) dv = 1 \quad \forall x \in X, v' \in \mathbb{S}^2.$$

In order to state the boundary conditions for Eq. (2.22) in the spatial boundary ∂X we introduce for any $x \in \partial X$ the hemispheres

$$\partial V_{(-)}(x) = \{v \in \mathbb{S}^2 : v \cdot \vec{n}(x) < 0\} \quad \partial V_{(+)}(x) = \{v \in \mathbb{S}^2 : v \cdot \vec{n}(x) > 0\} \quad (2.23)$$

where $\vec{n}(x)$ is the unit outer normal at $x \in \partial X$. In other words, at a point $x \in \partial X$, $\partial V_{(-)}(x)$ and $\partial V_{(+)}(x)$ are the hemispheres correspond to the inward and outward pointing directions respectively. The impact of time-homogeneous ambient illumination will be discussed later and we can assume that at time $t = 0$ there is no light inside the domain and consider a known source at the boundary given by a function defined on $X \times \mathbb{S}^2 \times \mathbb{R}$ which is inward pointing, i.e. $\text{supp } g \subseteq \{(x, v, t) \in \partial X \times \mathbb{S}^2 \times (0, \infty) : v \in \partial V_{(-)}(x)\}$. In the case of an instantaneous Lidar pulse at time t_0 released in direction v_0 from a source located at x_D we may think of g as smooth and compactly supported such that $g(x, v, t) \approx \delta(x - x_D)\delta(v - v_0)\delta(t - t_0)$. If we ignore reflections at the boundary this translates to

$$H(x, v, 0) = 0 \quad \forall x \in X, v \in \mathbb{S}^2 \quad (2.24a)$$

$$H(x, v, t) = g(x, v, t) \quad \forall x \in \partial X, v \in \partial V_{(-)}(x) \quad (2.24b)$$

More generally, reflections at the boundary could be expressed through

$$H(x, v, t) = g(x, v, t) + \int_{\partial V_{(+)}(x)} H(x, v', t) f_s(x, v' \rightarrow v) |\vec{n}(x) \cdot v'| dv' \quad (2.25)$$

for $x \in \partial X, v \in \partial V_{(-)}(x)$ and $t > 0$. Although it plays a similar role as f_p we don't require f_s to be a probability density but rather that it integrates to at most 1 (although we could obviously normalise it by introducing a normalising coefficient). The Eq. (2.25), also known as the *rendering equation*, can be understood in such a way that it describes the outgoing intensity as the reflected intensity of accumulated incoming light at the boundary represented by the integral. The factor $|\vec{n}(x) \cdot v'|$ accounts for the fact that a piece of solid angle from a light source that illuminates a surface will spread over a smaller area when it's perpendicular to the surface than when it comes in almost parallel. Henceforth we will assume that $f_s = 0$ (the primary purpose of which is to keep the notation simple), i.e. that light is absorbed by the boundary. This isn't always necessary and we will explain how our main theoretical results can be extended to more general non-absorbing boundaries. More details on the derivation and intuition behind the RTE

can be found in [5].

Integral representation and Neumann series Eq. (2.22) can be recast into an equivalent equation that only has integrals which will make it easier to analyse, at least for our purposes, and will also be used for solving the RTE. This is also known as the *volume rendering equation* due to its similarity with Eq. (2.25). We may write

$$H(x, v, t) = \mathcal{I}[g](x, v, t) + \mathcal{K}[H](x, v, t) \quad (2.26)$$

where the operators \mathcal{I} and \mathcal{K} are defined by

$$\mathcal{I}[g](x, v, t) := e^{-\int_0^{\ell_v^-(x)} \sigma_{a+s}(x-sv) ds} g(x - \ell_v^-(x)v, v, t - \ell_v^-(x)) \quad (2.27)$$

$$\begin{aligned} \mathcal{K}[H](x, v, t) &:= \int_0^t e^{-\int_0^s \sigma_{a+s}(x-rv) dr} 1_{(0, \ell_v(x))}(s) \\ &\quad \int_{\mathbb{S}^2} H(x - sv, v', t - s) \sigma_s(x - sv) f_p(x - sv, v' \rightarrow v) dv' ds \end{aligned} \quad (2.28)$$

where $1_A(x)$ is the indicator function on the set A and $\ell_v^\pm(x)$ are the boundary impact times from x in direction $\pm v$, i.e. we have

$$\ell_v^\pm(x) := \inf\{t > 0 : x \pm tv \notin X\}. \quad (2.29)$$

Note that \mathcal{I} takes arguments that are functions defined on the boundary while \mathcal{K} acts on functions that depend on interior points of the domain. Under suitable conditions, which are true for all materials that are of interest to us, we can invert $\text{Id} - \mathcal{K}$ and obtain the following *Neumann series* for $t > 0$

$$H(x, v, t) = (\text{Id} - \mathcal{K})^{-1} \mathcal{I}[g](x, v, t) = \sum_{j=0}^{\infty} \mathcal{K}^j \mathcal{I}[g](x, v, t) = \sum_{j=0}^{\infty} H_j(x, v, t) \quad (2.30)$$

which decomposes the contribution into the orders of scattering. Note that the expression for the series in case of non-trivial reflecting boundaries, i.e. when $f_s \neq 0$, is similar but ends up being slightly more involved than the one from [9] due to the mixed nature of

boundary conditions.

Albedo operators and their Schwartz kernels Since the measurements are typically taken at the boundary of the domain their description needs to consider the so-called albedo operators. If we define

$$\partial B_{(\pm)} := \{(x, v) \in \partial X \times \mathbb{S}^2 : v \in \partial V_{(\pm)}(x)\} \quad (2.31)$$

then these are given by the restriction of the solution from Equation (2.22) to the spatial boundary in the outward direction $\partial B_{(+)} \times (0, \infty)$, i.e.

$$\mathcal{A}[g] = (\text{Id} - \mathcal{K})^{-1} \mathcal{I}[g] \Big|_{\partial B_{(+)} \times (0, \infty)} = H \Big|_{\partial B_{(+)} \times (0, \infty)}. \quad (2.32)$$

For more general and trace results along with rigorous justifications for the operators \mathcal{A} we refer to [23, 24] and [28]. We will also consider the decomposition into the Neumann series components, i.e.

$$\mathcal{A}_j[g] = H_j \Big|_{\partial B_{(+)} \times (0, \infty)} \implies \mathcal{A}[g] = \sum_{j=0}^{\infty} \mathcal{A}_j[g] = \sum_{j=0}^{\infty} H_j \Big|_{\partial B_{(+)} \times (0, \infty)}$$

which is very similar to the developments in [10]. A more detailed treatment of the transport equation is given in [33, Ch. XXI] to which we refer for further details. Finally, we want to consider the Schwartz kernel, we refer to [45] for a proof regarding the existence and uniqueness in a distributional setting, $\mathfrak{a}(x, v, y, v', s)$ of \mathcal{A} . They have the property that for smooth test functions $g : \partial B_{(-)} \times (0, \infty) \rightarrow \mathbb{R}$ and $g_D : \partial B_{(+)} \times (0, \infty) \rightarrow \mathbb{R}$ we have

$$\begin{aligned} & \int_{\partial B_{(+)} \times (0, \infty)} \mathcal{A}[g](x, v, t) g_D(x, v, t) dx dv dt \\ &= \int_{\partial B_{(+)} \times \partial B_{(-)} \times (0, \infty)^2} \mathfrak{a}(x, v, y, v', s) g(y, v', t-s) g_D(x, v, t) dx dv dy dv' dt ds. \end{aligned} \quad (2.33)$$

Schwartz kernels \mathfrak{a}^j for the operators \mathcal{A}_j may be defined in a similar way as in Equation (2.33) and will be used in the following developments.

Stochastic process for photon transport A notable similarity between the solutions of Equations Eq. (2.22) and Eq. (2.11) is that, after suitable normalisation, both may be interpreted as probability densities. The solution $u_\delta(\cdot, t)$ of the dispersion model at time t describes the probability distribution of airborne particles moving according to the drift-diffusion Eq. (2.17) within the domain. This relationship between particles moving according to a Markov process and PDEs such as Eq. (2.11) is a special instance of a much deeper connection between Markov processes and operator semi-groups which naturally appear in solutions of relations that can be written as $(\frac{\partial}{\partial t} - A)u = 0$, note that both the RTE Eq. (2.22) and Eq. (2.11) is obviously of that form, where A is an (unbounded) operator on a suitably defined domain. In the case of radiative transfer $H(\cdot, \cdot, t)$ relates to the position and direction of photons which move through the domain based on the following probabilistic rules (see also [11, 82, 83]):

1. Upon release from $x_D \in \partial X$ in direction v_0 the photon moves along the line $v_0 r + x_D$ until it interacts with a particle at a (random) time \mathbf{t} such that

$$\mathbb{P}(\mathbf{t} < t) = 1 - e^{-\int_0^t \sigma_{a+s}(v_0 r + x_D) dr}$$

i.e. an inhomogeneous exponential distribution consistent with Beer-Lambert's law. For technical reasons we can extend σ_{a+s} with 0 outside of X and \mathbf{t} may not necessarily be finite.

2. If $\mathbf{t} < \infty$, i.e. the photons interacts with a particle at $\mathbf{x} = v_0 \mathbf{t} + x_D \in X$, then two things can happen:

- The photon is absorbed with probability $\frac{\sigma_a(\mathbf{x})}{\sigma_{a+s}(\mathbf{x})}$ and nothing else happens at any subsequent times $t > \mathbf{t}$.
 - The photon is scattered with probability $\frac{\sigma_s(\mathbf{x})}{\sigma_{a+s}(\mathbf{x})}$ and turns towards a new direction \mathbf{v} chosen according to the probability density $f_p(\mathbf{x}, v_0 \rightarrow \cdot)$.
3. Similarly, if $\mathbf{t} = \infty$ then the photon interacts with the boundary and is either absorbed or scattered according to f_s .

4. In case of scattering the procedure is repeated after replacing the initial release point x_D with the scattering location \mathbf{x} and v_0 with the new direction \mathbf{v} .

In a similar way the Schwartz kernels from Equation (2.33) can be thought of as densities conditional on a photon being detected which can also be thought of as a distribution over trajectories of the associated particles. In the case of atmospheric dispersion, we don't need to concern ourselves with the associated Brownian paths because the way a particle reaches a certain position doesn't influence its optical properties and therefore has no effect on our measurements. In the case of (differential) absorption, which is described by Beer-Lambert's law, the intensity is a function of the path integral and therefore dependent on a photon's trajectory.

It shouldn't come as much of a surprise that in general, we won't be able to reconstruct all optical parameters from two Lidar measurements unless they are significantly constrained. We argue that in a way this is a similar problem to the issues related to turbulence. The wind velocity at every point in time and space alongside the rate of gas release would in theory be sufficient to reconstruct atmospheric transport to an arbitrary degree of detail [129]. Of course, such measurements are unrealistic and to the best of our knowledge, there is no reliable model for predicting turbulent transport based on crude meteorological data such as wind velocity, pressure, humidity, temperature, etc. measured at a single point inside the domain. Similarly unrealistic optical data, i.e. noise-free measurements covering the whole domain at every point in time, would be sufficient for perfect reconstruction of (time-dependent) images as well. Consequently, our image parameterisation is chosen such that the signal strength in the optical measurement is likely to be sufficient in order to reconstruct point estimates of the parameters necessary for the evaluation of an approximation to the atmospheric forward problem in a lower resolution.

Unlike in the case of atmospheric dispersion, there is no intuitive notion of averages since the source is pulsed and the scattering/absorption parameters depend not only on the number of particles but also their type (e.g. Mie-scattering for spheres [68]) which typically won't change over time. Nevertheless, we can compensate for missing information by employing entropy maximisation but before we are ready to explain the

details of our method we must develop a better understanding of the information content within our measurement which will allow us to find suitable parameters/constraints that will be the target of our inverse problem.

Chapter 3

Point estimation

3.1 DIAL type measurements with wide FOVs

In the presence of a gas plume the perturbations in the scattering and attenuation coefficients of the ambient medium that describe the optical properties of the image should be proportional to the gas concentration u . In other words, there are wavelength-dependent constants $C_a > 0$ and $C_s > 0$ such that $\forall x \in X$

$$\begin{aligned}\sigma_{a,\text{ambient}}(x) + C_a u(x) &= \sigma_a(x) \\ \sigma_{s,\text{ambient}}(x) + C_s u(x) &= \sigma_s(x)\end{aligned}\tag{3.1}$$

where $u(x)$ is the gas concentration at $x \in X$. If the plume is a mixture of unknown gases and aerosols, then having access to C_s and C_a is not realistic. In the case of DIAL, where we seek to map the concentration of a *particular* species of gas that is believed to be one of the constituents, it is only assumed that we know the absorption behaviour of the gas of interest for two carefully selected wavelengths. The value for $\sigma_{a,\text{ambient}}(x)$ or even C_a corresponding to the entire mixture may still be unknown at these wavelengths due to absorption of particles other than the gas of interest. All other quantities in Eq. (2.22), including the absorption behaviour of particles other than the gas of interest, are assumed unknown but *identical* at both of these wavelengths which is justified if the spectrum of the gas of interest has narrow absorption features (see Fig. 1.2). If we denote the chosen

wavelengths by “on” and “off” then this can be expressed as $\forall x \in X$

$$\begin{aligned}\sigma_{s(\text{on})}(x) &= \sigma_{s(\text{off})}(x) \\ f_{p(\text{on})}(x, \cdot) &= f_{p(\text{off})}(x, \cdot)\end{aligned}\tag{3.2}$$

and the same holds for any possibly occurring surface reflections due to f_s in Eq. (2.25). Further we assume that the absorption difference $\sigma_{a(\text{on})} - \sigma_{a(\text{off})}$ for these two wavelengths satisfies

$$C_{\text{ambient}}[\theta] + C_{\text{DIAL}}u[\theta](x) = \sigma_{a(\text{on})}(x) - \sigma_{a(\text{off})}(x) =: \alpha \in K(\phi | x_D).\tag{3.3}$$

for some good kernel ϕ , $x_D \in \partial X$ and a *known* $C_{\text{DIAL}} = C_{a(\text{on})} - C_{a(\text{off})} > 0$ and possibly unknown and θ -dependent constant $C_{\text{ambient}}[\theta] = C_{\text{DIAL}}u[\theta] \geq 0$ which corresponds to a not necessarily known amount of homogeneously distributed gas in the ambient atmosphere. Henceforth we will write σ_a, σ_s and f_p for the non-differential (offline) optical parameters, i.e. omit writing “off” explicitly, and use α for the differential absorption. In view of Eqs. (3.2) and (3.3) the complete set of optical parameters can then be expressed as $(\alpha, \sigma_a, \sigma_s, f_p)$. In cases where the concentration is much bigger than what’s typically present in the atmosphere, we may assume $C_{\text{ambient}} \approx 0$. Note that although u in general depends on time, such as in Eqs. (2.16) and (2.18), the quantity in Eq. (3.3) decidedly doesn’t depend on t because the time scales relevant for our optical measurements are many orders of magnitude smaller than those on which any meaningful change in the gas concentration u might occur.

3.1.1 Wide FOVs & angularly averaged measurements

It is worth noticing that our approach in principle generalises to scenarios where the ambient gas concentration is inhomogeneous but known, which might be the case if an area is being actively monitored and accurate measurements that have been averaged over long periods are available, while for problems where a plume is released into an unknown inhomogeneous atmosphere our dispersion based prior knowledge cannot account for most

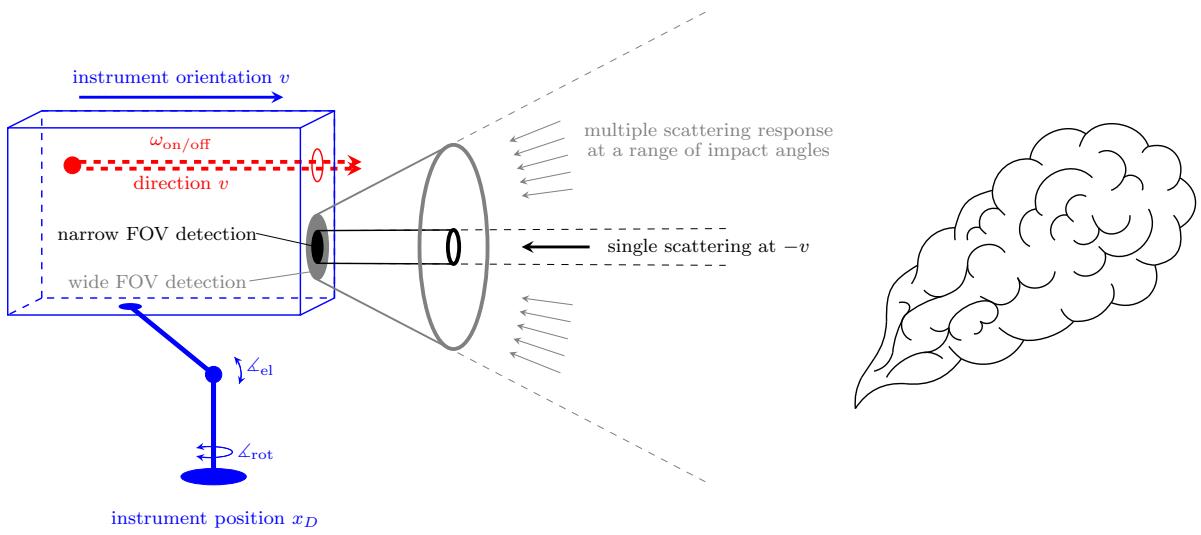


Figure 3.1: Online and offline laser pulses, denoted by $\omega_{\text{on/off}}$ (red), are released at the source located at x_D in direction v . A detector, also located at x_D , with a narrow FOV (black) captures the single scattering response incident at $-v$ whereas a wide FOV captures light from a range of directions that have scattered multiple times. After a measurement in direction v has been taken the instrument is re-oriented by adjusting \angle_{el} or \angle_{rot} (blue). This procedure is carried out for azimuthal and polar angles at a fixed instrument location x_D .

of the variability in the absorption difference α and our method based on Eq. (3.3) becomes unsuitable. These assumptions are quite different than assuming that the ambient quantities in Eq. (3.1) are fully known and all unknown perturbations are caused by the plume since we only assume that the difference in absorption is caused by the plume in a known way while all other quantities are free, although we will restrict them to some extent later on, as long as they are equal on both wavelengths. For the single scattering approximation and narrow FOVs this of course results in the usual DIAL method. When more complicated measurements are to be used there is no closed-form solution for the gas concentration anymore and, depending on the measurement setup and assumptions on the optical parameters it might not be possible to reconstruct all unknown quantities of Eq. (2.22) and careful analysis of the required assumptions and expected errors must be performed.

In practice, we measure light that reaches a detector that is orders of magnitude smaller than ∂X and can be essentially thought of as a point measurement at the boundary (see

Fig. 3.1). However, the solutions for Eq. (2.22), as well as the traces, are developed in such a way that, in order to have a mathematically consistent description of the measurement, we can a priori only restrict the albedo operator to some (small) region around a fixed point and think of a point measurement as a limit when the area of that region tends to zero. For an instantaneous pulse released at time t_0 in direction v_0 from a source and detector located at x_D that takes an angularly averaged measurement at time s within a field of view modelled by a continuous function g_m that can be seen as taking $g \rightarrow \delta(x - x_D)\delta(v - v_0)\delta(t - t_0)$ and $g_D \rightarrow \delta(x - x_D)\delta(t - s)g_m(x, v)$ in the situation of Eq. (2.33). The following result ensures that the resulting limit makes sense.

Lemma 3.1.1 (Angularly averaged distribution kernels). *Consider the Schwartz kernels $\mathfrak{a}^{(k)}$ for the summands of the albedo operator as in Eq. (2.33) and assume that the optical parameters σ_a, σ_s and f_p are bounded and continuous on the convex domain X . Let*

$$Z := \{(x, y, v, s) \in \partial X \times \partial X \times \mathbb{S}^2 \times (0, \infty) : \|x - y\| < s < 2\ell_v^+(y) \text{ and } v \in \partial V_{(-)}(y)\}$$

Then for any $k \geq 1$ and smooth function $g_m : \partial X \times \mathbb{S}^2 \rightarrow \mathbb{R}$

$$(x, y, v, s) \mapsto \mathfrak{m}^k(x, y, v, s) := \int_{\partial V_{(+)}(x)} \mathfrak{a}^k(x, v', y, v, s) g_m(x, v') dv' \quad (3.4)$$

as well as

$$(x, y, v, s) \mapsto \mathfrak{m}(x, y, v, s) := \sum_{k=0}^{\infty} \mathfrak{m}^k(x, y, v, s) \quad (3.5)$$

are continuous functions when restricted to Z . In particular, for any $x_D \in \partial X$ the expression $\mathfrak{m}(\cdot \mid x_D) : \partial V_{(-)}(x_D) \times (0, \infty) \rightarrow \mathbb{R}$

$$(v, s) \mapsto \mathfrak{m}(v, s \mid x_D) := \mathfrak{m}(x_D, x_D, v, s) \quad (3.6)$$

is well-defined and continuous as long as $x_D + \frac{s}{2}v \in X$, i.e. $\mathfrak{m}(\cdot \mid x_D) \in C(Z_{(-)}(x_D))$

where the domain is given by the open set

$$Z_{(-)}(x_D) := \{(v, s) \in \partial V_{(-)}(x_D) \times (0, \infty) : s < 2\ell_v^+(x_D)\}. \quad (3.7)$$

Using Eq. (3.4) we may similarly define for each order of scattering $k \in \mathbb{N}$ the functions $\mathbf{m}^k(\cdot | x_D) \in C(Z_{(-)}(x_D))$.

Remark. The quantities in Eqs. (3.4) to (3.6) obviously depend on the optical parameters σ_a, σ_s and f_p as well as the FOV modelled via g_m . As mentioned earlier, we want to consider measurements at two wavelengths “on/off” and therefore \mathbf{m}_{off} and \mathbf{m}_{on} corresponding to optical parameters $(\sigma_a, \sigma_s, f_p)$ and $(\sigma_a + \alpha, \sigma_s, f_p)$ respectively. The optical forward model can thus be regarded as a map $\mathcal{M} : \Xi \rightarrow C(Z_{(-)}(x_D)) \times C(Z_{(-)}(x_D))$ defined by

$$(\alpha, \sigma_a, \sigma_s, f_p) =: \xi \mapsto \mathcal{M}(\xi) = (\mathbf{m}_{\text{on}}(\cdot | x_D), \mathbf{m}_{\text{off}}(\cdot | x_D)) \quad (3.8)$$

where Ξ is a subset of bounded continuous and non-negative functions. Since we are primarily interested in the differential absorption component α it will be convenient to decompose $\xi = (\alpha, \sigma_a, \sigma_s, f_p) = (\alpha, \xi_{\text{off}})$

$$\Xi_{\text{off}}(\alpha) := \{(\sigma'_a, \sigma'_s, f'_p) \in C(X) \times C(X) \times C(X \times \mathbb{S}^2 \times \mathbb{S}^2) : (\alpha, \sigma'_a, \sigma'_s, f'_p) \in \Xi\} \quad (3.9)$$

so that $(\alpha, \xi_{\text{off}}) \in \Xi$ if and only if $\xi_{\text{off}} \in \Xi_{\text{off}}(\alpha)$. In order to fully connect the optical forward problem with the finite-dimensional estimation problem that will be considered throughout this work we also want to consider $\mathcal{G} : \Theta \rightarrow \Xi$

$$\theta \mapsto \mathcal{G}(\theta) = \xi[\theta] = (\alpha[\theta], \sigma_a[\theta], \sigma_s[\theta], f_p[\theta]) \quad (3.10)$$

which maps finite-dimensional vectors $\theta \in \Theta \subseteq \mathbb{R}^{N_\theta}$ to optical parameters $\xi[\theta] \in \Xi$. Combination of Eqs. (3.8) and (3.10) yields $\mathcal{M} \circ \mathcal{G}(\theta) = (\mathbf{m}_{\text{on}}[\theta], \mathbf{m}_{\text{off}}[\theta])$, i.e. Eq. (3.6) as a function of θ . More explicit examples will be defined later but will generally be obtained through relationships such as Eqs. (3.1) and (3.3) which are connected to θ via dispersion

processes $u[\theta]$ characterised by Eqs. (2.11) and (2.18).

In Lidar applications such as DIAL the instrument will typically be re-oriented along with the direction v of the initial pulse (see Fig. 3.1) and strictly speaking g_m (or ∂X) should depend on the initial direction to account for this. The presented results can be extended to that setting by considering Eqs. (3.4) and (3.5) for fixed g_m , corresponding to the sensor FOV for some v , and consider Eq. (3.6) as a function of time only by fixing v alongside x_D . Forming $m(v, s | x_D)$ as the “concatenation” of the resulting measurements for each v will preserve continuity in (v, s) under reasonable assumptions, i.e. when the sensor aperture changes continuously with the initial direction, and can be obtained from similar arguments based on dominated convergence as were used in the proof of Lemma 3.1.1. The domain must be sufficiently regular so that the albedo operator, corresponding Schwartz kernels and outer normal vectors are well defined. Technical difficulties could, for example, arise in situations where x_D is located at the edge of a convex polytope.

Proof. See Appendix A.1.1. □

3.1.2 Semi-parametric form of the RTE-based inverse problem

We need to rigorously address the question of whether the collected data is sufficient for a reliable estimation of the dispersion, or equivalently differential absorption, parameters. Standard RTE-related uniqueness conditions such as those given in [9, 28] for a variety of measurement types, require a detector almost everywhere on the boundary which may be realistic in some cases but clearly not when working with Lidar systems on such a large scale. The inverse transport problem with angularly resolved sources and angularly averaged measurements which is similar to our setting, apart from lacking temporal resolution, was considered in [81] where uniqueness and stability results for the optical parameters are given. In [49] the authors present similarity relations, i.e. negative results, for the case of arbitrary heterogeneous materials. In our case, we don't seek to recover all optical parameters and additionally have a strong constraint based on the dispersion process which means that the right question to ask is whether, and if so to what extent, the dispersion constraint regularises the inverse problem and whether this is enough to

recover the dispersion from single-detector data distorted by noise. In this section, we focus on the uniqueness of the reconstruction.

The following theorem shows that, under the condition from Eq. (3.2) and fixed scattering parameters, a time-resolved differential absorption measurement taken at the point x_D of the light source is good enough to recover the absorption difference $\sigma_{\text{on}(a)} - \sigma_{\text{off}(a)} = \alpha \in K(\phi | x_D)$ for a good kernel ϕ , i.e. α is a well behaved smooth function that admits a representation akin to that of Eq. (2.18). Note since we only require that the midpoints be inside X we strictly speaking have $K(\phi) \not\subseteq C_c^\infty(X)$, i.e. functions in $K(\phi)$ may extend over the boundary of X , which would be the case in situations such as when the gas is near the ground. The condition $\alpha \in K(\phi | x_D)$ states that x_D is outside the support of α which is weaker than requiring $\alpha \in C_c^\infty(X) \cap K(\phi)$. Intuitively, if $\alpha \in K(\phi | x_D)$ and X is convex, then an observer standing at x_D can “see” each component of α without having to fully turn around.

Simple uniqueness results As $K(\phi)$, and thus $K(\phi | x_D)$, only contains functions that can be represented by a finite tuple of real-valued parameters, Lemma 3.1.2 can be seen as a uniqueness result for discretised parameters. However, we don’t require an explicit choice of ϕ , i.e. any good kernel as per Definition 2.1.1 would work, the number of kernel functions used, or any other, possibly dispersion related, constraints that would restrict/regularise the problem further. Indeed, $K(\phi | x_D)$ is infinite-dimensional and looking for $\alpha \in K(\phi | x_D)$ without further regularisation is still an ill-posed problem.

Lemma 3.1.2. (*Uniqueness for given scattering*) *Assume that X is convex as well as otherwise sufficiently regular, $x_D \in \partial X$ and that \mathbf{m}_{off} is as in Lemma 3.1.1 for continuous optical parameters σ_a, σ_s and f_p . If $\alpha_1, \alpha_2 \in K(\phi | x_D)$ for a good kernel ϕ and $\sigma_s(x) > 0$ for all $x \in X$, then*

$$\alpha_1 = \alpha_2 \iff \mathbf{m}_{\text{on},1}(\cdot | x_D) = \mathbf{m}_{\text{on},2}(\cdot | x_D)$$

where $\mathbf{m}_{\text{on},1}$ and $\mathbf{m}_{\text{on},2}$ correspond to optical parameters $(\sigma_a + \alpha_1, \sigma_s, f_p)$ and $(\sigma_a + \alpha_2, \sigma_s, f_p)$ respectively. In other words, fixing σ_a, σ_s and f_p the forward map Eq. (3.8) is injective

when restricted to $\alpha \in K(\phi | x_D)$ for some good kernel ϕ as in Definition 2.1.1.

Proof. See Appendix A.1.2 □

In general, the scattering parameters won't be fixed a priori and it is necessary to consider an inverse problem where σ_a, σ_s and f_p are unknown as well. In the case of narrow (or separately measured) FOVs the validity of a single scattering approximation means that the terms necessary to prove Lemma 3.1.2 are always accessible from the measured data which therefore provides sufficient information about the optical parameters in order to reconstruct the absorption field. This is of course the well-known foundation for standard DIAL. The following result shows that if only a wide FOV is available, then the issue could essentially be further reduced to the influence of a subset of the optical parameters.

Lemma 3.1.3. *Assume as before that X is convex as well as otherwise sufficiently regular, $x_D \in \partial X$ and that \mathbf{m}_{off} is as in Lemma 3.1.1 for continuous optical parameters σ_a, σ_s and f_p . Further assume that f_p is known, $\sigma_a, \sigma_s \in K(\phi | x_D)$ for a good kernel ϕ with a known proportionality constant between them while the constant terms, which correspond to the homogeneous ambient quantities in the dispersion and are denoted by w_0 in Definition 2.1.1, are known. Then $\mathbf{m}_{\text{off}}(\cdot | x_D)$ uniquely determines σ_a and σ_s .*

Proof. See Appendix A.1.2 □

Combining Lemmas 3.1.2 and 3.1.3 we obtain that the measured data is arguably good enough to recover an additional scattering-related parameter even if it's assumed to have a virtually unconstrained and very general form. However, since we measure light from where it was released (see Fig. 3.1) the measurements will be greatly impacted by the phase function or, more precisely, its backscattering coefficient. Although this could potentially be compensated by means of additional kernel functions, it unfortunately becomes infeasible to use this naturally occurring parameterisation when solving the inverse problem computationally given how it requires evaluations of the RTE for potentially complex and high-dimensional parameters. In the next paragraph, we will present a way around this issue that allows us to use a high-dimensional parameterisation that (partially) accounts

for the dominant scattering behaviour but doesn't present an impractical computational challenge.

A semi-parametric relaxation From a practical perspective, one of the key features associated with DIAL is the ability to recover the absorption field without having to concern oneself with the scattering parameters. On a more technical level however, the reconstructed parameters are the differential absorption field α as well as \mathbf{m}_{off} . Assuming single scattering with a free line of sight and thus ignoring the boundary of the domain, we have, for $\mathbf{m}_{\text{on/off}}^1(\cdot | x_D)$ as in Lemma 3.1.1,

$$\mathbf{m}_{\text{off}}^1(v, 2t | x_D) \propto \frac{1}{t^2} e^{-2 \int_0^t \sigma_{a+s}(x_D + rv) dr} \sigma_s(x_D + tv) f_p(x_D + tv, v \rightarrow -v) \quad (3.11a)$$

$$\mathbf{m}_{\text{on}}^1(v, 2t | x_D) \propto \frac{1}{t^2} e^{-2 \int_0^t \sigma_{a+s}(x_D + rv) + \alpha(x_D + rv) dr} \sigma_s(x_D + tv) f_p(x_D + tv, v \rightarrow -v). \quad (3.11b)$$

The reconstruction of $\mathbf{m}_{\text{off}}^1$ is, of course, trivial as it's measured directly but at the same time also all that is required in order to account for arbitrary scattering behaviour when the detector uses a narrow FOV. Note that the measurements, regardless of the FOV (and model), can be written as $(\mathbf{y}_{\text{on}}(v, t), \mathbf{y}_{\text{off}}(v, t))$ subject to the constraints

$$\mathbf{y}_{\text{off}}(v, t) = \mathbf{m}_{\text{off}}(v, t | x_D) \psi(v, t) \quad (3.12a)$$

$$\mathbf{y}_{\text{on}}(v, t) = \mathbf{m}_{\text{on}}(v, t | x_D) \psi(v, t) \quad (3.12b)$$

$$\psi(v, t) = 1 \quad (3.12c)$$

where $\mathbf{m}_{\text{on/off}}$ have the same underlying optical parameters and ψ is (for now) a redundant constant. In classical DIAL the constraint in Eq. (3.12c) is dropped which results in what is essentially a modified forward model that, due to the multiplicative relationship between Eqs. (3.11a) and (3.11b), can capture all effects resulting from different σ_a , σ_s and f_p . It is easily seen that the value for the differential absorption α is still uniquely determined after this relaxation. The same is not true for the other optical parameters or ψ . Since ψ

is a free parameter after the relaxation, we may divide by $\mathfrak{m}_{\text{off}}(\cdot \mid x_D)$ and obtain

$$\bar{\mathfrak{h}}_{\text{off}}(v, t) = \tilde{\psi}(v, t) \quad (3.13a)$$

$$\bar{\mathfrak{h}}_{\text{on}}(v, t) = \frac{\mathfrak{m}_{\text{on}}(v, t \mid x_D)}{\mathfrak{m}_{\text{off}}(v, t \mid x_D)} \tilde{\psi}(v, t) \quad (3.13b)$$

$$= e^{-2 \int_0^{t/2} \alpha(x_D + rv) dr} \tilde{\psi}(v, t) \quad (3.13c)$$

with Eq. (3.13c) only being true for single scattering. Comparing Eqs. (3.11a) and (3.11b) with Eqs. (3.13a) and (3.13c) it becomes apparent that $\tilde{\psi}$ essentially acts as a scaled back-scattering coefficient in the case of narrow FOVs and a single scattering model.

Formally we may think of $\tilde{\psi}$ as a density corresponding to photon detection. More specifically, if we denote by Δt the bin width, H_D the energy of the light source (expressed as the expected number of photons per pulse) and A_D the detector size, then we have

$$\mathbb{P}(\text{Photon from direction } v \text{ observed in bin centred at } t) \approx H_D^{-1} \Delta t A_D \tilde{\psi}(v, t).$$

Furthermore, we also have

$$\mathbb{P}(\text{Differential absorption} \mid \text{Detection from direction } v \text{ at time } t) \approx 1 - \frac{\mathfrak{m}_{\text{on}}(v, t \mid x_D)}{\mathfrak{m}_{\text{off}}(v, t \mid x_D)}.$$

Before we give an interpretation of the optical parameters in the re-parameterised model, we give a uniqueness statement similar to those in Lemmas 3.1.2 and 3.1.3.

In view of Eq. (3.13c) it becomes apparent that $\tilde{\psi}$ is proportional to the local back-scattering rate $\sigma_s(x_D + tv) f_p(x_D + tv, -v \cdot v)$. We may (loosely) associate $\tilde{\psi}$ with a similar quantity also in a more general RTE setting. Indeed, the proof of Lemma 3.1.1 relied on the fact that a photon cannot be observed unless there was at least one scattering event resulting in a sufficiently large change of direction, the magnitude of which depends on the order of scattering under consideration and is maximal for single scattering. Assuming that the phase function f_p is (approximately) homogeneous in the spatial coordinate, which will be true if the same particles are responsible for the majority of scattering events throughout the domain, then the intensity along each path that corresponds to

a low order detection will contain a factor that resembles the backscattering coefficient $f_p(x_D + tv, -v \cdot v)$. In such situations our high-dimensional nuisance parameters $\tilde{\psi}$ can therefore be viewed as a parameter that accounts for the back- or, more generally, large-angle scattering rate.

Theorem 3.1.4. *Assume as before that X is convex as well as otherwise sufficiently regular, $x_D \in \partial X$ and that \mathbf{m}_{off} is as in Lemma 3.1.1 for continuous optical parameters σ_a, σ_s and f_p . Assume that $\sigma_a, \sigma_s \in K(\phi_{\mathfrak{s}} | x_D)$ for a good kernel $\phi_{\mathfrak{s}}$ and that f_p is known. Further assume that $\alpha \in K(\phi_{\mathfrak{a}} | x_D)$ for a good kernel $\phi_{\mathfrak{a}}$ such that for any $h > 0$*

$$\frac{\phi_{\mathfrak{s}}(1 - \frac{\varepsilon}{h})}{\int_0^\varepsilon \phi_{\mathfrak{a}}(1 - z) dz} \xrightarrow{\varepsilon \rightarrow 0} 0. \quad (3.14)$$

Further assume that σ_a, σ_s and α have common mid-points as well as widths while the weights are equal up to proportionality (with a constant shared between all summands) and the constant terms, which correspond to the homogeneous ambient quantities in the dispersion and are denoted by w_0 in Definition 2.1.1, are known and strictly positive in the case of σ_s . Then the absorption $\frac{\mathbf{m}_{\text{on}}(\cdot | x_D)}{\mathbf{m}_{\text{off}}(\cdot | x_D)}$ uniquely determines σ_a, σ_s and α .

Proof. Similar to Lemma 3.1.2. Condition Eq. (3.14) ensures that the scattering parameters do not “interfere” with single scattering as the uniqueness is based on the observation that single scattering “precedes” multiple scattering at the edge of a puff. It effectively requires that the scattering kernel has diminishing mass in its tails where its contribution to the signal will primarily be single/back-scattering and is meant to be compensated by $\tilde{\psi}$ and ensures that we can treat reconstructions “within” the plume in the same way as those at the “edges”. In fact, if the plume was given by a single kernel function this condition would not be necessary. Since the regularity imposed by the dispersion model makes the plume behave much more like a single structure as opposed to many independent puffs, we believe that this technical condition is best thought of as an artifact of our proof. Note that regardless of $\phi_{\mathfrak{a}}$ one can always find a kernel $\phi_{\mathfrak{s}}$ that is arbitrarily similar to $\phi_{\mathfrak{a}}$ in the L_1 sense but satisfies Eq. (3.14). The details can be found in Appendix A.1. \square

The essence of Theorem 3.1.4 is that the parameters associated with the relaxed model

as given in Eq. (3.13a) and Eq. (3.13b) can be identified from the data if the scattering and absorption particles follow the same atmospheric dispersion process. Knowledge of the ambient quantities based on previous measurements or experience is in principle not necessary. The non-differential absorption constant of σ_a is in fact completely irrelevant for the reconstruction of the differential absorption α but reconstruction of a homogeneous ambient scattering field requires additional constraints/assumptions and any ad-hoc reconstruction from noisy measurements is, not least because of this, arguably less realistic than a choice based on prior knowledge. It should be noted that conditions such as the required proportionality of parameters are somewhat necessary since we cannot reconstruct anything from differential absorption data in the absence of the trace gas. In particular, any obstacles before the plume can only be modeled as a *known* baseline concentration of scattering particles. We should also note that the conditions imposed as Eq. (3.14) are such that the resulting function spaces can produce a more accurate approximation in those scenarios where the application of wide FOV measurements is more appropriate but also highlight an inherent limitation of wider FOVs and the RTE model. If exact reconstruction is sought then we rely on the separation of single and multiple scattering which is possible only locally at times where we measure signals scattered at plume boundaries. Such data is fundamentally unstable as not only the tails of a distribution can be heavily perturbed by even small changes in the parameters but the signal at those times/distances will also be particularly weak, noisy and possibly dominated by ambient photons.

Nonetheless Theorem 3.1.4 can be used as a best-case reference as to which parameters we can hope to resolve by wide FOVs and although exact reconstruction is problematic, it turns out that other aspects of the measurement are more stable under perturbations of the scattering parameters than the tails and wide FOV measurements can still bear useful information with regards to the differential absorption. The necessary results are developed in Section 3.2 and Chapter 4.

Path spaces and entropy Recall that in the case of single scattering, i.e. sufficiently narrow FOVs, the differential absorption is independent of f_p , and not being able to

reconstruct it does not affect the reconstructed gas concentration. This is of course not true in a more general RTE setting and yet Theorem 3.1.4 requires knowledge of f_p . It is often assumed that the scattering $v' \rightarrow v$ only depends on the cosine of the angle $v' \cdot v$ in which case we may write $f_p(x, v' \rightarrow v) = f_p(x, v' \cdot v)$. As argued earlier, the nuisance parameter $\tilde{\psi}$ can account for variability in the backscattering behaviour, which is arguably the most important aspect, but even if f_p is assumed spatially homogeneous, i.e. $f_p(x, v \cdot v') = g_p(v \cdot v') = f_p(y, v \cdot v')$ for any $x, y \in X$ and some spatially homogeneous g , it may still have unaccounted irregularities on $[-1 + \gamma, 1]$ where $\gamma > 0$ serves (informally) as a cut-off for large angles. In practice, the most commonly modeled component of f_p is the forward peak, e.g. by means of the Henyey-Greenstein phase function [61]. If the distribution of direction depends only on the cosine of inward and outward direction $v \cdot v'$, this is somewhat similar to modelling the first angular moments. As mentioned earlier, the introduction of $\tilde{\psi}$ implies that the modelled intensity of measured light is independent of the RTE scattering parameters, particularly σ_s . Note that up to first order we can approximate $g \approx \lambda \delta_{\{v \cdot v' = 1\}} + (1 - \lambda)g_0 =: g_\delta$ by putting

$$\lambda = \int_{\mathbb{S}^2} v \cdot v' g(v \cdot v') dv'$$

and picking g_0 such that the analogous integral is 0. It is not hard to see that solving the RTE Eq. (2.22) with $f_p \rightarrow g_\delta$ is the same as setting $f_p \rightarrow g_0$ as well as $\sigma_s \rightarrow (1 - \lambda)\sigma_s$. Note that this observation is the foundation for other commonly used phase function approximations that use δ -distributions for forward peaks such as the δ -Eddington and related methods [74, 148]. The inverse idea is also true, i.e. one may also select a baseline g_0 with a forward peak, which effectively results in an implicit choice of a baseline that gets homogenised as σ_s increases.

Since the above developments indicate that the optical parameters that control the absorption $\frac{m_{on}(\cdot | x_D)}{m_{off}(\cdot | x_D)}$ can be used to model forward peaked behaviour, Theorem 3.1.4 can be interpreted as a set of restrictions on σ_a and σ_s which allow partial reconstruction of the phase function or at least certain aspects thereof (see Fig. 3.2).

Given that we cannot evaluate the forward model without a full set of optical parameters,

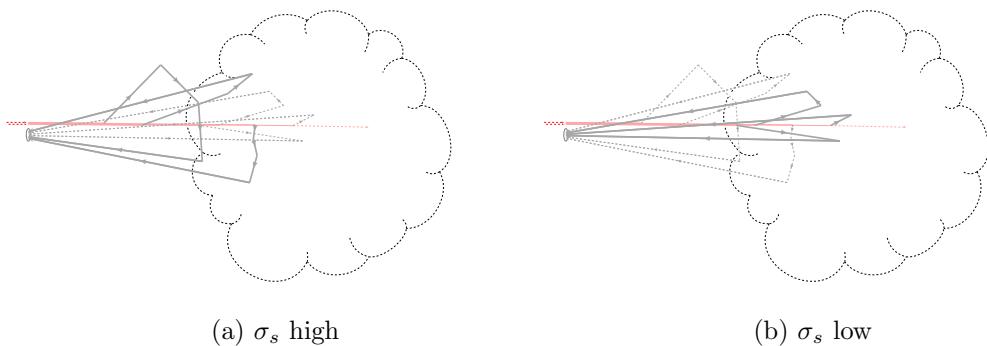


Figure 3.2: An increase in the scattering function σ_s means light is more likely to scatter resulting in less low-order scattering at distances corresponding to deep inside the plume. As the non-parametric component controls the response strength it acts as a normalising constant while σ_s controls the diffusivity and thus the likely trajectories of the measured photons. A high amount of scattering particles as shown in image (a) favours high-order scattering (solid line) over forward peaked trajectories that reach deeper inside the cloud (dashed lines) whereas the opposite can be observed in figure (b).

we must find a way to deal with the lack of information regarding f_p or at least find a suitable candidate for g_0 . In fact, it only makes sense to use wider FOVs, or even multiple FOVs where the narrow component can be separated, if we handle the resulting ill-posedness without falling victim to regularisation-induced errors/biases that are larger than narrow FOV reconstruction errors. In theory, we can select g_0 that maximises the entropy on the sphere \mathbb{S}^2 subject to given constraints which should be based on prior information/beliefs. It is worth noticing however that we found the choice of g_0 to be of diminishing importance when it comes to the actual performance of our algorithms. An indication as to why this phenomenon will likely be observed more generally, at least in the situations of relevance for our measurements and methods, is given in the next section.

3.2 Signal quality and noise considerations

The results that were shown in Lemmas 3.1.2 and 3.1.3 as well as Theorem 3.1.4 are similar to what is available for the classical DIAL approaches based on single scattering but weaker in certain ways. Combining Lemmas 3.1.2 and 3.1.3 essentially yields the

result of classical DIAL with the additional, unrealistic assumption such as that the phase function be known. The arguably more useful result in Theorem 3.1.4 uses the same data as classical dial but only provides what is perhaps best thought of as regularity conditions under which a subset of the optical parameters is uniquely determined by differential absorption only. Notably, each of our proofs relied on extracting the effect of single scattering in order to identify the perturbation α in the absorption function and we must ask ourselves whether there is any benefit in taking wider FOVs and multiple scattering into account. If we can separate first and higher-order scattering contributions, e.g. through partially resolving the incident angles, then a good reconstruction algorithm won't make things worse since we could just ignore all but the narrow FOV with the single scattering measurement. In situations where such a separation is not an option, things get more complicated and angularly averaged point measurements from wide FOVs can be better or worse than narrow ones, depending on the situation.

3.2.1 Poisson noise model

To make more formal statements, we must first introduce a noise model for our optical measurement. If the light source releases two instantaneous impulses at time $t = t_{\text{on/off}}$ in direction v consisting of $\text{Poisson}(H_D)$ many photons, where $H_D > 0$ is proportional to the pulse energy, then the fully time-resolved observation at the detector in the interval $[t_{\min}, t_{\max}]$ will take the form of a Poisson point process with intensity measure $H_D \mathbf{m}_{\text{on/off}}(v, \cdot | x_D)$ respectively. In practice, it is arguably more realistic to have measurements that are aggregated into N_t bins, rather than infinitely time-resolved, which would mean that for each direction v_i for $i \in 1:N_v$, in which a light pulse is released, we observe two independent random arrays $\mathbf{m}_{i,j}^{\text{on}}, \mathbf{m}_{i,j}^{\text{off}}$ with independent entries such that, at least approximately,

$$\begin{aligned}\mathbf{m}_{i,j}^{\text{on}} &\sim \text{Poisson}(\Delta t H_D A_D \mathbf{m}_{\text{on}}(v_i, t_j | x_D) \psi_{i,j}) \\ \mathbf{m}_{i,j}^{\text{off}} &\sim \text{Poisson}(\Delta t H_D A_D \mathbf{m}_{\text{off}}(v_i, t_j | x_D) \psi_{i,j})\end{aligned}\tag{3.15}$$

where Δt is the bin width, A_D the detector size, t_j is the centre of bin j for $j \in 1:N_t$ and a mid-point quadrature rule was used to integrate $\mathbf{m}_{\text{off}}(v_i, \cdot | x_D)$ over the bin and detector surface centred at $(v_i, t_j) \in Z_{(-)}(x_D)$. The values $\psi_{i,j}$ are a discrete version of the semi-parametric component introduced in Section 3.1.2. It should be noted that Eq. (3.15) ignores contributions of ambient light which change the mean but not the fact that the data has a Poisson distribution. The effect this has on image reconstruction quality is discussed later in this section. Realistically, there are likely other detector-related sources of distortion in our data apart from the Poisson nature of the optical measurement but for weak signals, i.e. very low photon counts per bin, or, perhaps more realistically, low differential absorption we can expect that the optical noise dominates and our assumption in Eq. (3.15) becomes a good approximation for the measurement. Recall that we want to use such data $(\mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}})$ to find the differential absorption

$$\alpha(x) = \sigma_{a(\text{on})}(x) - \sigma_{a(\text{off})}(x)$$

where the spatial absorption difference $\alpha = \alpha[\theta]$ is parameterised by $\theta \in \Theta$ and has the form Eq. (3.3). Under the conditions of Theorem 3.1.4 and in accordance with Eqs. (3.1) and (3.3) the non-differential parameters σ_a and σ_s are “overlapping” with α and will also be assumed as parameterised by θ through a map \mathcal{G} while f_p is fixed. This puts us in the situation described by Eqs. (3.8) and (3.10) such that we may write

$$\mathcal{M} \circ \mathcal{G}(\theta) = (\mathbf{m}_{\text{on}}[\theta](\cdot | x_D), \mathbf{m}_{\text{off}}[\theta](\cdot | x_D)). \quad (3.16)$$

Consequently $\mathbf{m}_{\text{on/off}} = \mathbf{m}_{\text{on/off}}[\theta]$ are functions of θ and we may form the log-likelihood $L(\psi, \theta | \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}})$ for data $(\mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}})$ as in Eq. (3.15)

$$\begin{aligned} L(\psi, \theta | \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}) &= \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \mathbf{m}_{i,j}^{\text{on}} \log(\psi_{i,j} \mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D)) \\ &\quad - H_D \Delta t A_D \psi_{i,j} \mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D) \\ &+ \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \mathbf{m}_{i,j}^{\text{off}} \log(\psi_{i,j} \mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)) \\ &\quad - H_D \Delta t A_D \psi_{i,j} \mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D) \end{aligned} \tag{3.17}$$

For that parameter, we may compute the (efficient) Fisher information matrix

$$J(\theta) = \mathbb{E} \left(-\nabla_{\theta}^2 \max_{\psi \in \Psi} L(\theta | \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}) \right) \tag{3.18}$$

where Ψ is a (typically non-trivial and very large) set of feasible values for the array ψ which can in principle be used to enforce a set of constraints such as Eq. (3.12c). Assuming that there is $\theta^* \in \Theta$ such that $\alpha[\theta^*]$ is the true difference in absorption, then we know that under suitable regularity conditions [104], which in our later more general developments may admittedly fail to hold, the maximum likelihood estimator $\hat{\theta}_{\text{MLE}}$ satisfies

$$\hat{\theta}_{\text{MLE}} \sim \text{AN}(\theta^*, J(\theta^*)^{-1}), \tag{3.19}$$

where **AN** denotes asymptotic normality. In certain situations Eq. (3.19) can be a reasonable approximation for the distribution of $\hat{\theta}_{\text{MLE}}$ but even if we ignore that fact we easily see that for a simple situation, such as the one considered in Section 3.2.2, Eq. (3.19) is essentially the square of the sensitivity of the noiseless measurement scaled by the standard deviation, a measure for the amount of noise, making it arguably a reasonable measure of signal quality in $\mathbf{m}_{\text{on/off}}$ for a particular set of optical parameters.

3.2.2 Trivial edge cases

Before considering more general scenarios we can develop a basic intuition by looking at the situation of known scattering as in Lemma 3.1.2 and in particular $\psi_{i,j} = 1$ fixed. This restriction will be dropped later in this work. In that case we have

$$J(\theta) = \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} J_{\text{on}}(v_i, t_j | x_D)[\theta] + J_{\text{off}}(v_i, t_j | x_D)[\theta]$$

where we abbreviated

$$J_{\text{on/off}}(v_i, t_j | x_D)[\theta] = H_D \Delta t A_D \frac{\nabla_\theta \mathbf{m}_{\text{on/off}}[\theta](v_i, t_j | x_D) \nabla_\theta \mathbf{m}_{\text{on/off}}[\theta](v_i, t_j | x_D)^\top}{\mathbf{m}_{\text{on/off}}[\theta](v_i, t_j | x_D)}. \quad (3.20)$$

Narrow vs. wide vs. multiple FOVs Note that we can think of averaged measurements from wider FOVs, corresponding to choices of the function g_m in Lemma 3.1.1 with a larger support, as the sum of single and multiple scattering photons which are independent **Poisson** distributed random variables with intensities given by $\Delta t H_D \mathbf{m}_{\text{on/off}}^1$ and $\Delta t H_D \sum_{j=2}^{\infty} \mathbf{m}_{\text{on/off}}^j$ respectively. First, consider a constant absorption difference, i.e. we have

$$C[\theta] = \sigma_{a(\text{on})}(x) - \sigma_{a(\text{off})}(x)$$

alongside some known $\sigma_s, \sigma_a > 0$ on X . We easily see that

$$\mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D) = \mathbf{m}_{\text{off}}(v_i, t_j | x_D) \exp(-t_j C[\theta]).$$

for any $(i, j) \in 1:N_v \times 1:N_t$ and $\theta \in \Theta$. If the scattering parameters, and thus \mathbf{m}_{off} , are fixed then \mathbf{m}_{off} is independent of θ and we have $J_{\text{off}} = 0$ while

$$J_{\text{on,wide}}(v_i, t_j | x_D)[\theta] = J_{\text{on,narrow}}(v_i, t_j | x_D)[\theta] \frac{\mathbf{m}_{\text{off,wide}}(v_i, t_j | x_D)}{\mathbf{m}_{\text{off,narrow}}(v_i, t_j | x_D)} \quad (3.21)$$

where $J_{\text{on,wide}}$ and $J_{\text{on,narrow}}$ are as in Eq. (3.20) and correspond to wide and narrow FOVs respectively. As $\mathbf{m}_{\text{off,wide}} \geq \mathbf{m}_{\text{off,narrow}}$ the information in each measurement point

increases, the overall quality of the signal would increase considerably in optically thick environments. This is of course expected as in homogeneous environments the number of gas molecules “encountered” by each photon depends only on the length of the path and a Lidar therefore collects all information that is relevant in order to evaluate the differential absorption between the online and offline measurements.

The case where the gas is spread evenly in the domain is in a way ideal because we don’t have to worry about how light came to the detector and averaging over a wider FOV has no downsides. This situation can be interpreted as the extreme case of a very large gas plume where its size is taken to ∞ . Its counterpart can be seen as the situation where the same, unknown amount $C[\theta]\text{vol}(X)$ of gas is taken and accumulated in a very small area around a point $\frac{t_j}{2}v_i + x_D$. As the area is taken to 0 we end up with $\nabla_\theta \mathbf{m}_{\text{on}}^k(v_i, t_j | x_D) \rightarrow 0$ for $k \geq 2$ since most of the time multiply scattered light won’t pass through that small area where the gas is located. Single scattering on the other hand is much more localised and remains sensitive to arbitrarily small objects. As an immediate consequence, we obtain

$$J_{\text{on,wide}}(v_i, t_j | x_D)[\theta] \approx J_{\text{on,narrow}}(v_i, t_j | x_D)[\theta] \frac{\mathbf{m}_{\text{off,narrow}}(v_i, t_j | x_D)}{\mathbf{m}_{\text{off,wide}}(v_i, t_j | x_D)} \quad (3.22)$$

which is exactly the reverse scaling compared to Eq. (3.21). For times larger than t_j the multiply scattered measurement component does retain some sensitivity but the overall situation is considerably worse than in the case of what can be thought of as an infinitely spread out gas plume to the point where averaging over wider FOVs will likely make things worse rather than yield an improvement in any meaningful way.

Effect of ambient light In practice we may also have to consider photons that reach the detector from light sources that are unrelated to our instrument and are out of our control, such as sunlight. Assuming that these sources are approximately constant over a sufficiently large period of time we can think of them as an independent Poisson distributed quantity with known intensity H_a that is added on top of \mathbf{m}^{on} and \mathbf{m}^{off} . Ambient light will in all situations of interest be emitted by a source with continuous spectrum and therefore

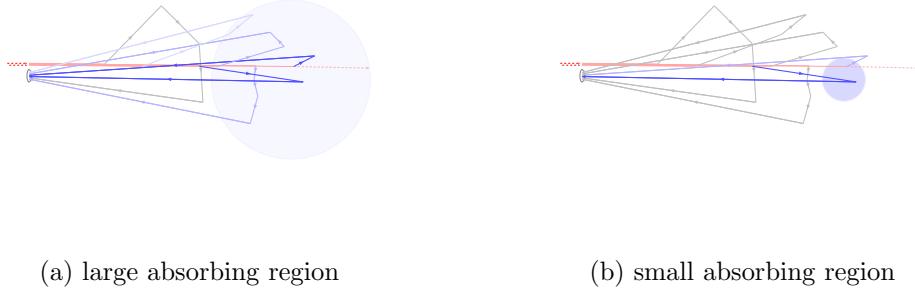


Figure 3.3: Figures (a) and (b) show why the absorption in the wide FOV is heavily dependent on ambient scattering. Only blue trajectories are sensitive to the patch of interest. Dark blue paths are strongly affected whereas photons along grey paths have a noise-like effect on the measurement.

be unaffected by differential absorption which only affects the online wavelength because the spectrum is singular, i.e. (at least approximately) given by a single wavelength. The amount of ambient photons scales with bin width, detector size and width of the aperture so this changes the mean of the data from $H_D \Delta t A_D \mathbf{m}_{\text{on/off}}$ to $H_D \Delta t A_D \mathbf{m}_{\text{on/off}} + \Delta t A_D H_a$ for some H_a that depends on the FOV, but it doesn't change its underpinning distribution. In that situation the quantities from Eq. (3.20) decrease to

$$\left(1 + \frac{H_a}{H_D \mathbf{m}_{\text{on/off}}[\theta](v_i, t_j | x_D)}\right)^{-1} J_{\text{on/off}}(v_i, t_j | x_D)[\theta] \quad (3.23)$$

which can be substantial when H_a is of the same order as the detector response $H_D \mathbf{m}_{\text{on/off}}$, or even larger. In reality, the factor in Eq. (3.23) depends on a lot of variables other than the FOV such as the spectral filter or the impulse strength H_D . $\mathbf{m}_{\text{on/off}}$ can also be highly inhomogeneous and its magnitude in the relevant data points, i.e. where the plume is located, may scale with the square of its distance from the detector. Quantifying the adversarial effect of an ambient light source is therefore not something that can be done in general but will depend on the use case. Nonetheless, the effects are always worse for larger FOVs because the aperture acts as a filter not only for multiple scattering but also for ambient illumination. Apart from having better resolution for fine image features, this observation is perhaps the primary reason why it is preferable to use multiple FOVs whenever possible. Indeed, if instead of averaging the data we are able to separate single

and multiple scattering and the Fisher information of the measurement is uniformly, i.e. for any set of optical parameters and any $\alpha[\theta]$, larger than that corresponding to just a narrow FOV and single scattering data. This should come as no surprise since with multiple FOV data we can perform any reconstruction procedure that would be possible with a single FOV alone.

3.2.3 Weak absorption errors

In general, neither of the above scenarios is realistic and the true concentration field lies most likely somewhere in between a fully homogeneous and strictly local distribution of gas and for approximately homogeneous σ_s one should only expect improvements by including photons from wider FOVs for fairly large plumes. However, if the σ_s is not homogeneous and scattering is caused mostly by particles in the plume, e.g. smoke or water vapour, we end up with a *very* different behaviour in the multiple scattering component. Indeed, assuming that ambient scatterers are rare and concentrated mostly around the plume of shape $\alpha = \sigma_{a(\text{on})} - \sigma_{a(\text{off})}$, then most photons that reach the detector will have travelled through the absorbing medium of interest and we end up with a situation much more like Eq. (3.21), even in the case of highly localised plumes. Unlike with the two trivial edge cases considered in Section 3.2.2, for non-trivial images it is not possible to eliminate the scattering parameters from the quantities that can be used in order to quantify the uncertainty/errors of the reconstructed absorption parameters. In the following, we will study a more relevant scenario where the plume shape is known and the concentration is low. Although not entirely unrealistic this may seem rather restrictive at first but we will later discuss how this transfers to considerably more general plumes of unknown shape. We start with the following simple result which relates **Poisson** and **Normal** random variables when the **Poisson** intensities are relatively large. It further establishes a natural connection with our semi-parametric formulation as we explicitly consider the differential absorption in the form of $\frac{\mathbf{m}_{i,j}^{\text{off}}}{\mathbf{m}_{i,j}^{\text{on}}}$.

Lemma 3.2.1. *Let $\mathbf{m}_{i,j}^{\text{on}}, \mathbf{m}_{i,j}^{\text{off}}$ as well as A_D, H_D and Δt be as in Eq. (3.15) and define for each direction v_i and time bin corresponding to mid-point t_j the random variables $\mathbf{y}_{i,j}$*

and $\mathbf{z}_{i,j}$ via

$$\begin{aligned}\mathbf{y}_{i,j} &:= \log \left(\frac{\mathbf{m}_{i,j}^{\text{off}}}{\mathbf{m}_{i,j}^{\text{on}}} \right) \\ \mathbf{z}_{i,j} &:= \frac{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}}{2}.\end{aligned}\tag{3.24}$$

Further assume that the differential absorption is considerably smaller than the signal, i.e. that we have

$$\frac{\mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D) - \mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D)}{\mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)} \ll 1$$

as well as

$$\Delta t H_D A_D \mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D) \gg 1.$$

Then $\mathbf{y}_{i,j}$ as well as $\mathbf{z}_{i,j}$ are mutually independent and their distribution can be approximated as

$$\mathbf{y}_{i,j} \sim \text{Normal} \left(\frac{\mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)}{\mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D)} - 1, \frac{2}{\Delta t H_D A_D \mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)} \right)\tag{3.25}$$

$$\mathbf{z}_{i,j} \sim \text{Normal} \left(\Delta t H_D A_D \mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D), \frac{\Delta t H_D A_D \mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)}{2} \right).\tag{3.26}$$

Proof. Note that $\mathbf{y}_{i,j}$ and $\mathbf{z}_{i,j}$ are not independent but the arrays \mathbf{y} and \mathbf{z} created from objects as in Eq. (3.24) have independent entries due to the assumed independence of the data in each bin and direction. The normal approximation for the $\mathbf{z}_{i,j}$ is standard while the result for $\mathbf{y}_{i,j}$ requires [36] after a Taylor approximation of the logarithm $\log(x) \approx x - 1$ for $x \approx 1$. Note that the effect of (differential) absorption is neglected in the variance terms. \square

It is easily seen that \mathbf{y} is essentially a discretised and noisy version of the same quantity (subject to a bijective transform) as the one analysed in Theorem 3.1.4 which considered its dependence on the dispersion parameters. Recall that due to Eqs. (3.11a) and (3.11b)

for single scattering we have

$$\frac{\mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)}{\mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D)} - 1 = e^{2 \int_0^{t_j/2} \alpha[\theta](x_D + rv_i) dr} - 1 \approx 2 \int_0^{t_j/2} \alpha[\theta](x_D + rv_i) dr \quad (3.27)$$

where the final approximation's accuracy depends on the magnitude of the absorption $\alpha[\theta]$ or, more precisely, the corresponding line integrals. For the remainder of this section we assume that α is known up to a constant, i.e. we have knowledge of a function u_0 such that we can write

$$\alpha[\theta](x) = \theta_0 u_0(x) \quad (3.28)$$

where $\theta_0 > 0$ is unknown. For the validity of our theoretical developments u_0 does not have to be restricted any further but a more specific choice will be considered later on.

Following the notation introduced in Eqs. (2.2) and (2.3) we decompose the parameter into $\theta = (\theta_0, \theta_{1:d}) \in \Theta \subseteq [0, \infty) \times \mathbb{R}^d$, i.e. the absorption component θ_0 (which is of primary interest in this setting) along with a d -dimensional nuisance component which controls the scattering behaviour. This is identical to the cases discussed in the previous section except that the plume may now take an arbitrary but known shape given by u_0 . Note that Eq. (3.27) can be seen as a first order Taylor approximation around $\theta_0 = 0$ which generalises to $\mathbf{m}_{\text{on/off}}$ in the RTE setting. Since \mathbf{m}_{on} is identical to \mathbf{m}_{off} in the absence of differential absorption, i.e. when $\theta_0 = 0$, we have $\mathbf{m}_{\text{on}}[(0, \theta_{1:d})] = \mathbf{m}_{\text{off}}[\theta]$ and we obtain

$$\frac{\mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)}{\mathbf{m}_{\text{on}}[\theta](v_i, t_j | x_D)} - 1 \approx - \left(\frac{\frac{\partial}{\partial \theta_0} \mathbf{m}_{\text{on}}[(0, \theta_{1:d})](v_i, t_j | x_D)}{\mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)} \right) (\theta_0 - 0). \quad (3.29)$$

The online measurement \mathbf{m}_{on} decreases as a function of θ_0 which means that the right hand side of Eq. (3.29) is non-negative for any $\theta_0 \geq 0$. A more detailed treatment will not be needed at this point but can be found in Appendix A.1 alongside more explicit formulas that can be used to derive the analogous integral expression of Eq. (3.27) from Eq. (3.29), specifically Eqs. (A.3) and (A.5). A noteworthy difference is that the single scattering absorption from Eq. (3.27), assuming Eq. (3.28) for known u_0 , is independent

of the nuisance component $\theta_{1:d}$. This is not true in the RTE setting and naturally has an effect on our ability to retrieve information regarding the parameter θ_0 of interest. For brevity, we will introduce

$$\bar{L}_{i,j}[\theta] = \frac{\Delta t H_D A_D \mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)}{2} \quad (3.30)$$

$$\bar{q}_{i,j}[\theta] = \frac{-\frac{\partial}{\partial \theta_0} \mathbf{m}_{\text{on}}[(0, \theta_{1:d})](v_i, t_j | x_D)}{\mathbf{m}_{\text{off}}[\theta](v_i, t_j | x_D)} \quad (3.31)$$

and note that even though θ is unknown, the quantity $\bar{L}_{i,j}[\theta]$ can be estimated accurately from $\mathbf{z}_{i,j}$ and therefore (under the conditions of Lemma 3.2.1) will be assumed as given/observed. As before H_D is the pulse energy, A_D the detector area and Δt the bin width. We have the following.

Theorem 3.2.2 (Optimal detection of low concentrations). *Let the random variables $\mathbf{y}_{i,j}$ be as in Lemma 3.2.1 and $\bar{L}_{i,j}[\theta]$ and $\bar{q}_{i,j}[\theta]$ as in Eq. (3.30) and Eq. (3.31). Further let $\hat{q}_{i,j} > 0$ be an arbitrary collection of positive scalars and assume that $\bar{L}_{i,j}[\theta]$ is known for $\theta \in \Theta$. If we define the set*

$$\Theta_0 = \{\theta \in \Theta : \theta_0 = 0\}$$

then we have for all $\theta \in \Theta$

$$\frac{\sum_{i,j} \hat{q}_{i,j} \bar{L}_{i,j}[\theta] \mathbf{y}_{i,j} - \theta_0 \hat{q}_{i,j} \bar{q}_{i,j}[\theta] \bar{L}_{i,j}[\theta]}{\sqrt{\sum_{i,j} \hat{q}_{i,j}^2 \bar{L}_{i,j}[\theta]}} \sim \text{Normal}(0, 1). \quad (3.32)$$

In particular, Eq. (3.32) can be used to test the hypothesis $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta \setminus \Theta_0$ where the rejection regions take the form

$$\frac{\sum_{i,j} \hat{q}_{i,j} \bar{L}_{i,j}[\theta] \mathbf{y}_{i,j}}{\sqrt{\sum_{i,j} \hat{q}_{i,j}^2 \bar{L}_{i,j}[\theta]}} \geq R > 0 \quad (3.33)$$

with R is a quantile of the standard normal distribution that depends only on the signifi-

cance level. The power of the test is given by

$$\mathbb{P} \left(\text{Normal}(0, 1) > R - \theta_0 \frac{\sum_{i,j} \hat{q}_{i,j} \bar{q}_{i,j}[\theta] \bar{L}_{i,j}[\theta]}{\sqrt{\sum_{i,j} \hat{q}_{i,j}^2 \bar{L}_{i,j}[\theta]}} \right). \quad (3.34)$$

and is maximised for $\bar{q}_{i,j}[\theta] = \hat{q}_{i,j}$ which is independent of θ in the case of narrow FOV measurements and yields the uniformly most powerful test for the above hypothesis.

Proof. The only non-trivial claim is that the test for narrow FOV data is uniformly most powerful, the rest is essentially an immediate consequence of Lemma 3.2.1 following simple rearrangements. Note that in the case of single scattering the quantities in Eq. (3.30) as well as Eq. (3.31) are assumed known and $\bar{q}_{i,j}$ doesn't depend on the parameter θ . The probability density for the sample $\mathbf{y}_{i,j}$ is given by

$$\prod_{i,j} \sqrt{\frac{\bar{L}_{i,j}[\theta]}{2\pi}} \exp \left(-\frac{1}{2} \sum_{i,j} \mathbf{y}_{i,j} \bar{L}_{i,j}[\theta] \mathbf{y}_{i,j} + \theta_0^2 \bar{q}_{i,j} \bar{L}_{i,j}[\theta] \bar{q}_{i,j} \right) \exp \left(\theta_0 \sum_{i,j} \mathbf{y}_{i,j} \bar{L}_{i,j}[\theta] \bar{q}_{i,j} \right)$$

from which the factorisation theorem implies that $T(\mathbf{y}) := \sum_{i,j} \mathbf{y}_{i,j} \bar{L}_{i,j}[\theta] \bar{q}_{i,j}$ is a sufficient statistic for θ_0 (see Theorem 6.2.6 in [22]) and the likelihood ratio is an increasing function of $T(\mathbf{y})$. The optimality of the rejection regions as in Eq. (3.33) is a direct consequence of the Karlin-Rubin Theorem (see e.g. Theorem 8.3.17 in [22]) because $\theta_0 \geq 0$. \square

The procedure presented in Theorem 3.2.2 is in a rather strong sense optimal when it comes to detecting low-concentration plumes with narrow FOV data. As such, an improvement with respect to that particular task is achieved as soon as we can construct a test from wide FOV data that outperforms its narrow FOV equivalent. We also see that the power is determined by θ_0 and

$$\Pi_L(\bar{q}[\theta] \rightarrow \hat{q}) := \frac{\sum_{i,j} \hat{q}_{i,j} \bar{q}_{i,j}[\theta] \bar{L}_{i,j}[\theta]}{\sqrt{\sum_{i,j} \hat{q}_{i,j}^2 \bar{L}_{i,j}[\theta]}} \quad (3.35)$$

which is nothing but a scaled inner product of \hat{q} with $\bar{q}[\theta]$ weighted by $\bar{L}[\theta]$ or, more precisely, the scalar projection of $\bar{q}[\theta]$ onto \hat{q} w.r.t. that inner product. In what follows we

want to study the behaviour of $\Pi_L(\bar{q}[\theta] \rightarrow \hat{q})$ from Eq. (3.35) as a function of the weights \hat{q} and nuisance parameters $\theta_{1:d}$.

In view of Theorem 3.2.2 and Eq. (3.31) we see that the ideal choice of \hat{q} , i.e. the one maximising the power of a test that has the form of Eq. (3.33), depends on $\theta_{1:d}$. Typically we won't have access to the scattering/nuisance parameters but (at best) we may have a priori guess, denoted by $\hat{\theta}_{1:d}$. Putting $\hat{q}_{i,j} = \bar{q}_{i,j}[\hat{\theta}]$ in the context of Theorem 3.2.2 the test power becomes a function of $(\theta, \hat{\theta}) \in \Theta \times \Theta$ rather than $(\theta, \hat{q}) \in \Theta \times (0, \infty)^{N_t \times N_v}$. We will usually have $\dim(\hat{\theta}) \ll \dim(\hat{q})$ resulting in reduced dimensionality of the input by restriction of \hat{q} to a lower dimensional manifold of $(0, \infty)^{N_t \times N_v}$. However, the expression $\Pi_L(\bar{q}[\theta] \rightarrow \bar{q}[\hat{\theta}])$ for all inputs $(\theta, \hat{\theta}) \in \Theta \times \Theta$ is still rather cumbersome to analyse directly. Since the Cauchy-Schwartz inequality implies $\Pi_L(\bar{q}[\theta] \rightarrow \bar{q}[\cdot]) \leq \Pi_L(\bar{q}[\theta] \rightarrow \bar{q}[\theta])$ for all $\theta \in \Theta$ we may use the upper bound as a reference point and look at $\Pi_L(\bar{q}[\theta] \rightarrow \bar{q}[\hat{\theta}])$ for varying degrees of accuracy in the guess $\hat{\theta}$, e.g. by limiting the magnitudes of $|\theta_i - \hat{\theta}_i|$ for $i \in 1:d$. The latter can be expressed by means of a θ -dependent set $C(\theta)$ along with the knowledge that $\theta \in [0, \infty) \times C(\theta)$, i.e. $\theta_{1:d} \in C(\theta)$. Note that we made explicit the distinction between $C(\theta)$, which is a set-valued function and can be regarded as *situation specific* prior knowledge, and Θ , which is regarded a set of generally feasible values. The set $C(\theta)$ can thus be used to express partial knowledge of the parameter, e.g. regarding certain components, and will allow us to analyse the behaviour of $\Pi_L(\bar{q}[\theta] \rightarrow \bar{q}[\hat{\theta}])$ through selected lower dimensional restrictions.

Given a set $C(\theta)$ of constraints, we could select an arbitrary $\hat{\theta} \in C(\theta)$ and set $\hat{q}_{i,j} = \bar{q}_{i,j}[\hat{\theta}]$ resulting in a power no worse than had we chosen

$$\Pi_{C,L}[\theta] := \inf_{\hat{\theta} \in C(\theta)} \Pi_L(\bar{q}[\theta] \rightarrow \bar{q}[\hat{\theta}]). \quad (3.36)$$

Note that θ can be seen as a ground truth parameter and $\Pi_{C,L}[\theta]$ is indicative of the worst case test power in that situation provided we can constrain the parameter to within $C(\theta)$. If $C(\theta)$ is compact then the infimum is attained at some value $\hat{\theta}_C(\theta)$ and we have $\Pi_L(\bar{q}[\theta] \rightarrow \bar{q}[\hat{\theta}_C(\theta)]) = \Pi_{C,L}[\theta]$. In other words, Eq. (3.36) is obtained via the restriction of Eq. (3.35) to the graph $\{(\theta, \bar{q}[\hat{\theta}_C(\theta)]) : \theta \in \Theta\}$ of the “curve” $\theta \mapsto \bar{q}[\hat{\theta}_C(\theta)]$ where $\hat{\theta}_C(\theta)$

is the worst possible parameter in a vicinity of θ set forth by C . We will make this more concrete later on with examples provided in Table 3.1.

The quantity subject to minimisation in Eq. (3.36) depends on the FOV and becomes trivial when the FOV is assumed sufficiently narrow to collect only single scattering, thus light with known trajectories and $\bar{q}_{i,j}[\theta] = q_{i,j}^1$ for all $\theta \in \Theta$ and some θ -independent $q_{i,j}^1$. In the following we want to compare wide FOV data in the RTE setting to classical DIAL based on narrow FOV data and the single scattering model Eq. (3.13c). To make the distinction between them more clear we will refer to $\Pi_{C,L}[\theta]$ only for a given/fixed wide FOV and refer to the narrow FOV equivalent as

$$\Pi_L^1[\theta] := \sqrt{\sum_{i,j} q_{i,j}^1 L_{i,j}^1[\theta] q_{i,j}^1}$$

where $q_{i,j}^1$ and $L_{i,j}^1[\theta]$ are as in Eqs. (3.30) and (3.31) but with $m_{\text{on/off}}$ replaced by the single scattering contribution $m_{\text{on/off}}^1$ defined in Lemma 3.1.1. Note that $L^1[\theta]$ depends on θ and $L_{i,j}^1[\theta] \leq \bar{L}_{i,j}[\theta]$ since all else being equal a wide FOV collects more light than the narrow one. By studying the ratio

$$\Phi_C[\theta] := \frac{\Pi_{C,L}[\theta]}{\Pi_L^1[\theta]} - 1 \tag{3.37}$$

for varying ground truth parameters θ and constraints C we can gain insight into the relative performance of wide and narrow FOV data by means of a simple scalar indicator without having to assume knowledge regarding the parameters $\theta_{1:d}$. The insight is primarily of qualitative nature as the signal strength L^1 and \bar{L} (for narrow and wide FOVs respectively), in particular the entry-wise ratios $\frac{\bar{L}_{i,j}[\theta]}{L_{i,j}^1[\theta]}$, will have a major impact on the behaviour of Φ_C whose precise quantitative characteristics therefore depend on additional variable quantities such as the solid angle subtended by the wide FOV.

In general, we will have parameters that result in the minimisation in Eq. (3.36) having intractable integrals that cannot be dealt with analytically. Numerical evaluation of \bar{L} and \bar{q} is fairly expensive as it requires relatively accurate solutions of the RTE in order to obtain $m_{\text{on/off}}$ and some care must be taken in order to construct a computationally

feasible yet meaningful case study. In the following we consider a simple example where u_0 is given by a spherically symmetric kernel, i.e.

$$u_0(x) = C_{\text{DIAL}} \phi_0(x)$$

where ϕ_0 resembles a good kernel from Definition 2.1.1 but has Gaussian tails and is chosen such that the majority of its “mass” is contained within a ball with radius 20m which is placed at a distance of 100m away from the detector. The wide FOV’s aperture is chosen to reflect that, i.e. is selected such that it captures objects of approximately 20m at a 100m distance. We consider $\theta = (\theta_0, \theta_{1:3})$, i.e. the nuisance part of θ has three components, and parameterise the optical parameters as follows:

- The scattering coefficient σ_s is given by

$$\sigma_s(x) = \theta_1 + \theta_2 \phi_0(x). \quad (3.38)$$

Here θ_1 can be thought of as modelling the amount of ambient scattering while θ_2 controls scattering that is aligned with the “plume” u_0 .

- The phase function $f_p(x, v' \rightarrow v) = f_{\text{HG}}(v \cdot v' \mid \theta_3)$ is a Henyey-Greenstein density with parameter θ_3 and independent of $x \in X$.
- We assume that there is no non-differential absorption, i.e. $\sigma_a = 0$.

Given ϕ_0 (and $C_{\text{DIAL}} = 1$) the mapping \mathcal{G} from Eq. (3.10) can be expressed as

$$\mathcal{G}(\theta) = (\theta_0 \phi_0(\cdot), 0, \theta_1 + \theta_2 \phi_0(\cdot), f_{\text{HG}}(\cdot \mid \theta_3))$$

which in turn defines $\mathbf{m}_{\text{on/off}}[\theta](v, t \mid x_D)$ through Eqs. (3.8), (3.10), and (3.16) and makes possible the explicit evaluation of $\Phi_C[\theta]$ via Eqs. (3.30), (3.31), and (3.35) to (3.37) for each constraint scenario described in Table 3.1. In order to minimise complexity the measurement kernels $\mathbf{m}_{\text{on/off}}[\theta](v, t \mid x_D)$ are evaluated for a single direction $v = \frac{b^0 - x_D}{\|b^0 - x_D\|}$, where as usual $x_D = (0\text{m}, 0\text{m}, 0\text{m})$ is the detector location and $b^0 = (0\text{m}, 100\text{m}, 0\text{m})$ is the

centre of u_0 , and 50 equally spaced grid points corresponding to path lengths t between 120m and 330m (single scattering at b^0 corresponds to a 200m long path). We used an MC-based solver using between 10^6 and 10^7 photons per order of scattering. More details regarding the exact method can be found in Chapter 5. The primary reason behind setting $\sigma_a = 0$ is that evaluation of $\mathbf{m}_{\text{on/off}}[\theta]$ for all combinations of $\theta_{1:d}$ scales exponentially with its dimension d . In principle we could have chosen a parameterisation for $\sigma_a = 0$ akin to Eq. (3.38), say $\sigma_a(x) = \theta_4 + \theta_5\phi_0(x)$. In such a situation ambient non-differential absorption θ_4 would have no effect on \bar{q} and only affect \bar{L} and L^1 (by reducing the signal strength exponentially) in such a way that leaves the ratio $\frac{\bar{L}_{i,j}[\theta]}{L_{i,j}^1[\theta]}$ invariant. This doesn't guarantee invariance of Φ_C w.r.t. θ_4 (as it contains a ratio of weighted sums rather than individual ratios) but it arguably stands to reason that the effect of ambient absorption on Φ_C will be less pronounced compared to other parameters which impact both \bar{q} as well as $\frac{\bar{L}_{i,j}[\theta]}{L_{i,j}^1[\theta]}$. Non-differential absorption from a plume-like perturbation does not have those properties, i.e. \bar{q} would be a function of θ_5 , but one can show that much like θ_4 it primarily impacts the magnitude of the signal and its effect on \bar{q} is of a lesser order than that caused by θ_2 (see Eqs. (A.18) and (A.19)). Since we are interested in comparing wide FOV and narrow FOV test power relative to each other (rather than absolute test power which strongly depends on absolute signal strength \bar{L} and L^1) we chose to set $\sigma_a = 0$ in order to make computations more manageable.

The graphics in Figs. 3.4 and 3.5 show cross sections of $\theta \mapsto \Phi_{C,L}[\theta]$ at $\theta_3 = 0$ for different choices C summarised in Table 3.1. Note that we fixed the phase function in the ground truth so the plots become 2-dimensional and homogeneous scattering was chosen as it maximises the back scattering coefficient, thus L^1 , as well as "diffusivity" and therefore can be thought of as particularly "noisy" (see also Fig. 3.3) putting the wider FOV at a comparative disadvantage. Note that in Eq. (3.36) we consider parameters up to $\theta_3^{\max} = 0.7$ so the phase function is not fixed throughout. The behaviour is qualitatively similar for other cross sections. Since θ_1 and $\theta_2\phi_0$ capture scattering per unit length it

Scenario	Domain of Φ_C	Constraint C
Fig. 3.4 (a)	$\theta_{1:d} \in [0, \theta_1^{\max}] \times [0, \theta_2^{\max}] \times \{0\}$	$C(\theta) = [0, \theta_1^{\max}] \times [0, \theta_2^{\max}] \times [0, \theta_3^{\max}]$
Fig. 3.4 (b)	$\theta_{1:d} \in [0, \theta_1^{\max}] \times [0, \theta_2^{\max}] \times \{0\}$	$C(\theta) = [0, \theta_1^{\max}] \times [0, \theta_2^{\max}] \times \{\theta_3\}$
Fig. 3.4 (c)	$\theta_{1:d} \in [0, \theta_1^{\max}] \times [0, \theta_2^{\max}] \times \{0\}$	$C(\theta) = \{\theta_1\} \times \{\theta_2\} \times [0, \theta_3^{\max}]$
Fig. 3.4 (d)	$\theta_{1:d} \in [0, \theta_1^{\max}] \times [0, \theta_2^{\max}] \times \{0\}$	$C(\theta) = \{\theta_1\} \times \{\theta_2\} \times \{\theta_3\}$
Fig. 3.5 (a)	$\theta_{1:d} \in [0, \theta_1^{\lim}] \times [0, \theta_2^{\max}] \times \{0\}$	$C(\theta) = [0, \theta_1^{\lim}] \times [0, \theta_2^{\max}] \times [0, \theta_3^{\max}]$
Fig. 3.5 (b)	$\theta_{1:d} \in [0, \theta_1^{\lim}] \times [0, \theta_2^{\max}] \times \{0\}$	$C(\theta) = \{\theta_1\} \times \{\theta_2\} \times \{\theta_3\}$

Table 3.1: System parameters corresponding to different situations as shown in Figs. 3.4 and 3.5. Φ_C can easily be evaluated for the set $[0, \theta_1^{\max}] \times [0, \theta_2^{\max}] \times [0, \theta_3^{\max}]$ with the available data but the domain in the above table is shown such that it corresponds to the respective plots. Note that $\theta_{1:d} = (\theta_1, \theta_2, \theta_3)$ is the argument of Φ_C (which does not depend on θ_0) whereas $\theta_1^{\max}, \theta_2^{\max}$ and θ_1^{\lim} are a priori selected scalar hyper-parameters.

makes sense to consider the dimensionless and arguably more intuitive quantities

$$\mathcal{S}_{\text{ambient}}(\theta) := \theta_1 \times 100\text{m} \quad (3.39)$$

$$\mathcal{S}_{\text{plume}}(\theta) := \theta_2 \int_{-\infty}^{+\infty} \phi_0 \left(x_D + r \frac{b^0 - x_D}{\|b^0 - x_D\|} \right) dr \quad (3.40)$$

for which it is easily seen that $\theta_{1:2} \mapsto (\mathcal{S}_{\text{ambient}}(\theta), \mathcal{S}_{\text{plume}}(\theta))$ is a bijective linear transform. In Table 3.1 the limits $\theta_1^{\max}, \theta_1^{\lim}$ were chosen such that $100\text{m} \times \theta_1^{\max} \approx 5.9$ as well as $100\text{m} \times \theta_1^{\lim} \approx 2.0$ while θ_2^{\max} ensures $\mathcal{S}_{\text{plume}}(\theta) \leq 13.7$ on the domain of Φ_C . This means that the expected number of scattering events due to ambient particles on the segment between source/detector at x_D and the plume centre is at most 5.9 (or 2.0 in the case of θ_1^{\lim}). Similarly, θ_2^{\max} limits scattering due to the plume on a straight trajectory to 13.7 and, much like θ_1^{\max} , was chosen ‘‘empirically’’ such that the ranges in the plots are large enough in order to show the transition between different behaviour that can be seen for varying θ .

The plots suggest that the quantities that matter for good ‘‘worst-case alignment’’ are precisely those that are considered reconstructable in Theorem 3.1.4. Fig. 3.4 and Fig. 3.5 clearly indicate that amongst the optical nuisance parameters, the distribution of scattering particles has the biggest effect on the quality of our measurement. Indeed, as Fig. 3.5 shows even knowledge of the complete set of scattering parameters does not provide much better alignment w.r.t. Π_L as a simple constraint that limits the amount

of ambient scattering particles. Although we only considered a single parameter and a simplified scenario, the same idea extends to more general θ which is to say that photons from wider FOVs are most useful when we are interested in aspects of the differential absorption field α that are relatively well aligned with σ_s , i.e not too small or far away from most of the scattering particles. Indeed, the simple scenario can more generally be thought of as an indication of how much information wider FOVs preserve in the presence of unknown scattering parameters with regard to a single component of an element from a good kernel space.

Similar results (after suitable modification of Eq. (3.31) and Theorem 3.2.2) are to be expected for position and size parameters for which an alignment of the scattering parameters with respect to the corresponding gradients is required. Furthermore, the ambient scattering may be replaced by scattering particles from the plume surrounding a smaller feature. In other words, the information within wide FOVs declines as soon as the plume feature can only be resolved with a higher resolution than the bulk of the plume. This in effect limits the usefulness of wide FOVs when a high-resolution image is sought (compare also Fig. 3.3 and Fig. 3.6). We note that the behaviour for high ambient scattering is in a way similar, albeit for a different type of measurement, to the findings of decreased sensitivity at a larger optical thickness from [93, 94]. The approach taken here differs insofar that image features aren't strictly local but rather we consider more smoothly varying aspects of the gas concentration profile which as a whole will be rather stable even for plumes of large optical thickness.

Although ambient noise has been ignored in this comparison it is not hard to see that its effect will be very similar to that of ambient scattering particles, i.e. it doesn't matter whether we observe photons that didn't originate from the controlled source or light that never reached the region of interest (gray trajectories in Fig. 3.3 and Fig. 3.6). In much the same way as with ambient scattering, it becomes an issue once the number of peak signal photons observed is of a similar magnitude as the ambient noise.

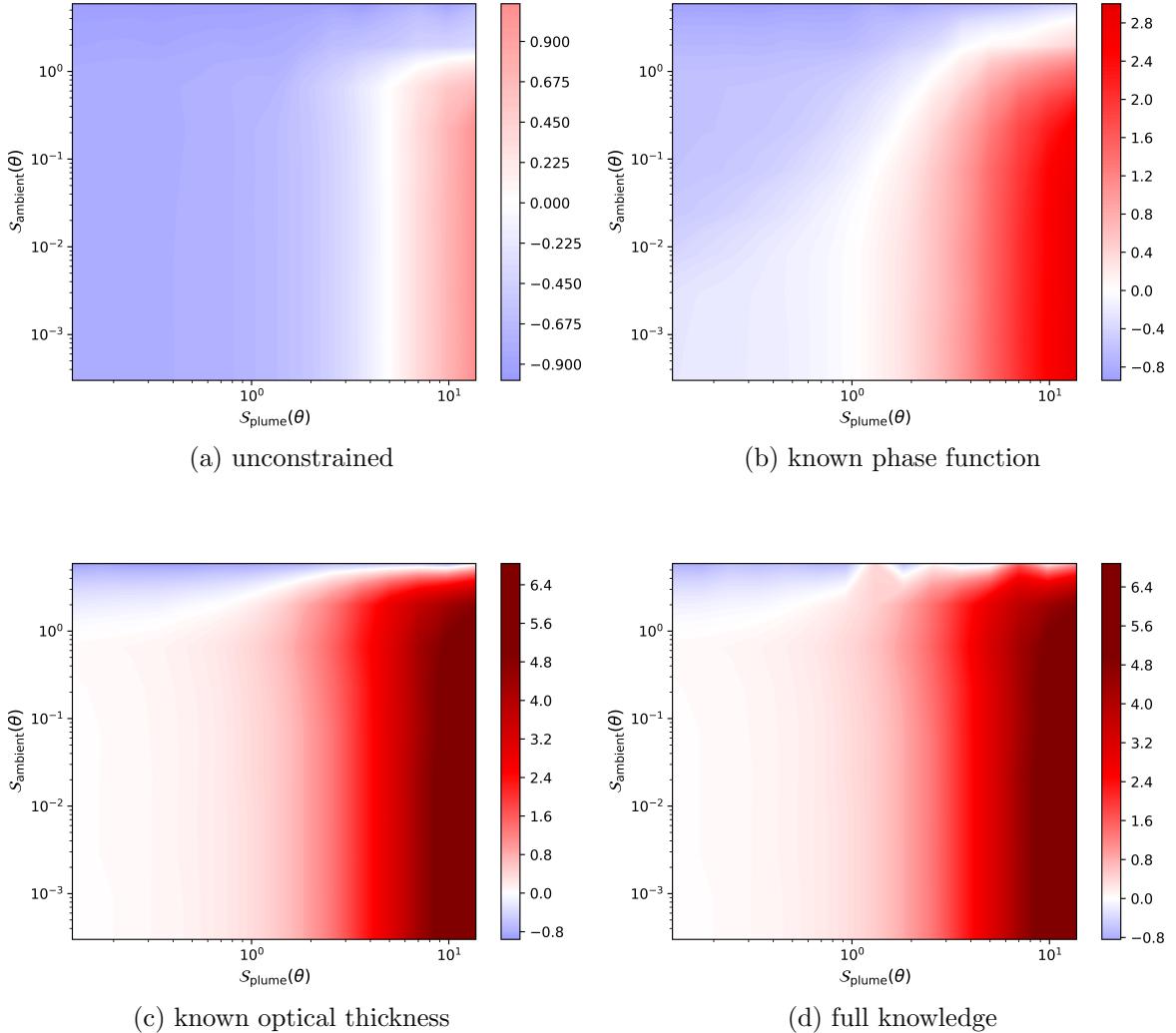


Figure 3.4: The red regions indicate better detection with wide FOV data. Figure (a) shows that if nothing other than loose bounds for the distribution of ambient particles is known, then wide FOVs will improve the reconstruction only for optically thick plumes. Figures (b), where the optical thickness is assumed unknown while the HG-parameter θ_3 is fixed to the correct value, and (c), where only the phase function is unknown while σ_s is fixed via θ_1 and θ_2 , show that knowledge regarding the distribution of scattering particles is considerably more valuable than knowledge of the phase function. The former is in this example virtually equivalent to full knowledge of the scattering particles, i.e. $C(\theta) = \{\theta_{1:d}\}$, which is shown in Figure (d).

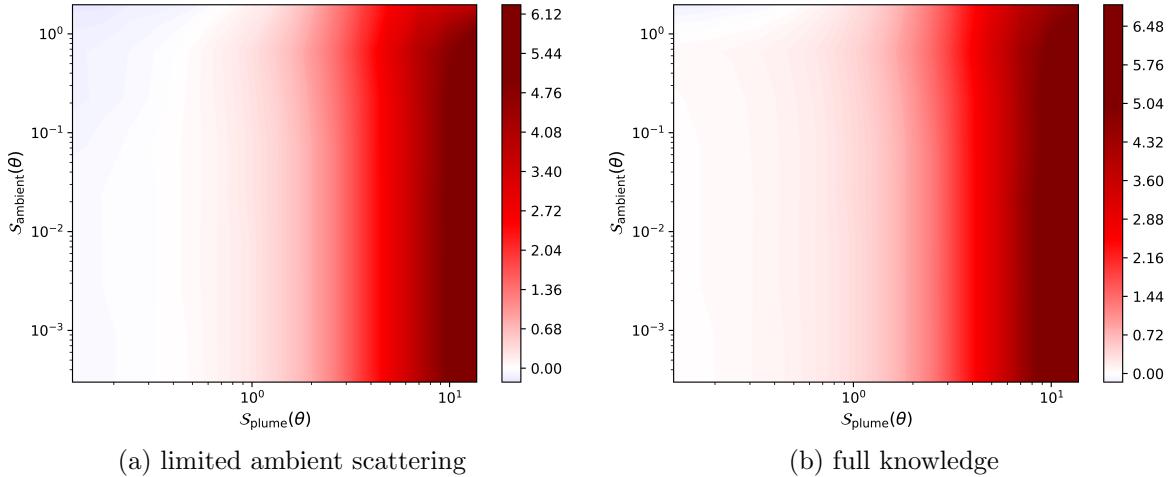


Figure 3.5: The red regions indicate better detection with wide FOV data. Figures (a) and (b) show very similar results indicating that full knowledge of the scattering parameters does not yield considerably better results than merely a limit on ambient scattering corresponding to $\mathcal{S}_{\text{ambient}}(\theta) \lesssim 2$.

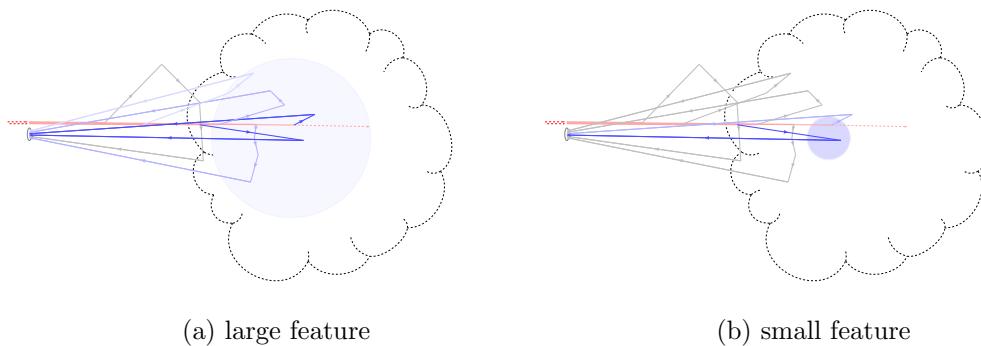


Figure 3.6: Despite most of the scattering happening inside the plume (dashed region), most trajectories don't cross the small kernel in figure (b). As such the marginal distributions of parameters reconstructed from wide FOV data corresponding to small features will suffer from essentially the same issues as those shown in Fig. 3.4 for high ambient scattering.

Chapter 4

Consistency & asymptotic behaviour

In the previous chapter, we have studied the behaviour of the (incomplete) forward operator \mathcal{Q} defined on suitable function spaces Ξ by the relation

$$\mathcal{Q} : \Xi \rightarrow C(Z_{(-)}(x_D)), \quad (\alpha, \sigma_a, \sigma_s, f_p) = \xi \mapsto \frac{\mathfrak{m}_{\text{on}}(\cdot \mid x_D)}{\mathfrak{m}_{\text{off}}(\cdot \mid x_D)} \quad (4.1)$$

where $(\mathfrak{m}_{\text{on}}, \mathfrak{m}_{\text{off}}) = \mathcal{M}(\xi)$ are as in Eq. (3.8). To that end provided conditions under which it is injective when the optical model involves multiple scattering in accordance with the RTE (2.22). Although certainly desirable, statements such as Theorem 3.1.4, even if they were accompanied by continuity of the inverse operator, cannot be used to argue the use of wide FOV measurements in favour of filtering multiple scattering and only keeping what is detected by a narrow FOV. In fact, the requirements imposed on the (non-differential) optical parameters in order to obtain such a result are more restrictive than what is needed if the forward model is given by single scattering. On the other hand, the detection-related results give an indication that there may be instances where the wide FOV is useful. In the following, we essentially want to investigate what happens when the assumptions of Theorem 3.1.4 are violated or dropped, i.e. the consequences of relaxing them.

4.1 Continuity & convergence of the estimators

We begin with a result that will be of some use throughout the chapter. It is a straightforward extension of standard results in elementary statistics to non-identifiable statistical models.

Lemma 4.1.1. *Suppose that $(\mathbf{X}_j)_{j \in \mathbb{N}}$ is a sequence of Φ -valued random elements¹, (Θ, d_Θ) a compact Polish space, $L : \Theta \times \Phi \rightarrow \mathbb{R}$ and $L_0 \in C(\Theta)$. If*

1. $L(\cdot, x) \in C(\Theta)$ for any $x \in \Phi$
2. $\|L(\cdot, \mathbf{X}_j) - L_0\|_\infty \rightarrow 0$ in probability for $j \rightarrow \infty$

Then, assuming all quantities are measurable, the set $\Theta^ := \arg \max_{\theta \in \Theta} L_0(\theta)$ of maximisers of L_0 is non-empty and for the non-empty random sets $\hat{\Theta}_j$ of maximisers for $L(\cdot, \mathbf{X}_j)$ we have*

$$\sup_{\hat{\theta} \in \hat{\Theta}_j} d_\Theta(\hat{\theta}, \Theta^*) \rightarrow 0 \quad (4.2)$$

in probability for $j \rightarrow \infty$. In particular, if $\Theta^ = \{\theta^*\}$ then any sequence of maximisers $\hat{\theta}(\mathbf{X}_j)$ of $L(\cdot, \mathbf{X}_j)$ converges to θ^* in probability.*

Proof. First note that all sets of maximisers are non-empty due to continuity of $L(\cdot, x)$ as well as $L_0(\cdot, x)$. We must show that for any $\varepsilon > 0$ there is $N_\varepsilon \in \mathbb{N}$ such that

$$j \geq N_\varepsilon \implies \mathbb{P} \left(\sup_{\hat{\theta} \in \hat{\Theta}_j} d_\Theta(\hat{\theta}, \Theta^*) > \varepsilon \right) < \varepsilon. \quad (4.3)$$

Define $\Theta_\varepsilon := \{\theta \in \Theta : d_\Theta(\theta, \Theta^*) \geq \varepsilon\} = \bigcap_{\theta^* \in \Theta^*} \{\theta \in \Theta : d_\Theta(\theta, \theta^*) \geq \varepsilon\}$. Then Θ_ε is a closed subset of a compact set, thus compact, and there is some $\delta_\varepsilon > 0$ such that

$$\sup_{\theta \in \Theta_\varepsilon} L_0(\theta) = \sup_{\theta \in \Theta} L_0(\theta) - \delta_\varepsilon < \sup_{\theta \in \Theta} L_0(\theta) = \sup_{\theta \in \Theta^*} L_0(\theta) \quad (4.4)$$

¹here a Φ -valued random element \mathbf{X} is measurable map $\mathbf{X} : \Omega \rightarrow \Phi$ from a probability space $(\Omega, \mathcal{B}_\Omega, \mathbb{P})$ to a measurable space (Φ, \mathcal{B}_Φ) . See also [76, p. 83] for some examples.

and all suprema are attained by continuity and compactness. Due to the second assumption there is N_ε such that for all $j \geq N_\varepsilon$

$$\mathbb{P} \left(\|\mathbf{L}(\cdot, \mathbf{X}_j) - \mathbf{L}_0\|_\infty \geq \frac{\delta_\varepsilon}{2} \right) < \varepsilon. \quad (4.5)$$

We claim that this choice satisfies Eq. (4.3). In order to see this we start by using the inequality $|b - a| \geq b - a \implies a \geq b - |b - a|$ any $a, b \in \mathbb{R}$ and the definition of $\hat{\Theta}_j$ to get

$$\sup_{\theta \in \Theta_\varepsilon} \mathbf{L}(\cdot, \mathbf{X}_j) \leq \sup_{\theta \in \Theta_\varepsilon} \mathbf{L}_0(\theta) + \|\mathbf{L}(\cdot, \mathbf{X}_j) - \mathbf{L}_0\|_\infty \quad (4.6)$$

$$\sup_{\theta \in \hat{\Theta}_j} \mathbf{L}(\cdot, \mathbf{X}_j) = \sup_{\theta \in \Theta} \mathbf{L}(\cdot, \mathbf{X}_j) \geq \sup_{\theta \in \Theta} \mathbf{L}_0(\theta) - \|\mathbf{L}(\cdot, \mathbf{X}_j) - \mathbf{L}_0\|_\infty \quad (4.7)$$

for any $j \in \mathbb{N}$. After combining Eqs. (4.6) and (4.7) with Eq. (4.4) we obtain

$$\|\mathbf{L}(\cdot, \mathbf{X}_j) - \mathbf{L}_0\|_\infty < \frac{\delta_\varepsilon}{2} \implies \sup_{\theta \in \Theta_\varepsilon} \mathbf{L}(\cdot, \mathbf{X}_j) < \sup_{\theta \in \Theta} \mathbf{L}_0(\theta) - \frac{\delta_\varepsilon}{2} < \sup_{\theta \in \hat{\Theta}_j} \mathbf{L}(\cdot, \mathbf{X}_j)$$

and, as $\mathbf{L}(\cdot, \mathbf{X}_j)$ is by construction constant on $\hat{\Theta}_j$, this yields

$$\|\mathbf{L}(\cdot, \mathbf{X}_j) - \mathbf{L}_0\|_\infty < \frac{\delta_\varepsilon}{2} \implies \hat{\Theta}_j \cap \Theta_\varepsilon = \emptyset \iff \sup_{\hat{\theta} \in \hat{\Theta}_j} d_\Theta(\hat{\theta}, \Theta^*) < \varepsilon.$$

The above then immediately implies

$$\mathbb{P} \left(\sup_{\hat{\theta} \in \hat{\Theta}_j} d_\Theta(\hat{\theta}, \Theta^*) \geq \varepsilon \right) \leq \mathbb{P} \left(\|\mathbf{L}(\cdot, \mathbf{X}_j) - \mathbf{L}_0\|_\infty \geq \frac{\delta_\varepsilon}{2} \right)$$

and what needed to be shown follows from Eq. (4.5). \square

As many of the phenomena regarding our measurement are best described in a continuous setting, we start by introducing a continuous version of the noisy data considered in Equation (3.15).

Definition 4.1.2 (Poisson Point Processes (PPP)). Let $Z \subseteq \mathbb{R}^n$ and $(\Omega, \mathcal{B}_\Omega, \mathbb{P})$ a probability space. We call a measure M point measure if $M(A) \in \mathbb{N} \cup \{0, \infty\}$ for any measurable set $A \subseteq Z$ and denote the set of all such measures by $\mathcal{N}_p(Z)$. A random

element $\mathbf{M} : \Omega \rightarrow \mathcal{N}_p(Z)$ is called Poisson Point Process if there is a Borel measure Λ on Z , called intensity measure, such that

- If $\Lambda(A) < \infty$ then $\mathbf{M}(A, \cdot)$ is a Poisson random variable with parameter $\Lambda(A)$
- If $A_1, \dots, A_n \subseteq Z$ are disjoint and measurable, then $\mathbf{M}(A_1, \cdot), \dots, \mathbf{M}(A_n, \cdot)$ are independent random variables.

We abbreviate this by writing $\mathbf{M} \sim \text{PPP}(\Lambda)$. For further details we refer to [78].

In order to introduce noise into our model on an arbitrary small scale we will consider $\mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}$ to be independent and distributed according to $\text{PPP}(\Lambda_{\text{on/off}})$. We start with the scenario where both ψ as well as $\mathcal{M}(\xi) = (\mathfrak{m}_{\text{on}}, \mathfrak{m}_{\text{off}})$ are functions defined on $Z_{(-)}(x_D)$ as introduced in Eq. (3.7) and $\Lambda_{\text{on/off}}$ are given by

$$d\Lambda_{\text{on/off}}(v, t) = \psi(v, t) \mathfrak{m}_{\text{on/off}}(v, t | x_D) dv dt \implies \frac{d\Lambda_{\text{on}}}{d\Lambda_{\text{off}}} = \frac{\mathfrak{m}_{\text{on}}(\cdot | x_D)}{\mathfrak{m}_{\text{off}}(\cdot | x_D)} = \mathcal{Q}(\xi) \quad (4.8)$$

but in the later stages of this chapter also consider instances where the relationship in Eq. (4.8) is not exact. We will henceforth consider data that is restricted to *compact* measurement domains $Z_{[-]}(x_D)$ satisfying

$$Z_{[-]}(x_D) \subseteq Z_{(-)}(x_D) \subseteq \partial V_{(-)}(x_D) \times (0, \infty) \quad (4.9)$$

and data $\mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}$ will be restricted to those sets. Note that

$$f_{ZX} : Z_{(-)}(x_D) \rightarrow X, \quad (v, t) \mapsto x_D + \frac{t}{2} v \quad (4.10)$$

is always continuous and injective, bijective if X is convex, so that one can identify $Z_{[-]}(x_D)$ with a compact set $X_{[-]}(x_D) := f_{ZX}(Z_{[-]}(x_D)) \subseteq X$.

Example. After possibly a rotation of the coordinate system we may (w.l.o.g.) assume that the surface normal at x_D is given by $\vec{n}(x_D) = (-1, 0, 0)$ so that for $a_D, b_D \in (0, \frac{\pi}{2})$

the sets $V_{\llbracket - \rrbracket}^{\text{box}}(x_D)$ of directions defined by

$$V_{\llbracket - \rrbracket}^{\text{box}}(x_D) := \left\{ \begin{pmatrix} \sin(a) \cos(b) \\ \sin(a) \sin(b) \\ \cos(a) \end{pmatrix} : (b, a) \in [-b_D, +b_D] \times \left[\frac{\pi}{2} - a_D, \frac{\pi}{2} + a_D \right] \right\} \quad (4.11)$$

are compact subsets of $V_{(-)}(x_D)$ and assuming that $x_D + \frac{t}{2}v \in X$ for all $(v, t) \in V_{\llbracket - \rrbracket}(x_D) \times (0, T_{\max}]$ we can put for $0 < T_{\min} \leq T_{\max}$

$$Z_{\llbracket - \rrbracket}^{\text{box}}(x_D) := V_{\llbracket - \rrbracket}^{\text{box}}(x_D) \times [T_{\min}, T_{\max}] \quad (4.12)$$

which can be identified with the “box” $[-b_D, +b_D] \times \left[\frac{\pi}{2} - a_D, \frac{\pi}{2} + a_D \right] \times [T_{\min}, T_{\max}]$ and would correspond to an acquisition of Lidar returns with ToF $t \in [T_{\min}, T_{\max}]$ while \angle_{el} and \angle_{rot} in Fig. 3.1 are varied accordingly.

The function f_{ZX} from Eq. (4.10) is, much like the polar coordinate transform, very well behaved so that the mapping theorem [78, sec. 2.3] ensures a change of coordinates preserves the PPP distribution and, although this technically constitutes a slight loss of generality and our theory is more generally applicable, we may intuitively think of $Z_{\llbracket - \rrbracket}(x_D)$ as a product of compact intervals. Since $X_{\llbracket - \rrbracket}(x_D) = f_{ZX}(Z_{\llbracket - \rrbracket}(x_D))$ is a proper subset of the domain X it will always be assumed that the data domain is sufficiently large to result in non-trivial estimators. In particular, if the optical parameters are taken from a kernel space $K(\phi | x_D)$ then $Z_{\llbracket - \rrbracket}(x_D)$ is assumed to contain all directions that point from the detector to a kernel centre and times at which the kernels can be reached. This holds true if X is convex and b_j as in Eq. (2.5) are constrained to $X_{\llbracket - \rrbracket}(x_D)$. Practically this can be implemented as part of the parameter space Θ and the corresponding map $\mathcal{G} : \Theta \rightarrow \Xi$ which maps $\theta \in \Theta$ to a tuple of optical parameters $(\alpha, \sigma_a, \sigma_s, f_p) = \xi \in \Xi$ as defined in Eq. (3.10).

In this case, we have for the analogous quantity from Eq. (3.17)

$$\begin{aligned} \mathsf{L}_\infty(\xi, \psi \mid \mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}) &= - \int_{Z_{[-]}(x_D)} (\mathfrak{m}_{\text{on}}(v, t \mid x_D) + \mathfrak{m}_{\text{off}}(v, t \mid x_D)) \psi(v, t) dt dv \\ &\quad + \int_{Z_{[-]}(x_D)} \log(\psi(v, t) \mathfrak{m}_{\text{on}}(v, t \mid x_D)) d\mathbf{M}^{\text{on}}(v, t) \\ &\quad + \int_{Z_{[-]}(x_D)} \log(\psi(v, t) \mathfrak{m}_{\text{off}}(v, t \mid x_D)) d\mathbf{M}^{\text{off}}(v, t) \end{aligned} \quad (4.13)$$

which becomes a function of θ when $\mathcal{G}(\theta) = \xi$. Note that this is a consistent generalisation of the previously considered model as the binning that will occur in practice is nothing but integration over disjoint subsets which results in a Poisson distribution while the second assumption in Definition 4.1.2 ensures their independence.

There is an important connection between Lemma 4.1.1 and continuity of an inverse problem. If \mathbf{X}_j is a sequence of data with decreasing levels of (random) noise then consistency of the estimator means that small errors in the data will, with high probability, result in small errors in the estimator. We have the following result which connects Lemma 4.1.1 with Definition 4.1.2.

Proposition 4.1.3 (Consistency for continuous parameters). *Consider a set of continuous functions $\Xi \subseteq C(X) \times C(X) \times C(X) \times \{f_p\}$ and $\Psi \subseteq C(Z_{[-]}(x_D))$ such that $\Xi \times \Psi$ is compact with respect to the product topology induced by uniform convergence. Further assume that $\xi = (\alpha, \sigma_a, \sigma_s, f_p) \in \Xi$ is the ground truth for intensity measures $\Lambda_{\text{on/off}}$ satisfying Eq. (4.8). If the conditions of Theorem 3.1.4 are met for each $\xi' \in \Xi$ then for i.i.d. observations $\mathbf{M}_k^{\text{on}}, \mathbf{M}_k^{\text{off}} \sim \text{PPP}(\Lambda_{\text{on/off}})$ the maximum likelihood estimator*

$$(\hat{\alpha}, \hat{\sigma}_a, \hat{\sigma}_s, \hat{f}_p, \hat{\psi})_n = (\hat{\xi}, \hat{\psi})_n \in \arg \max_{(\xi', \psi') \in \Xi \times \Psi} \frac{1}{n} \sum_{k=1}^n \mathsf{L}_\infty(\xi', \psi' \mid \mathbf{M}_k^{\text{on}}, \mathbf{M}_k^{\text{off}}) \quad (4.14)$$

exists and the differential absorption convergences uniformly in probability to the ground truth value. More precisely, for any $\delta > 0$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\|\hat{\alpha}_n - \alpha\|_\infty > \delta) = 0.$$

In situations where Ξ is parameterised by a finite-dimensional set, i.e. $\Xi = \mathcal{G}(\Theta)$ for $\Theta \subseteq \mathbb{R}^d$ compact while \mathcal{G} continuous and invertible, then any sequence of implied MLEs $\hat{\theta}_n$ will converge to the ground truth $\theta = \mathcal{G}^{-1}(\xi)$ in probability.

Proof. The proof makes use of the fact that Ξ is compact so the optimisation is carried out over a compact set of continuous functions and therefore all integrals in Eq. (4.13) are well defined and the objective is continuous. Such sets can easily be constructed based on the Arzela-Ascoli Theorem (see [125], p. 395). As continuous functions attain a (possibly non-unique) maximum on compact sets we know that the estimators are well-defined, albeit their value is possibly not unique. It remains to show the convergence result. We first note that the loss function is bounded by

$$C \left(1 + \frac{1}{n} \sum_{k=1}^n \mathbf{M}_k^{\text{on}}(Z_{\llbracket - \rrbracket}(x_D)) + \mathbf{M}_k^{\text{off}}(Z_{\llbracket - \rrbracket}(x_D)) \right) \quad (4.15)$$

for some $C > 0$ that depends on the feasible set only. Note that that must be a universal constant that bounds all elements in Ξ or else there would be an unbounded and thus not uniformly convergent sub-sequence. By dominated convergence and continuity of the objective function for finite n it follows that

$$\frac{1}{n} \sum_{k=1}^n \mathsf{L}_{\infty}(\xi | \mathbf{M}_k^{\text{on}}, \mathbf{M}_k^{\text{off}}) \rightarrow \mathbb{E} (\mathsf{L}_{\infty}(\xi | \mathbf{M}_1^{\text{on}}, \mathbf{M}_1^{\text{off}})) \quad (4.16)$$

and the limit is continuous in (ξ, ψ) . Furthermore (see [78]) we have

$$\begin{aligned} \mathbb{E} (\mathsf{L}_{\infty}(\xi, \psi | \mathbf{M}_1^{\text{on}}, \mathbf{M}_1^{\text{off}})) &= - \int_{Z_{\llbracket - \rrbracket}(x_D)} \mathfrak{m}_{\text{on}}(v, t | x_D) + \mathfrak{m}_{\text{off}}(v, t | x_D) \psi(v, t) dt dv \\ &\quad + \int_{Z_{\llbracket - \rrbracket}(x_D)} \log(\psi(v, t) \mathfrak{m}_{\text{on}}(v, t | x_D)) d\Lambda_{\text{on}}(v, t) \\ &\quad + \int_{Z_{\llbracket - \rrbracket}(x_D)} \log(\psi(v, t) \mathfrak{m}_{\text{off}}(v, t | x_D)) d\Lambda_{\text{off}}(v, t) \end{aligned}$$

where $\Lambda_{\text{on/off}}$ denotes the ground truth's intensity measures. It is easily seen that the integrand on the right-hand side of Eq. (4.16) is minimised for each point at the ground truth parameter which means the entire integral must be too. If we can show that Eq. (4.16) is

uniform in probability and the limit is uniquely minimised at the ground truth, then the result follows from Lemma 4.1.1. However, uniform convergence in probability is implied by compactness and the fact that continuous functions on compact spaces are uniformly continuous, see e.g. [107, Lemma 2.4]. \square

Assumptions such as in Theorem 3.1.4 are the basis for the above result and as it turns out not just rather optimistic but also necessary for a result of that strength to remain true. As mentioned earlier, the proof of that result relies primarily on our ability to locally extract single scattering at the “edges” of each kernel and “extrapolate”. Much of the remainder of this chapter is devoted to the analysis of the estimator’s behaviour under relaxed conditions and specifically how it affects consistency as well as the rate of convergence which is not specified in Proposition 4.1.3.

Parallel planes - A simple example In order to develop intuition of what to expect from an estimator such as Eq. (4.14) and make analytical computations easier, we consider a simplified scenario akin to that of cloud imaging [20] and assume a detector located at the origin $x_D = 0 \in \mathbb{R}^3$ along with a single pulse sent towards $v = (1, 0, 0)$. We will also assume that the optical parameters are functions of x_1 . More precisely we consider the following.

Definition 4.1.4 (Parallel planes geometry). Let $x_D = 0 \in \mathbb{R}^3$ and $v_0 = (1, 0, 0) \in \mathbb{S}^2$ and assume that the optical parameters σ_s, σ_a and α are functions of x_1 , i.e. for $f \in \{\sigma_s, \sigma_a, \alpha\}$ and $x, y \in \mathbb{R}^3$ we have

$$\langle x, v_0 \rangle = \langle y, v_0 \rangle \implies f(x) = f(y)$$

and $f_p(x, v' \rightarrow v) = \frac{1}{4\pi}$ is homogeneous in both its spatial and angular argument. A setup like this will be referred to as *parallel planes geometry* and we will write $\mathbf{m}_{\text{on/off}}(t)$ instead of $\mathbf{m}_{\text{on/off}}(t, v_0)$ as well as $f(x) = f(\langle x, v_0 \rangle)$ for any $f \in \{\sigma_s, \sigma_a, \alpha\}$, i.e. use the same symbol for the unique one-dimensional functional that determines the corresponding 3-dimensional optical parameter. We will also write for the boundaries $0 \leq s_f = \inf \text{supp } f$ for the x_1 coordinate of the topological boundaries of $f \in \{\sigma_s, \sigma_a, \alpha\}$. Finally, we also want

the parameters to be bounded on compact sets, continuous on (s_{σ_s}, ∞) , right continuous at s_{σ_s} with σ_s additionally having to be (strictly) positive for values larger than s_{σ_s} , i.e. we require $t > s_{\sigma_s} \implies \sigma_s(t) > 0$.

Note that these parameters are not from $K(\phi \mid x_D)$ and may violate some of the usual assumptions made regarding the optical parameters but it is easy to see that for sufficiently regular parameters the RTE has a (continuous) solution, i.e. no scattering in some areas does not cause any issues in that regard. The corresponding intensities $\mathbf{m}_{\text{on/off}}(t)$ are still well defined and continuous when restricted $t \geq s_{\sigma_s}$ and $\mathbf{m}_{\text{off}}(t) > \mathbf{m}_{\text{on}}(t)$ for all $t > \max\{s_{\sigma_s}, s_\alpha\}$. The class of parallel planes is not meant to serve as a different layer of regularity but rather make the case for the existence of “bad behaviour” and when it is to be expected on more general function classes.

Proposition 4.1.5 (Boundary consistency). *For a parallel planes geometry let $\mathbf{m}_{\text{on/off}}(t)$ and $\tilde{\mathbf{m}}_{\text{on/off}}(t)$ be a pair of measurements corresponding to optical parameters $\alpha, \sigma_s, \sigma_a$ and $\tilde{\alpha}, \tilde{\sigma}_s, \tilde{\sigma}_a$. Define $s_0 := \max\{s_{\sigma_s}, s_\alpha\}$ and assume that $s_{\sigma_s} = s_{\tilde{\sigma}_s}$. If there is $\varepsilon > 0$ such that for all $t \in (s_0, s_0 + \varepsilon)$ we have*

$$\frac{\mathbf{m}_{\text{on}}(t)}{\mathbf{m}_{\text{off}}(t)} = \frac{\tilde{\mathbf{m}}_{\text{on}}(t)}{\tilde{\mathbf{m}}_{\text{off}}(t)} \quad (4.17)$$

and thus $\Lambda_{\text{on/off}} = \tilde{\Lambda}_{\text{on/off}}$ is a possibility in the situation of Eq. (4.8), then

$$\int_0^{s_0} \alpha(t) dt = \int_0^{s_0} \tilde{\alpha}(t) dt. \quad (4.18)$$

If in addition $(\sigma_s, \sigma_a) = (\tilde{\sigma}_s, \tilde{\sigma}_a)$ on $[0, s_0]$ then

$$\alpha(s_0) = \lim_{t \searrow s_0} \alpha(t) = \lim_{t \searrow s_0} \tilde{\alpha}(t) = \tilde{\alpha}(s_0). \quad (4.19)$$

In summary, if the parameters result in identical measurements then the differential absorption parameter coincides at the boundary of its support.

Proof. If $s_\alpha \geq s_{\sigma_s}$ then Eq. (4.17) implies $s_{\tilde{\alpha}} = s_\alpha$ but Eq. (4.18) becomes a trivial

statement as both sides are 0. Otherwise, both sides are positive and we have

$$2 \int_0^{s_{\sigma_s}} \alpha(t) dt = -\log \left(\lim_{r \searrow 0} \frac{\mathfrak{m}_{\text{on}}(s_{\sigma_s} + r)}{\mathfrak{m}_{\text{off}}(s_{\sigma_s} + r)} \right) = -\log \left(\lim_{r \searrow 0} \frac{\tilde{\mathfrak{m}}_{\text{on}}(s_{\sigma_s} + r)}{\tilde{\mathfrak{m}}_{\text{off}}(s_{\sigma_s} + r)} \right) = 2 \int_0^{s_{\sigma_s}} \tilde{\alpha}(t) dt$$

which proves Eq. (4.18) for the case $s_\alpha \leq s_{\sigma_s}$. For $s_\alpha \leq s_{\sigma_s}$

$$\lim_{r \searrow 0} \left(\frac{\mathfrak{m}_{\text{on}}(s_0 + r)}{2r\mathfrak{m}_{\text{off}}(s_0 + r)} - \frac{e^{-2 \int_0^{s_0} \alpha(t) dt}}{2r} \right) = \lim_{r \searrow 0} \frac{\mathfrak{m}_{\text{off}}^1(s_0 + r) \left(e^{-2 \int_0^{s_0+r} \alpha(t) dt} - e^{-2 \int_0^{s_0} \alpha(t) dt} \right)}{2r\mathfrak{m}_{\text{off}}(s_0 + r)}$$

the right hand side of which equals $-\alpha(s_0)$. This follows essentially the same arguments that led to Lemma 3.1.2 and exploit that higher order scattering $\mathfrak{m}_{\text{on/off}}^{2+} = \sum_{k=2}^{\infty} \mathfrak{m}_{\text{on/off}}^k$ vanishes for $r \searrow 0$ along with

$$\begin{aligned} e^{-2 \int_0^{s_0} \alpha(t) dt} \mathfrak{m}_{\text{off}}^{2+}(s_0 + r) - \mathfrak{m}_{\text{on}}^{2+}(s_0 + r) &\leq \mathfrak{m}_{\text{off}}^{2+}(s_0 + r) 2r \sup_{t \in [s_\alpha, s_0 + \varepsilon]} \alpha(t) \\ e^{-2 \int_0^{s_0} \alpha(t) dt} \mathfrak{m}_{\text{off}}^1(s_0 + r) - \mathfrak{m}_{\text{on}}^1(s_0 + r) &= \mathfrak{m}_{\text{off}}^1(s_0 + r) \left(e^{-2 \int_0^{s_0+r} \alpha(t) dt} - e^{-2 \int_0^{s_0} \alpha(t) dt} \right) \end{aligned}$$

for r sufficiently small. However, the same must (by assumption) also be true for $\tilde{\mathfrak{m}}_{\text{on/off}}(t)$ and the claim follows due to the same arguments as the parameters are bounded on compact sets. It remains to show the result for $s_{\sigma_s} < s_\alpha$ in which case we also required a necessary condition for $\mathfrak{m}_{\text{off}}^{1/2+} = \tilde{\mathfrak{m}}_{\text{off}}^{1/2+}$ and $\mathfrak{m}_{\text{off}} = \mathfrak{m}_{\text{on}}$ on $[s_{\sigma_s}, s_\alpha]$. We argue that this implies

$$\begin{aligned} \lim_{r \searrow 0} \frac{\mathfrak{m}_{\text{off}}(s_0)}{2r\mathfrak{m}_{\text{off}}^1(s_0)} \left(1 - \frac{\mathfrak{m}_{\text{on}}(s_0 + r)}{\mathfrak{m}_{\text{off}}(s_0 + r)} \right) &= \frac{\mathfrak{m}_{\text{off}}(s_0)}{\mathfrak{m}_{\text{off}}^1(s_0)} \lim_{r \searrow 0} \frac{\mathfrak{m}_{\text{off}}^1(s_0 + r) \left(1 - e^{-2 \int_0^{s_\alpha+r} \alpha(t) dt} \right)}{2r\mathfrak{m}_{\text{off}}(s_0 + r)} \\ &= \lim_{r \searrow 0} \frac{\left(1 - e^{-2 \int_0^{s_\alpha+r} \alpha(t) dt} \right)}{2r} \\ &= \lim_{r \searrow 0} \frac{\left(1 - e^{-2 \int_0^{s_\alpha+r} \tilde{\alpha}(t) dt} \right)}{2r} \\ &= \frac{\tilde{\mathfrak{m}}_{\text{off}}(s_0)}{\tilde{\mathfrak{m}}_{\text{off}}^1(s_0)} \lim_{r \searrow 0} \frac{\tilde{\mathfrak{m}}_{\text{off}}^1(s_0 + r) \left(1 - e^{-2 \int_0^{s_\alpha+r} \alpha(t) dt} \right)}{2r\tilde{\mathfrak{m}}_{\text{off}}(s_0 + r)} \\ &= \lim_{r \searrow 0} \frac{\tilde{\mathfrak{m}}_{\text{off}}(s_0)}{2r\tilde{\mathfrak{m}}_{\text{off}}^1(s_0)} \left(1 - \frac{\tilde{\mathfrak{m}}_{\text{on}}(s_0 + r)}{\tilde{\mathfrak{m}}_{\text{off}}(s_0 + r)} \right). \end{aligned}$$

and, given that $s_\alpha = s_0$, the claim follows from the equality between lines two and three which states $\alpha(s_\alpha) = \tilde{\alpha}(s_\alpha)$. To see that this is true, note that the left-hand side of the first line is by assumption equal to the expression in the last line. The equality in line one as well as that between four and five is due to the same arguments that we used to show Lemma 3.1.2 which is to say we use that higher order scattering is absorbed only on a set of vanishing measure for $r \rightarrow 0$. The relation between lines one and two as well as three and four is due to the right-continuity of the measurement. This proves the claim by virtue of transitivity. \square

Note that the assumptions made in Proposition 4.1.5 are rather different from those Theorem 3.1.4 and our other uniqueness related claims in that the differential absorption field cannot be extrapolated from its behaviour locally at the support boundary. The following results show that in a sense this is also the best we can hope for if we allow the parameters to vary freely in the x_1 coordinate.

Proposition 4.1.6. *Let $\mathbf{m}_{\text{on/off}}(t)$ as before for a parallel plane geometry and denote by $s_{\sigma_s}, s_{\sigma_a}, \alpha$ its corresponding parameters. There exists a choice of parameters such that*

(R1) *the parameters have identical support and thus $s_{\sigma_s} = s_{\sigma_a} = s_\alpha =: s_0 > 0$*

(R2) *there is an open set U containing s_0 such that σ_a, σ_s and α are non-decreasing and continuously differentiable on U*

and a sequence $\mathbf{m}_{\text{on/off},j}(t)$ with corresponding parameters $\alpha_j, \sigma_{s,j}, \sigma_{a,j}$ and $j \in \mathbb{N}$ as well as some $\varepsilon > 0$ which depends only on σ_s, σ_a and α such that the corresponding sequence of differential absorptions measurements satisfies

$$\lim_{j \rightarrow \infty} \sup_{t \in [s_0, s_0 + \varepsilon]} \left| \frac{\mathbf{m}_{\text{on}}(t)}{\mathbf{m}_{\text{off}}(t)} - \frac{\mathbf{m}_{\text{on},j}(t)}{\mathbf{m}_{\text{off},j}(t)} \right| = 0 \quad (4.20)$$

while at the same time for any $s \in (s_0, s_0 + \varepsilon)$

$$\liminf_{j \rightarrow \infty} \left| \int_0^s \alpha_j(t) dt - \int_0^s \alpha(t) dt \right| > 0. \quad (4.21)$$

Moreover the sequences α_j , $\sigma_{s,j}$, $\sigma_{a,j}$ can be chosen uniformly convergent on $[s_0, s_0 + \varepsilon]$. Note that by Proposition 4.1.5 $\alpha_j(s_0) \rightarrow \alpha(s_0) = 0$ for such a sequence so this is indeed the worst case possible as no average apart from a single point evaluation coincides.

Proof. We prove this by giving an example. Let $\sigma_{s,j} = \frac{1}{j}\sigma_s$ and $\sigma_{a,j} = \frac{1}{j}\sigma_a$ for all $j \in \mathbb{N}$, which leaves the fixed parameter σ_s to be specified later. These are clearly viable sequences for which the path distribution degenerates to single scattering, i.e. for $s \geq s_0$

$$-\lim_{j \rightarrow \infty} \log \left(\frac{\mathbf{m}_{\text{on},j}(s)}{\mathbf{m}_{\text{off},j}(s)} \right) = \log \left(\frac{\mathbf{m}_{\text{off}}^1(s)}{\mathbf{m}_{\text{on}}^1(s)} \right) = 2 \int_0^s \lim_{j \rightarrow \infty} \alpha_j(t) dt \quad (4.22)$$

as $j \rightarrow \infty$ and the convergence, as a function in the variable s , is uniform on compact sets for any uniformly convergent sequence α_j . As before α will be specified later. We claim that for suitable α and σ_s there exists $\varepsilon > 0$ such that $\frac{\mathbf{m}_{\text{on}}(t)}{\mathbf{m}_{\text{off}}(t)}$ is decreasing and continuously differentiable on $(s_0, s_0 + \varepsilon)$ which means that choosing

$$\alpha_j(t) = -\frac{1}{2} \frac{d}{dt} \log \left(\frac{\mathbf{m}_{\text{on}}(t)}{\mathbf{m}_{\text{off}}(t)} \right) \quad (4.23)$$

for all $t \in [s_0, s_0 + \varepsilon)$ and $j \in \mathbb{N}$ is a viable choice in the sense that it is non-negative and continuous. Geometrically it is clear that for any $r > 0$

$$e^{-2 \int_0^r \alpha_j(s_0+t) dt} = \frac{\mathbf{m}_{\text{on}}(s_0+r)}{\mathbf{m}_{\text{off}}(s_0+r)} > e^{-2 \int_0^r \alpha(s_0+t) dt}$$

because any ray that isn't perfectly straight spends more time outside $[s_0, \infty)$ than single scattering and α is non-decreasing. This implies Eq. (4.21). Since α_j converge uniformly Eq. (4.20) follows from Eq. (4.22) and all we need to show is that Eq. (4.23) is well defined and, in particular, non-negative. Note that by definition σ_s is required to be strictly positive on (s_0, ∞) , which makes $\mathbf{m}_{\text{on/off}}$ positive and continuous. As such, Eq. (4.23) is well defined and continuous if and only if $\mathbf{m}_{\text{on/off}}$ are continuously differentiable. To see the differentiability consider Eq. (A.3) and form differential quotients

$$\lim_{r \searrow 0} \frac{\mathbf{m}_{\text{on/off}}(t) - \mathbf{m}_{\text{on/off}}(t+r)}{r}.$$

After a change of variables $s_i \mapsto s_i(1 + r/t)$ in Eq. (A.3), which scales each segment of a closed polygon and transforms one of length t to one of length $t+r$ in a way that preserves angles, the indicator that was included as part of Eq. (A.4), which is the only point of concern, can be factored out of the differential quotient. All parameters are continuously differentiable on $(0, \infty)$ with bounded derivatives on any compact subset of the open set U which contains the support boundary point s_0 and thus an open interval $(s_0, s_0 + \varepsilon)$ for some $\varepsilon > 0$. This means that we can differentiate the integrand and its derivatives are bounded by a value that is growing at most exponentially fast in the order of scattering. Due to Eq. (A.7) this means $\mathbf{m}_{\text{on/off}}$ is continuously differentiable. As this was true for any choice satisfying conditions *(R1)* and *(R2)* it is enough to show that there is a choice which makes Eq. (4.23) non-negative. A simple choice is to select some $\delta > 0$ and let $\alpha(t) = (t - s_0)^{1+\delta} 1_{(s_0, \infty)}(t)$ while $\sigma_a(t) = \sigma_s(t) = (t - s_0)^{2+2\delta} 1_{(s_0, \infty)}(t)$ but this only needs to be true locally around s_0 . Then there are functions $g_1, g_2 \geq 0$ such that

$$\frac{\mathbf{m}_{\text{on}}(s_0 + r)}{\mathbf{m}_{\text{off}}(s_0 + r)} = \frac{1 - \frac{2}{2+\delta} r^{2+\delta} + g_1(r)}{1 + g_2(r)}$$

and $g'_1(r), g'_2(r) \in O(r^{1+2\delta})$ which means the derivative is locally dominated by $-2r^{1+\delta}$ and therefore implies that Eq. (4.23) is indeed non-negative at least for r sufficiently small. \square

Remark. The proof constructs optical parameters that satisfy a similar condition as Eq. (3.14) in the sense that the non-differential parameters σ_a, σ_s decay faster than α at the boundary while regularity assumptions *(R1)* and *(R2)* are similar to the impositions on the kernel functions. Crucially the functions are not assumed known up to a finite number of degrees of freedom.

A minor modification to the optical parameters constructed in the proof of Proposition 4.1.6 can be used to show that recovery of the differential absorption field from the ratio of $\mathbf{m}_{\text{on/off}}$ is an ill-posed problem. To better appreciate the next result and the developments to come, note that the log-likelihood \mathcal{L}_∞ from Eq. (4.13) can be decomposed

into

$$\begin{aligned} \mathcal{L}_{\text{Poi}}(\xi, \psi \mid \mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}) &= \int_{Z_{\llbracket - \rrbracket}(x_D)} \log((\mathfrak{m}_{\text{on}} + \mathfrak{m}_{\text{off}})\psi) d\mathbf{M}^{\text{on}} + \mathbf{M}^{\text{off}} \\ &\quad - \int_{Z_{\llbracket - \rrbracket}(x_D)} (\mathfrak{m}_{\text{on}} + \mathfrak{m}_{\text{off}})\psi dt dv \end{aligned} \quad (4.24)$$

$$\begin{aligned} \mathcal{L}_{\text{Bin}}(\xi \mid \mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}) &= \int_{Z_{\llbracket - \rrbracket}(x_D)} \log\left(\frac{\mathcal{Q}(\xi)}{1 + \mathcal{Q}(\xi)}\right) \frac{d\mathbf{M}^{\text{on}}}{d\mathbf{M}^{\text{on}} + \mathbf{M}^{\text{off}}} d\mathbf{M}^{\text{on}} + \mathbf{M}^{\text{off}} \\ &\quad + \int_{Z_{\llbracket - \rrbracket}(x_D)} \log\left(\frac{1}{1 + \mathcal{Q}(\xi)}\right) \frac{d\mathbf{M}^{\text{off}}}{d\mathbf{M}^{\text{on}} + \mathbf{M}^{\text{off}}} d\mathbf{M}^{\text{on}} + \mathbf{M}^{\text{off}} \end{aligned} \quad (4.25)$$

and expressed as

$$\mathcal{L}_\infty(\xi, \psi \mid \mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}) = \mathcal{L}_{\text{Poi}}(\xi, \psi \mid \mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}) + \mathcal{L}_{\text{Bin}}(\xi \mid \mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}).$$

The expression \mathcal{L}_{Bin} in Eq. (4.25) is notably independent of ψ and a function of $\mathcal{Q}(\xi)$ only. The variable ψ in Proposition 4.1.3 was constrained to a compact set Ψ so that all estimators exist and are well defined. In view of our previous developments where we considered Eq. (3.12b) after dropping the restriction from Eq. (3.12c) one should think of Ψ as a very large set of highly variable functions. If $\psi \in [0, \infty)^{Z_{\llbracket - \rrbracket}(x_D)}$ is a “free” function-valued variable then Eq. (4.24) is strictly speaking unbounded but, ignoring some technicalities, we can consider discrete approximations, i.e. we consider index sets $\mathbb{D}_1 \subseteq \mathbb{D}_2 \subseteq \mathbb{D}_3 \dots \subseteq \mathbb{N}^2$ and put

$$\psi(\cdot) = \sum_{(i,j) \in \mathbb{D}_l} \psi_{i,j}^l \text{bin}_{i,j}^l(\cdot) \quad (4.26)$$

where for all $l \in \mathbb{N}$ the $\psi_{i,j}^l \in [0, \infty)$ are scalars, $\text{bin}_{i,j}^l(\cdot)$ are “voxel-like” indicator functions with centres $(v_i, t_j) \in Z_{\llbracket - \rrbracket}(x_D)$ normalised such that for all $(i, j) \in \mathbb{D}_l$ and

$$\int_{Z_{\llbracket - \rrbracket}(x_D)} \text{bin}_{i,j}^l(v, t) dv dt = 1, \quad \text{bin}_{i,j}^l(\cdot) \xrightarrow{l \rightarrow \infty} \delta_{v_i, t_j}(\cdot). \quad (4.27)$$

Eq. (4.27) can be interpreted as having increasingly accurate discrete measurements with distribution as in Eq. (3.15). When $\bigcup_{l \in \mathbb{N}} \{(v_i, t_j) : (i, j) \in \mathbb{D}_l\}$ is dense in $Z_{\llbracket - \rrbracket}(x_D)$ then

Eq. (4.26) approximates any continuous function and the maximiser $\hat{\psi}$ of the likelihood $L_\infty(\xi, \psi | M^{\text{on}}, M^{\text{off}})$ w.r.t. ψ subject to Eq. (4.26) satisfies for increasingly fine binning

$$\hat{\psi}_{i,j}^l \xrightarrow{l \rightarrow \infty} \frac{M^{\text{on}}(\{(v_i, t_j)\}) + M^{\text{off}}(\{(v_i, t_j)\})}{m_{\text{on}}(v_i, t_j | x_D) + m_{\text{off}}(v_i, t_j | x_D)} \implies \hat{\psi} \xrightarrow{l \rightarrow \infty} \frac{M^{\text{on}} + M^{\text{off}}}{m_{\text{on}} + m_{\text{off}}}. \quad (4.28)$$

The discrete analog for this result is obtained in Eq. (5.2) and a similar expression will be derived later in Eq. (5.17) in the presence of ambient noise. Given Eq. (4.28) we see that in the limit we can write, up to an additive constant,

$$L_\infty(\xi, \hat{\psi} | M^{\text{on}}, M^{\text{off}}) = L_{\text{Bin}}(\xi | M^{\text{on}}, M^{\text{off}})$$

which means that ξ , and by extension θ when $\xi = \mathcal{G}(\theta)$, is estimated based on information contained in \mathcal{Q} . Moreover, the estimation is based on $\frac{dM^{\text{off}}}{dM^{\text{on}} + M^{\text{off}}}$ and $\frac{dM^{\text{on}}}{dM^{\text{on}} + M^{\text{off}}}$ conditional on $M^{\text{on}} + M^{\text{off}}$. Seeing that

$$\frac{\mathcal{Q}(\xi)}{1 + \mathcal{Q}(\xi)} \xrightarrow{\text{PPP}} \frac{dM^{\text{on}}}{dM^{\text{on}} + M^{\text{off}}} \quad \text{and} \quad \frac{1}{1 + \mathcal{Q}(\xi)} \xrightarrow{\text{PPP}} \frac{dM^{\text{off}}}{dM^{\text{on}} + M^{\text{off}}},$$

i.e. the quantities on the right can be regarded as noisy realisations of functions of $\mathcal{Q}(\xi)$, it becomes clear that some form of continuity of \mathcal{Q}^{-1} is crucial for consistent estimation of the optical parameters. In Proposition 4.1.3 this was guaranteed through the assumptions made Theorem 3.1.4. The following result indicates that such restrictions are to some degree a necessity in the RTE setting even if we just seek to recover the differential absorption α .

Corollary 4.1.7. *Let $X = \{x \in \mathbb{R}^3 : x_1 > 0\}$ and consider $Z_{[-]}(x_D)$ as in Eq. (4.12) for a wide FOV detection modality, i.e. we observe multiple scattering consistent with the Radiative Transfer Equation (2.22) restricted to a compact set $Z_{[-]}(x_D)$. We can find $Z_{[-]}(x_D)$ with non-empty interior and a subset Ξ of bounded and continuous functions such that $\mathcal{Q} : \Xi \rightarrow C(Z_{[-]}(x_D))$ as defined in Eq. (4.1)² does not have a continuous inverse in the differential absorption component. More precisely, let $\lambda = \mathcal{Q}(\xi)$ for some*

²for well behaved optical parameters $\frac{m_{\text{on}}}{m_{\text{off}}} = 1$ in case $m_{\text{off}} = 0$, see also Lemma 4.2.3

$(\alpha, \sigma_a, \sigma_s, f_p) = \xi \in \Xi$ and define $A(\lambda)$ as the level set of \mathcal{Q} in the differential absorption component which is to say

$$A(\lambda) = \left\{ \tilde{\alpha} \in C(X) : \exists \tilde{\xi}_{\text{off}} \in \Xi_{\text{off}}(\tilde{\alpha}) \text{ s.t. } \lambda = \mathcal{Q}((\tilde{\alpha}, \tilde{\xi}_{\text{off}})) \right\}.$$

where $\Xi_{\text{off}}(\cdot)$ is as in Eq. (3.9). In other words, $A(\lambda)$ is the set of all differential absorption fields $\tilde{\alpha}$ such that there is a corresponding parameter tuple $(\tilde{\alpha}, \tilde{\sigma}_a, \tilde{\sigma}_s, \tilde{f}_p)$ for which we have $\mathcal{Q}(\tilde{\xi}) = \lambda = \mathcal{Q}(\xi)$. For any metric d_α on $C(X)$ there is $\lambda = \mathcal{Q}(\xi)$ and $\delta > 0$ such that for any $\varepsilon > 0$ there is a $\lambda_\varepsilon = \mathcal{Q}(\xi_\varepsilon)$ which satisfies

$$\|\lambda_\varepsilon - \lambda\|_\infty < \varepsilon \quad \text{and} \quad \sup_{(\alpha_1, \alpha_2) \in A(\lambda_\varepsilon) \times A(\lambda)} d_\alpha(\alpha_1, \alpha_2) > \delta.$$

There is a choice for the set Ξ such that the above statement remains valid under the additional constraint

$$\xi_{\text{off}} = (\sigma_a, \sigma_s, f_p) \in \Xi_{\text{off}}(\alpha) \implies \text{supp } \alpha = \text{supp } \sigma_s = \text{supp } \sigma_a$$

which is as a weaker version of the proportionality requirement made in Theorem 3.1.4.

Proof. Use the sequence from Proposition 4.1.6 but spherically symmetric instead of parallel planes, so that the response in each direction is identical, in order to get the result not just in time but also direction. Neither α_1 nor α_2 depend on ε but only on λ and, given that $\alpha_1 \neq \alpha_2$, we may select $\delta \in (0, d_\alpha(\alpha_1, \alpha_2)) \neq \emptyset$. \square

Remark. Corollary 4.1.7 is in stark contrast to the single scattering model Eqs. (3.11a) and (3.11b) where \mathcal{Q} is independent of ξ_{off} and only a function of α . A statement such as the above is trivial for any map that isn't injective in the differential absorption component as it has no inverse at all. The above statement is slightly weaker in that it allows for an inverse to exist but if it does it certainly will not be continuous with respect to any metric on $C(X)$ which includes rather weak notions of convergence such as those related to Eq. (4.21). In fact, \mathcal{Q} may very well not be injective in the α component and this is no contradiction to Theorem 3.1.4 as that space is quite heavily restricted in its degrees of

freedom. Also note that the sequence of functions depends on the domain $C(Z_{[-]}(x_D))$ as well as the FOV.

Upon closer inspection, the above result may seem even more of a problem than it would seem initially as even the conventional ideas regarding regularisation, which would force the optical parameters to be smooth, seem futile as an examination of the example constructed in Proposition 4.1.6 shows that the issue is caused by smooth deviations of the optical parameters. However, if we only care about certain “features”, which loosely speaking can be associated with different choices of d_α , then Corollary 4.1.7 does no harm as long as we can ensure that the possible error δ cannot become too large. Moreover, the statement makes no claim about discrepancies that are measured by maps which aren’t metrics and we will show that there are aspects of the differential absorption field, such as those discussed in Proposition 4.1.5, which are always stable.

4.2 Convergence Rates

Up until this point we have primarily considered consistency of the (infinite-dimensional) maximum likelihood estimator. A necessary requirement for this is of course a uniqueness statement such as Theorem 3.1.4. The essence of Section 4.1 is that:

- (C1) If the space of optical parameters is sufficiently restricted then the inverse operator is continuous for (increasingly accurate) PPP data as shown in Proposition 4.1.3.
- (C2) If we optimise over a space of more general nature then we cannot consistently recover the differential absorption field, even if we were to settle for averages as shown in Proposition 4.1.6.
- (C3) If this is combined with Proposition 4.1.5 we see that *exact* recovery is possible precisely where we can extract single scattering.

The realisation in (C3), especially when coupled with Corollary 4.1.7 which for narrow FOVs is not true in such generality, of course, begs the question of why we would want

to use a wide FOV and, in view of Section 3.2 where we claimed wider FOVs can outperform the narrow equivalent, seems somewhat contradictory to the very idea of this work. As it turns out there are aspects of an image, such as generalisations of the parameter $1_{(0,\infty)}(s_\alpha)$ which is the presence of a gas in finite time for parallel plane geometry, that are stable under conditions akin to those imposed in Fig. 3.5. Before we can address this more rigorously through an analysis of the convergence rate we will make a simplification/relaxation to our model which, like we have done with the introduction of ψ by dropping Eq. (3.12c), enlarges the parameter space and thereby simplifies computations and analyses at the price of possibly worse convergence rates.

Definition 4.2.1 (β -errors). Let $\mathbf{M}^{\text{on}} \sim \text{PPP}(\Lambda_{\text{on}})$ and $\mathbf{M}^{\text{off}} \sim \text{PPP}(\Lambda_{\text{off}})$ on $Z_{\llbracket - \rrbracket}(x_D)$ and, as before, Θ a finite-dimensional set and $\mathfrak{m}_{\text{on/off}}[\theta]$ defined through Eqs. (3.8), (3.10), and (3.16). Any function β' that satisfies

$$\frac{d\Lambda_{\text{on}}}{d\Lambda_{\text{off}}}(v, t) = \beta'(v, t) \frac{\mathfrak{m}_{\text{on}}[\theta](v, t)}{\mathfrak{m}_{\text{off}}[\theta](v, t)} =: \mathfrak{q}[\theta, \beta'](v, t) \quad (4.29)$$

for all $(v, t) \in Z_{\llbracket - \rrbracket}(x_D)$ is called β -deviation of $\mathfrak{m}_{\text{on/off}}[\theta]$ with respect to $\Lambda_{\text{on/off}}$. If $\mathfrak{m}_{\text{on/off}}[\theta]$ is positive everywhere then a β -deviation exists, is uniquely determined by Eq. (4.29) and will be denoted by $\beta(\cdot | \theta)$. The β -formulation of the loss is in that case defined as

$$\begin{aligned} \mathsf{B}_\infty(\theta, \beta' | \mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}) &= \int_{Z_{\llbracket - \rrbracket}(x_D)} \log \left(\frac{\mathfrak{q}[\theta, \beta']}{1 + \mathfrak{q}[\theta, \beta']} \right) d\mathbf{M}^{\text{on}} \\ &\quad + \int_{Z_{\llbracket - \rrbracket}(x_D)} \log \left(\frac{1}{1 + \mathfrak{q}[\theta, \beta']} \right) d\mathbf{M}^{\text{off}}. \end{aligned} \quad (4.30)$$

If $\mathfrak{m}_{\text{on/off}}[\theta]$ is positive everywhere then for any $\theta \in \Theta$ the expected value of Eq. (4.30) is minimised at $\beta(\cdot | \theta)$. Also note that Eq. (4.30) is reminiscent of the log-likelihood for a binomial process and closely resembles Eq. (4.25) but is parameterised differently.

From a mathematical perspective, the β -formulation is essentially a re-parameterisation of $\mathcal{Q}(\Xi)$ for some $\Xi \neq \mathcal{G}(\Theta)$ that is *not* a finite-dimensional space of optical parameters. Instead of modelling $\mathcal{Q}(\Xi)$ through Ξ and evaluation of \mathcal{Q} , which is computationally

rather expensive, we may want to approximate $\mathcal{Q}(\Xi)$ differently. If, as usual, $\theta \in \Theta$ a finite-dimensional parameter and $\mathcal{G} : \theta \mapsto \xi[\theta] = (\alpha[\theta], \sigma_a[\theta], \sigma_s[\theta], f_p[\theta])$ as in Eq. (3.10), then we may form

$$\mathcal{Q} \circ \mathcal{G}(\theta) = \frac{\mathfrak{m}_{\text{on}}[\theta]}{\mathfrak{m}_{\text{off}}[\theta]}.$$

If we define the set $B_\Xi(\theta)$ of “ratios” at $\theta \in \Theta$ as

$$B_\Xi(\theta) = \left\{ \frac{\mathcal{Q}(\xi)}{\mathcal{Q} \circ \mathcal{G}(\theta)} : \xi \in \Xi \right\} \quad (4.31)$$

then $\beta' \in B_\Xi(\theta)$ if and only if there is $\xi \in \Xi$ such that

$$\mathcal{Q}(\xi) = \beta' \frac{\mathfrak{m}_{\text{on}}[\theta]}{\mathfrak{m}_{\text{off}}[\theta]} = \beta' \mathcal{Q} \circ \mathcal{G}(\theta).$$

If we are given PPP data $\mathbf{M}^{\text{on}}, \mathbf{M}^{\text{off}}$ and are confident *a priori* that Eq. (4.8) is true at some $\xi \in \Xi$ of a possibly large set Ξ , then $B_\Xi(\theta)$ are the “correct” parameters to optimise Eq. (4.30) over. Due to complexity constraints, such a relationship will rarely be true exactly and the main idea that we will follow is to restrict the β' -component in the loss to sets that are convenient in the sense that they are easy to handle computationally but also useful in the sense that they preserve as many structural properties as possible. In other words, we seek to reverse engineer $B_\Xi(\theta)$ as best as possible in a computationally and/or analytically convenient form. Consequently we may think of $\mathfrak{q}[\beta', \theta]$ from Eq. (4.30) as spanning a (slightly) larger function space compared to $\mathcal{Q}(\xi)$ from Eq. (4.25). In addition to computation constraints, a relaxation of Eq. (4.31) results in models and estimators that account for errors in the relationship Eq. (4.8). This will obviously limit our ability to recover accurate estimates of θ but we will show that, under certain conditions, aspects of the parameter are uniquely preserved even under weakened assumptions. The following definitions will be useful in order to describe different physical situations and help in defining suitable function spaces for the β' argument in Eq. (4.30).

Definition 4.2.2. Let $u \in C(X) \setminus \{0\}$ be such that $x_D \notin \text{supp } u$ and recall $Z_{[-]}(x_D)$

from Eq. (4.9). For such u we define the non-empty sets

$$\partial_u D := \{(v, t) \in Z_{[-]}(x_D) : \exists \varepsilon > 0 \text{ with } \eta(x_D + sv) > 0 \ \forall s \in (t/2, t/2 + \varepsilon)\} \quad (4.32)$$

$$\partial_u^{0,+} D := \{(v, t+r) \in Z_{[-]}(x_D) : (v, t) \in \partial_u D \text{ and } r \geq 0\} \quad (4.33)$$

$$\partial_u^+ D := \{(v, t+r) \in Z_{[-]}(x_D) : r \geq 0 \text{ and } (v', t) \in \partial_u D \text{ for some } v' \in \mathbb{S}^2\} \quad (4.34)$$

and put $\partial_u^{0,-} D := Z_{[-]}(x_D) \setminus \partial_u^{0,+} D$ as well as $\partial_u^- D := Z_{[-]}(x_D) \setminus \partial_u^+ D$. Note that when considered as points in X via Eq. (4.10) $\partial_u^- D$ is the intersection of a sphere with $Z_{[-]}(x_D)$ and under suitable regularity conditions $\partial_u D$ is a 2-dimensional manifold in $Z_{[-]}(x_D)$ and corresponds to the earliest times when single scattering could “detect” a plume u with a pulse in a given direction. This is true when $u \in K(\phi \mid x_D)$ as in Definition 2.1.1. In this case $\partial_u D$ does not depend on the kernel weights but only on the position and width parameters.

Using the above sets we can define the sets of continuous functions

$$B_{\infty,0}(\alpha) := \{\beta \in C(Z_{[-]}(x_D)) : \beta = 1 \text{ on } \partial_\alpha^{0,-} D\} \quad (4.35)$$

$$B_\infty(\alpha) := \{\beta \in C(Z_{[-]}(x_D)) : \beta = 1 \text{ on } \partial_\alpha^- D\}. \quad (4.36)$$

If we parameterise $\alpha = \alpha[\theta]$ as usual with a finite-dimensional parameter θ , then restricting the β -deviation to a space like $B_{\infty,0}(\alpha[\theta])$ (approximately) corresponds to an assumption where scattering occurs (primarily) inside the plume, such as Fig. 3.5. Assuming that the β -deviation lies in $B_\infty(\alpha[\theta])$ is, for the most part, no constraint at all as absorption cannot occur on times where no light can reach the plume. The set considered in Eq. (4.31) will typically lie somewhere in between $B_{\infty,0}(\alpha[\theta])$ and $B_\infty(\alpha[\theta])$. There are two important distinctions regarding the structure of admissible perturbations with implications to convergence behaviour.

Ergodic time averages If the measurement is taken over long periods of time we may reasonably assume that the low-dimensional “average” of the atmospheric quantities has generated the data and assume that the differential absorption α is accurately represented

by functions in $K(\phi \mid x_D)$. We begin with a technicality.

Lemma 4.2.3. *Let \mathcal{Q} be the operator from Eq. (4.1) that maps optical parameters to Radon-Nikodym derivatives $\frac{\mathbf{m}_{\text{on}}}{\mathbf{m}_{\text{off}}}$ and $\Xi \subseteq K(\phi \mid x_D)^3 \times C(X \times \mathbb{S}^2 \times \mathbb{S}^2)$. Then, writing $(\alpha, \sigma_s, \sigma_a, f_p) = \xi \in \Xi$ and assuming phase functions f_p are strictly positive (with uniform lower bound), we have*

$$\mathcal{Q}(\xi)(v, t) = \begin{cases} e^{-2 \int_0^t \alpha(x_D + rv) dr} & \text{if } \mathbf{m}_{\text{off}}(v, t) = 0 \\ \frac{\mathbf{m}_{\text{on}}(v, t)}{\mathbf{m}_{\text{off}}(v, t)} & \text{if } \mathbf{m}_{\text{off}}(v, t) > 0 \end{cases} \quad (4.37)$$

is a continuous operator $\Xi \rightarrow C(Z_{[-]}(x_D))$. In particular, if Ξ is a compact set then so is the range, i.e. $\{\mathcal{Q}(\xi) : \xi \in \Xi\} \subseteq C(Z_{[-]}(x_D))$ is also compact.

Proof. $\mathcal{Q}(\xi)$ as in Eq. (4.37) is well defined and an element of $C(Z_{[-]}(x_D))$. This shows $\mathcal{Q} : \Xi \rightarrow C(Z_{[-]}(x_D))$ but we claim that this mapping is also continuous as a non-linear operator. For that take $\xi_n = (\alpha_n, \sigma_{s,n}, \sigma_{a,n}, f_{p,n})$ converging uniformly in each component to some $\xi = (\alpha, \sigma_s, \sigma_a, f_p)$. The function \mathbf{m}_{off} is always strictly positive on the interior of $\partial_{\sigma_s}^{0,+} D$. If ξ makes the measured intensity positive everywhere then so does ξ_n , for n large enough, and the claim follows after analysing the numerator and denominator separately. If not then $\partial_{\sigma_{s,n}}^{0,+} D \rightarrow \partial_{\sigma_s}^{0,+} D$ and outside of the kernel component, i.e. in the ambient component, $\sigma_{s,n} \rightarrow 0$ resulting in single scattering which is consistent with how $\mathcal{Q}(\xi)$ was defined when $\mathbf{m}_{\text{off}} = 0$. \square

Remark. The above statement is true on more general function spaces, with essentially identical proof, as long as the sets are reasonably regular.

The following result, despite describing a very simple α , gives some insight into the convergence behaviour of Eq. (4.30) based estimators. An analogous statement for more general function spaces is easily obtained via simple modifications of the proof.

Theorem 4.2.4 (Single kernel consistency). *Assume the differential absorption takes the form $\alpha[\theta^*] = w_0 \phi(h_0^{-1} \|\cdot - b_0\|_2) \in K(\phi \mid x_D) \setminus \{0\}$ for a (known) good kernel ϕ and let $\Theta \subseteq (0, \infty)^2 \times X$ be a compact set that contains the true value $\theta^* = (w_0, h_0, b_0)$. Further,*

assume that the injective and continuous function

$$\mathcal{G} : \Theta \rightarrow K(\phi \mid x_D)^3 \times C(X \times \mathbb{S}^2 \times \mathbb{S}^2) \quad \theta \mapsto (\alpha[\theta], \sigma_a[\theta], \sigma_s[\theta], f_p[\theta])$$

which maps θ to optical parameters is such that for all $\theta \in \Theta$ $\text{supp } \alpha[\theta] \subseteq \text{supp } \sigma_s[\theta]$. For data $\mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}} \sim \text{PPP}$ as in Proposition 4.1.3 generated by the set of ground truth optical parameters $(\alpha[\theta^*], \sigma_a^0, \sigma_s^0, f_p^0) = (\alpha[\theta^*], \xi_{\text{off}}^0) \notin \mathcal{G}(\Theta)$, meaning that the differential absorption α is captured by the range of \mathcal{G} but the other parameters may not be. We assume that f_p^0 is strictly positive, $\text{supp } \alpha[\theta^*] \subseteq \text{supp } \sigma_s^0$ and consider the β -loss based estimator

$$(\hat{\beta}_n, \hat{\theta}_n) = \arg \max_{(\theta, \beta') \in \bar{B}_{\Theta}} \frac{1}{n} \sum_{j=1}^n \mathsf{B}_{\infty}(\theta, \beta' \mid \mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}}) \quad (4.38)$$

for different constraints \bar{B}_{Θ} which are assumed to contain the β -deviation $\beta(\cdot \mid \theta^*)$ at θ^* and we write $\hat{\theta}_n = (\hat{w}_n, \hat{h}_n, \hat{b}_n)$.

1. If \bar{B}_{Θ} is compact, all feasible points (θ, β') satisfy $\beta' \mathcal{Q} \circ \mathcal{G}[\theta] < 1$ on the interior of $\partial_{\alpha[\theta]}^{0,+} D$ and

$$\bar{B}_{\Theta} \subseteq \{(\theta, \beta) : \theta \in \Theta \text{ and } \beta \in B_{\infty}(\alpha[\theta])\} \quad (4.39)$$

then the distance to the plume can be consistently estimated, i.e.

$$|(\|x_D - \hat{b}_n\|_2 - \hat{h}_n) - (\|x_D - b_0\|_2 - h_0)| \in o_{\mathbb{P}}(1). \quad (4.40)$$

If further Eq. (4.39) is strengthened by replacing B_{∞} with $B_{\infty,0}$ then the shape is uniquely determined, i.e.

$$\|\hat{b}_n - b_0\|_2 + \|\hat{h}_n - h_0\|_2 \in o_{\mathbb{P}}(1). \quad (4.41)$$

2. Consider compact sets $\Xi_{\text{off}, \theta} \subseteq C(X)^2 \times C(X \times \mathbb{S}^2 \times \mathbb{S}^2)$, put $\Xi_{\text{off}} := \bigcup_{\theta \in \Theta} \Xi_{\text{off}, \theta}$ and

suppose \bar{B}_Θ is defined as

$$\bar{B}_\Theta = \left\{ \left(\theta, \frac{\mathcal{Q}[(\alpha[\theta], \xi_{\text{off}})]}{\mathcal{Q} \circ \mathcal{G}[\theta]} \right) : \theta \in \Theta \text{ and } \xi_{\text{off}} \in \Xi_{\text{off}, \theta} \right\}.$$

If \bar{B}_Θ and Ξ_{off} are compact, $\xi^0 \in \Xi_{\text{off}, \theta^*}$ and for all $\theta \in \Theta$ such that $(\sigma_a, \sigma_s, f_p) \in \Xi_{\text{off}, \theta}$ the functions σ_a and σ_s are proportional to $\alpha[\theta]$, all feasible f_p are strictly positive and $\sigma_s \neq 0$ then the estimator is consistent, i.e.

$$\|\hat{\theta}_n - \theta^*\|_2 \in o_{\mathbb{P}}(1). \quad (4.42)$$

Proof. Note that B_∞ is continuous w.r.t. the topology of uniform convergence in the β -component and \bar{B}_Θ is compact then, due to continuity and boundedness, this means the optimisation is valid and has well-defined solutions $\hat{\theta}_n$ in all instances. Note that feasible f_p are bounded below (for a θ -independent bound), or else a sequence in Ξ_{off} would have to converge to a phase function that is 0 somewhere which is not feasible by construction. As in Proposition 4.1.3, and writing $\Lambda_{\text{on+off}} = \Lambda_{\text{on}} + \Lambda_{\text{off}}$ we can show

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n B_\infty(\cdot \mid \mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}}) &\xrightarrow{n \rightarrow \infty} \int_{Z_{\llbracket - \rrbracket}(x_D)} \log \left(\frac{\mathfrak{q}[\theta, \beta']}{1 + \mathfrak{q}[\theta, \beta']} \right) \frac{d\Lambda_{\text{on}}}{d\Lambda_{\text{on+off}}} \\ &\quad + \log \left(\frac{1}{1 + \mathfrak{q}[\theta, \beta']} \right) \frac{d\Lambda_{\text{off}}}{d\Lambda_{\text{on+off}}} d\Lambda_{\text{on+off}} \end{aligned} \quad (4.43)$$

which is an expression maximised at the feasible point $(\theta^*, \beta(\cdot \mid \theta^*))$ and any other pair $(\bar{\theta}, \bar{\beta})$ such that $\bar{\beta}$ is a β -deviation of $\mathfrak{m}_{\text{on/off}}[\bar{\theta}]$ with respect to $\Lambda_{\text{on/off}}$ which means $\bar{\beta} \mathcal{Q} \circ \mathcal{G}[\bar{\theta}] = \mathcal{Q}[(\alpha[\theta^*], \xi_{\text{off}}^0)]$ at least on $\partial_{\alpha[\theta^*]}^{0,+} D$. Clearly, if $\bar{\theta}$ places $\alpha[\bar{\theta}]$ too far away then in a vicinity of $(v', \|x_D - b_0\|_2 - h_0) \in \partial_{\alpha[\theta^*]} D$ for the choice $v' = (b_0 - x_D)\|b_0 - x_D\|_2^{-1}$ we have $\bar{\beta} = 1$, because $\bar{\beta} \in B_\infty(\alpha[\bar{\theta}])$ yet at the same time any β -deviation must be less than 1 on $(v', \|x_D - b_0\|_2 - h_0 + \delta)$ for $\delta > 0$ sufficiently small. Similarly, if $\bar{\theta}$ places $\alpha[\bar{\theta}]$ too close then in a vicinity of $(v', \|x_D - b_0\|_2 - h_0) \in \partial_{\alpha[\theta^*]} D$ for $v' = (b_0 - x_D)\|b_0 - x_D\|_2^{-1}$

it must be true that for some $\delta > 0$

$$\frac{1}{1 + \mathbf{q}[\bar{\theta}, \bar{\beta}]} = \frac{1}{1 + \bar{\beta} \mathcal{Q} \circ \mathcal{G}[\bar{\theta}]} = \frac{1}{1 + \mathcal{Q}[(\alpha[\bar{\theta}], \bar{\xi}_{\text{off}})]} \geq \frac{1}{2} + \delta \quad (4.44)$$

and the third expression is only relevant in the second instance of the above theorem for some $\bar{\xi}_{\text{off}} \in \Xi_{\text{off}, \bar{\theta}}$. This is valid because all optical parameters are bounded due to compactness of Ξ and our condition on feasible σ_s , the bounds will determine δ . However, this must be equal to $1/2$ on $\partial_{\alpha[\theta^*]}^- D$ to be a minimiser. An application of Lemma 4.1.1 shows Eq. (4.40). Under the stronger assumption that $\beta \in B_{\infty, 0}(\alpha[\theta])$ we can apply the same reasoning at any $(v', t') \in \partial_{\alpha[\theta^*]} D$ which shows that any minimiser must match b_0 and h_0 . To obtain consistency in the w -component note that for any viable $\xi_{\text{off}} \in \Xi_{\text{off}, \bar{\theta}}$ where $\bar{\theta} = (\bar{w}, b_0, h_0)$ we have

$$\frac{1 - \mathcal{Q}[(\alpha[\bar{\theta}], \xi_{\text{off}})] \left(\frac{b_0 - x_D}{\|b_0 - x_D\|_2}, \|b_0 - x_D\|_2 - h_0 + t \right)}{2 \int_0^t \phi \left(1 - \frac{r}{h_0} \right) dr} \rightarrow \bar{w} \quad (4.45)$$

as $t \rightarrow 0$. The argument is identical to that in the proof of Lemma 3.1.2. But it must be true that $\bar{\beta} \mathcal{Q} \circ \mathcal{G}[\bar{\theta}] = \mathcal{Q}[(\alpha[\theta^*], \xi_{\text{off}}^0)]$ if $(\bar{\beta}, \bar{\theta})$ is an optimiser of Eq. (4.43) meaning $\bar{w} = w_0$ which, again after applying Lemma 4.1.1, shows Eq. (4.42). \square

Remark. For a fully parametric model $\Xi_{\text{off}, \theta} = \{(\sigma_s[\theta], \sigma_a[\theta], f_p[\theta])\} \implies \frac{\mathcal{Q}[(\alpha[\theta], \xi_{\text{off}})]}{\mathcal{Q} \circ \mathcal{G}[\theta]} = 1$ which leads to results like Proposition 4.1.3. The introduction of the β -parameter in the first instance is very similar in spirit to the relaxation from Eq. (3.12c) but with respect to the differential component instead of the absolute magnitude of the signal. There are several ways in which the above assumptions can be weakened/modified:

1. The differential absorption needs not be a single puff. More generally $\beta \in B_\infty(\alpha[\theta])$ and $\beta \in B_{\infty, 0}(\alpha[\theta])$ are sufficient to determine $\partial_{\alpha[\theta^*]}^+ D$ and $\partial_{\alpha[\theta^*]}^{0,+} D$ respectively. The former is simply the distance to the plume measured from the detector whereas the latter is its silhouette and determines the shape if no kernel function is obscured.
2. In order to obtain consistency one does not have to require that σ_s or σ_a are zero outside the support of α but it is enough that they are known outside, e.g. up to

an additive constant modelling ambient scattering. Consistency cannot be obtained exactly as in the proof but the same arguments that were used to show Lemma 3.1.2 and Theorem 3.1.4.

3. As seen in the proof, recovery of w hinges on a growth condition regarding $1 - \beta$ at $\partial_{\alpha[\theta^*]}^{0,+} D$, see Eq. (4.44) in the proof, more than anything else. For more general functions α this is closely related to the condition Eq. (3.14) used in the proof of the main uniqueness statement Theorem 3.1.4.
4. The compactness assumptions are not very restrictive, especially given all other requirements. In the most intuitive case where the non-differential optical parameters are defined through a compact (infinite-dimensional) set of “reference functions” which are shifted/scaled in accordance with $\alpha[\theta]$ (they have to be aligned by assumption), resulting in the sets $\Xi_{\text{off},\theta}$, all of the above assumptions will be met.

In general we cannot expect \sqrt{n} rates of convergence, i.e.

$$\sqrt{n} \|\hat{\theta}_n - \theta^*\|_2 \in O_{\mathbb{P}}(1) \quad (4.46)$$

as would be typical for parametric estimators or indeed (infinitesimally) narrow and by extension multiple FOVs. Any asymptotically normal estimator satisfies Eq. (4.46). To see why this may be the case we give a sketch of the idea. We may consider an estimator like Eq. (4.38) where the β -parameters are restricted by two functions $l_B(\cdot \mid \theta)$ and $u_B(\cdot \mid \theta)$ which are assumed to satisfy $l_B(v, t \mid \theta) < u_B(v, t \mid \theta)$ on $B_{\infty,0}(\alpha[\theta])$, equal otherwise, and we optimise over

$$\bar{B}_{\Theta} \subseteq \{(\theta, \beta) : \theta \in \Theta \text{ and } l_B(\cdot \mid \theta) \leq \beta \leq u_B(\cdot \mid \theta)\}. \quad (4.47)$$

Assuming the shape is fixed correctly and the β -deviation lies in the interior, then there is *some* influence by non-parametric factors at distances beyond the boundary of the absorber α . Disregarding compactness and other criteria related to when such estimators may or may not be well defined, it is clear that a sufficiently small change in α can be

fully accounted for by functions from Eq. (4.47) except maybe in a small vicinity around $\partial_{\alpha[\theta]} D$. If $u_B(\cdot, v \mid \theta) - l_B(\cdot, v \mid \theta)$ grows slowly then θ can potentially be consistent. In order to ensure Eq. (4.46) in a general setting, we cannot allow $l_B < u_B$ even though Corollary 4.1.7 and the results preceding it tell us that this isn't something that we can entirely avoid for wider FOVs. The issue lies in the fact that one can replace the data $\mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}}$ with arbitrarily small restrictions to sets surrounding $\partial_{\alpha[\theta^*]} D$ and still observe the same asymptotic behaviour as the estimator will eventually, i.e. for large enough n , be entirely driven by the observations in those regions as the remainder can be fully accounted for by β . This procedure can make the smallest possible constants in Eq. (4.46) arbitrarily large as those must depend on the expected number of points in the data for each j leading inevitably to contradictions.

However, Eq. (4.31) degenerates to a singleton and we can recover the best possible version of Eq. (4.46) for the detection problem as the following result shows. Note that if $\alpha = 0$ then there is no way to reconstruct any parameters related to the shape of the plume because all of them will result in identical differential absorption making them non-unique and so we only consider the weight component of α .

Proposition 4.2.5 (Detection rate). *For $\Theta = [0, w_{\max}] \times [h_{\min}, h_{\max}] \times M_X$, with $h_{\min} > 0$ and compact $M_X \subseteq X$. We assume PPP data but for vanishing differential absorption, i.e. $\alpha = 0$, and the map \mathcal{G} as in Theorem 4.2.4. As before we consider the estimator*

$$(\hat{\theta}_n, \hat{\xi}_{\text{off},n}) = \arg \max_{(\theta, \xi_{\text{off}}) \in \Theta \times \Xi_{\text{off}}} \frac{1}{n} \sum_{j=1}^n \mathsf{B}_\infty \left(\theta, \frac{\mathcal{Q}[(\alpha[\theta], \xi_{\text{off}})]}{\mathcal{Q} \circ \mathcal{G}[\theta]} \mid \mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}} \right)$$

and write $\hat{\theta}_n = (\hat{w}_n, \hat{h}_n, \hat{b}_n)$. If $Z_{[-]}(x_D)$ is convex, the true value σ_s is positive on M_X , Ξ_{off} is compact and sufficiently regular then $\hat{w}_n \in O_{\mathbb{P}}(n^{-1/2})$. For regularity, we require that B_∞ is twice continuously differentiable in not just the parameter but also the (v, t) -argument. It is sufficient that all optical parameters are in $C_c^k(X)$ and uniformly bounded w.r.t. the semi-norms from Eq. (2.1) and large enough n and k .

Remark. We don't need σ_s to be positive everywhere as long as we have restricted b_0 and h_0 suitably. This is necessary regardless of the model used given that α can be anything

as long as it only affects null sets of the intensity measure of our data $\mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}}$. In other words, we must avoid “hiding” absorption in a place where no light is scattered. The shape does not have to be known either, indeed we could have chosen $\alpha(x) = w_0 \mathbf{p}(x)$ and run the optimisation over probability densities \mathbf{p} . If any feasible sequence q has non-vanishing support the result remains valid. Convexity of $Z_{\llbracket - \rrbracket}(x_D)$ is required for the results in [127].

Proof. By construction, the maximiser exists and is well-defined. Write

$$\begin{aligned} D_{n,\infty}(\theta, \xi_{\text{off}}) &= \frac{\partial}{\partial w} \frac{1}{n} \sum_{j=1}^n B_\infty \left(\theta, \frac{\mathcal{Q}[(\alpha[\theta], \xi_{\text{off}})]}{\mathcal{Q} \circ \mathcal{G}[\theta]} \mid \mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}} \right) \\ H_{n,\infty}(\theta, \xi_{\text{off}}) &= \frac{\partial^2}{\partial w^2} \frac{1}{n} \sum_{j=1}^n B_\infty \left(\theta, \frac{\mathcal{Q}[(\alpha[\theta], \xi_{\text{off}})]}{\mathcal{Q} \circ \mathcal{G}[\theta]} \mid \mathbf{M}_j^{\text{on}}, \mathbf{M}_j^{\text{off}} \right) \end{aligned}$$

for the derivatives in the w -component which are well defined for sensible choices of \mathcal{G} and Ξ_{off} . Consider the event that $\{\hat{w}_n > 0\}$. On that event we have, due to the product structure of the feasible set, $0 = D_{n,\infty}(\hat{\theta}_n, \hat{\xi}_{\text{off},n})$. By Taylor’s Theorem, it must also be true for some $\bar{w}_n \in [0, \hat{w}_n]$ that

$$0 = D_{n,\infty}((0, \hat{h}_n, \hat{b}_n), \hat{\xi}_{\text{off},n}) + H_{n,\infty}((\bar{w}_n, \hat{h}_n, \hat{b}_n), \hat{\xi}_{\text{off},n})(\hat{w}_n - 0). \quad (4.48)$$

We claim that conditional on the total on- and off-counts $\mathbf{m}_\Sigma^{\text{on}} := \sum_{j=1}^n \mathbf{M}_j^{\text{on}}(Z_{\llbracket - \rrbracket}(x_D))$ and $\mathbf{m}_\Sigma^{\text{off}} := \sum_{j=1}^n \mathbf{M}_j^{\text{off}}(Z_{\llbracket - \rrbracket}(x_D))$

$$\mathbb{E} \left[\sup_{(h,b,\xi_{\text{off}}) \in [h_{\min}, h_{\max}] \times M_X \times \Xi_{\text{off}}} |D_{n,\infty}((0, h, b), \xi_{\text{off}})| \mid \mathbf{m}_\Sigma^{\text{on}}, \mathbf{m}_\Sigma^{\text{off}} \right] \leq c_1 \frac{\sqrt{\mathbf{m}_\Sigma^{\text{on}} + \mathbf{m}_\Sigma^{\text{off}}}}{n} \quad (4.49)$$

where $c_1 > 0$ is some constant depending only on the feasible set and the expectation is taken over the positional component of the PPP (see e.g. [32, Ch. 7] for details regarding such conditioning). This is true due to [127, Thm. 4] because Ξ_{off} is assumed to contain only sufficiently smooth functions and differentiability is obtained by the same arguments as those used in the proof of Proposition 4.1.6. For any choices (h, b, ξ_{off}) and vanishing

differential absorption we have

$$\mathcal{Q}[(0, \xi_{\text{off}})] = \mathcal{Q} \circ \mathcal{G}[(0, h, b)] = 1 \implies \mathbb{E} [\mathsf{D}_{n,\infty}((0, h, b), \xi_{\text{off}})] = 0.$$

After bounding the expectation of $\sqrt{\mathbf{m}_\Sigma^{\text{on}} + \mathbf{m}_\Sigma^{\text{off}}}$, e.g. with concentration inequalities such as [123, prop. 7], we obtain

$$|\mathsf{D}_{n,\infty}((0, \hat{h}_n, \hat{b}_n), \hat{\xi}_{\text{off},n})| \leq \sup_{(h,b,\xi_{\text{off}}) \in [h_{\min}, h_{\max}] \times M_X \times \Xi_{\text{off}}} |\mathsf{D}_{n,\infty}((0, h, b), \xi_{\text{off}})| \in O_{\mathbb{P}}(n^{-1/2})$$

which follows, after taking the expectation of Eq. (4.49), from a simple application of Markov's inequality or sharper bounds as can be found in [79].

In order to properly apply Theorem 4 in [127] we can associate the stochastic integrals in $\mathsf{D}_{n,\infty}$ conditional on $\mathbf{m}_\Sigma^{\text{on}}, \mathbf{m}_\Sigma^{\text{off}}$ as a single sample of size $\mathbf{m}_\Sigma^{\text{on}} + \mathbf{m}_\Sigma^{\text{off}}$ drawn i.i.d. in the form $(\mathbf{x}_k, \mathbf{r}_k) \in Z_{[-]}(x_D) \times [0, 1]$ where $\mathbf{x}_k \sim \Lambda_{\text{off}}/\Lambda_{\text{off}}(Z_{[-]}(x_D))$ and \mathbf{r}_k is independent of \mathbf{x}_k and distributed as independent Bernoulli(1/2). We then integrate

$$\frac{\partial}{\partial w} \log \left(\frac{\mathfrak{q}}{1+\mathfrak{q}} \right) f(r) + \log \left(\frac{1}{1+\mathfrak{q}} \right) f(1-r)$$

for a smooth function f such that $f(1) = 1 = 1 - f(0)$ which thereby becomes identical to the expression of interest and is of the right form for [127, Thm. 4]. Based on Eq. (4.48) we are left with

$$|\mathsf{H}_{n,\infty}((\bar{w}_n, \hat{h}_n, \hat{b}_n), \hat{\xi}_{\text{off},n})(\hat{w}_n - 0)| \in O_{\mathbb{P}}(n^{-1/2}).$$

To prove the claim it is enough to show that there is $\delta > 0$ such that

$$\liminf_{n \rightarrow \infty} \mathbb{P} \left(\mathsf{H}_{n,\infty}((\bar{w}_n, \hat{h}_n, \hat{b}_n), \hat{\xi}_{\text{off},n}) \geq \delta \right) = 1$$

which in essence replaces the non-singularity requirement in classical results for parametric estimators. Based on arguments as in Theorem 4.2.4 we know $\hat{w}_n \rightarrow 0$ in probability which

means $\bar{w}_n \rightarrow 0$ in probability. Since

$$\mathsf{H}_{n,\infty}((\bar{w}_n, \hat{h}_n, \hat{b}_n), \hat{\xi}_{\text{off},n}) \geq \inf_{(h,b,\xi_{\text{off}}) \in [h_{\min}, h_{\max}] \times M_X \times \Xi_{\text{off}}} \mathsf{H}_{n,\infty}((\bar{w}_n, h, b), \xi_{\text{off}})$$

and by the uniform law of large numbers and the continuity of the infimum (see [72, Lemma 1]) we know the quantity on the right converges in probability to

$$\begin{aligned} & \inf_{(h,b,\xi_{\text{off}}) \in [h_{\min}, h_{\max}] \times M_X \times \Xi_{\text{off}}} \mathsf{H}_{n,\infty}((\bar{w}_n, h, b), \xi_{\text{off}}) \\ & \xrightarrow{n \rightarrow \infty} \inf_{(h,b,\xi_{\text{off}}) \in [h_{\min}, h_{\max}] \times M_X \times \Xi_{\text{off}}} \mathbb{E}(\mathsf{H}_{1,\infty}((0, h, b), \xi_{\text{off}})) > 0 \end{aligned}$$

in probability. The conditions on the intensity Λ_{off} via σ_s and $h_{\min} > 0$ ensure that the expectation on the right is indeed strictly positive. \square

Note that this should not come as too much of a surprise given that we have Theorem 3.2.2 which states a less rigorous version of a similar result. The above is, however, part of a more general realisation in that the errors of an estimator $\hat{\theta}$ for a parameter θ can be thought of as

$$\alpha[\hat{\theta}] \approx \hat{\delta}_s \alpha[\theta] + \hat{\varepsilon} \quad (4.50)$$

where $1 - \hat{\delta}_s$ is a bounded *relative* error induced by uncertainty in the scattering parameters which needs not converge fast (or even at all) to 0 whereas $\hat{\varepsilon}$ is an error induced by optical noise and typically enjoys $n^{-1/2}$ rates of convergence. When the concentration is sufficiently large both terms matter. For the detection problem, which becomes the only viable option once $\|\hat{\varepsilon}\| \approx \|\alpha[\theta]\|$ (to be understood informally), then the optical noise term becomes dominant. In other words, $\alpha \rightarrow 0$ amplifies optical noise to the point where it becomes the only thing that influences the reconstruction.

Non-ergodic limits If instead of taking the limit in the detector response, i.e. let the pulse energy tend to infinity, then the optical error disappears instantly and we cannot take advantage of the ergodic properties of atmospheric dispersion. If our goal is to

reconstruct steady-state parameters then the turbulence is non-zero noise and we cannot expect to reconstruct the parameters exactly regardless of our measurement type. The measured value for the differential absorption then takes the form

$$\alpha[\theta](\cdot \mid \mathbf{T}) \approx \delta_{\mathbf{T}} \alpha[\theta] \quad (4.51)$$

for some $\delta_{\mathbf{T}}$ which is bounded in regions where α is bounded from below. This means that the best we can hope for realistically is qualitatively similar to what we saw in Eq. (4.50). The important distinction here is that turbulence is assumed to be spatially ergodic meaning that for $\text{vol}(S) \gg 0$ we have

$$\left| \frac{1}{\text{vol}(S)} \int_S \alpha[\theta](x \mid \mathbf{T}) - \alpha[\theta](x) dx \right| \ll \frac{1}{\text{vol}(S)} \int_S |\alpha[\theta](x \mid \mathbf{T}) - \alpha[\theta](x)| dx. \quad (4.52)$$

which is not necessarily the case for the scattering-induced error δ_s .

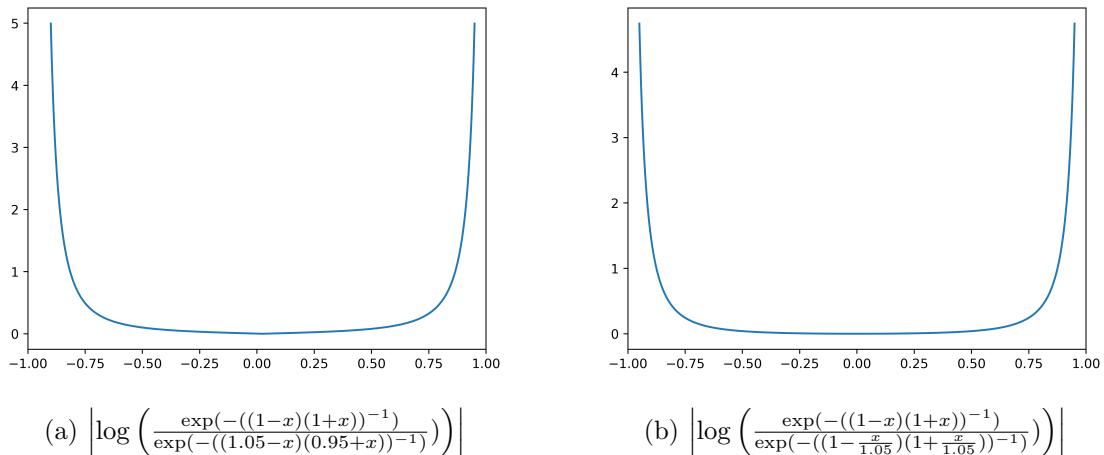


Figure 4.1: Relative errors for smooth compactly supported functions. The graphs indicate that small perturbations have a super-exponential effect at the support boundaries, which are not stable.

The observation that for a probability density p on $[-1, +1]$, e.g. a normalised good kernel, and a sequence $p_n \rightarrow p$ uniformly (or for transport problems more realistically in

a Wasserstein metric) we might have

$$0 = \liminf_{n \rightarrow \infty} \lim_{x \rightarrow \pm 1} \frac{p_n(x)}{p(x)} < \limsup_{n \rightarrow \infty} \lim_{x \rightarrow \pm 1} \frac{p_n(x)}{p(x)} = \infty \quad (4.53)$$

means that all *exact* reconstructions, which were intrinsically linked to the tail behaviour, are not possible anymore. in fact, the only aspects that one can reasonably expect to be preserved infinitely well in wider FOVs, are the ones arguably affected most by a turbulence-induced perturbation. Consequently, $1 - \hat{\delta}_s$ should in practically relevant situations certainly not be expected to approach 0 although Proposition 4.2.5 remains valid for the same reason as before.

Despite Eq. (4.52) it should be clear that realistically we cannot expect to reconstruct even the integral of α over the entire domain *exactly* since there will always be non-zero, or indeed infinite, averaging times required in order to obtain perfectly accurate measurements. More importantly, however, a narrow FOV measurement can suffer from the effect seen in Fig. 4.1 just as much, if not more so than a wide FOV measurement because it detects a larger ratio of light from the “edges”.

- If $\sigma_s \rightarrow \infty$ “faster” than the energy of a pulse, then the lack of signal at greater depths may result in narrow FOV reconstructions having potentially larger errors due to observations being concentrated in a region with small volume and Eq. (4.52) is exploited to a greater degree in wide FOVs whose averages suffer less from the instabilities such as Eq. (4.53).
- If σ_s isn’t too large, so that averages over a large region are taken into account, then Eq. (4.52) might be exploited to an extent where δ_s becomes the dominant effect.

From a purely mathematical point of view one can certainly design distributions for the turbulence along with optical parameters which will favour wide or narrow FOVs by construction and without a solid understanding of the underlying physical phenomena and mathematical equations it seems somewhat futile to make statements with the purpose of being realistic, rigorous and quantitatively precise.

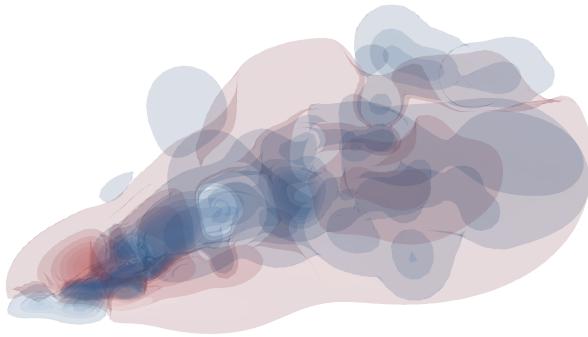


Figure 4.2: Qualitative structure of the turbulence: In the blue regions the turbulent plume has a higher concentration than its smooth counterpart, red indicates the opposite. If the plume is highly scattering and a pulse is released towards the gas from a location outside then the response will be much larger, at least initially, for directions “aiming” at the blue parts. The opposite is true for the red parts.

4.3 Modelling discrepancies & estimation bias

In order to paint a full picture of the situation at hand we must also consider the previously ignored nuisance component ψ as introduced by the relaxation of Eq. (3.12c) where it was essentially argued that having the free parameter ψ helps if we seek to fit a model with a small number of kernel functions.

Turbulence & model misspecification Recall that the single scattering returns from direction v at time $2t$, i.e. Eqs. (3.11a) and (3.11b), contain a factor $\sigma_s(x_D + vt)$. If we model the (true) scattering parameter $\sigma_s(\cdot \mid \mathbf{T})$ responsible for the observation results, which is an instantaneous snapshot of a turbulent plume, as a low-dimensional function $\sigma_s[\theta]$ then, in the region where photons are detected, the value for ψ that matches the low-dimensional $\mathfrak{m}_{\text{on/off}}[\theta]$ to the observed intensity

- may be unbounded on a set $U_\infty \subset Z_{[-]}(x_D)$
- may tend to 0 on a set $U_0 \subset Z_{[-]}(x_D)$
- and there is a “surface” such that ψ is bounded from above and below on the side that doesn’t contain x_D

In view of Eq. (4.53) this should not come as too much of a surprise and one would expect that $U_\infty \cup U_0 \approx \partial_{\sigma_s} D$ which forms the aforementioned “surface”. If we enforce Eq. (3.12c), i.e. require $\psi = 1$ then $\mathbf{m}_{\text{on/off}}$ will not be able to match the observed photon counts. If the distribution of the counts is assumed to be Poisson as in Eq. (3.15) then this results in a poor model fit. To give a more concrete example, it would be *extremely* unlikely to observe a close to zero count when the expected intensity is in the thousands as would be the case on U_0 . Estimation of θ would then effectively rely on U_0 being cancelled out by U_∞ so that on average the count is correct. These large discrepancies between fully parametric estimators, when the additional variance from model mismatch is ignored entirely, have been found (numerically) to result in low-variance estimators with large errors. In other words, large bias due to under-fitting. Ignoring incorrectly specified phase functions and assuming that $\sigma_s[\theta]$ will capture the average of $\sigma_s(\cdot \mid \mathbf{T})$, we may think of the issue described above as an underestimation of variance by the Poisson distribution. A popular remedy to such issues is replacing the Poisson distribution with a **NegativeBinomial** which is obtained by placing a **Gamma** distribution on the intensity and marginalising.

The heuristic decomposition of $Z_{[-]}(x_D)$ from above indicates that ψ can be virtually “anything” and we should not specify a constraint too strict, or **NegativeBinomial** with too small variance, so we avoid under-fitting. Indeed, if we don’t allow sufficiently many degrees of freedom in ψ then the additional degrees of freedom available due to the differential absorption component will be abused so that $\psi \mathbf{m}_{\text{on}} \approx \mathbf{M}^{\text{on}}$ instead of modelling the actual difference in absorption. This is especially true if the absorption is low, the modelling discrepancy is large and the counts are high. In that instance, misfit in the observed counts carries a much larger penalty than an error in the differential absorption. In such situations we won’t be able to enjoy results such as Proposition 4.2.5 unless we require for the estimate $\hat{\psi}$ that

$$\hat{\psi}(\mathbf{m}_{\text{on}} + \mathbf{m}_{\text{off}}) \approx \mathbf{M}^{\text{off}} + \mathbf{M}^{\text{on}} \iff \hat{\psi} \approx \frac{\mathbf{M}^{\text{off}} + \mathbf{M}^{\text{on}}}{\mathbf{m}_{\text{on}} + \mathbf{m}_{\text{off}}}. \quad (4.54)$$

The technicalities regarding the above approximation, where ψ would have to mimic a

measure, are not important for this work and, due to of the intricate nature of infinite-dimensional likelihoods, won't be discussed in detail and we refer the heuristic discussion leading to Eq. (4.28) for some intuition. If we had chosen a method based on over-dispersing the data, such as the **NegativeBinomial** approach, where the marginal is taken over a "flat" distribution which assumes what is perhaps best described as infinite variance in $\mathbf{M}^{\text{off}} + \mathbf{M}^{\text{on}}$, then this becomes equivalent to the proposed maximisation over ψ in the sense that both approaches result in a loss function like Eq. (4.30). This observation is at its core a consequence of Lemma 4.4.1.

It should be clear that the non-parametric components β and ψ have rather different effects on the likelihood. This should not come as too much of a surprise as, although they stem both from underlying uncertainties in the same optical quantities, after the introduction of ψ the remaining quantities all involve averages. Once the troublesome deviations in $\mathbf{M}^{\text{off}} + \mathbf{M}^{\text{on}}$ have been taken care of, one must make a choice regarding how much of the set Eq. (4.31) we want to model and in what way this should be done. The approach taken for ψ was direct, i.e. instead of using high-dimensional optical parameters we model the effect directly, and would correspond to using sets such as Eqs. (4.35) and (4.36). However, unlike with ψ , a closer inspection of the structural dependency, based on Proposition 4.1.6, reveals that "worst case behaviour" is attained if the scattering parameters are varied proportionally to the differential absorption field α . In fact, this was the idea behind the construction of the sequence that led to Corollary 4.1.7 and makes sense given that all variability is now due to α . The loss function is therefore primarily impacted by the intensity values in those areas that are subject to large differential absorption and consequently have been averaged over larger regions and benefit from Eq. (4.52), thus reducing the impact of turbulence-induced errors. More specifically, assuming that we have $\alpha \in K(\phi | x_D)$:

- Decreasing σ_s proportionally to α , i.e. setting $\alpha \propto \sigma_s$ and $\sigma_s \rightarrow 0$ degenerates the RTE to single scattering and in the process ensures $\frac{m_{\text{on}}}{m_{\text{off}}} < \frac{d\Lambda_{\text{on}}}{d\Lambda_{\text{off}}}$ at least on a subset of $Z_{[-]}(x_D)$
- Increasing σ_a proportionally to α , i.e. setting $\alpha \propto \sigma_a$ and $\sigma_a \rightarrow \infty$ achieves $\frac{m_{\text{on}}}{m_{\text{off}}} \rightarrow 1$

which is the largest possible value.

In other words, the smooth error component $\hat{\delta}_s$ from Eq. (4.50) can be modelled rather well by finite-dimensional parameters. From a computational perspective, this makes things rather convenient as a non-parametric approach to a set like Eq. (4.31), direct or implicitly through optical parameters, would pose challenges regarding discretisation as well as parameter fitting. As mentioned previously, instabilities such as Eq. (4.53) will arguably prevent a perfectly unbiased estimation either way, at least if multiple scattering isn't isolated by means of multiple FOVs, and assuming Eq. (4.52) it stands to reason that a purely parametric model will not suffer from much bigger biases in the quantities of interest than if we were to include a high dimensional set to capture the β -deviation accurately. This is also supported by our previous detection-related observation that trajectories are primarily affected by the distribution of scattering particles.

Discretisation & data collection errors In practice, it is not realistic to have data that fully resolve the domain $Z_{\llbracket - \rrbracket}(x_D)$ which can be thought of as the spherical coordinates of a cone, or segment thereof, in \mathbb{R}^3 corresponding to initial directions of the released light and time-of-flight of the detected photons. Ignoring scattering from surfaces (as we did throughout) we can expect that the PPP has an intensity measure which is continuous w.r.t. the Lebesgue measure on $Z_{\llbracket - \rrbracket}(x_D)$. As far as discretisation is concerned it therefore seems reasonable to consider a partition as induced by functions $\text{bin}_{i,j}^l$ from Eq. (4.27) of the domain $Z_{\llbracket - \rrbracket}(x_D)$ on which non-zero photon counts are expected but this is arguably not entirely appropriate as there are two quite distinct mechanisms responsible for the discrete nature of our data. It is more reasonable to assume that:

- The time-of-flight component of our data will be binned by the detector and aggregate all photons detected within an interval $t_i \pm \frac{\Delta t}{2}$ into a single count.
- The released laser beam is coherent with directions incremented discretely and therefore the angular component is singular rather than an average over all released directions.

The primary difference between temporal and angular discretisation is perhaps seen when it comes to single scattering. Any photon that is scattered once is entirely unaffected by anything that falls entirely in between “adjacent” release directions if directions are singular whereas a partitioning of $Z_{\llbracket - \rrbracket}(x_D)$ would include an average across all possible paths. This means that differential absorption can, at least in theory, “hide” between the pulses. From this observation alone it is evident that naive partitioning will result in a structurally different discretisation error than what one would have in practice. At short distances and/or large plumes this may not be an issue but given that atmospheric remote sensing problems can involve rather long distances between plume and detector, the angular discretisation error may not be negligible in practice. Naturally the feasible set Θ and $\mathcal{G} : \Theta \rightarrow K(\phi | x_D)$ should be chosen such that the angular discretisation is not a (major) concern which means that the plume should intersect multiple lines-of-sight which can be achieved by limiting the smallest admissible kernel width.

4.4 Uncertainty quantification for estimates in Θ

Although a parametric approach presents a reasonable way as far as point estimation of $\theta \in \Theta$ is concerned and can account for the effects on multiple scattering data caused by smooth deviations in the optical parameters, the matter of quantifying uncertainties is more intricate. Before we address this issue in more detail we start with an observation stating that the likelihood factors into a θ and a ψ dependent component which will make computations considerably easier down the line.

Lemma 4.4.1. *Let $\mathbf{X}_k \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda_X)$ and $\mathbf{Y}_k \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda_Y)$ be independent sequences of random variables. Then*

$$\sqrt{n} \begin{pmatrix} \lambda_X + \lambda_Y - \frac{1}{n} \sum_{k=1}^n \mathbf{X}_k + \mathbf{Y}_k \\ \frac{\lambda_X}{\lambda_X + \lambda_Y} - \frac{\sum_{k=1}^n \mathbf{X}_k}{\sum_{k=1}^n \mathbf{X}_k + \mathbf{Y}_k} \end{pmatrix} \xrightarrow{n \rightarrow \infty} \mathbf{Z} \quad (4.55)$$

where \mathbf{Z} is a Normal random vector with zero mean, non-degenerate variances and independent components.

Proof. The components are chosen such that they are maximum likelihood estimators for parameters $\lambda_Q = \frac{\lambda_X}{\lambda_X + \lambda_Y}$ estimators and $\lambda_S = \lambda_X + \lambda_Y$. The claim of the theorem is thus equivalent to the Fisher information matrix (with respect to a (ϕ, ψ) parameterised likelihood) being diagonal. This is true because the log-likelihood can be written as

$$\sum_{k=1}^n -\lambda_S + \log(\lambda_S)(\mathbf{X}_k + \mathbf{Y}_k) + \log(\lambda_Q)\mathbf{X}_k + \log(1 - \lambda_Q)\mathbf{Y}_k \quad (4.56)$$

from which Eq. (4.55) is easily seen to coincide with scaled and centred maximum likelihood estimators for parameters λ_S and λ_Q respectively. \square

Remark. Note that due to the infinite divisibility of the Poisson distribution the statement of Lemma 4.4.1 holds not only for sums but also large intensity parameters of a single observation. When the observations are not identically distributed but only independent, then the factorisation which was used in the proof is still valid.

The continuous case for PPP data $\mathbf{M}^{\text{on/off}}$ as it was considered in this chapter Lemma 4.4.1 admits an analogous factorisation as in Eq. (4.56). It can be understood as stating an estimate of the intensity measure for $\mathbf{M}^{\text{on}} + \mathbf{M}^{\text{off}}$ such as given by Eq. (4.54) will be asymptotically independent of the maximum likelihood estimator for the Radon-Nikodym derivative whose noisy counterpart is given by $\frac{d\mathbf{M}^{\text{on/off}}}{d\mathbf{M}^{\text{on}} + \mathbf{M}^{\text{off}}}$. In our case θ is only used to parameterise the Radon-Nikodym derivatives and uncertainties caused by the Poisson nature of the data can be quantified independently of the high dimensional nuisance parameter ψ which is far simpler as the following standard result indicates.

Lemma 4.4.2. *Let $\Theta \subseteq \mathbb{R}^{N_\theta}$ be compact and $\mathbf{x}_k \stackrel{\text{iid}}{\sim} \mathbf{p}_{\theta^*}$ for $k \in \mathbb{N}$ and some probability distribution \mathbf{p}_{θ^*} from a parametric family $(\mathbf{p}_\theta)_{\theta \in \Theta}$ with log-likelihood $L(\theta | \cdot)$. Assume that $\theta \mapsto \mathbf{p}_\theta$ is injective, $L(\theta | \cdot)$ twice continuously differentiable and θ^* lies in the interior of the feasible set Θ and denote the MLE as*

$$\hat{\theta} = \arg \max \frac{1}{n} \sum_{k=1}^n L(\theta | \mathbf{x}_k).$$

If $\mathbb{E}(\nabla_\theta^2 L(\theta^ | \mathbf{x}_1))$ is non-singular and $L(\theta | \mathbf{x}_1)$ is dominated in C^2 , i.e. the semi-norms*

Eq. (2.1) are dominated uniformly for all $\theta \in \Theta$ by an integrable function $g(\mathbf{x}_1)$, then

$$\sqrt{n}(\hat{\theta} - \theta^*) \rightarrow \text{Normal}(0, I_p^{-1}) \quad (4.57)$$

where $I_p = \mathbb{E}(-\nabla_\theta^2 \mathcal{L}(\theta^* | \mathbf{x}_1)) = \mathbb{V}(\nabla_\theta \mathcal{L}(\theta^* | \mathbf{x}_1))$ is the Fisher-Information matrix and \mathbb{V} denotes the (co-)variance operator.

Remark. It is not strictly necessary that p is the data likelihood for a result like this to be true. If suitable uniqueness, boundedness and differentiability requirements are fulfilled we can obtain

$$\sqrt{n}(\hat{\theta} - \bar{\theta}) \rightarrow \text{Normal}(0, \mathbb{E}(\nabla_\theta^2 \mathcal{L}(\bar{\theta} | \mathbf{x}_1))^{-1} \mathbb{V}(\nabla_\theta \mathcal{L}(\bar{\theta} | \mathbf{x}_1)) \mathbb{E}(\nabla_\theta^2 \mathcal{L}(\bar{\theta} | \mathbf{x}_1))^{-1})$$

where $\bar{\theta}$ is the maximiser of $\theta \mapsto \mathbb{E}(\mathcal{L}(\theta | \mathbf{x}_1))$. A special case where \mathcal{L} gets an additional quadratic penalty term is presented below. Eq. (4.57) yields confidence intervals once I_p is estimated by replacing the ground truth parameter with the MLE.

Proof. The proof can be found in most texts on introductory statistics, see e.g. [107, Thm. 3.1]. The requirement for the semi-norms $s_{n,m}(\mathcal{L}(\cdot | \mathbf{x}_1))$ is such that all limits and derivatives can be interchanged and the needed limits results are obtained from dominated convergence. \square

Given that we have gotten rid of ψ we are, ignoring the effects of turbulence-induced perturbations and assuming the data is indeed Poisson distributed, left with a situation much like is the case in Lemma 4.4.2 which justifies, at least to some degree, the use of a quadratic approximation and suggests a way to estimate uncertainties of a purely parametric estimator $\hat{\theta}$. If $n^{-1} \sum_{k=1}^n \mathcal{L}(\theta | \mathbf{x}_k)$ is augmented by a quadratic regularisation term, i.e. we consider

$$\tilde{\theta} = \arg \max \frac{1}{n} \sum_{k=1}^n \mathcal{L}(\theta | \mathbf{x}_k) - \frac{1}{2} (\theta - \theta^{\text{reg}})^\top V_{\text{reg}} (\theta - \theta^{\text{reg}}) \quad (4.58)$$

then we can make use of the approximation

$$\frac{1}{n} \sum_{k=1}^n \mathsf{L}(\theta \mid \mathbf{x}_k) \approx -\frac{1}{2}(\theta - \hat{\theta})^\top I_p(\theta - \hat{\theta}) \quad (4.59)$$

and put

$$\hat{\theta}^{\text{reg}} = (I_p + V_{\text{reg}})^{-1} V_{\text{reg}} \theta^{\text{reg}} + (I_p + V_{\text{reg}})^{-1} I_p \hat{\theta} \quad (4.60)$$

in order to obtain an approximation for the regularised loss function

$$\frac{1}{2}(\theta - \hat{\theta}^{\text{reg}})^\top [I_p + V_{\text{reg}}](\theta - \hat{\theta}^{\text{reg}}) + \text{const.} \quad (4.61)$$

where the constant term is independent of θ but may depend on the data and regularisation parameters. The quantity $\hat{\theta}^{\text{reg}}$ is an a priori estimate for $\tilde{\theta}$ based on $\hat{\theta}$ and, given the MLE, can be used to estimate the bias induced by a choice V_{reg} and θ^{reg} . We will make use of this later on when it comes to the selection of regularisation parameters. Further should take note that

$$\mathbb{V}(\tilde{\theta}) \approx (I_p + V_{\text{reg}})^{-1} I_p (I_p + V_{\text{reg}})^{-1} \quad (4.62)$$

which is a consequence of Lemma 4.4.2 and elementary results regarding linear transforms of random variables. The quality of Eqs. (4.60) to (4.62) is dependent on how close the likelihood is to a quadratic and whether or not an asymptotic expansion is appropriate which is to say will be more accurate for large and low-noise samples.

Contribution of turbulence The estimation of differential absorption is not without faults and shortcomings. If the problem is approached by looking for a low-dimensional parameter θ that determines the differential absorption and at the same time α is not static but driven by some (random) dynamical motion then the quantification of uncertainties becomes a much more intricate issue. It is arguably not even clear what in such a situation the “correct” estimate for the parameter θ is, let alone what the amount of uncertainty

would be. To get an idea of why this is the case we may look at the following toy example. Suppose

$$\alpha(x \mid \mathbf{T} = \mathbf{b}(t)) = \frac{e^{-\frac{1}{2}[(x_1 - \mathbf{b}_1(t))^2 + (x_2 - \mathbf{b}_2(t))^2 + (x_3 - \mathbf{b}_3(t))^2]}}{(2\pi)^{\frac{3}{2}}} \quad (4.63)$$

where $\mathbf{b}(t)$ is some a stochastic process such that $\mathbf{b}(t) = \mathbf{b}(s)$ if there is $n \in \mathbb{N}$ such that $t, s \in [n, n+1)$ and $\mathbf{b}(j)$ are independent standard Normal random variables for all $j \in \mathbb{N}$. In this instance, the long-term average would be

$$\alpha_0(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \alpha(x \mid \mathbf{T} = \mathbf{b}(t)) dt = \frac{1}{(4\pi)^{\frac{3}{2}}} e^{-\frac{\|x\|_2^2}{4}}$$

which is a Gaussian kernel with twice the variance, half of which is contributed by random jumps. If the reconstruction $\alpha[\theta]$ is parameterised by a single Gaussian kernel and assumed to be taken in an instant at time t , one would expect that the estimated centre $b[\hat{\theta}](t)$ is close to $\mathbf{b}(t)$, that is subject to uncertainty in the measurement. If a procedure such as Lemma 4.4.2, or any of the developments thereafter, is used in order to obtain (approximate) confidence intervals, then those give rise to bounds for $\mathbf{b}(t)$. In other words, the intervals are conditional on a realisation of the turbulent dynamics $\mathbf{b}(t)$ and will not overlap with the long-term centre at 0 with the same probability. There is in general no “correct” answer to the dilemma of whether the value $b[\hat{\theta}](t)$ is to be seen as an estimate of the realisation $\mathbf{b}(t)$ or rather its long-term average. This depends on whether one is interested in the current state or the long-term behaviour and consequently the same is true for the width of a potential confidence interval around $b[\hat{\theta}](t)$ which would have to be adjusted if the long-term average centre is the targeted parameter.

The above situation can easily be extended to changes in width, weight and more general parameters, which we shall denote by $\mathbf{T}(t)$, with analogous consequences. Where matters get more complicated is when, unlike with a jumping Gaussian kernel, the low-dimensional parameter and the resulting range $\{\alpha[\theta] : \theta \in \Theta\}$ cannot capture the instantaneous realisation $\alpha(x \mid \mathbf{T}(t))$. This, of course, is far more realistic than our previous toy example and if the instantaneous absorption $\alpha(x \mid \mathbf{T}(t))$ was the target then one would have to not

only decide in what sense $\alpha[\theta]$ should resemble $\alpha(x \mid \mathbf{T}(t))$ but modelling the resulting discrepancies between the reconstruction and target becomes impossible with Θ alone. As such, we also cannot represent a set of functions, i.e. a confidence region, with the purpose of containing $\alpha(x \mid \mathbf{T}(t))$. In short, if we restrict ourselves to Θ then uncertainty quantification is in a sense even less feasible than point estimation assuming the target is $\alpha(x \mid \mathbf{T}(t))$. This is not entirely surprising given that the utilisation of a low-dimensional set of parameters is inspired by the notably simpler structure of long-term averages.

Even though the function $\alpha(x \mid \mathbf{T}(t))$ cannot be fully captured by the low-dimensional parameters, point estimates are qualitatively unaffected by this because any estimate will be different from the conditional ground truth $\alpha(x \mid \mathbf{T}(t))$ due to (additional) noise. However, if we accept the premise that the target isn't the instant snapshot of turbulence but the long-term average, then the previously constructed asymptotic distribution is centred around the incorrect value. An adjustment of those intervals will, one way or another, require an input that encodes the behaviour of the turbulence process $\mathbf{T}(t)$ and its effect on $\alpha(x \mid \mathbf{T}(t))$. The distribution $\mathbb{P}(\mathbf{T}(t) \in \cdot)$ likely cannot be described by θ alone and at this point we have no ready solution to estimate additional parameters from the optical measurement that would allow us to do so. Nonetheless, we can proceed under the assumption that, by means of experience or otherwise obtained (atmospheric) data we have gained access to $\mathbb{P}(\mathbf{T}(t) \in \cdot)$ and are therefore able to evaluate expectations of $f(\alpha(x \mid \mathbf{T}(t)))$, at least approximately. This isn't entirely dissimilar to how plume models are typically set up in practice except that the atmospheric data is used as an input to obtain the low-dimensional structures themselves rather than their (random) deviations. We start with the array $\mathbf{y} = (\mathbf{y}_{i,j})_{(i,j) \in 1:N_v \times 1:N_t}$ of independent random variables as defined in Lemma 3.2.1 along with the approximation Eq. (3.27) which is valid in situations low-absorption. For sufficiently narrow FOVs and measurements taken at time s (which is unrelated to bin t_j) we can put

$$\mathbf{R}_{i,j}[\theta] = \frac{\int_0^{t_j/2} \alpha[\theta](x_D + rv_i) dr}{\int_0^{t_j/2} \alpha(x_D + rv_i \mid \mathbf{T}(s)) dr} \quad (4.64)$$

and use the single scattering approximation Eq. (3.27) leading to

$$\mathbb{E}(\mathbf{y}_{i,j} \mid \mathbf{T}(s)) = 2 \int_0^{t_j/2} \alpha(x_D + rv_i \mid \mathbf{T}(s)) dr = \frac{2}{\mathbf{R}_{i,j}[\theta]} \int_0^{t_j/2} \alpha[\theta](x_D + rv_i) dr$$

An application of the law of total variance [15, Thm. 9.5.4] would result in

$$\mathbb{V}(\mathbf{y}_{i,j} \mathbf{R}_{i,j}[\theta]) = \mathbb{E}[\mathbb{V}(\mathbf{y}_{i,j} \mathbf{R}_{i,j}[\theta] \mid \mathbf{T}(s))] + \mathbb{V}(\mathbb{E}[\mathbf{y}_{i,j} \mathbf{R}_{i,j}[\theta] \mid \mathbf{T}(s)]). \quad (4.65)$$

Strictly speaking both $\mathbf{y}_{i,j}$ and $\mathbf{R}_{i,j}$ are potentially infinite (or undefined) with positive probability. As such they aren't integrable, don't have a mean or variance and consequently we can neither take expectations nor does Eq. (4.65) apply. For $\mathbf{y}_{i,j}$ Lemma 3.2.1 gives a way to obtain asymptotic approximations for which such operations indeed make sense. If we assume that θ is a parameter such that $\alpha[\theta]$ is the long-term average of $\alpha(\cdot \mid \mathbf{T}(s))$ then

$$\int_0^{t_j/2} \alpha(x_D + rv_i \mid \mathbf{T}(s)) \rightarrow \int_0^{t_j/2} \alpha[\theta](x_D + rv_i) dr \implies \mathbf{R}_{i,j}^{-1}[\theta] \rightarrow 1 \quad (4.66)$$

and $\mathbf{R}_{i,j}[\theta] \rightarrow 1$ can be argued. While this can be used to make some sense of the first summand in Eq. (4.65) which would only involve computing expectations of $\mathbf{R}_{i,j}[\theta]$, without any assumption beyond Eq. (4.66) there is no way to give a reasonable interpretation of the variance of $\mathbf{R}_{i,j}[\theta]$. If $0 < \delta \leq \alpha(\cdot \mid \mathbf{T}(s))$ with probability one for some fixed δ , which would be the case if there is a non-zero amount of differential absorption in the ambient atmosphere, then $\mathbf{R}_{i,j}[\theta]$ is a bounded function of \mathbf{T} and the distribution of \mathbf{T} is all we need. Given that we are primarily interested in situations where a scattering agent is aligned with the plume, we can exploit that the variance of $\mathbf{y}_{i,j}$ will be large at locations where absorption tends to zero so that this idea, i.e. adding a small amount of ambient absorption, can be used to regularise $\mathbf{R}_{i,j}$ in cases where it may otherwise be problematic to evaluate.

Assuming that all variances exist at least in an asymptotic sense, we can extend Eq. (4.65)

to vectors via

$$\mathbb{V}(\mathbf{y} \odot \mathbf{R}[\theta]) \approx \text{diag}\left(\frac{2}{\mathbf{z}}\right) + \mathbf{y}\mathbf{y}^\top \odot \mathbb{V}(\mathbf{R}[\theta]) := I_{\mathsf{T}} \quad (4.67)$$

where \mathbf{z} is defined in Lemma 3.2.1 along with \mathbf{y} , \odot is entry-wise multiplication and the division is to be understood entry-wise as well. All arrays in Eq. (4.67) are assumed “flattened”, i.e. vectors with $N_v N_t$ many components rather than two-dimensional arrays. Note that this means $\mathbb{V}(\mathbf{y} \odot \mathbf{R}[\theta])$ as well as I_{T} are matrices whose rows and columns are indexed by $(i, j) \in 1:N_v \times 1:N_t$. By assumption, the data in each bin given an image is independent which translates to $\mathbf{y} | \mathbf{T}(s)$ having independent entries and is the reason behind the diagonal matrix in Eq. (4.67). In the second summand $\mathbb{E}(\mathbf{y} | \mathbf{T}(s)) \approx \mathbb{E}(\mathbf{y} \odot \mathbf{R}) \approx \mathbf{y}$ was used, which is the observed value. Of course, other approximations are possible, e.g. by using an estimate $\hat{\theta}$ of θ and

$$\mathbb{E}(\mathbf{y}_{i,j} | \mathbf{T}(s)) \approx \mathbb{E}(\mathbf{y}_{i,j} \mathbf{R}_{i,j}) \approx 2 \int_0^{t_j/2} \alpha[\hat{\theta}](x_D + rv_i) dr. \quad (4.68)$$

Lastly, we use the (approximate) log-likelihood of the unconditional distribution to obtain

$$\sqrt{n}(\hat{\theta} - \theta) \sim \text{Normal}(0, \nabla_\theta g(\theta)^\top I_{\mathsf{T},0} \nabla_\theta g(\theta)) \quad (4.69)$$

where $I_{\mathsf{T},0}$ is as Eq. (4.67) but with Eq. (4.68) in place of \mathbf{y} . Here $g : \Theta \rightarrow [0, \infty)^{N_v N_t}$ was also “flattened” into a vector and is defined as

$$g_{i,j}(\theta) = 2 \int_0^{t_j/2} \alpha[\theta](x_D + rv_i) dr$$

and $\nabla_\theta g(\theta)$ is thus a matrix with rows index by $(i, j) \in 1:N_v \times 1:N_t$ and N_θ columns, i.e. the classical Jacobian matrix. Adjusted confidence bands can then easily be obtained by using the normal distribution in Eq. (4.69) but replacing θ with the MLE $\hat{\theta}$.

For non-turbulent images Eq. (4.69) is slightly different from the covariance matrix that would be obtained with Eq. (4.57). The difference stems primarily from (extensive) use of the delta-method which preserves asymptotic properties, see e.g. [34, Sec. 2.2.2]. In

practice, i.e. numerically, we found little to no difference between the two estimates. The main motivation behind the use of Eq. (4.69) is that the turbulent perturbations appear more directly when a single scattering approximation is used. This is a nice property to have given that it is by no means a simple task to estimate quantities such as $\nabla(\mathbf{R}[\theta])$ even if we have access to a distribution for the turbulence process \mathbf{T} .

Chapter 5

Algorithms & experiments

5.1 Computational modelling and reconstruction

5.1.1 Variational form for regularised DIAL

As it turns out, the relaxed semi-parametric form, when paired with an expression as in Eq. (3.17), becomes very convenient for optimisation purposes. Consider the loss function

$$\begin{aligned} \text{Loss}(\psi, \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}) = & \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} (\mathfrak{m}_{\text{on}}[\theta](v_i, t_j) + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j)) \psi_{i,j} \\ & - \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \mathbf{m}_{i,j}^{\text{on}} \log(\mathfrak{m}_{\text{on}}[\theta](v_i, t_j)) - \mathbf{m}_{i,j}^{\text{off}} \log(\mathfrak{m}_{\text{off}}[\theta](v_i, t_j)) \quad (5.1) \\ & - \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} (\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}}) \log(\psi_{i,j}) + R(\theta, \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}) \end{aligned}$$

for ψ free which can therefore be used to “absorb” the known system parameter $H_D \Delta t A_D$ formed as a product of pulse strength H_D bin width Δt and detection area A_D . It is the sum of Eq. (3.17) and a regularisation term R , which will be discussed in greater detail later on, that, alongside \mathcal{G} from Eq. (3.10), enforces dispersion-based constraints such as alignment according to the wind direction or continuity and smoothness constraints (i.e. no “holes” or “corners”) in between the kernel components.

Optimisation of Eq. (5.1) can now be carried out sequentially. Note that R by assumption

only depends on θ and (possibly) the data so optimising in the nuisance parameter ψ for fixed θ , i.e. taking

$$\psi^*(\theta) = \arg \min_{\psi} \text{Loss}(\theta, \psi \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}),$$

it is easily seen that we have

$$\psi_{i,j}^*(\theta) = \frac{\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}}}{\mathfrak{m}_{\text{on}}[\theta](v_i, t_j) + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j)}. \quad (5.2)$$

We may plug $\psi_{i,j}^*(\theta)$ into Eq. (5.1) which yields, after rearranging and removal of additive constants, the expression

$$\begin{aligned} \text{Loss}(\theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}; \psi_{i,j}^*) &= \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} -\mathbf{m}_{i,j}^{\text{on}} \log \left(\frac{\mathfrak{m}_{\text{on}}[\theta](v_i, t_j)}{\mathfrak{m}_{\text{on}}[\theta](v_i, t_j) + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j)} \right) \\ &\quad + \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} -\mathbf{m}_{i,j}^{\text{off}} \log \left(\frac{\mathfrak{m}_{\text{off}}[\theta](v_i, t_j)}{\mathfrak{m}_{\text{on}}[\theta](v_i, t_j) + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j)} \right) \\ &\quad + \mathsf{R}(\theta, \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}) \end{aligned} \quad (5.3)$$

which has essentially the same structure as the negative log-likelihood of a collection of binomial random variables with trial lengths $\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}$ and probabilities

$$P_{i,j}(\theta) = \frac{\mathfrak{m}_{\text{on}}[\theta](v_i, t_j)}{\mathfrak{m}_{\text{on}}[\theta](v_i, t_j) + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j)}. \quad (5.4)$$

As such Eq. (5.3) may be optimised by what can be regarded as Fisher-Scoring for a binomial distribution which will avoid computation of second derivatives. In other words, we iterate

$$\theta^{(k+1)} = \theta^{(k)} - \zeta \mathbf{H}(\theta^{(k)})^{-1} \nabla_{\theta} \text{Loss}(\theta^{(k)} \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}; \psi_{i,j}^*) \quad (5.5)$$

where ζ is a step size and $H(\theta)$ is given by

$$H(\theta) = \nabla_\theta^2 R(\theta) + \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} (\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}) \frac{\nabla_\theta P_{i,j}(\theta) \nabla_\theta P_{i,j}(\theta)^\top}{P_{i,j}(\theta)(1 - P_{i,j}(\theta))} \quad (5.6)$$

Ambient noise term The expression in Eq. (5.1) is not appropriate when there is noticeable ambient illumination present in the detected photon counts. We note that a pulsed laser releasing pulses of duration in the nanosecond range, which is necessary if it is supposed to resolve objects in the order of a few meters within a kilometer radius, “is inactive” most of the time unless the pulsing frequency is multiple kHz. Indeed one kilometer is roughly 3.3×10^{-6} light-seconds which in effect means that after a few microseconds, it is rather unlikely to observe a photon that was actively released from the controlled light source. The detector will therefore be exposed to primarily ambient illumination unless a pulse is released immediately after light from the previous was collected. At least with currently available technology this pulsing frequency is rather unrealistic for a laser of sufficient strength for an application such as DIAL and consequently the detector can, at least in theory, be used to measure the intensity of ambient light during those periods where active counts are extremely unlikely to occur. If we assume that the odds of observing a large number of ambient photons during the inactivity period are high, i.e. the ambient intensity isn’t very low (in which case we can arguably neglect it and use Eq. (5.1) directly), then we can effectively assume that the expected number of ambient photons is *known* a priori and needs no estimation. A reasonable assumption for the distribution of ambient photons is that they are **Poisson** distributed and, given the respective intensities, stochastically independent of the photons released from the laser.

This means that we have for the observed data $\mathbf{m}_{i,j}^{\text{on}}$ and $\mathbf{m}_{i,j}^{\text{off}}$

$$\begin{aligned} \mathbf{m}_{i,j}^{\text{on}} &\sim \text{Poisson}(H_a + \mathbf{m}_{\text{on}}(v_i, t_j)\psi_{i,j}) \\ \mathbf{m}_{i,j}^{\text{off}} &\sim \text{Poisson}(H_a + \mathbf{m}_{\text{off}}(v_i, t_j)\psi_{i,j}) \end{aligned} \quad (5.7)$$

for some $H_a > 0$. The resulting likelihood-based loss function takes the form

$$\begin{aligned} \text{Loss}(\psi, \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) = & \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} (\mathfrak{m}_{\text{on}}[\theta](v_i, t_j) + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j)) \psi_{i,j} \\ & - \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \mathbf{m}_{i,j}^{\text{on}} \log(H_a + \mathfrak{m}_{\text{on}}[\theta](v_i, t_j) \psi_{i,j}) \\ & - \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \mathbf{m}_{i,j}^{\text{off}} \log(H_a + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j) \psi_{i,j}) \\ & + \mathsf{R}(\theta, \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}) \end{aligned} \quad (5.8)$$

and $\psi \mapsto \text{Loss}(\psi, \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a)$ is still convex but considerably more tedious to minimise in closed form. Note that $\psi_{i,j} > 0$ is enforced ‘‘automatically’’ in Eq. (5.1) but $\psi_{i,j} \geq 0$ does not have to hold in order for Eq. (5.8) to be well defined. One may, of course, add such a constraint without making the resulting optimisation problem a whole lot more complicated but despite being a more consistent choice in the sense that no reasonable physical model would suggest less light being observed in the presence of an active light source and a scattering medium that deflects the released signal towards the detector, there is a strong *statistical* argument against such a constraint.

In general we may consider the optimisation problem subject to θ -dependent constraint $\psi \in \Psi(\theta) \supseteq [0, \infty)^{N_v \times N_t}$, ideally such that $\Psi(\theta)$ is convex for all θ . Proceeding in the usual fashion by taking derivatives w.r.t. $\psi_{i,j}$ we obtain

$$\begin{aligned} \frac{\partial}{\partial \psi_{i,j}} \text{Loss}(\psi, \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) = & \mathfrak{m}_{\text{on}}[\theta](v_i, t_j) + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j) \\ & - \frac{\mathfrak{m}_{\text{on}}[\theta](v_i, t_j) \mathbf{m}_{i,j}^{\text{on}}}{H_a + \mathfrak{m}_{\text{on}}[\theta](v_i, t_j) \psi_{i,j}} - \frac{\mathfrak{m}_{\text{off}}[\theta](v_i, t_j) \mathbf{m}_{i,j}^{\text{off}}}{H_a + \mathfrak{m}_{\text{off}}[\theta](v_i, t_j) \psi_{i,j}}. \end{aligned} \quad (5.9)$$

If as before we denote by $\psi_{i,j}^*(\theta)$ the global optimum of Eq. (5.8) subject to constraints,

then it is easily seen that $\psi_{i,j}^*(\theta) > 0$ if and only if

$$0 > \frac{\partial}{\partial \psi_{i,j}} \text{Loss}(0, \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) = \mathbf{m}_{\text{on}}[\theta](v_i, t_j) + \mathbf{m}_{\text{off}}[\theta](v_i, t_j) - \frac{\mathbf{m}_{\text{on}}[\theta](v_i, t_j) \mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{\text{off}}[\theta](v_i, t_j) \mathbf{m}_{i,j}^{\text{off}}}{H_a}. \quad (5.10)$$

which clearly depends on the variable θ of interest and, for a non-trivial closed set such as $\Psi = [0, \infty)^{N_v \times N_t}$ the function $\psi_{i,j}^*(\theta)$ is therefore not necessarily differentiable in θ (even if $\mathbf{m}_{\text{on/off}}[\theta](v_i, t_j)$ are differentiable) and iterations such as Eq. (5.5) are potentially problematic. A generally viable approach when Ψ does not depend on θ can be implemented by fixing feasible initial values ψ^0, θ^0 and sequentially updating

$$\psi^{k+1} = \arg \min_{\psi \in \Psi} \text{Loss}(\psi, \theta^k \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) \quad (5.11)$$

$$\theta^{k+1} = \arg \min_{\theta \in \Theta} \text{Loss}(\psi^{k+1}, \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a). \quad (5.12)$$

Note that for sets of the form $[a, \infty)^{N_v \times N_t}$, with $a \leq 0$, a solution to Eq. (5.11) can be found in closed form by using that either ψ^{k+1} solves the quadratic equation associated with the system of equations

$$\frac{\partial}{\partial \psi_{i,j}} \text{Loss}(\psi_{i,j}^{k+1}, \theta^k \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) = 0,$$

or, if a solution violates the constraint, using the fact that $\psi_{i,j}$ will be feasible if and only if the equivalent of Eq. (5.10) holds at the lower bound in which case the minima are attained at the lower bound. The problems Eq. (5.12) are low dimensional and can be solved via iterations such as Eq. (5.5).

Smooth approximation for low absorption Although generally applicable and guaranteed to converge to a local minimum, any sequential procedure akin to Eqs. (5.11) and (5.12) can quickly become expensive. Solving for θ^{t+1} is structurally similar to minimisation of Eq. (5.3) which, in the absence of ambient noise, needs to be performed only once and thus finds the optimum without the need for multiple optimisations w.r.t. θ .

These iterations can be costly as, although low-dimensional, they involve evaluations of the RTE as well as associated θ -derivatives. This begs the question of whether we can develop a similarly efficient method in the presence of ambient noise.

If we were to replace $\mathbf{m}_{i,j}^{\text{on}}$ and $\mathbf{m}_{i,j}^{\text{off}}$ by “corrected” values $\mathbf{m}_{i,j}^{\text{on}} - H_a$ and $\mathbf{m}_{i,j}^{\text{off}} - H_a$ then, although an unbiased procedure, we may end up with negative photon counts. This in turn means that the loss may become concave in ψ and unbounded from *below*, especially if H_a is relatively large. We will show that this is indeed possible and fix the aforementioned issue later on by means of an adjusted loss function. To emphasise that care must be taken with such an approach we can consider an intuitive “solution” that applies a threshold to the data, e.g. take

$$\begin{aligned}\bar{\mathbf{m}}_{i,j}^{\text{on}} &= \max(\mathbf{m}_{i,j}^{\text{on}} - H_a, 0) \\ \bar{\mathbf{m}}_{i,j}^{\text{off}} &= \max(\mathbf{m}_{i,j}^{\text{off}} - H_a, 0),\end{aligned}\tag{5.13}$$

and replace \mathbf{m}^{on} and \mathbf{m}^{off} by $\bar{\mathbf{m}}^{\text{on}}$ and $\bar{\mathbf{m}}^{\text{off}}$ respectively and continuing with the equivalent of Eq. (5.5) in order to minimise a modified version of Eq. (5.3).

Before we proceed with an explanation as to why these types of strategies that involve thresholding are usually not going to perform well, we want to recall the low-absorption assumption

$$\left| \frac{\mathbf{m}_{\text{on}} - \mathbf{m}_{\text{off}}}{\mathbf{m}_{\text{on}} + \mathbf{m}_{\text{off}}} \right| \ll 1.\tag{5.14}$$

Such an assumption is easily verified by checking whether $\sum_{i,j} \mathbf{m}_{i,j}^{\text{on}} \approx \sum_{i,j} \mathbf{m}_{i,j}^{\text{off}}$. Instead of solving $\frac{\partial}{\partial \psi_{i,j}} \text{Loss}(\psi, \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) = 0$ we may consider, again subject to any possible constraint on ψ ,

$$0 = \mathbf{m}_{\text{on}}[\theta](v_i, t_j) + \mathbf{m}_{\text{off}}[\theta](v_i, t_j) - \frac{(\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}})(\mathbf{m}_{\text{off}} + \mathbf{m}_{\text{off}})[\theta](v_i, t_j)}{2H_a + (\mathbf{m}_{\text{off}} + \mathbf{m}_{\text{off}})[\theta](v_i, t_j)\psi_{i,j}}\tag{5.15}$$

where we used that under the condition set in Eq. (5.14) $\mathbf{m}_{\text{off}} \approx \frac{1}{2}(\mathbf{m}_{\text{off}} + \mathbf{m}_{\text{off}}) \approx \mathbf{m}_{\text{on}}$ is a

good approximation. The solution, subject to $\Psi = [0, \infty)^{N_v \times N_t}$, of Eq. (5.15) is given by

$$\psi_{i,j}^{\#,0}(\theta) = \begin{cases} 0 & \text{if } \frac{\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}}}{2} \leq H_a \\ \frac{\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}} - 2H_a}{\mathbf{m}_{\text{off}}[\theta](v_i, t_j) + \mathbf{m}_{\text{on}}[\theta](v_i, t_j)} & \text{if } \frac{\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}}}{2} > H_a \end{cases} \quad (5.16)$$

and is guaranteed to behave just like $\psi_{i,j}^*(\theta)$ from Eq. (5.2) but can be used in the situation of non-negligible ambient intensities. If we were to drop the non-negativity requirement $\Psi = [0, \infty)^{N_v \times N_t}$ then the solutions reduce to the second line, i.e.

$$\psi_{i,j}^\#(\theta) = \frac{\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}} - 2H_a}{\mathbf{m}_{\text{off}}[\theta](v_i, t_j) + \mathbf{m}_{\text{on}}[\theta](v_i, t_j)}. \quad (5.17)$$

Before we proceed with a performance comparison of $\psi_{i,j}^\#$, $\psi_{i,j}^{\#,0}$ and Eq. (5.13) based approaches we note that if in Eq. (5.1) we consider the parameter $(\mathbf{m}_{\text{on}} + \mathbf{m}_{\text{off}})\psi$ then from Lemma 4.4.1 it is easily seen that the maximum likelihood estimator for θ is asymptotically independent of that parameter owing to the fact that it is estimated based on $P_{i,j}(\theta)$ as defined in Eq. (5.4). Such approximate independence guaranteed is a very desirable property given that we consider $\psi_{i,j}$ a nuisance parameter which should ideally have no effect, and in particular not spoil, any estimate for the quantities of interest related to absorption. If we were to plug capped values such as those obtained from Eq. (5.13) then not only is the independence violated but, depending on how limits are taken, it may even fail to be consistent in situations that otherwise enjoy good convergence rates thanks to its ability to be asymptotically biased. Indeed, for $X \sim \text{Normal}(0, \tau)$ we know

$$\mathbb{E}(X \mid X \geq 0) = O(\sqrt{\tau})$$

which would suggest that thresholding effectively increases both online and offline counts and thus leads to an underestimation of differential absorption and if $\psi\mathbf{m}_{\text{off}} \ll \sqrt{H_a}$ the relative bias will approach 100% regardless of photon counts! This of course should immediately disqualify it from consideration. The estimator as in Eq. (5.16), which enforces non-negativity of the nuisance parameter, suffers from a similar issue. Instead of apply-

ing a threshold component-wise, it does the same to the sum of online and offline signals which can equally result in poor performance.

The estimators $\psi_{i,j}^\#(\theta)$ are, much like its counterpart $\psi_{i,j}^*(\theta)$ from Eq. (5.2), a linear statistic of $\mathbf{m}^{\text{on}} + \mathbf{m}^{\text{off}}$ and thus preserve some of the statistical properties even though the factorisation used to obtain the independence is not valid anymore. Assuming that likelihood-based estimation of θ through $P_{i,j}(\theta)$ is desirable we seek to find a closely related and well-behaving surrogate that can be used in place of the likelihood. To that end first note that $\mathbf{m}^{\text{off}} - \mathbf{m}^{\text{on}}$ has independent components indexed by $(i, j) \in 1:N_t \times 1:N_v$ such that

$$\mathbb{E}(\mathbf{m}_{i,j}^{\text{off}} - \mathbf{m}_{i,j}^{\text{on}}) = (\mathbf{m}_{\text{off}}[\theta](v_i, t_j) - \mathbf{m}_{\text{on}}[\theta](v_i, t_j))\psi_{i,j}$$

and for large enough \mathbf{m}_{off} and \mathbf{m}_{on} has approximately a **Normal** distribution whose variance can be estimated, and rather accurately so, by $\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}}$. This yields

$$\sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \frac{([\mathbf{m}_{i,j}^{\text{off}} - \mathbf{m}_{i,j}^{\text{on}}] - \psi_{i,j}[\mathbf{m}_{\text{off}}[\theta](v_i, t_j) - \mathbf{m}_{\text{on}}[\theta](v_i, t_j)])^2}{2(\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}})}.$$

Plugging in the small absorption estimator $\psi_{i,j}^\#$ from Eq. (5.17) leads to surrogate loss function (ignoring R which is unaffected) of the form

$$\sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \frac{([\mathbf{m}_{i,j}^{\text{off}} - \mathbf{m}_{i,j}^{\text{on}}] - (1 - 2P_{i,j}(\theta))[\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}} - 2H_a])^2}{2(\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}})} \quad (5.18)$$

which mimics the likelihood and is appropriately bounded for any $H_a \geq 0$. Dividing both the numerator and denominator in Eq. (5.18) by $(\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}})^2$ shows that the introduction of $H_a > 0$ affects $\frac{\mathbf{m}_{i,j}^{\text{off}} - \mathbf{m}_{i,j}^{\text{on}}}{\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}}}$ and in turn the estimator for θ via the term $\frac{H_a}{\mathbf{m}_{i,j}^{\text{on}} + \mathbf{m}_{i,j}^{\text{off}}}$ whose typical deviation is of the order $H_a^{-1/2}$ or smaller and, unlike thresholding based approach, is (virtually) free of bias at least for large photon counts. Unless we have $\psi \mathbf{m}_{\text{off}} \gg \sqrt{H_a}$ this will, as is to be expected, have an effect on the estimation but on the other hand effects of that magnitude must be accepted as they would be encountered even in an idealised hypothetical scenario where the ground truth for ψ was provided by

an oracle. If we defined the ‘‘corrected’’ data as

$$\mathbf{m}_{i,j}^{\text{on},\#} = \mathbf{m}_{i,j}^{\text{on}} - H_a \quad \text{and} \quad \mathbf{m}_{i,j}^{\text{off},\#} = \mathbf{m}_{i,j}^{\text{off}} - H_a \quad (5.19)$$

then computing the derivatives approximates the gradient of Eq. (5.8) via

$$\sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \nabla_{\theta} P_{i,j}(\theta) \frac{\mathbf{m}_{i,j}^{\text{on},\#} + \mathbf{m}_{i,j}^{\text{off},\#}}{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}} (4P_{i,j}(\theta)\mathbf{m}_{i,j}^{\text{off},\#} - 4(1 - P_{i,j}(\theta))\mathbf{m}_{i,j}^{\text{on},\#})$$

while the expected value of the second derivative conditional on the sum $\mathbf{m}^{\text{off}} + \mathbf{m}^{\text{on}}$ takes the form

$$\sum_{i=1}^{N_v} \sum_{j=1}^{N_t} 4 \frac{(\mathbf{m}_{i,j}^{\text{off},\#} + \mathbf{m}_{i,j}^{\text{on},\#})^2}{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}} \nabla_{\theta} P_{i,j}(\theta) \nabla_{\theta} P_{i,j}(\theta)^{\top}$$

and can be taken as an approximate Hessian. We claim that Eq. (5.18) can be optimised with a modification of Eq. (5.5). More specifically we may iterate

$$\theta^{(k+1)} = \theta^{(k)} - \zeta \mathbf{H}^{\#}(\theta^{(k)})^{-1} \mathbf{D}^{\#}(\theta^{(k)}) \quad (5.20)$$

where ζ is as before a step size, $\mathbf{H}^{\#}(\theta)$ is given by the positive semi-definite matrix

$$\mathbf{H}^{\#}(\theta) = \nabla_{\theta}^2 \mathbf{R}(\theta) + \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} 4 \frac{(\mathbf{m}_{i,j}^{\text{off},\#} + \mathbf{m}_{i,j}^{\text{on},\#})^2}{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}} \nabla_{\theta} P_{i,j}(\theta) \nabla_{\theta} P_{i,j}(\theta)^{\top} \quad (5.21)$$

$$\approx \nabla_{\theta}^2 \mathbf{R}(\theta) + \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \frac{(\mathbf{m}_{i,j}^{\text{off},\#} + \mathbf{m}_{i,j}^{\text{on},\#})^2}{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}} \frac{\nabla_{\theta} P_{i,j}(\theta) \nabla_{\theta} P_{i,j}(\theta)^{\top}}{P_{i,j}(\theta)(1 - P_{i,j}(\theta))} \quad (5.22)$$

while the gradient can be approximated via

$$\mathbf{D}^{\#}(\theta) = \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \nabla_{\theta} P_{i,j}(\theta) \frac{\mathbf{m}_{i,j}^{\text{off},\#} + \mathbf{m}_{i,j}^{\text{on},\#}}{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}} (4P_{i,j}(\theta)\mathbf{m}_{i,j}^{\text{off},\#} - 4(1 - P_{i,j}(\theta))\mathbf{m}_{i,j}^{\text{on},\#}) \quad (5.23)$$

$$+ \nabla_{\theta} \mathbf{R}(\theta) \\ \approx \sum_{i=1}^{N_v} \sum_{j=1}^{N_t} \nabla_{\theta} P_{i,j}(\theta) \frac{\mathbf{m}_{i,j}^{\text{off},\#} + \mathbf{m}_{i,j}^{\text{on},\#}}{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}} \left(\frac{\mathbf{m}_{i,j}^{\text{off},\#}}{1 - P_{i,j}(\theta)} - \frac{\mathbf{m}_{i,j}^{\text{on},\#}}{P_{i,j}(\theta)} \right) + \nabla_{\theta} \mathbf{R}(\theta). \quad (5.24)$$

Note that Eqs. (5.21) and (5.23) contain derivatives of the Gaussian approximation (which is bounded above and well behaved) whereas Eqs. (5.22) and (5.24) are consistent with Eq. (5.5) in the sense that for $H_a \rightarrow 0$ they are equivalent. Eqs. (5.22) and (5.24) are obtained from Eqs. (5.21) and (5.23) with a Taylor expansion at $\frac{1}{2}$

$$p^{-1} \approx 2 - 4 \left(p - \frac{1}{2} \right) = 4(1-p) \quad \text{and} \quad (1-p)^{-1} \approx 2 + 4 \left(p - \frac{1}{2} \right) = 4p$$

for the $\mathbf{m}_{i,j}^{\text{off},\#}$ and $\mathbf{m}_{i,j}^{\text{on},\#}$ terms respectively. The resulting scheme is similar to replacing $\mathbf{m}_{i,j}^{\text{off}}$ and $\mathbf{m}_{i,j}^{\text{on}}$ with quantities that were “corrected” for ambient noise but it takes into account the resulting increase in variance through the weighting factors $\frac{\mathbf{m}_{i,j}^{\text{off},\#} + \mathbf{m}_{i,j}^{\text{on},\#}}{\mathbf{m}_{i,j}^{\text{off}} + \mathbf{m}_{i,j}^{\text{on}}}$.

5.1.2 Computational complexity

We argue that performing the iterations as described in Eqs. (5.5) and (5.20) is straightforward and inexpensive once we have a method for computing $P_{i,j}$ as well as the gradients $\nabla_\theta P_{i,j}$. To see why this is true, denote by $C_{P(\theta)}$ the computational cost of evaluating $P_{i,j}(\theta)$ for all $(i, j) \in 1:N_v \times 1:N_t$ and by $C_{\nabla_\theta P(\theta)}$ the analogous cost for obtaining the derivatives of $P_{i,j}(\theta)$. If we write N_θ for the dimension of θ , i.e. $N_\theta := \dim(\theta)$, then the cost of one iteration (for a given step size $\zeta > 0$) can be decomposed into:

- $O(C_{\nabla_\theta P(\theta)} + C_{P(\theta)})$ in order to evaluate $P_{i,j}(\theta)$ as well as $\nabla_\theta P_{i,j}(\theta)$ at each data entry $(i, j) \in 1:N_v \times 1:N_t$
- $O(N_v N_t N_\theta^2)$ in order to assemble the approximate Hessian matrix Eq. (5.22) and score vector Eq. (5.24)
- $O(N_\theta^3)$ in order to solve an instance of the linear system arising in Eq. (5.20) and update $\theta^{(q)} \rightarrow \theta^{(q+1)}$

Note that the above implicitly assumes that the cost of obtaining $\mathsf{R}(\theta)$ and its derivatives is negligible. In each instance considered in this work the function $\mathsf{R}(\theta)$ is a simple quadratic, corresponding to a Gaussian prior, and this assumption is easily verified to be $O(N_\theta^2)$. Given that we need at least N_θ many data points to have a reasonable chance of

reconstructing θ , we essentially have to require that $N_v N_t \geq N_\theta$. Moreover it is clear that $C_{\nabla_\theta P(\theta)} \geq N_v N_t N_\theta$ since the related quantities have dimension N_θ . From its definition Eq. (5.4) we see that evaluation of $P_{i,j}(\theta)$ for a fixed $i \in 1:N_v$ and all $j \in 1:N_t$ is obtained from a solution of the time dependent RTE Eq. (2.22), or rather its angularly averaged albedo kernel Eq. (3.5), with source term $\delta(x - x_D)\delta(v - v_i)\delta(t)$. Although a single scattering approximation can, at least for sufficiently simple optical parameters, be obtained in time $O(N_t)$, i.e. $O(1)$ cost per time bin, and result in $C_{\nabla_\theta P(\theta)} \in O(N_v N_t N_\theta)$, it is not hard to see that for moderately large $N_\theta \lesssim 100$ the cost of evaluating the relevant products and exponentials of line integrals of the optical parameters will exceed N_θ quite handily. Indeed, if the image is an element of $K(\phi | x_D)$ consisting of at least N_θ basis functions then a naive evaluation, i.e. not using non-trivial approximations such as found in [54, 150], will already require more than $N_\theta N_t$ floating point operations.

The floating computations needed to update θ on the other hand are very simple and it is not hard to see that the leading constant in the assembly of $H^\#(\theta)$ is in fact equal to 1 by virtue of its symmetry and will likely not present a challenge on modern computer hardware.

Given how even single scattering approximations for non-trivial optical parameters will dominate the cost of a parameter update, it becomes clear that only $P_{i,j}(\theta)$ evaluations and the related gradients contribute to the computational cost in any meaningful way and we may focus our attention to the complexity of $P_{i,j}(\theta)$ evaluations. Henceforth we will assume that the optical parameters satisfy the assumptions of Theorem 3.1.4 (with the exception of Eq. (3.14)) and additionally:

P-1 The cost of computing $\phi(z)$ is $O(1)$ at any $z \in \mathbb{R}$ and thus for $x \in X$ and any $(b, h, w) \in X \times (0, \infty)^2$ the cost of computing

$$\mathbf{E}_\phi(x | b, h, w) := w\phi\left(\frac{\|x - b\|_2}{h}\right),$$

i.e. the cost of an evaluation of an element of the good kernel space at x , is $O(1)$.

P-2 For any pair $(x, v, t) \in X \times \mathbb{S}^2 \times (0, \infty)$ and any $(b, h, w) \in X \times (0, \infty)^2$ the cost of

computing

$$\mathbb{I}_\phi(x, v, t \mid b, h, w) := w \int_0^t \phi \left(\frac{\|[x + vr] - b\|_2}{h} \right) dr,$$

i.e. the line integral between x and $x + vt$, is $O(1)$.

P-3 For any pair $(x, v, t) \in X \times \mathbb{S}^2 \times (0, \infty)$ and any $(b, h, w) \in X \times (0, \infty)^2$ the cost of computing the derivatives w.r.t.

$$\begin{aligned} d\mathbb{E}_\phi(x \mid b, h, w) &:= \nabla_{(b,h,w)} w \phi \left(\frac{\|x - b\|_2}{h} \right) \\ d\mathbb{I}_\phi(x, v, t \mid b, h, w) &:= \int_0^t \nabla_{(b,h,w)} w \phi \left(\frac{\|[x + vr] - b\|_2}{h} \right) dr \end{aligned}$$

i.e. point evaluation and component-wise line integral corresponding to the previous two items but for the derivative of ϕ w.r.t. its parameters, is $O(1)$.

P-4 The phase function f_p is independent of x , i.e. homogeneous in space and only dependent in the incident and outgoing directions, and can be evaluated as well as sampled from at cost $O(1)$.

The following result shows that these assumptions are true in the arguably rather important case of a multivariate Gaussian ϕ which, although strictly speaking is not compactly supported and therefore not a good kernel, is used in all of our simulations.

Lemma 5.1.1. *Let $\phi(x) = e^{-x^2/2}$, then the assumptions of Items **P-1** to **P-3** are fulfilled for $X \subseteq \mathbb{R}^3$ provided that the exponential and error function (see e.g. [147, p. 341]) can be evaluated in $O(1)$. In particular,*

$$\mathbb{I}_\phi(x, v, t \mid b, h, w) = we^{-\frac{\|x-b\|_2^2 - \langle x-b, v \rangle^2}{2h^2}} \int_0^t e^{-\frac{(s+\langle x-b, v \rangle)^2}{2h^2}} ds. \quad (5.25)$$

Proof. We have that

$$\begin{aligned} \|[x + vs] - b\|_2^2 &= \|x - b\|_2^2 + 2s\langle x - b, v \rangle + s^2 \\ &= \|x - b\|_2^2 + (s + \langle x - b, v \rangle)^2 - \langle x - b, v \rangle^2 \end{aligned}$$

which shows the claimed identity for $\mathbf{I}_\phi(x, v, t \mid b, h, w)$. The one-dimensional integral can be transformed into a difference of error functions, denoted by erf, through the elementary change of variables

$$\int_0^t e^{-\frac{(s+\langle x-b, v \rangle)^2}{2h^2}} ds = \frac{\sqrt{2\pi}h}{2} \left(\operatorname{erf}\left(\frac{t + \langle x - b, v \rangle}{\sqrt{2}h}\right) - \operatorname{erf}\left(\frac{\langle x - b, v \rangle}{\sqrt{2}h}\right) \right).$$

The gradient of \mathbf{E}_ϕ is straight forward and so is the first factor of Eq. (5.25) and it remains to deal with the integral within \mathbf{I}_ϕ . However, as this can be expressed in terms of error functions whose derivatives are expressed in terms of ϕ the claim follows easily from the chain rule. \square

Remark. The assumptions of Lemma 5.1.1 are reasonable in the sense that there rapidly converging expansions for both the exponential and error functions and they are implemented in standard libraries for scientific computing such as SciPy [143].

In view of Eqs. (2.27), (2.28), and (2.30) it is clear that evaluation of RTE quantities such as Eq. (3.5) requires point evaluation of the scattering intensity σ_s and phase function f_p as well as line integration of the differential absorption α and total extinction σ_{a+s} parameters.

For the purposes of our computations, the optical parameters besides the phase function are elements of $K(\phi \mid x_D)$ so that they can be evaluated as a finite sum of \mathbf{E}_ϕ and \mathbf{I}_ϕ related operations as well as evaluations of the phase function. We will use the following short-hand notation

Definition 5.1.2. For any $k \in \mathbb{N}$ we define the sets $\Gamma_k(x, v, t) \subseteq (\mathbb{R}^3)^{k+2}$

$$\Gamma_k(x, v, t) = \left\{ \vec{p} \in \{x\} \times X^k \times \{x\} : \frac{\vec{p}_2 - \vec{p}_1}{\|\vec{p}_2 - \vec{p}_1\|_2} = v \text{ and } \sum_{j=1}^{k+1} \|\vec{p}_{j+1} - \vec{p}_j\|_2 = t \right\} \quad (5.26)$$

which is the set of all $k+2$ -tuples that have first and last element equal to x . In the case of a “complete” photon path $x = x_D$ will always be the source/detector location, while the other k are meant to represent the locations where a scattering event occurred but we will also consider segments. As such, the subsequent definitions are meant to be valid for

$\vec{p} \in (\mathbb{R}^3)^{k+2}$. The shortest, i.e. piece-wise linear, curve connecting those points is denoted by $\gamma(\vec{p})$ or more formally we have for $\gamma(\vec{p}) \in (\mathbb{R}^3)^{k+2}$

$$\gamma(\vec{p}) = \bigcup_{j=1}^{k+1} \left\{ \lambda \vec{p}_j + (1 - \lambda) \vec{p}_{j+1} : 0 < \lambda < 1 \right\}. \quad (5.27)$$

For notational convenience we also define for $\vec{p} \in (\mathbb{R}^3)^{k+2}$ the set of scattering triplets

$$\zeta(\vec{p}) = \bigcup_{j=2}^{k+1} \left\{ \left(\frac{\vec{p}_j - \vec{p}_{j-1}}{\|\vec{p}_j - \vec{p}_{j-1}\|_2}, \vec{p}_j, \frac{\vec{p}_{j+1} - \vec{p}_j}{\|\vec{p}_{j+1} - \vec{p}_j\|_2} \right) \right\} \quad (5.28)$$

where each element of the set is a 3-tuple consisting of incident direction, scattering location and outgoing direction (in that order). Let $f : X \times \mathbb{S}^2 \times \mathbb{S}^2 \rightarrow \mathbb{R}$ and $\vec{p} \in (\mathbb{R}^3)^{k+2}$, then we write

$$\mathbf{E}_f(\vec{p}) := \prod_{(v', x, v) \in \zeta(\vec{p})} f(x, v', v)$$

if $f : X \rightarrow \mathbb{R}$, i.e. f does not depend on direction, then the above is to be understood as the product where each factor is evaluated only at the X -valued argument. For such f we additionally define the line integral

$$\mathbf{I}_f(\vec{p}) := \int_{\gamma(\vec{p})} f(z) dz = \sum_{j=2}^{k+2} \int_0^{\|\vec{p}_j - \vec{p}_{j-1}\|_2} f \left(\vec{p}_{j-1} + z \frac{\vec{p}_j - \vec{p}_{j-1}}{\|\vec{p}_j - \vec{p}_{j-1}\|_2} \right) dz.$$

If $f \in K(\phi \mid x_D)$ then the above can be expressed via \mathbf{E}_ϕ and \mathbf{I}_ϕ respectively which allows us to define $d\mathbf{E}_f$ as well as $d\mathbf{I}_f$ via the expressions provided under Item **P-3**.

With the above notation we can express the measurement kernel $\mathbf{m}(\cdot \mid x_D)$ as

$$\mathbf{m}(v_0, t \mid x_D) = \sum_{k=0}^{\infty} \int_{\Gamma_k(x_D, v_0, t)} e^{-\mathbf{I}_{\sigma_{a+s}}(\vec{p})} \mathbf{E}_{\sigma_s f_p}(\vec{p}) \mu_{k|x_D, v_0, t} d\vec{p} \quad (5.29)$$

for suitable measures $\mu_{k|x_D, v_0, t}$ which can be seen as absolutely continuous w.r.t. the product of $k-1$ Lebesgue measures on X , i.e. on \mathbb{R}^3 , which is to say standard integration over X^{k-1} .

With accurate measurements one could use a direct inversion method from narrow FOV data which means the present work is only of interest when the data is noisy. In such situations it may not be necessary to evaluate the forward map to a high degree of accuracy and we might be able to get away with using random approximations $\hat{P}_{i,j}(\theta)$ and derivatives $\nabla_\theta \hat{P}_{i,j}(\theta)$ thereof obtained from Monte Carlo ray-tracing very similar to what was described in [49]. This effectively turns our iterations from Eq. (5.20) into a stochastic optimisation procedure. In order to evaluate the necessary integrals related to the RTE and its Neumann expansion we will, in addition to sampling the phase function, also need to sample distances that represent the time in between two subsequent scattering events. In what follows we will describe what can be thought of as an importance sampling-based MC method for evaluating Eq. (5.29). Of course, we ideally seek to sample from a density that is proportional to the integrand as this would result in a variance of 0 and exact integration. This is unfortunately intractable but we can sample from a density proportional to a factor within the integrand. We will need the following result.

Lemma 5.1.3. *For $x \in X$ and $v \in \2 we say that the random variables \mathbf{a}, \mathbf{t} have distribution $p_{\sigma_s}(\cdot | x, v)$ if they are independent and \mathbf{a} is a Bernoulli random variable*

$$\mathbb{P}(\mathbf{a} = 1) = 1 - \mathbb{P}(\mathbf{a} = 0) = 1 - e^{-I_{\sigma_s}((x, x + \ell_v^+(x)v))}$$

while \mathbf{t} is a continuous random variable with cumulative distribution function

$$F_{\sigma_s}(t | x, v) = \frac{1 - e^{-I_{\sigma_s}((x, x + tv))}}{1 - e^{-I_{\sigma_s}((x, x + \ell_v^+(x)v))}}$$

for $t \in (0, \ell_v^+(x))$. If σ_s has N_ϕ components then a sample from $p_{\sigma_s}(\cdot | x, v)$ can be obtained in time $O(N_\phi \log(\varepsilon))$ where ε is the required accuracy relative to the diameter of the (w.l.o.g. bounded) domain and \mathbf{t} has density function

$$\frac{E_{\sigma_s}(x + tv)e^{-I_{\sigma_s}((x, x + tv))}}{1 - e^{-I_{\sigma_s}((x, x + \ell_v^+(x)v))}} 1_{(0, \ell_v^+(x))}(t). \quad (5.30)$$

Proof. Combining Items **P-1** and **P-2** implies that one can evaluate F_{σ_s} at cost $O(N_\phi)$.

Sampling from one-dimensional distributions via the inverse sampling method requires inverting a monotone function which is equivalent to root finding and can be performed with $O(\log(\varepsilon))$ evaluations, where ε is the required accuracy relative to the diameter of the domain (which bounds $\ell_v^+(x)$). To see that the density has the desired form we must only observe that

$$\sigma_s(x + vt) = \mathbb{E}_{\sigma_s}(x + tv) = \frac{d}{dt} \mathbb{I}_{\sigma_s}((x, x + tv)) = \frac{d}{dt} \int_0^t \sigma_s(x + vs) ds$$

with the first and last terms obviously being equal while the other identities stem from Definition 5.1.2. \square

The following result summarises the approach in our setting.

Proposition 5.1.4. *Consider the sequences of $(\mathbf{a}_k, \mathbf{t}_k)_{k \in \mathbb{N}}$ and $(\mathbf{v}_k)_{k \in \mathbb{N}}$ of $(0, \infty) \times \{0, 1\}$ - and \mathbb{S}^2 -valued random variables whose distribution, for \mathbf{p}_{σ_s} as in the previous lemma, is uniquely determined by*

$$\mathbf{a}_k, \mathbf{t}_k \mid \mathbf{t}_1, \dots, \mathbf{t}_{k-1}, \mathbf{v}_{k-1}, \dots, \mathbf{v}_0 \sim \mathbf{p}_{\sigma_s}(\cdot \mid \mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \quad (5.31)$$

$$\mathbf{v}_k \mid \mathbf{t}_1, \dots, \mathbf{t}_k, \mathbf{v}_{k-1}, \dots, \mathbf{v}_0 \sim f_p(\mathbf{x}_k, \mathbf{v}_{k-1} \rightarrow \cdot) \quad (5.32)$$

while the initial direction satisfies $\mathbf{v}_0 = v_0$ almost surely for some fixed $v_0 \in \mathbb{S}^2$ and \mathbf{x}_k are defined via the recursion

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{t}_k \mathbf{v}_{k-1} \quad (5.33)$$

with $\mathbf{x}_0 = x_D$ almost surely. Further let for $t \geq \sum_{j=1}^{k-1} \mathbf{t}_j + \|x_D - \mathbf{x}_{k-1}\|_2$

$$T_k(t) := \frac{1}{2} \frac{(t - \sum_{j=1}^{k-1} \mathbf{t}_j)^2 - \|x_D - \mathbf{x}_{k-1}\|_2^2}{t - \sum_{j=1}^{k-1} \mathbf{t}_j - \mathbf{v}_{k-1} \cdot (x_D - \mathbf{x}_{k-1})}$$

$$X_k(t) := \mathbf{x}_{k-1} + T_k(t) \mathbf{v}_{k-1}$$

$$V_k(t) := \frac{x_D - X_k(t)}{\|x_D - X_k(t)\|_2}$$

$$W_k(t) := e^{-\mathbf{I}_{\sigma_a}((\mathbf{x}_{0:k-1}, X_k(t), x_D))} e^{-\mathbf{I}_{\sigma_s}((\mathbf{x}_{k-1}, X_k(t), x_D))} \mathbf{E}_{f_p \sigma_s}((\mathbf{x}_{k-1}, X_k(t), x_D))$$

and $W_k(t) = 0$ in case of $t < \sum_{j=1}^{k-1} \mathbf{t}_j + \|x_D - \mathbf{x}_{k-1}\|_2$ or $X_k(t) \notin X$. Then, for $\mathfrak{m}^k(\cdot | x_D)$ as in Lemma 3.1.1,

$$\mathfrak{m}^k(v_0, t | x_D) = \mathbb{E} \left[\frac{g_{\mathfrak{m}}(x_D, V_k(t)) W_k(t) |\vec{n}(x_D) \cdot V_k(t)|}{\|x_D - X_k(t)\|_2^2 (1 - \mathbf{v}_{k-1} \cdot V_k(t))} \prod_{j=1}^{k-1} \mathbf{a}_j \right] \quad (5.34)$$

Proof. For $k = 1$, i.e. single scattering there is no sampling performed, and $T_1(t) = t/2$ is the distance of the only possible scattering location. For the case of multiple scattering we first recall Eq. (A.5) and that if the change of variables is performed on the “last” variables then we obtain

$$\mathfrak{m}^{k+1}(v_0, t | x_D) = \int_{(0,\infty)^k \times (\mathbb{S}^2)^k} e^{-\mathbb{I}_{\sigma_{a+s}}(x_{0:k+2})} \mathbf{E}_{\sigma_s f_p}(x_{0:k+2}) R(x_{0:k+2}) dv_1 \dots dv_k dt_1 \dots dt_k$$

where we set

$$R(x_{0:k+2}) = \frac{g_{\mathfrak{m}}(x_D, v_{k+1}) |\vec{n}(x_D) \cdot v_{k+1}|}{\|x_D - x_k\|_2^2 (1 - v_k \cdot v_{k+1})} \mathbf{1}_{(0,t)} \left(\sum_{j=1}^{k-1} t_j + \|x_D - x_k\|_2 \right) \mathbf{1}_{X^{k+1}}(x_{1:k+1})$$

and v_{k+1} as well as x_0, \dots, x_{k+1} are implicitly defined analogously to $X_k(t)$ and $V_k(t)$ from before and depend on the independent variables $v_1 \dots v_k$ and $t_1 \dots t_k$ in the same fashion. Also note that the joint density of \mathbf{a}, \mathbf{t} as in Lemma 5.1.3 for $(x, v) \in X \times \mathbb{S}^2$ is given by

$$\frac{\mathbf{E}_{\sigma_s}(x + tv) e^{-\mathbb{I}_{\sigma_s}((x, x + tv))}}{1 - e^{-\mathbb{I}_{\sigma_s}((x, x + \ell_v^+(x)v))}} \left(1 - e^{-\mathbb{I}_{\sigma_s}((x, x + \ell_v^+(x)v))} \right)^a e^{-(1-a)\mathbb{I}_{\sigma_s}((x, x + \ell_v^+(x)v))}$$

and integration is taken w.r.t. the Lebesgue measure dt and the counting measure on $\{0, 1\}$. At $a = 1$ the denominator cancels and, when paired with the density of Eq. (5.32) as well as the weight adjustment W_k , the result is

$$\int_{(0,\infty)^k \times (\mathbb{S}^2)^k} e^{-\mathbb{I}_{\sigma_{a+s}}(x_{0:k+2})} \mathbf{E}_{\sigma_s f_p}(x_{0:k+2}) \tilde{R}(x_{0:k+2}) dv_1 \dots dv_k dt_1 \dots dt_k$$

where \tilde{R} is a weight function. In view of Eq. (5.34), $R = \tilde{R}$ would follow if the indicator functions appearing in R are a match to those in \tilde{R} . The indicator functions w.r.t. the X -valued variables, i.e. the factor that ensures we only integrate over points within the

domain is enforced by design of \mathbf{p}_{σ_s} , with the exception of $X_k(t)$. The indicator function in Eq. (5.30) is such that a value can be sampled if and only if the resulting location is inside the boundary. The restrictions $t > \sum_{j=1}^{k-1} t_j + \|x_D - x_k\|_2$ and $X_k(t) \in X$ are enforced as part of the weighting function W_k which is of course chosen precisely in such a way that the expectation matches \mathbf{m}_k for each $k \in \mathbb{N}$. \square

Remark. In order to obtain $\mathbf{m}_{\text{off}}^k$ and \mathbf{m}_{on}^k one has to evaluate W_k for $\sigma_{a(\text{off})} = \sigma_a$ and $\sigma_{a(\text{on})} = \sigma_a + \alpha$ respectively. With T_k, X_k and V_k as in Proposition 5.1.4 we define for a collection of optical parameters $\xi = (\alpha, \sigma_s, \sigma_s, f_p)$

$$\mathbf{M}_{\text{off}}^k[t](\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a} \mid \xi) := \frac{g_{\mathbf{m}}(x_D, V_k(t)) W_k(t) |\vec{n}(x_D) \cdot V_k(t)|}{\|x_D - X_k(t)\|_2^2 (1 - \mathbf{v}_{k-1} \cdot V_k(t))} \prod_{j=1}^{k-1} \mathbf{a}_j \quad (5.35)$$

$$\mathbf{M}_{\text{on}}^k[t](\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a} \mid \xi) := \frac{g_{\mathbf{m}}(x_D, V_k(t)) W_k^{\text{on}}(t) |\vec{n}(x_D) \cdot V_k(t)|}{\|x_D - X_k(t)\|_2^2 (1 - \mathbf{v}_{k-1} \cdot V_k(t))} \prod_{j=1}^{k-1} \mathbf{a}_j \quad (5.36)$$

and $W_k^{\text{on}}(t) = e^{-\mathbf{I}_{\alpha}((\mathbf{x}_{0:k-1}, X_k(t), x_D))} W_k(t)$ accounts for differential absorption while W_k as in Proposition 5.1.4 for the offline quantities. Eqs. (5.35) and (5.36) are then such that $\mathbb{E}[\mathbf{M}_{\text{on/off}}^k[t](\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a} \mid \xi)] = \mathbf{m}_{\text{on/off}}^k(v_0, t \mid x_D)$. Note that the argument $\mathbf{t} = \mathbf{t}_{1:k}$ are sampled scattering times while t is the detection time, i.e. total path length. The pseudo-code summarising the suggested strategy can be found in Algorithm 1 in Appendix A.2. If paths are sampled with the strategy as outlined above then:

1. \mathbf{a}_k is an indicator as to whether a scattering event occurred at the k -th iteration of the Markov Chain. Clearly if $\mathbf{a}_k = 0$ then the trajectory becomes irrelevant for $k' \geq k$ and no more scattering times/distances need to be sampled.
2. If there is a boundary present then $\mathbf{a}_k = 0$ corresponds to a boundary scattering event and the integration needs to be adjusted accordingly.
3. The random \mathbf{a}_k can be replaced with their expectation. In the absence of a boundary, this does not increase the cost by a lot (instead of sampling only a part of the initially released photons we sample all of them in each iteration) but when there is boundary scattering it would result in a splitting of the trajectory each time and exponential

growth in the number of photon paths.

4. Similarly \mathbf{a}_k can be adjusted to account for σ_a (note that absorption in W_k is integrated over the whole trajectory whilst σ_s only over a fraction). In that situation $\mathbf{a}_k = 0$ if the photon got absorbed or didn't scatter.

In addition to those variations, there are several ways to treat different k and t . Even though it is clearly necessary to use different paths for varying initial direction v_i , one may choose to evaluate each time bin t_j based on the same sampled segments. Equally, one may re-use previously sampled segments, i.e. from scattering order $k - 1$, in order to evaluate scattering of order k . This reduces the computational complexity but also reduces the number of independently sampled trajectories which can result in higher MC integration errors

Computation of derivatives Generally, our optical parameters will be described by a parameter $\theta \in \Theta$ that is related to a model inspired by atmospheric dispersion processes. This is done by expressing the kernel parameters (b, h, w) of each component via θ , i.e. we have $(b, h, w) = (b[\theta], h[\theta], w[\theta])$. In order to compute derivatives of σ_a , σ_s or α we must, in addition to the assumption under Item **P-3**, only ensure that $\nabla_\theta b$, $\nabla_\theta h$ and $\nabla_\theta w$ exists for each set of kernel parameters. In the present case this is ensured by expressing the kernel parameters explicitly through the components of θ in a way that ensures differentiability and simple computations. More specifically, the components of θ either appear directly in the kernel parameters or form the knots of a linear spline although other choices are possible. Derivatives of Eq. (5.34) can be computed through score function estimators [49, 101]. In particular, even though we draw samples whose distribution depends on the parameters of interest, we do not need to compute derivatives of the random variables (as is the case with path-wise estimators). For notational simplicity, we only give a description of the procedure for generic densities p_θ as the necessary adjustments are elementary. For

sufficiently well-behaved \mathbf{p}_θ we have

$$\begin{aligned}\nabla_\theta \int \mathbf{p}_\theta(y) f(y) dy &= \int f(y) \nabla_\theta \mathbf{p}_\theta(y) + \mathbf{p}_\theta(y) \nabla_\theta f(y) dy \\ &= \int \frac{\nabla_\theta \mathbf{p}_\theta(y)}{\mathbf{p}_\theta(y)} \mathbf{p}_\theta(y) f(y) + \mathbf{p}_\theta(y) \nabla_\theta f(y) dy \\ &\approx n^{-1} \sum_{i=1}^n \frac{\nabla_\theta \mathbf{p}_\theta(y_i)}{\mathbf{p}_\theta(y_i)} f(y_i) + n^{-1} \sum_{i=1}^n \nabla_\theta f(y_i)\end{aligned}$$

where $y_i \sim \mathbf{p}_\theta$ are independent samples from \mathbf{p}_θ . and exhibit an overall similar behaviour to Eq. (5.34). Note that

$$\int \mathbf{p}_\theta(y) f(y) dy \approx n^{-1} \sum_{i=1}^n f(y_i)$$

but the θ -derivative of the right-hand side would *not* be a valid estimator and must be corrected via the score $\frac{\nabla_\theta \mathbf{p}_\theta}{\mathbf{p}_\theta}$ of \mathbf{p}_θ .

Computing the score for the distribution suggested in Proposition 5.1.4 only requires calls to `dE` and `dI` and can be performed on a per-segment basis thanks to

$$\frac{\nabla_\theta \mathbf{p}_\theta(x, z)}{\mathbf{p}_\theta(x, z)} = \frac{\nabla_\theta \mathbf{p}_\theta(x | z)}{\mathbf{p}_\theta(x | z)} + \frac{\nabla_\theta \mathbf{p}_\theta(z)}{\mathbf{p}_\theta(z)}$$

where in the case of photon paths $y = (x, z)$ would correspond to a separation of a full path y into segments x and z . For $(\alpha[\theta], \sigma_a[\theta], \sigma_s[\theta], f_p[\theta]) = \mathcal{G}(\theta)$ as in Eq. (3.10) we will apply this to densities \mathbf{p}_θ^k given by

$$\mathbf{p}_\theta^k(\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a}) := \prod_{j=1}^k \mathbf{p}_{\sigma_s[\theta]}(\mathbf{a}_j, \mathbf{t}_j | \mathbf{x}_{j-1}, \mathbf{v}_{j-1}) f_p[\theta](\mathbf{x}_j, \mathbf{v}_{j-1} \rightarrow \mathbf{v}_j) \quad (5.37)$$

determined by Eqs. (5.31) to (5.33) from Proposition 5.1.4.

Parallelisation and cost of evaluation We briefly summarise our previous analysis

1. Assuming that the optical parameters consist of N_ϕ kernel components and that paths for $k+1$ are evolved from k we have

- Evolving a photon path in order to evaluate scattering of order $k + 1$ based on k requires $O(\log(\varepsilon))$ calls to \mathbf{I}_{σ_s} which equals $O(N_\phi \log(\varepsilon))$ calls to \mathbf{I}_ϕ . For the Gaussian kernel ϕ each evaluation of \mathbf{I}_ϕ requires one evaluation of the exponential function as well as two calls to erf.
 - Storing the score function requires a call to \mathbf{dE} and \mathbf{dE} for each path and scattering order bringing the overall cost to $O(N_\phi(\log(\varepsilon) + N_\theta))$
2. If, given an initial direction v_i , the same segments are used for all t_j and Eq. (3.5) is evaluated up to order K_{\max}
- Evaluating the integrand in Eq. (5.34) and its derivative for order k requires $O(N_\phi N_\theta N_t)$ evaluations of \mathbf{I}_ϕ and \mathbf{dI}_ϕ related quantities. The score function can be found in $O(N_\phi N_\theta)$ based on the score of order $k - 1$.
 - This brings the cost of differentiation to $O(K_{\max} N_\phi N_\theta N_t)$ \mathbf{I}_ϕ and \mathbf{dI}_ϕ related evaluations.
3. The above needs to be performed for each initial direction and photon trajectory, resulting in a total of $O(K_{\max} N_\phi N_\theta N_v N_t N_{\vec{p}})$ \mathbf{I}_ϕ and \mathbf{dI}_ϕ related evaluations if $N_{\vec{p}}$ trajectories are used per initial direction. The $N_{\vec{p}} N_v$ computations across initial directions and trajectories are independent of each other and can easily be done in parallel.

An important aspect in the choice of the solver was not only the ability to evaluate it approximately at a low cost but also the aforementioned parallel nature of the problem. This was exploited in the presented simulations with the CUDA framework [108] along with the Numba compiler for Python [80]. We have also observed good scaling across multiple GPUs. Note that the only $O(N_\phi N_\theta)$ kernel parameters, which are needed to evaluate the image and θ -derivatives, have to be present on each compute unit while the score estimates can be split in accordance with the initial directions and independent trajectories. A different variation of MC-based path tracing will have slightly different, typically better, scaling across multiple cores and compute units but will have a higher overall cost.

5.1.3 Randomisation induced errors

As mentioned earlier, the accuracy of our integration scheme, much like any Monte Carlo-based method, will depend on the “smoothness” of the integrand which here is to be understood as its deviation from a constant.

- The factor $g_m(x_D, \cdot)$ captures the sensitivity of the detector and may or may not be a well-behaved function. This will primarily depend on the FOV itself, i.e. if we were to measure several FOVs and have high angular resolution then this will result in poor quality of the described approach. Such situations are particularly suitable for more sophisticated, bi-directional path tracing strategies, see for example [112]. As we only considered the case of two FOVs with one of them assumed to be infinitesimal (so it captures precisely the single scattering) this was not an issue.
- The factor $\|x_D - X_k(t)\|_2^{-2}$ is only an issue if photons are likely to scatter near the detector where, by assumption on $K(\phi | x_D)$, only ambient particles can be found. Given that measuring wide FOV data is not of much use in those situations where there is considerable ambient scattering away from the plume of interest, we do not have to worry too much about potential singularities caused by such events.
- The factor $(1 - \mathbf{v}_{k-1} \cdot V_k(t)) \rightarrow 0$ can only happen if \mathbf{v}_{k-1} points from the last sampled scatter point \mathbf{x}_{k-1} towards the detector at x_D . In order to deal with the issue of having singularities in the integrand it is therefore sufficient to sample the final direction \mathbf{v}_{k-1} conditional on not pointing towards the detector and perform the integration over the region of solid angle that causes the issue separately. This correction was tested and found to make no appreciable difference whilst adding considerable complexity and was not included as part of our simulation runs. A possible explanation as to why this might have been the case is given in the following.

Despite these potential pitfalls, the sampling used in this work is straightforward and no attempt to further reduce the variance with regards to the points mentioned above has been made. In situations with less well-behaved parameters or more complex environments this may not be good enough and rectification of the arising issues might be a

non-trivial matter.

An important aspect that we haven't addressed as of yet is that we do not require an accurate evaluation of $\mathbf{m}_{\text{on/off}}$ but rather we must only find

$$P_{i,j}(\theta) = \frac{\mathbf{m}_{\text{on}}[\theta](v_i, t_j)}{\mathbf{m}_{\text{on}}[\theta](v_i, t_j) + \mathbf{m}_{\text{off}}[\theta](v_i, t_j)} = 1 - \frac{1}{1 + \frac{\mathbf{m}_{\text{on}}[\theta](v_i, t_j)}{\mathbf{m}_{\text{off}}[\theta](v_i, t_j)}}.$$

It is not hard to see that for strictly positive random variables Z and W

$$\text{corr}(Z, W) = 1 \implies \frac{Z}{W} = \frac{\mathbb{E}(Z)}{\mathbb{E}(W)}$$

which would suggest that estimating ratios is done preferably with correlated random variables. In the present case we are free to sample in any way we desire and a simple but effective way to ensure positive correlation is to use the same paths for the \mathbf{m}_{on} and \mathbf{m}_{off} while only adjusting the integrand.

Lemma 5.1.5. *If $\hat{\mathbf{m}}_{\text{off}}$ and $\hat{\mathbf{m}}_{\text{on}}$ are estimated as in Proposition 5.1.4 while the same samples are used for each, then, with probability one, for the estimate $\hat{P}_{i,j}$ of $P_{i,j}$*

$$0 \leq \hat{P}_{i,j}(\theta) = \frac{\hat{\mathbf{m}}_{\text{on}}[\theta](v_i, t_j)}{\hat{\mathbf{m}}_{\text{on}}[\theta](v_i, t_j) + \hat{\mathbf{m}}_{\text{off}}[\theta](v_i, t_j)} \leq \frac{1}{2} \quad (5.38)$$

for all i and j regardless of sample size. Further, $\nabla_\theta \hat{P}_{i,j}$ is bounded with probability one.

Despite being essentially a trivial conclusion, the above implies that, no matter how inaccurate the evaluations of the individual measurements \mathbf{m}_{off} and \mathbf{m}_{on} are, we can always ensure that $\hat{P}_{i,j}$ obeys the physical constraint set out in Eq. (5.38). Note that using the same photon trajectories is also consistent with our assumption on equal scattering, which in particular requires that the distribution of photons at both wavelengths be identical conditional on no differential absorption of the online photon. If we consider the empirical distribution over paths induced by our sample then using the same segments for both online and offline intensities ensures that this condition is met for the *empirical* distributions and as such scattering is identical as sought. Note that this may also be a reason why the singularities of the integrand that were discussed earlier didn't have a

great impact. Indeed, if we sample an awkward direction pointing straight at the detector, then that large spike results in what is essentially a reduction of the sample size to 1. If the dominating trajectory is somewhat “typical” then this does not affect the loss function in a very meaningful way and the optimisation can proceed normally.

It should be noted that the sums of random vectors and matrices such as $H(\theta)$ or $\nabla_\theta \text{Loss}$ used in our optimisation behave much like approximations used in randomised numerical linear algebra. The reduction of the image to dispersion-related parameters means that $N_\theta \ll N_v N_t$ and, assuming that we ensure boundedness of the integrand that is to be evaluated via MC path tracing, we may use the matrix Bernstein inequality [140] to obtain concentration estimates for quantities such as the spectral norm error $\|\hat{H}(\theta) - H(\theta)\|$. For further details we refer to [98, 139]. Estimates for the extreme eigenvalues of the spd matrix $\hat{H}(\theta)$ in the form of matrix Chernoff inequalities are also available [30, 140].

From a practical perspective the simplest way to, *a posteriori*, assess randomisation-induced errors is arguably by inspection of multiple re-starts.

- Beginning with an arbitrary initial parameter $\theta^{\text{init}} \in \Theta$ we can find θ^{ssct} by using a single scattering approximation Eqs. (3.11a) and (3.11b) instead of the full RTE.
- Using $\theta^{\text{ssct}} = \theta^{(0)}$ we can find the solution θ^{RTE} by iterating Eq. (5.20) until a convergence criterion is met.

If the above two-step procedure is repeated for random choices of θ^{init} , then variability in θ^{ssct} is purely down to the initial values and can be attributed to the non-convex nature of the objective and the possible presence of multiple local minima. There can even be multiple global minima if the objective function is not the log-likelihood, e.g. when there is some amount of model error as will often be the case. Typically the θ^{ssct} can be treated as good initial values, all of which are (approximately) equal, as there are rigorous statements that would suggest the preservation of certain aspects such as the distance to the gas. The variability of the final solution θ^{RTE} is then primarily down to randomisation.

Early stopping & regularisation The optimisation scheme and error tolerance in our optimisation algorithm is chosen such that the iterations essentially stop a little early. This, although a long-standing idea [102, 133] and known in some cases to improve gradient descent-based estimators by making a bias-variance trade-off [120, 151], may become the source of additional errors. While we haven't carried out a formal analysis of this strategy it seems intuitive that when the wide FOV fitting is initialised with the corresponding single scattering approximation, then premature stopping will result in the same bias, albeit to a lesser extent, as if the wide FOV was treated like a narrow one. In other words, the effects of early stopping are not necessarily desirable. Assume for the moment, we will make this more concrete as Eq. (5.44) in the next section, that θ can be partitioned into $(\theta_s, \theta_\alpha)$ so that $\sigma_s = \theta_s u[\theta]$ with u depending only on θ_α and consider the single scattering initialisation

$$\hat{\theta}^{\text{ssct}} \in \arg \max_{\theta: \theta_s=0} \text{Loss}(\psi_{i,j}^\#(\theta), \theta \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a)$$

and the $P_{i,j}$ are defined at $\theta_s = 0$ via the limit $\theta_s \searrow 0$ resulting in single scattering. If the non-trivial part $\theta_\alpha^{\text{ssct}}$ is not at the boundary of the feasible region then

$$\nabla_{\theta_\alpha} \text{Loss}(\psi_{i,j}^\#(\theta^{\text{ssct}}), \theta^{\text{ssct}} \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) = 0.$$

Issues may now arise because when the randomised scheme as described in Proposition 5.1.4 is used, then for any feasible θ (in probability)

$$\lim_{\theta_s \searrow 0} \frac{\partial}{\partial \theta_s} \text{Loss}(\psi_{i,j}^\#((\theta_s, \theta_\alpha)), (\theta_s, \theta_\alpha) \mid \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}}, H_a) = \frac{\partial}{\partial \theta_s} R((0, \theta_\alpha), \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}})$$

and it follows that the only thing which is preventing the θ -derivative of the loss function from becoming zero at the initialisation point, thus potentially getting stuck there, is the regularisation R term from Eq. (5.8). This is true simply because $a_1 \rightarrow 0$ as $\theta_s \searrow 0$ and the stochastic approximation of the loss eventually degenerates to its deterministic single scattering component. An immediate way out would of course be to pick R such that it is bounded away from zero for $\theta_s \searrow 0$ but this can be problematic in case the wide and

narrow FOV aren't collected separately and there is a fairly small amount of scattering. If we are going to select R adaptively, i.e. as a function of the data, in such a way that the solution doesn't get stuck in single scattering with high probability, then it would be good to have an a priori estimate as to what kind of effect the choice will have in case the photons have in fact mostly undergone single scattering. For the case of a ridge-like penalty, which was initially suggested as a method for problems suffering from multicollinearity in linear regression [62, 63], we may use Eq. (4.60) in order to measure the potential incurred bias. This can be evaluated either based on a single scattering approximation or a single call to the RTE solver for a potentially more accurate assessment. It remains to specify a (quadratic) form for R which ideally should counteract the effects of stopping before full convergence. To better understand the effect of higher-order scattering on absorption strength, note that a photon that scatters once but not in the direction of the detector can only be observed if it scatters again. Scattering again happens with maximal probability if the photon scatters in the direction of maximum optical thickness, i.e. towards the center of the plume where typically absorption will be higher as well. Thus, a photon initially shot at the edge of the plume will scatter into the middle, thereby increasing the absorption corresponding to those directions, whereas photons aimed at the middle will flare out, hence reducing the absorption relative to the straight path. This effect, which causes some of the aforementioned collinearity, can be mimicked by bloating the plume a little, pushing it backward so that $\partial_\alpha D$ remains intact, and increasing the absorption strength so that the bloating, which reduces optical thickness and thus absorption, is counteracted at the boundary of the plume in the central direction. As such, we want to apply a shrinkage penalty to a subset of the parameters θ_h and θ_0 as in Eq. (5.44) corresponding to plume width and absorption strength in addition to penalising values for θ_s that are very small. The latter involves a quadratic term that is not centred at 0 but rather a value at which scattering, i.e. $a_1 = 1$, will occur in the solver with non-negligible probability. Owing to symmetries there is no reason to shrink parameters corresponding to positions that don't affect $\partial_\alpha D$. In addition to this, we will consider a data-independent smoothing penalty such as in [6, ch. 4.4] to stabilise the estimation in general, i.e. also, and especially, for single scattering (see Appendix A.2).

The effect of early/biased stopping is arguably still somewhat noticeable in our estimates. It was accepted primarily as a way to make the optimisation stop in the first place as it is based on noisy gradient estimates and it would be non-trivial to obtain steps that are convergent with a small tolerance. Variance reduction methods which can accelerate the convergence of stochastic optimisation methods related to the idea of control-variates [55] are known and based on averaging of gradients [35, 73, 126] and Hessian matrices [105]. Although they may have desirable properties, with the present code basis these methods would require too many evaluations and therefore weren't deemed feasible at this stage. Our path sampling strategy, which uses the same number of samples in each direction, is sub-optimal because it would make sense to estimate those directions and time bins that are “important” more accurately. Intuitively it seems plausible that the measurements in direction v_i and time t_j with low photon count will not have a large effect on the fitted values and therefore one would not expect that a very good estimate $\hat{P}_{i,j}$ for $P_{i,j}$ is required. In Eq. (5.18) we ultimately seek to find a (weighted) least squares solution to an over-determined system of (non-linear) equations. For linear systems it is well known that leverage score weighted importance sampling of rows leads to subspace embeddings and probabilistic bounds for the quality of the least squares solution (see e.g. [96, 149]). The computation of the importance weights is in our case not a computational challenge but obtaining the matrix itself is computationally expensive. One possibility would be to estimate the leverage from single scattering or, alliteratively, based on an estimate obtained with an equal amount of paths per direction, i.e. similar in spirit to what was proposed in [30] in the context of matrix approximation by row sampling. Although our random approximations don't rely on sub-sampling of rows/equations and the problem is non-linear, each path of length t_j starting in direction v_i contributes to the accuracy of $\hat{P}_{i,j}$ and its derivatives. Similarly, the steps in the optimisation procedure are such that two tall matrices are multiplied and essentially a spectral approximation of the product is sought which would also benefit from such importance sampling. The latter operations are of course linear which is no surprise given that they are similar to Gauss-Newton schemes in appearance and derivation.

5.1.4 Image approximation errors

Up to this point we have primarily discussed randomisation and errors caused by inaccurate/noisy optical measurements. In order to paint a full picture we should also consider the error due to inexact approximation of the plume $u[\theta]$ as in Eq. (2.18) and, by extension, the optical parameters $\xi[\theta] = (\alpha[\theta], \sigma_a[\theta], \sigma_s[\theta], f_p[\theta])$ that satisfy assumptions as outlined in Eqs. (3.1) to (3.3). For simplicity we will ignore the (constant) ambient terms whose approximation error obviously depends on how much they deviate from a constant, and focus on the perturbation of interest and the assumptions that are made in the subsequent experiments in Section 5.2. In that situation the perturbation caused by the plume can be expressed in the form

$$u[\theta](x, t) = \int_0^t u_\delta[\theta](x, s) ds = \int_0^t w[\theta](s) \phi\left(\frac{\|x - b[\theta](s)\|_2}{h[\theta](s)}\right) ds \quad (5.39)$$

$$\approx \sum_{j=1}^{N_\phi} w_j[\theta] \phi\left(\frac{\|x - b_j[\theta]\|_2}{h_j[\theta]}\right) \quad (5.40)$$

where $u_\delta[\theta]$ as before is an instantaneous puff that solves Eq. (2.11) for a point source $\rho[\theta] = \rho_0 \delta(\cdot - b^0)$ at $b^0 \in \mathbb{R}^3$, drift $\eta[\theta] : (0, t) \rightarrow \mathbb{R}^3$, isotropic diffusion coefficient $\kappa[\theta] : (0, t) \rightarrow (0, \infty)$ and $t > 0$ is the time since the release started. Note that although it describes time, the relevant scale is of a very different order than that relevant for the Lidar measurement and the plume is assumed time-independent for the duration of the optical measurement. We can therefore think of t as a fixed parameter satisfying $t \approx \infty$ for steady-state plumes¹. We recall the continuous functions $b[\theta] : (0, t) \rightarrow \mathbb{R}^3$ and $h[\theta] : (0, t) \rightarrow (0, \infty)$ are as in Eqs. (2.14) and (2.15)

$$b[\theta](t) = b^0 + \int_0^t \eta[\theta](s) ds \quad (5.41)$$

$$h[\theta](t) = \sqrt{\int_0^t \kappa_0[\theta](s) ds}, \quad (5.42)$$

¹Since we measure on compact domains t can always be assumed finite.

while $w[\theta](t) = \frac{\rho_0}{h[\theta](t)}$ and we recall from Eq. (2.19) that one can associate the approximation Eq. (5.40) with a discretisation

$$(w_j = w[\theta](t_j)\Delta_{\text{step}}, b_j = b[\theta](t_j), h_j = h[\theta](t_j))_{j \in 1:N_\phi} \quad (5.43)$$

of those curves for points $t_j = \frac{\Delta_{\text{step}}}{2} + j\Delta_{\text{step}}$, $j = 0, \dots, N_\phi - 1$ and $\Delta_{\text{step}} = \frac{t}{N_\phi}$. Note that the η_1 - and η_2 -components of Eq. (5.41) are assumed known and constant (in time) in order to resolve the ambiguities outlined described in Eq. (2.21). This is a significant *a priori* constraint, similar to the assumptions made in (Gaussian) plume models as initially mentioned in the introduction, specifically Fig. 1.4, and imply that $u_\delta[\theta]$ evolves along a 1-dimensional line in \mathbb{R}^3 . In order to better explain the assumption made in our numerical studies we give a more explicit characterisation of the operator \mathcal{G} from Eq. (3.10). We want to partition θ into

$$\theta = (\theta_s, \theta_0, \theta_{\text{source}}, \theta_{\text{up}}, \theta_h) \quad (5.44)$$

such that θ_{source} are the x_1 - and x_2 -coordinates of a ground based source at $b^0 = (\theta_{\text{source}}, 0)$, $\vec{\eta}_{\text{amb}}$ is the known/measured ambient advection, i.e. the wind, \vec{e}_3 is the unit vector in x_3 direction and we can express

$$b[\theta](s) = b^0 + s\vec{\eta}_{\text{amb}} + \vec{e}_3 \mathcal{L}[\exp(\theta_{\text{up}}) \mid t](s) \quad (5.45)$$

$$h[\theta](s) = \mathcal{L}[\exp(\theta_h) \mid t](s). \quad (5.46)$$

Here $\mathcal{L}[y \mid t]$ is a 1-dimensional linear spline computed from nodes $(0, 0)$ and $\left(\frac{i}{\dim(y)}t, y_i\right)$ for $i = 1, \dots, \dim(y)$ and the element-wise exponentiation ensures positivity for unconstrained parameter vectors. Note that if we put $\alpha[\theta] = \exp(\theta_0)u[\theta]$ and $\sigma_s[\theta] = \exp(\theta_s)u[\theta]$ then $\theta_\alpha = (\theta_0, \theta_{\text{source}}, \theta_{\text{up}}, \theta_h)$ from the previous section is such that u as well as the differential absorption α are a function of the sub-array θ_α . The parameter θ_0 plays the same role as in Section 3.2.3, except for being on a logarithmic scale, while θ_s controls the amount of scattering from the plume in a way similar to θ_2 in Eq. (3.38). If we additionally assume that f_p is fixed and, in order to keep the complexity more manageable,

also assume that $\sigma_a = 0$ for the reasons outlined in Section 3.2.3 then this completely characterises an operator \mathcal{G} and without loss of generality we can normalise u by setting $\rho_0 = t^{-1}$ in the point source from Eq. (2.12).

If \mathbf{u}_δ is a general function, e.g. a solution to Eq. (2.11) but for a source term given by multiple and well separated point releases, then the discrepancy between the simple puff $u_\delta[\theta]$ and \mathbf{u}_δ can get arbitrarily large regardless of how we select (ρ, η, κ) . We therefore limit the discussion to situations where the ground truth function

$$\mathbf{u}(x, t) = \int_0^t \mathbf{u}_\delta(x, s) ds \quad (5.47)$$

is also a gas release from a point source and consequently can also be thought of as evolving through space on a 1-dimensional curve but whose moments may not match Eqs. (5.41) and (5.42) exactly. It is worth noticing that although the theory that was developed in Chapters 3 and 4 deals with functions of the same form as the right hand side of Eq. (5.40), in many instances no such assumptions were needed or made for the results to be valid. The specific structure imposed in the experiments is not only required in order to ensure that the estimation is robust against measurement noise but also benefits the optimisation algorithm for reasons which will become more evident in the following. The error between a possible ground truth function \mathbf{u} from Eq. (5.47), not necessarily parameterised by θ , and $u[\theta](x, t)$ as in Eq. (5.39) has essentially three different contributors as discussed below.

Quadrature error: Perhaps the simplest and most obvious error occurs when going from Eq. (5.39) to Eq. (5.40) in which case we are approximating an integral by finite sums. The integration can be thought of in a point-wise manner or as an integration problem in the Banach space $L_1(\mathbb{R}^3)$. The KL-divergence between two Gaussian functions with mean $b_1, b_2 \in \mathbb{R}$ and variance h_1^2, h_2^2 , see e.g. [110, p. 33], is

$$\text{KL}(\text{Normal}(b_1, h_1^2) \mid \text{Normal}(b_2, h_2^2)) = \frac{1}{2} \left(\frac{(b_1 - b_2)^2}{h_2^2} + \frac{h_1^2 - h_2^2}{h_2^2} + \log \left(\frac{h_2^2}{h_1^2} \right) \right) \quad (5.48)$$

and thanks to Pinsker's inequality we know that the distance in L_1 norm is bounded by $[2\text{KL}(\text{Normal}(b_1, h_1^2) \mid \text{Normal}(b_2, h_2^2))]^{1/2}$. As such the quadrature error at a given discretisation step $t_j - t_{j+1}$ heavily depends on $h[\theta](t_j)$ and in order for it to be small we want $\|\eta(t_j)\|_1(t_j - t_{j+1}) \approx \|b[\theta](t_j) - b[\theta](t_{j+1})\|_1 \ll h[\theta](t_j)$. However, in the situation of Eq. (5.42) we have $h[\theta](t) \searrow 0$ as $t \searrow 0$. In order to ensure compatibility with numerical integration as in Eq. (5.40) as well as the RTE we have replaced $\rho[\theta] = \delta(\cdot - b^0) \approx \frac{1}{h_0} \phi\left(\frac{\cdot - b^0}{h_0}\right)$ which can be seen as a “bloating” of the source that is compatible with Gaussian solutions Eq. (2.13). In reality the source will of course never be a point so this arguably isn't a major restriction. For our numerical experiments this error was $\lesssim 3\%$ when measured in L_1 norm and, relative to the other error contributions discussed below as well as the optical noise, rather small (see also Table 5.3). Although not done in this work, by means of adaptive quadrature grids one could likely reduce this further with minimal increase in computational complexity.

Moment approximation error: Regardless of the underlying atmospheric transport model we can always form the time-dependent first and second moments

$$\mathbf{b}(s) = \int_{\mathbb{R}^3} x \mathbf{u}_\delta(x, s) dx \in \mathbb{R}^3 \quad (5.49)$$

$$\mathbf{h}^2(s) = \int_{\mathbb{R}^3} x x^\top \mathbf{u}_\delta(x, s) dx \in \mathbb{R}^{3 \times 3} \quad (5.50)$$

of an instantaneous release \mathbf{u}_δ as in Eq. (5.47). A reasonable assumption for Eqs. (5.49) and (5.50) is continuity so that in the case where θ_{up} contains the values of $b_3[\theta] = \vec{e}_3 \cdot b[\theta]$ at $N_{\text{spline}} = \dim(\theta_{\text{up}})$ knots of a piece-wise linear function as described in Eq. (5.45) the approximation error $|\mathbf{b}(s) - b[\theta](s)|$ will depend on:

- The accuracy of the assumed wind measurement $\vec{\eta}_{\text{amb}}$ in the x_1 - x_2 component²
- The number N_{spline} of degrees of freedom in the linear spline

Of course there is an inherent trade-off between accuracy and regularity here since an

²In situations where the measured wind directions are very variable one could extend the spline approximation to each component of the advection term b

increase in the degrees of freedom N_{spline} will also make the estimator more prone to overfitting. This becomes particularly problematic when considering that we want to iterate Eq. (5.20) for which we need to invert a matrix $\mathbf{H}^\#$ that is approximated by MC sampling and becomes (almost) singular for $N_{\text{spline}} \rightarrow \infty$. To see this we may again look at Eq. (5.48) and realise that as the spline knots get closer to each other the sensitivity to each value decreases or, more precisely, the derivative in the direction $(\theta_{\text{up}})_i - (\theta_{\text{up}})_{i+1}$ tends to zero. In order to have a stably converging algorithm some care must be taken to select $N_{\text{spline}} = N_{\text{spline}}(t, \vec{\eta}_{\text{amb}})$ depending on the wind and time since release, which together determine the “size” of the plume, so that the points corresponding to knots on $b[\theta]$ are sufficiently far apart. Since the primary focus of this work was on the optical part of the inverse problem, we selected the parameterisations for the experiments in Section 5.2 in a way so that the estimators become reasonably stable when some (mild) additional regularisation is employed. As we have found the errors in situations without modelling discrepancies to be of a similar order as in the perturbed case we are likely not incurring excessively large errors due to under-fitting. The situation regarding the piece-wise approximation for $\mathbf{h}^2(s)$ is similar but here we are assuming isotropic covariance matrices. Consequently we should expect an additional error due to anisotropy. Numerically it would not be particularly problematic to implement a more flexible structure for $h[\theta]$ but our theory heavily relies on the spherical nature of the kernel functions. This may however be an artifact of the techniques used in the proofs rather than an actual requirement.

Turbulence induced error: Lastly, the assumption that $u_\delta[\theta]$ solves Eq. (2.11) is in reality not going to hold true exactly and the transport process that generates \mathbf{u}_δ will induce an additional approximation error on top of the previous two. As mentioned in the introduction, we think of $u[\theta]$ as an average which is motivated by a distribution of maximum entropy subject to moment constraints. The error due to the use of $u_\delta[\theta]$ in place of \mathbf{u}_δ will generally depend on various factors that affect the turbulent atmospheric transport process in an intractable manner. The assumption made in this work is that this residual error will “average out” to a sufficient degree for $u_\delta[\theta]$ and $u[\theta]$ to be reasonable approximations of \mathbf{u}_δ and \mathbf{u} .

For the numerical examples in the next section we have implemented deviations \mathbf{u} that depend on a branching Ornstein-Uhlenbeck process with jumps which to some extent incorporates all the described sources of error. In particular, Δ_{step} used to integrate the paths was smaller, the centre lines and covariances Eqs. (5.49) and (5.50) deviate from the parametric forms in a non-trivial manner and the resulting \mathbf{u}_δ are multi-modal Gaussian mixtures. The procedure is outlined in Appendix A.2.2.

5.2 A numerical example

In the following, we give some numerical examples in order to highlight some aspects of the previously developed theory and give a more concrete idea of what can potentially be expected from the suggested method. The structure and parameterisations will follow that of the previous section and the measurement grid of size $N_{\angle_{\text{rot}}} \times N_{\angle_{\text{el}}} \times N_t$, i.e. $N_v = N_{\angle_{\text{rot}}} N_{\angle_{\text{el}}}$ is the number of directions, is obtained from a discretisation of the box corresponding to $Z_{[-]}^{\text{box}}(x_D)$ as introduced in Eqs. (4.11) and (4.12), see also Fig. 3.1.

5.2.1 Bias & convergence rates

As we ignore errors in the model due to discrepancies that cannot be captured by a finite-dimensional set of parameters and solve the optimisation problem with a stochastic solver which might stop too early in conjunction with a biased initial value, it is to be expected that the error (in case of wider FOVs) will in general not vanish with increasing energy or number of pulses used. In reality, it is arguably more important how big, rather than of what nature and composition, an error is. Given the additional sources of bias, one would however expect a greater benefit in situations where the error itself is rather large which is of course somewhat concerning unless the *additional* biases are sufficiently small. Analysing complicated images through simulations is computationally costly but in order to get a sense of how large these effects will potentially be, we can look at the less demanding single kernel reconstruction, i.e. the “building block” u_δ , of a plume. This corresponds to the situation from the previous section except that $b[\theta], h[\theta]$ are given as

Table 5.1: System parameters used in the single puff simulation

Resolution	$20 \times 20 \times 50$
Detector	3cm lens with 4% detection rate
Methane amount	15.6mol or 0.25kg
Distance	100m
FOV width at plume	30m
Wavelength (absorbing)	1645.55nm
Pulse Energy	250μJ ($\times 10$ and $\times 100$ in the convergence plots)
Ambient intensity	0.025W uniformly over hemisphere

points rather than by means of a spline as in Eqs. (5.45) and (5.46).

The reconstruction errors shown as part of Figs. 5.1 and 5.2 are subject to minimal discretisation artifacts and otherwise very small modeling errors. The reconstructions were run with 3 restarts, each with a different phase function given by a HG-density with parameter equal to 0, 0.35 and 0.70 respectively, while the simulations use 0.25. Aside from that the parameter θ_α of interest has five components, three positional and one for width and weight, with an additional parameter $\theta_s = \log(\int_X \sigma_s[\theta](x)dx)$ for the weight of σ_s which is otherwise proportional to the differential absorption α while $\sigma_a = 0$ throughout. This means the target of the optimisation $\theta = (\theta_\alpha, \theta_s)$ has six components. We also consider an optical thickness akin to Eq. (3.40)³ defined

$$\mathcal{S}(\theta_s) = \exp(\theta_s) \int_{-\infty}^{+\infty} \frac{1}{h[\theta]} \phi \left(x_D + r \frac{x_D - b^0}{\|x_D - b^0\|_2} \right) dr \quad (5.51)$$

which, due to low ambient scattering, makes up essentially to the total optical thickness of the image. For data from multiple FOVs Fig. 5.1 shows relatively stable reconstructions for the weight parameter of $\theta_0 = \log(\int_X \alpha[\theta](x)dx)$ with errors slightly above 10%. The primary benefits of multiple and wider FOVs are, as expected, observed for plumes of larger optical thickness. The initially worse performance of wider FOVs without separation is primarily to be attributed to the presence of ambient noise which is filtered by the narrow FOV and thus doesn't affect the reconstruction.

³The parameter θ_2 in Eq. (3.40) is, unlike θ_s , on a natural scale and “absorbed” the scaling by the kernel width h . The total optical thickness in that instances has an additional ambient component.

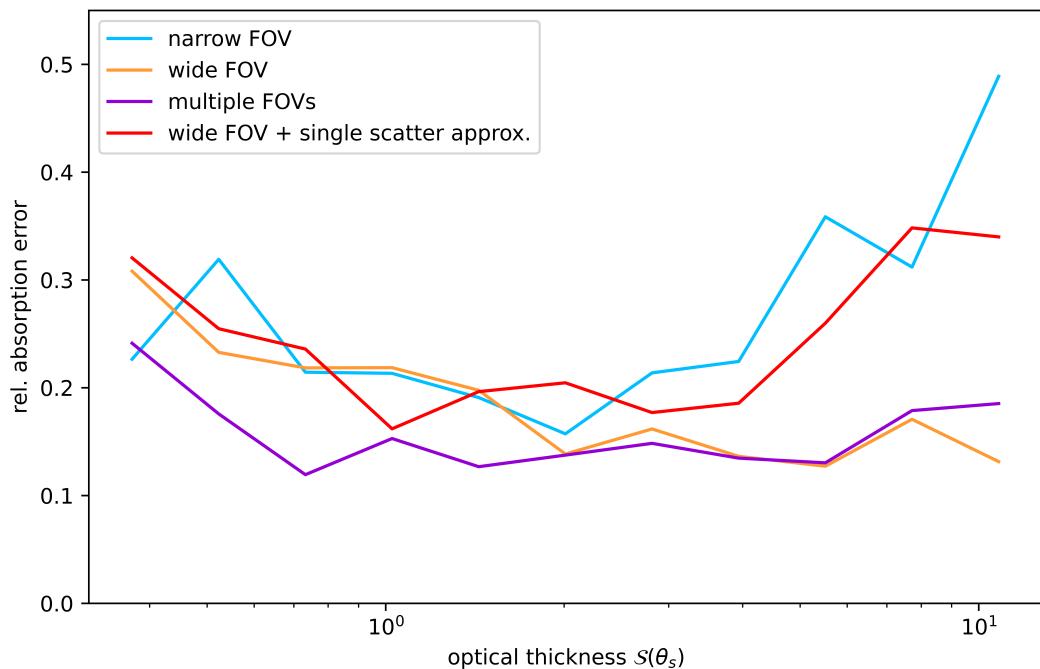


Figure 5.1: Relative errors $|\exp(\hat{\theta}_0 - \theta_0) - 1|$ (see also Eq. (5.44)) averaged over 25 independent samples. Simulation was done with $250\mu\text{J}$ pulses released in 20×20 directions and an instrument resolving 50 time bins spread across the region containing the bulk of the gas and system parameters as in Table 5.1. Here “single scatter approx.” refers to the use of Eq. (3.13c) instead of an RTE model.

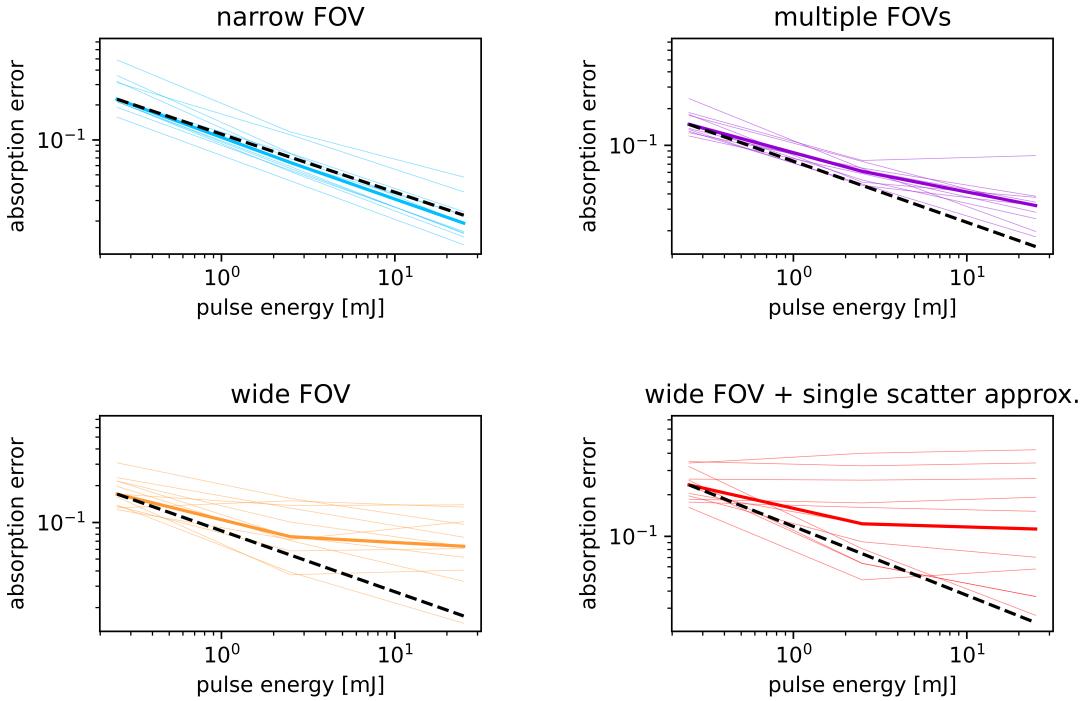


Figure 5.2: Convergence rates of absorption estimates. Each thin line plots relative absorption errors $|\exp(\hat{\theta}_0 - \theta_0) - 1|$ (see also Eq. (5.44)) averaged over 25 independent samples for fixed value of optical thickness $\mathcal{S}(\theta_s)$ against pulse energy. The thick line is their median. The dashed black line corresponds to the behaviour of an unbiased parametric estimator and is anchored at the median error of the values shown in Fig. 5.1. Here “single scatter approx.” refers to the use of Eq. (3.13c) instead of an RTE model.

As seen in Fig. 5.2 the error for the narrow FOV estimates behaves exactly as we would expect for a parametric estimator following a line with slope -0.5 very closely. Consequently, and as expected, the narrow FOV estimates end up outperforming the other estimators for sufficiently large energies. The straight lines observed in the single scattering approximation with wide FOV data indicate that at high energies the error is purely driven by bias with very little variance due to optical noise. The reconstruction with multiple FOVs is still competitive at errors below 10% and energies of 2.5 mJ but appears to bottom out somewhere between 2% and 5%. This is consistent with the choice of shrinkage penalty which was chosen to have an effect of 2.5% in the high energy limit whereas it is bounded by the estimated standard deviation in the case of lower energies.

Fig. 5.3 confirms that the observed errors in case of RTE-based estimators are also induced by asymptotically non-vanishing biases rather than variance incurred due to instabilities in the stochastic solver or uncertainty in the additional parameter θ_s that determines the scattering rate σ_s . Structurally the bias at lower optical thickness is dominated by the ridge term which causes underestimation of the absorption whereas at larger optical thickness it becomes increasingly difficult to fit the parameter θ_s which results in overestimation. It should be noted that the prior was increased along with the pulse energy, i.e. line 16 in Algorithm 2 was omitted in the higher energies, so this isn't entirely unexpected. Had this not been done, then the variance in case of RTE-based estimators would certainly be higher but it was found that the errors do not tend to improve as it is not easy to fit this parameter properly with our sampling/optimisation method.

5.2.2 Steady state plume & turbulence

In order to further validate our method on simulated data we consider a simulated reconstruction from $45 \times 15 \times 50$ Lidar scan of a 14 parameter dispersion, similar to the ones above but the location and width are now parameterised by a spline as in Eqs. (5.45) and (5.46) (see also Appendix A.2). The system parameters in Table 5.2 are such that some high-level features can be recovered but conventional (point-wise) reconstruction fails due to the low signal-to-noise ratio. As suggested by our developments in Section 3.2, we only consider instances where the scattering is caused largely by particles around the gas plume and the effectively required resolution is roughly granularity of absorbing gas within the scatterer.

In order to avoid contrived scenarios we have fixed the phase function f_p to different values for simulation and reconstruction respectively in the same way as in the previous subsection, i.e. a HG function with parameter 0.25 for the simulations and restarts with values in $\{0.70, 0.35, 0\}$. The fixed system parameters were selected in an attempt to mimic somewhat realistic conditions as if a DIAL instrument for methane measurement was used for the experiment.

In addition to the optical noise, we have also considered perturbations of the disper-

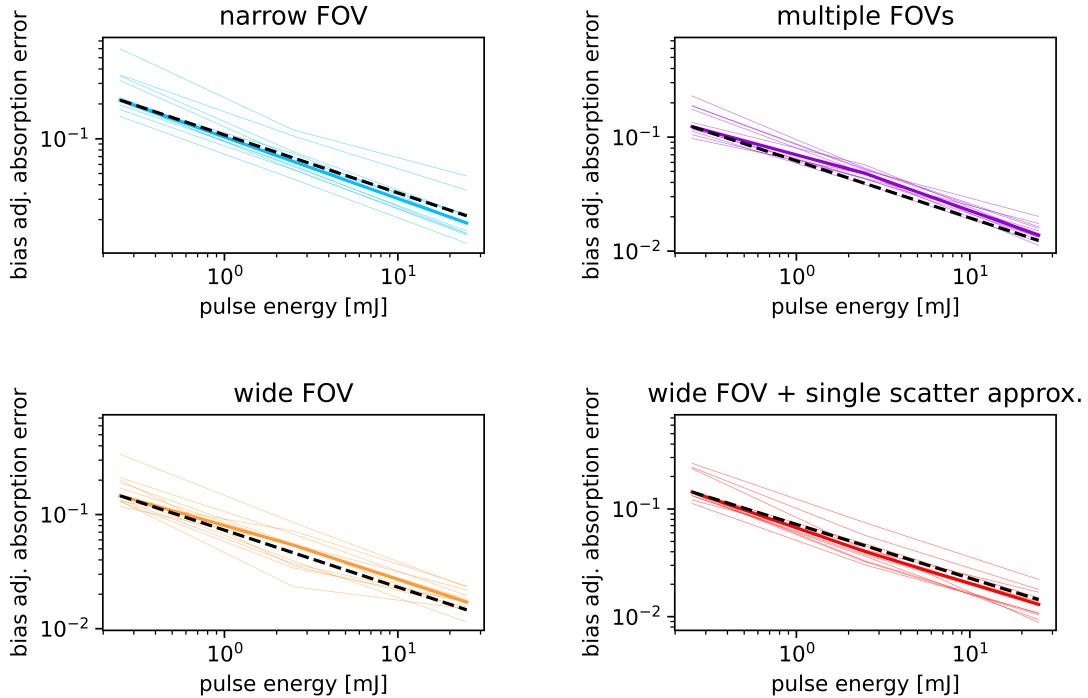


Figure 5.3: Bias adjusted convergence rates of absorption estimates. Each thin line plots bias adjusted relative absorption errors $|\exp(\hat{\theta}_0 - \bar{\theta}_0) - 1|$ averaged over 25 independent samples with sample mean $\bar{\theta}_0$ for fixed value of optical thickness $\mathcal{S}(\theta_s)$ against pulse energy. The thick line is their median. The dashed black line corresponds to the behaviour of a parametric estimator and is anchored at the initial bias adjusted median error. Here “single scatter approx.” refers to the use of Eq. (3.13c) instead of an RTE model.

Table 5.2: System parameters used in the steady-state simulation

Resolution	$45 \times 15 \times 50$
Detector	3cm lens with 1.78% detection rate
Methane amount	50mol or 0.8kg
Distance	125m
FOV width at plume	20m
Wavelength (absorbing)	1645.55nm
Pulse Energy	$200\mu\text{J}$
Ambient intensity	0.025W uniformly over hemisphere

sion model via a super-position of branching jump-diffusion processes. We recall that prior to the introduction of ψ the misfit between observed photon counts and parametric intensities could quickly get very large in situations with plenty of scattering once the shape of the turbulent plume doesn't match the parametric form. This issue, alongside good flexibility in the backscattering coefficient of the phase function, was mitigated by the non-parametric component ψ . Of course the parametric form of the differential absorption $\alpha[\theta]$ still won't match the turbulent plume but given that the differential data is affected by turbulence only through line integrals we can take advantage of properties such as Eq. (4.52) which suggest that the reconstructions may now be less affected by the perturbations. Nonetheless, we want to investigate the ability of our method to adapt to situations where we have discrepancies that aren't accounted for by the Poisson distribution. The “magnitude” of observed noise will in general depend on atmospheric conditions, which affect the amount of turbulence within the dispersion process, and the amount of (temporal) averaging done as part of the data acquisition. In particular, the plume will in practice be a dynamic object and for a stationary release rate its temporal average will, if taken over sufficiently long periods, resemble some low-dimensional simpler average model. For the purpose of these simulations we did not simulate a dynamic turbulent dispersion process, mainly due to computational constraints, and instead used a realisation of a random process as midpoints which combined with suitably chosen kernel widths will in expectation satisfy Eq. (2.11). The resulting method can be thought of as a snapshot of a process which aims to mimic the empirical observation that “Big whorls have little whorls which have lesser whorls...” in a turbulent dispersion process. For further details regarding the noise structure as well as the hyper-parameters and regularisers used in the fitting of the dispersion parameters we refer to Appendix A.2.

Table 5.3 shows different error metrics for various situations and measurements using a single pulse per direction with nFOV, wFOV and mFOV denoting narrow, wide and multiple FOVs respectively. The first column indicates whether the plume was perturbed additionally via the random process described in Appendix A.2.2. “Yes” therefore indicates a higher amount of noise in the measurement and higher errors are to be expected. The level of optical noise depends on the optical thickness in a non-trivial way. While the

number of photons increases with more scattering particles the increased optical thickness means that back-scattered light does not penetrate the plume as much which in turn results in a lower differential absorption. Qualitatively the behaviour will be roughly as in Fig. 5.1 which is to say the error will grow as the optical thickness tends to 0 or ∞ . In the former case, the intensity of measured light will tend to 0 resulting in growing relative errors. If the optical thickness grows infinitely large, the plume will turn into a hard target and eventually there will be no differential absorption. Consequently, we should expect that there is a value, depending on the FOV, for which the error is minimal.

Table 5.3: Relative reconstruction errors obtain as average from 20 simulation runs per row. Release amount corresponds to the relative error in the component $\exp(\theta_0)$ as in Eq. (5.44). Optical thickness is measured as before in Eq. (5.51) with multiplier $\exp(\theta_s)$ but here the average over all kernels is used.

Dispersion noise	Opt. thickness	L_1 -errors (nFOV—wFOV—mFOV)	Release amount errors (nFOV—wFOV—mFOV)
Yes	1.0	(46% — 39% — 37%)	(22% — 17% — 16%)
Yes	2.6	(52% — 42% — 40%)	(20% — 17% — 15%)
Yes	4.2	(59% — 40% — 37%)	(24% — 17% — 15%)
Yes	5.7	(66% — 34% — 33%)	(36% — 15% — 13%)
No	1.0	(35% — 30% — 29%)	(15% — 13% — 13%)
No	2.6	(37% — 23% — 23%)	(14% — 10% — 11%)
No	4.2	(39% — 20% — 19%)	(16% — 9% — 10%)
No	5.7	(61% — 21% — 21%)	(29% — 12% — 11%)

Table 5.3 shows a summary of the simulation with the aforementioned parameters and values for the optical thickness “averaged” across the plume centre-line. It suggests that the wider FOV tends to outperform a like-for-like reconstruction that uses a narrow FOV only. In fact, the average improvement with respect to L_1 and release rate is 35.2% and 33.7% respectively. Note that the reduction in error from an additional (statistically independent) narrow FOV measurement would be 29.3%, i.e. an accuracy much closer to what one might expect from two sets of measurements (instead of one). Of course, this does not take accurately into account how the turbulent component might behave nor is the improvement equally distributed but rather depends heavily on the optical thickness and the resulting photon counts in the wide FOV. We emphasise that the reported errors should not be taken as an absolute indication of what can be expected in practice as this

will be highly dependent on the system parameters (see Table 5.2). As long as the ambient intensity does not attain levels comparable to the signal response, i.e. we have a high enough pulse energy, we may use our findings as an indication of the relative improvement that the proposed measurements can have compared to classical narrow FOV methods. What can be said with a fair degree of confidence is that those types of improvements can, if at all, *only* be expected for errors of a certain magnitude. Running the simulation with $\times 100$ the energy per pulse results in a behaviour much like Fig. 5.2, i.e. the narrow FOV errors decay faster and eventually there is no point in using an RTE-based forward model and the corresponding data.

Figs. 5.4 and 5.5 show an instance that can be thought of as a best-case scenario with high improvement (line 4 in Table 5.3). The primary benefits are found in the reconstruction of the release rate as well as close to the source. This is to be expected as the release rate is the “smoothest” possible parameter while the source location also has a global effect on the image. The errors in the centre-lines increase with the distance from the source as the plume widens which results in less (multiple) scattering along with a lower sensitivity to the relevant parameters due to increased smoothness. The behaviour of the narrow FOV estimates as shown in Fig. 5.4 is the primary reason why a smoothing penalty for the spline, which parameterises the centre-lines, was necessary. Note that the improvements are observed in the instance with arguably the lowest absolute errors among turbulent plumes which suggests that the unknown scattering, which is only partly compensated by the non-parametric nuisance component, doesn’t cause any notable bias in the reconstruction. A similar pattern would be observed for the width of the plume which is not shown explicitly but large errors would have a negative effect on the L_1 errors in Table 5.3. It should be noted that multiple and wide FOV reconstructions occasionally suffer from outliers, often caused by getting stuck in single approximations, which can skew the averages towards larger errors. The re-starting approach along with the ridge terms largely takes care of these issues but it has to be acknowledged that the increased amount of randomness from the RTE solver paired with a lack of information regarding the scattering parameters appears to cause such phenomena with non-negligible regularity and at present we have no way to entirely avoid these artifacts. As such they should not,

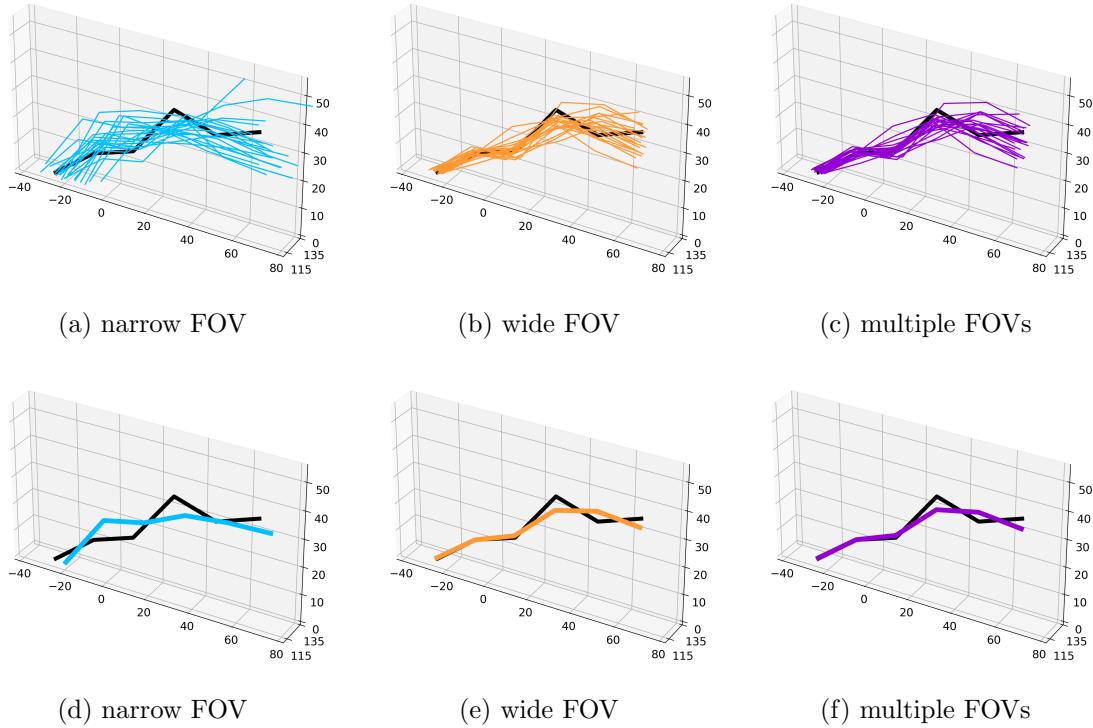


Figure 5.4: (a)–(c): Reconstructed plume centre-lines $b[\hat{\theta}]$ as in Eq. (5.45) (20 samples) for a turbulent plume and the scattering particle concentration corresponding to the instance from line 4 in Table 5.3. (d)–(f): Sample mean of reconstructed centre-lines. Ground truth centre-line $b[\theta]$ shown in black. The deviation of the sample mean from the ground truth towards the “end” of the plume is likely due to the regulariser used for smoothing the spline. Since the same smoothing parameters were used in all instances and uncertainty in the narrow FOV, compare (a) to (b) and (c), is higher in situations with large amounts of scattering, such as this one, its effect is expected to be more pronounced in (d).

and were not, disregarded from performance metrics.

An instance with less improvement is shown in Figs. 5.6 and 5.7. It should of course not come as much of a surprise given that it is also the instance with the smallest amount of multiple scattering recorded (the turbulent perturbations cause localised spots of higher optical thickness and sometimes more light in the wider FOV). As the worsening performance, relative to the narrow FOV data, of wide and multiple FOVs is caused by a lack of scattering it should be expected that the added uncertainty due to unknown parameters in the wider FOV will outweigh the potential benefits if the amount of scattering decreases further. In fact, if we were to consider the log-RMSE instead of the average of relative deviations for the release amount, then the performance of the narrow FOV would

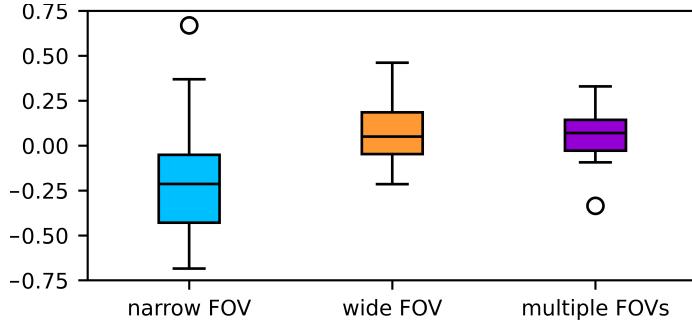


Figure 5.5: Relative deviation $\exp(\hat{\theta}_0 - \theta_0) - 1$ (see Eq. (5.44)) of reconstructed release rates $\exp(\hat{\theta}_0)$ corresponding to the same instance as shown in Fig. 5.4. Line 4 in Table 5.3 shows the average of absolute values.

be better than that of the wide FOV (albeit slightly). Such an instance would roughly correspond to the bottom left corner of Fig. 3.5 and although the difference seems small, narrow FOV measurements should be expected to outperform wide FOV in “extreme” cases just as much as wide FOVs may have benefits in the opposite situations.

Even though the condition from Eq. (3.14) has not been implemented specifically and injectivity of the forward operator is not guaranteed by Theorem 3.1.4, whose proof heavily relies on the tail behaviour of the kernels, the perturbation of the plume, which causes particularly large (relative) errors in the tail, seems to affect the narrow FOV data more so than the wide. Such an observation can be explained intuitively. The condition in Eq. (3.14) is a technical requirement relevant only in what can be thought of as a best-case scenario which, in particular, assumes accurate measurements. The narrow FOV, at least in instances of high optical thickness (see line 3 and 4 in Table 5.3), collects more light scattered at the “outside” of the plume and is therefore affected by the model errors in that region to a greater degree.

5.2.3 Uncertainty estimation

In view of the errors in Table 5.3 it seems plausible that, in addition to a point estimate, one might seek an estimate for the observed uncertainty as it most definitely isn’t negligible in the present case. To that end Figs. 5.8 to 5.11 show the reconstructed values along

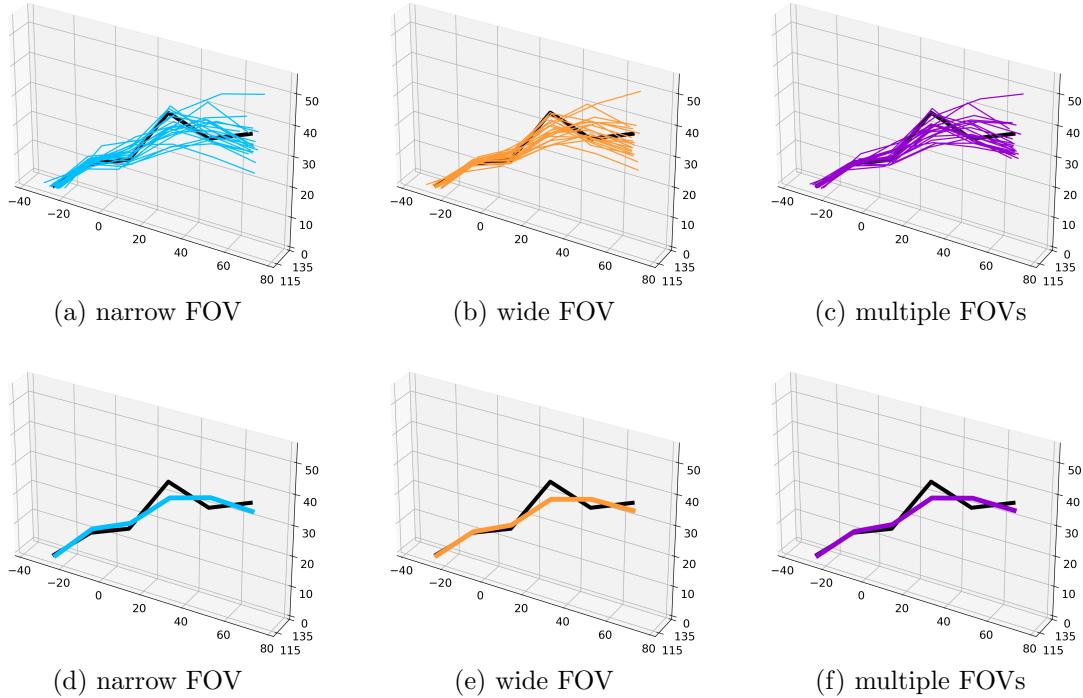


Figure 5.6: (a)–(c): Reconstructed plume centre-lines $b[\hat{\theta}]$ as in Eq. (5.45) (20 samples) for a turbulent plume and the scattering particle concentration corresponding to the instance from line 5 in Table 5.3. (d)–(f): Sample mean of reconstructed centre-lines. Ground truth centre-line $b[\theta]$ shown in black. The deviation of the sample mean from the ground truth towards the “end” of the plume is likely due to the regulariser used for smoothing the spline. Since the same smoothing parameters were used in all instances and data/uncertainty in the narrow FOV, compare (a), (b) and (c), is similar for each measurement in situations with low amounts of scattering, such as this one, the effect of smoothing is expected to be very similar.

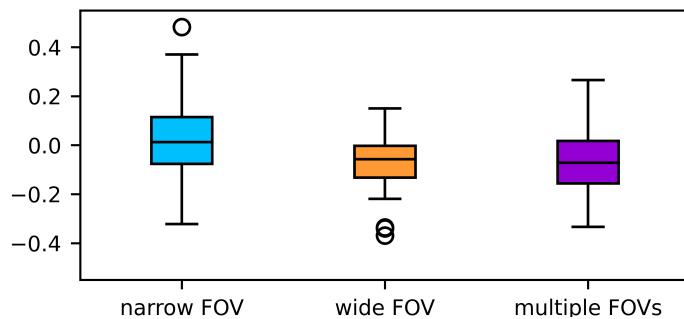
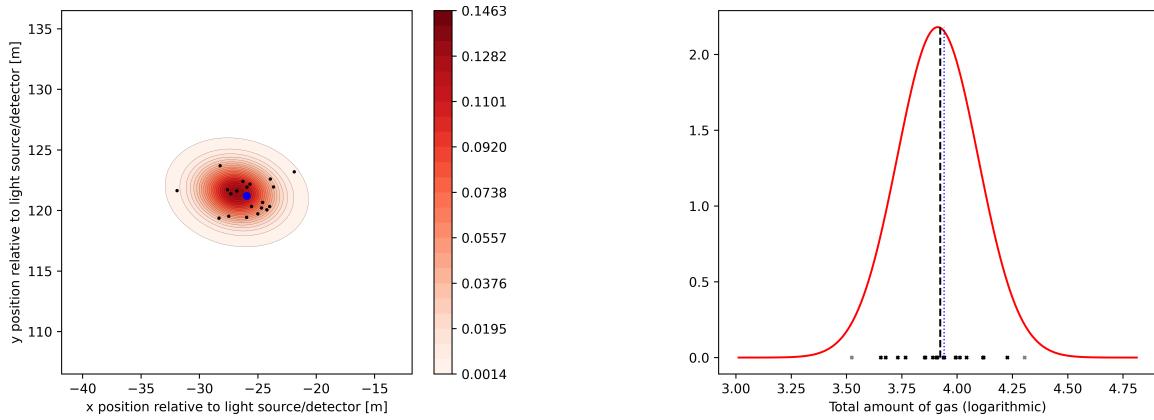


Figure 5.7: Relative deviation $\exp(\hat{\theta}_0 - \theta_0) - 1$ (see Eq. (5.44)) of reconstructed release rates $\exp(\hat{\theta}_0)$ corresponding to the same instance as shown in Fig. 5.6. Line 5 in Table 5.3 shows the average of absolute values.

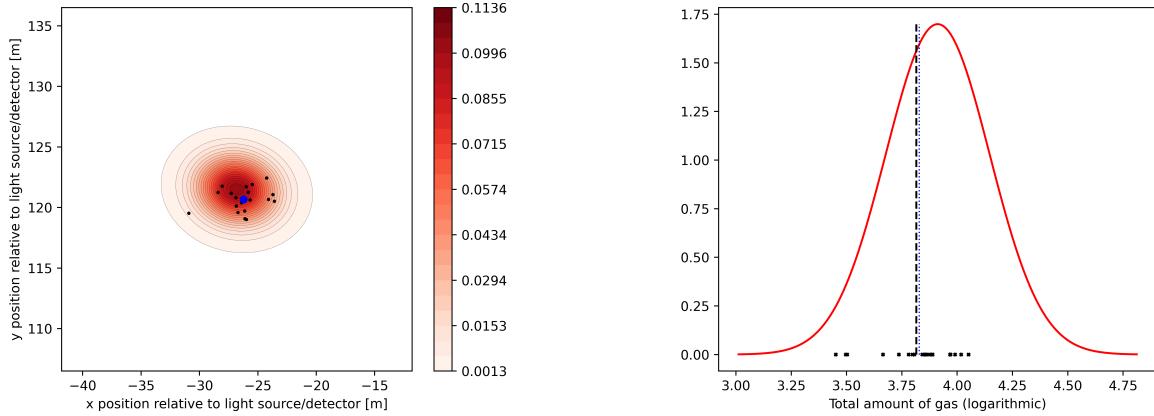
with (asymptotic) confidence regions centred around their ground truth values with the exception of the scalar parameter component θ_s that models the optical thickness of the plume. In the numerical experiments reconstructed values for θ_s are subject to large uncertainties and far less accurate than the rest. One may pick θ_s “conservatively”, which in this instance means larger than it probably is (see Appendix A.2 for a more formal interpretation) as it tends to result in overestimation of the confidence bounds. Turbulent or not, the covariances were estimated from Eq. (4.69) and only the first summand of the covariance matrix is taken if there is no turbulence.

It is admittedly rather difficult to assess the quality of the shown results given the small number of samples although it would, unsurprisingly, appear that the confidence estimates in case of low optical thickness, Figs. 5.8 and 5.10, and narrow FOVs are better than the ones obtained for wide FOVs which tend to be too large. For high optical thickness, Figs. 5.9 and 5.11, the narrow FOV errors are not accounted for correctly, which might be down to an inappropriate use of asymptotic distributions as those are only rigorously justified in the limit for small errors (given how this was observed for the non-turbulent case to a similar extent). The wide FOV agrees relatively well in the non-turbulent instances but again shows less variability than what the estimates would suggest if turbulence is present. This might be down to the use of a single scattering approximation because, given the additional smoothing of multiple scattering, wide FOVs might be less susceptible to those effects. It could however also be attributed to bias from the optimisation.

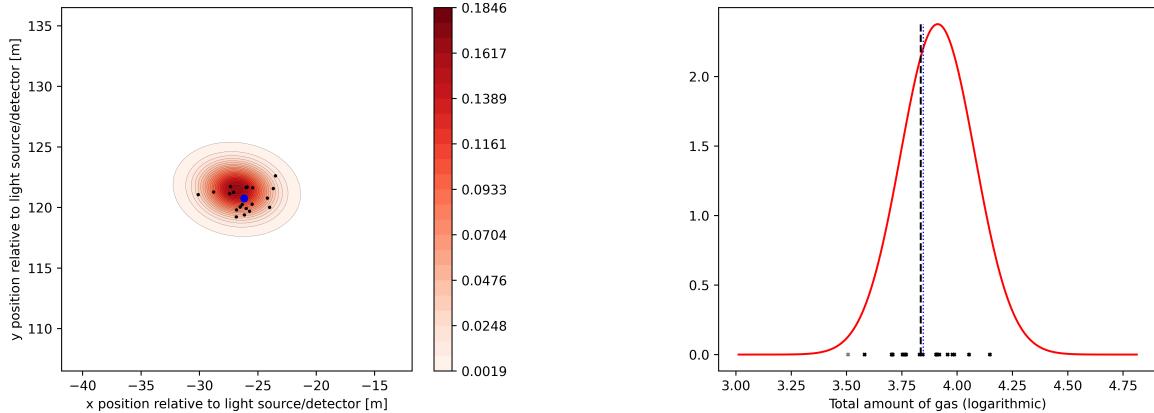
Practically speaking, the more relevant instance is the turbulent one. For that, it must be emphasised that the shown UQ results, and especially the way in which they were obtained, even if we have a way of accessing the distribution of the turbulence process, is not very practical. The bottleneck here is the computation of $\mathbb{V}(\mathbf{R}[\theta])$ from Eq. (4.67) and by extension $I_{T,0}$. The former was obtained through sampling which turns out to be considerably more expensive than solving the RTE-based inverse problem without accounting for turbulence. Although this doesn't slow down the computation of point estimates where turbulence effects are ignored, it results in overall compute times of several hours instead of a few minutes. This is due to the high complexity of the turbulent structure, which



(a) narrow FOV: Line 5 in Table 5.3

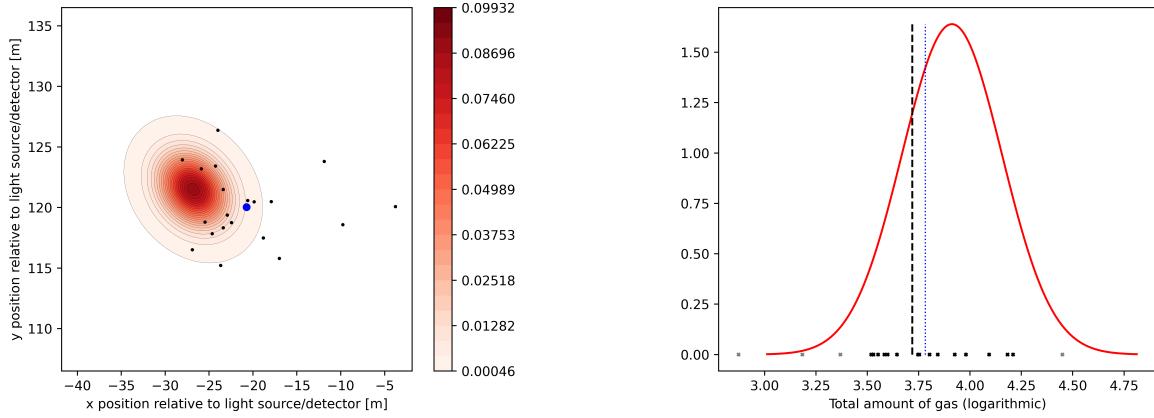


(b) wide FOV: Line 5 in Table 5.3

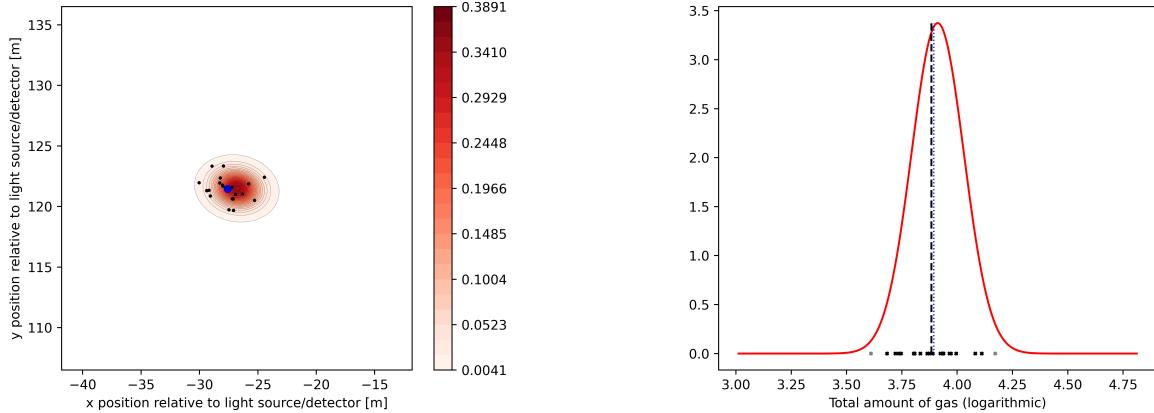


(c) multiple FOVs: Line 5 in Table 5.3

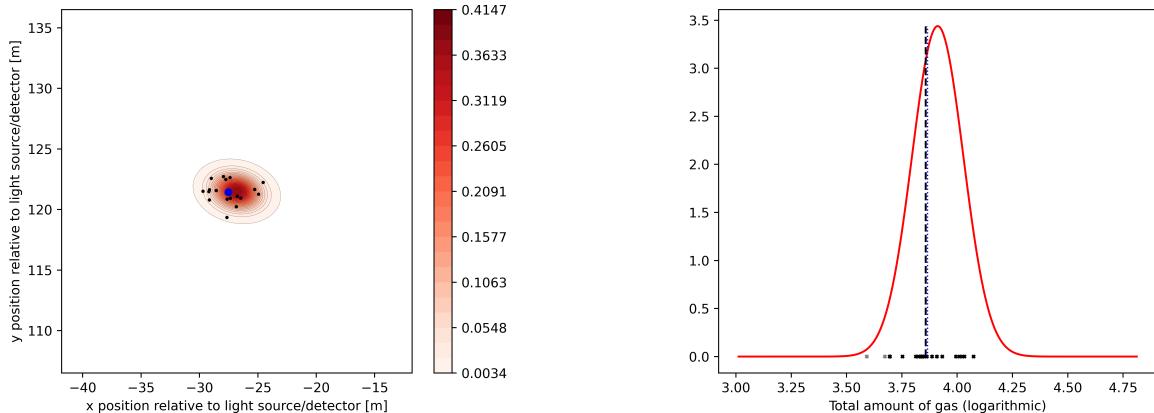
Figure 5.8: Reconstructed samples along with approximate confidence regions centred at the ground truth value for line 5. Left: Source location θ_{source} , samples in black and their mean in blue, with ellipses corresponding to two standard deviations in each coordinate. Right: Logarithmic amount of gas θ_0 with samples at the bottom, black dashed line is their mean and blue dotted the logarithm of the mean on the natural scale. Gray dots indicate samples with error exceeding two standard deviations. See also Eq. (5.44).



(a) narrow FOV: Line 8 in Table 5.3

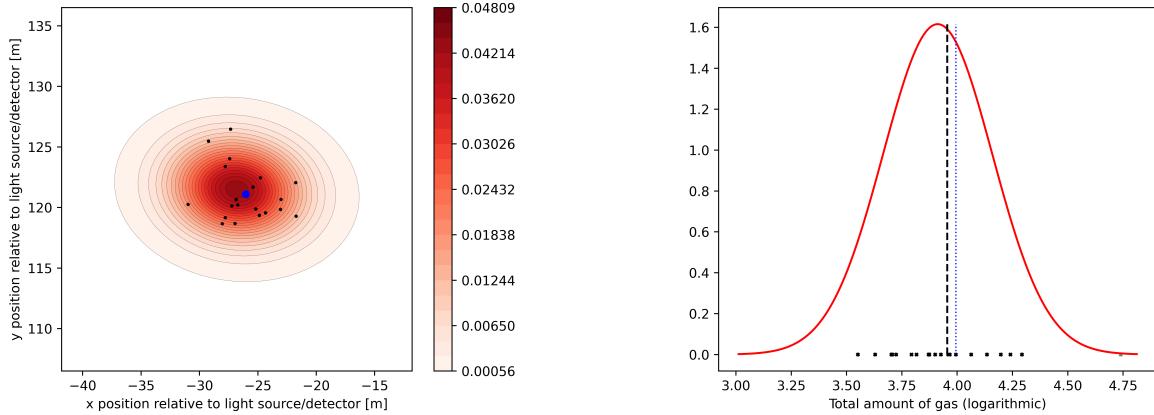


(b) wide FOV: Line 8 in Table 5.3

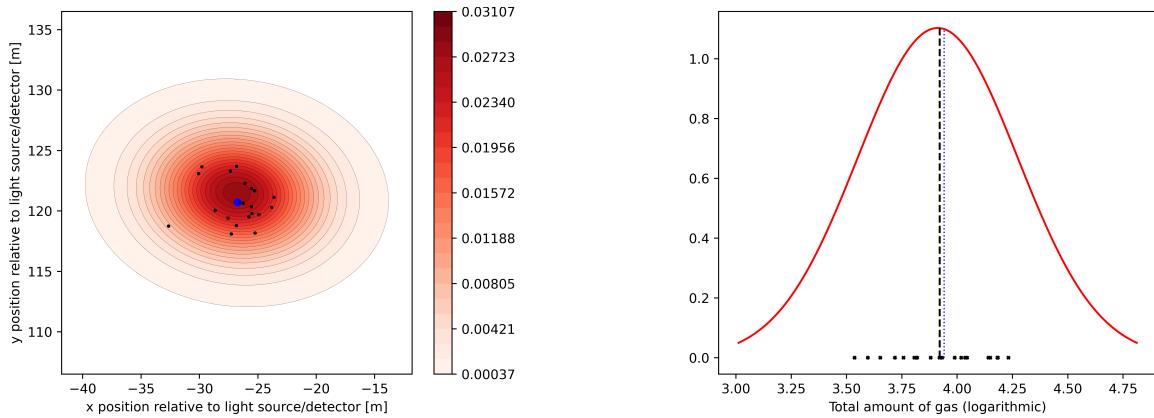


(c) multiple FOVs: Line 8 in Table 5.3

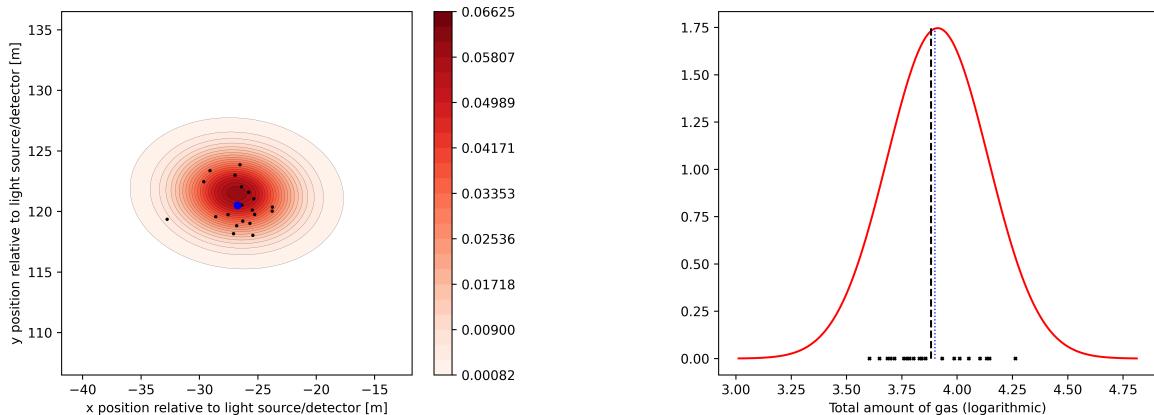
Figure 5.9: Reconstructed samples along with approximate confidence regions centred at the ground truth value for line 8. The graphs are like the ones in Fig. 5.8.



(a) narrow FOV: Line 1 in Table 5.3

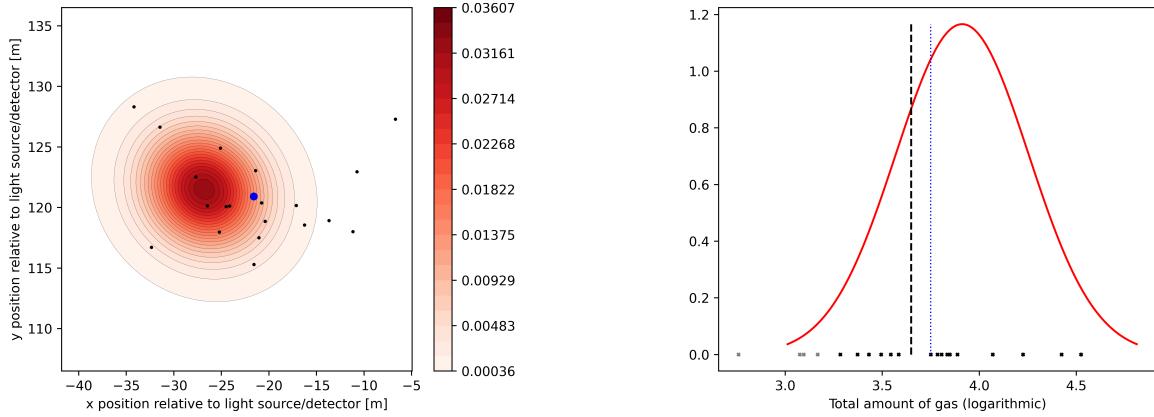


(b) wide FOV: Line 1 in Table 5.3

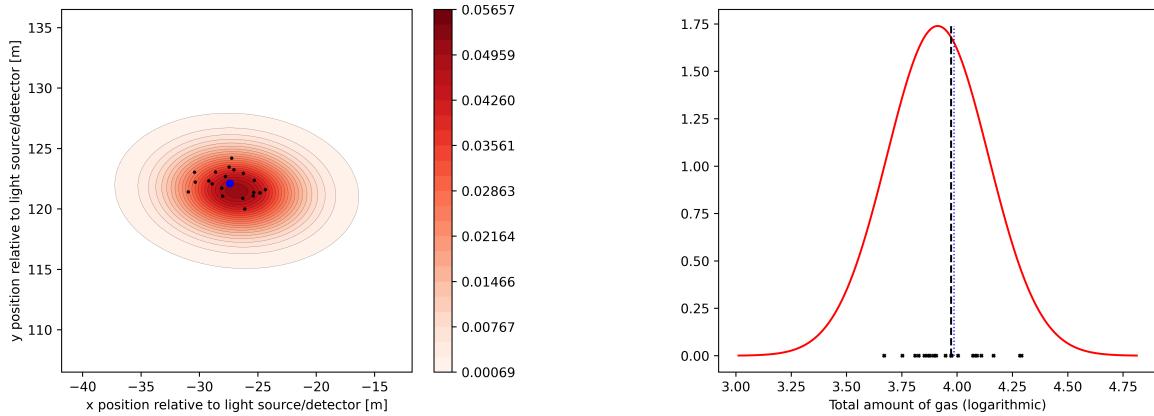


(c) multiple FOVs: Line 1 in Table 5.3

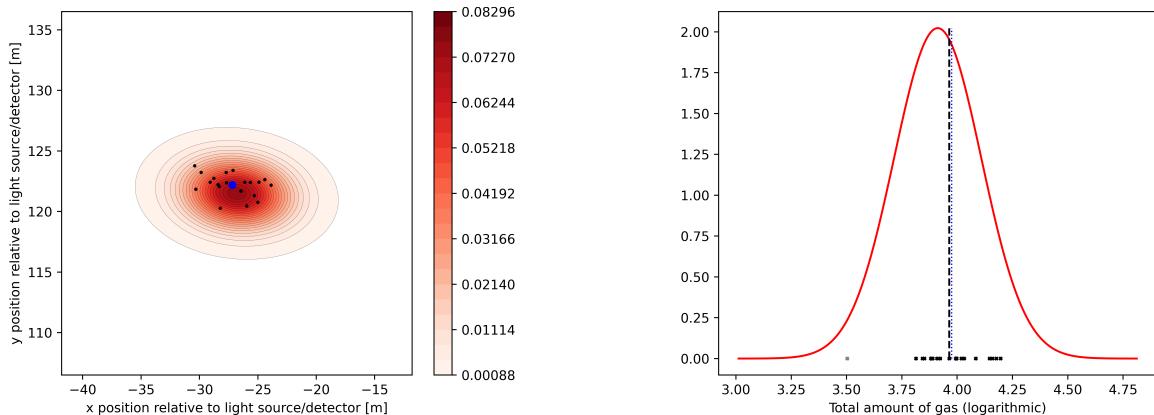
Figure 5.10: Reconstructed samples along with approximate confidence regions centred at the ground truth value for line 1. The graphs are like the ones in Fig. 5.8 but the confidence regions are adjusted based on Eq. (4.69).



(a) narrow FOV: Line 4 in Table 5.3



(b) wide FOV: Line 4 in Table 5.3



(c) multiple FOVs: Line 4 in Table 5.3

Figure 5.11: Reconstructed samples along with approximate confidence regions centred at the ground truth value for line 4. The graphs are like the ones in Fig. 5.8 but the confidence regions are adjusted based on Eq. (4.69).

requires many more kernel functions for its parameterisation than the simpler averages, and consequently high cost of evaluating integrals which makes even single scattering approximations rather slow. Indeed, the reason for using a single expansion around the true value in the examples is so that $\mathbb{V}(\mathbf{R}[\theta])$ needs only be computed once per experiment instead of having to perform an estimation at each sampled value. To make matters worse, our estimate of $\mathbb{V}(\mathbf{R}[\theta])$ is singular because the number of samples used was roughly an order magnitude less than the number of rows/columns in the matrix. In and of itself this isn't a problem since a non-zero diagonal matrix is added to it and we only seek to estimate the comparably small matrix $I_{\mathsf{T},0}$ formed by multiplication with relatively well behaved Jacobian matrices. The increased confidence regions in Figs. 5.10 and 5.11 relative to Figs. 5.8 and 5.9 certainly suggest that the aforementioned Jacobian matrices are not in the null space of the estimates for $\mathbb{V}(\mathbf{R}[\theta])$ which is the main risk associated with an estimation based on “insufficient” sampling.

Chapter 6

Conclusion & outlook

The main goal of this research was to examine DIAL in the context of radiative transfer theory with an emphasis on potential benefits that a measurement of multiply scattered photon returns may have on gas concentration image reconstruction. Following a theoretical investigation we have developed a variant of DIAL imaging approach and have demonstrated its computational viability and performance through computational implementation and simulation. The main outcomes of this research can be summarised as follows:

- We conducted a theoretical analysis of DIAL driven by the RTE including the construction of a class of images that can be used to formulate a well-posed inverse problem amenable to noisy data.
- We have demonstrated that light from wider FOVs is beneficial as far as the reconstruction of high-level image features based on smooth approximations in the presence of noise is concerned. Our analysis however shows that those benefits vanish at lower noise levels and the imaging performance deteriorates when higher resolution is sought. Effectively, our approach becomes useful when seeking a smooth approximation of the gas concentration image from low SNR data. This paradigm is particularly relevant for imaging plumes in instances where measurements are challenging because absorption is weak and the plumes cannot be observed for a long time without taking action, e.g. accidents.

- We showed that a separation of the data into two FOVs, capturing single and multiple scattering individually, could combine the benefits whilst minimising the shortcomings of the conventional single-scattering DIAL and the proposed RTE-based DIAL approaches.
- We implemented an algorithm that estimates the parameters of interest and argued that the low-dimensional structure implied by our regularisation constraints not only improves robustness to measurement noise but also makes the problem computationally tractable by allowing for the use of inexpensive Monte-Carlo evaluations of the 3-dimensional RTE forward model.

We began by formulating a problem in the context of DIAL that treats optical measurements from wide FOVs as valuable data rather than noise. This shift in perspective is the transformative aspect of our approach. As the noise in measurements used for atmospheric imaging is typically rather high, we proposed a regularisation that attempts to mimic PDE-based constraints for the image and is inspired by the simple form commonly associated with plume models. After introducing a semi-parametric model formulation analogous to that of DIAL but in the context of the RTE, we showed that under suitable conditions these measurements bear enough information to recover a low-dimensional image approximation. Based on the idealised problem of gas detection we further argued that data from wider FOVs can aid in determining gas concentration if high-level features of a plume are of interest. As the RTE-based forward model results in a notably more ill-posed reconstruction process, a more detailed analysis of the convergence behaviour was performed to provide an insight into the utility of the proposed data. It was shown that some key aspects of the image are more stable to reconstruct than others and that although the rate of decay in the error depends on the sought resolution, crucially this becomes bigger for decreasing gas concentration when optical noise becomes the dominant contributor to the error. The limitation with respect to accuracy was then used as a justification for the use of a purely parametric model which, although asymptotically biased, still accounts for many of the observable optical effects and yet it yields a model that is more efficient to deal with computationally. Lastly, we suggested a method based

on a Laplace approximation for conservative estimation of the amount of uncertainty in the reconstructed parameters, i.e. an upper bound. We tested the approach under the assumptions that the instrument response is Poisson distributed and a suitable prior distribution accounting for turbulent effects is known and showed that although the estimates weren't always accurate, they were able to avoid underestimation of the sample variance.

After having developed the theory and set expectations as to what can and what likely cannot be achieved with the proposed model and measurements, the main challenge is the numerical evaluation of the involved quantities. The structure of our regularisation naturally leads to model simplifications and a reduced dimensionality of the problem. Concentration phenomena commonly observed for such structures mean that computations can be accelerated through randomisation as the RTE, which is the primary bottleneck, does not have to be solved very accurately. The results of our simulations are in relatively good agreement with the theory. They suggest that the chosen parameterisation of the problem is able to capture scattering effects that cannot be explained by the dispersion parameters alone and that a plume-shaped concentration profile can be recovered from noise-corrupted data that is unsuitable for the use in direct inversion formulas of classical DIAL. As expected, the benefits are primarily observed in cases involving plumes with high optical thickness, which result in significant photon scattering.

The research conducted in this work leaves several open questions. It is worthwhile to highlight the following two.

- **Real-data experiment:** The true test of a theoretical work such as this, is without a doubt its validation in the field. Despite our best efforts to create data that mimic reality, the number of uncontrollable variables in practice will certainly exceed any simulation. When it comes to comparing the classical DIAL to our method, the following aspects are specifically relevant to validate:
 - The sensor efficiency we have assumed identical in the wide and narrow FOV which, along with the Poisson distribution for the noise. Such assumptions may not be appropriate in practice.

- The RTE model setup as considered in this work is certainly more restrictive than reality with respect to optical phenomena and the discrepancy between model and reality may be larger than we have considered.
- **Turbulence & Uncertainty Quantification:** As wide FOV data is less useful in the small-error limit, it seems natural to consider the image reconstruction alongside the corresponding UQ problem. The proposed method relies on fast convergence, i.e. at parametric rates, to an asymptotic distribution and thus on a technique which by design only applies to small-error limits. It would certainly be desirable to have a better account of the aforementioned model discrepancies with less reliance on negligible discrepancies between data and low-dimensional models. This should ideally be accompanied by a thorough investigation of turbulence, in and of itself a complex and challenging problem, which was expressed “generically” with no rigorous connection to the physical process.

Appendix A

Additional mathematical details

A.1 Supplementary proofs and remarks

A.1.1 Continuity of RTE solutions

The proof consists of a relatively straightforward change of variable argument but relies on a few explicit computations which makes the expressions complicated and the overall argument rather hard to follow. Most of the ideas and techniques used are not necessarily new and are given mostly for completeness.

Proof of Lemma 3.1.1. We give a proof for a constant $g_m = 1$, i.e. we drop it from the formulas, as the argument is virtually identical for other choices of g_m . The ballistic term m_0 vanishes on Z and the claim can be verified easily for $k = 1$. Let's start by fixing $k \geq 2$. We can express

$$\begin{aligned} & \int_{\partial V_{(+)}(x)} \mathcal{A}_k[g](x, v, t) dv \\ &= \int_{(0, \infty)^k \times (\mathbb{S}^2)^{k+1}} g\left(x_0, v_0, t - \sum_{l=1}^k s_l - \ell_{v_0}^-(x_1)\right) F_k(x; v_0, v_{1:k}, s_{1:k}) ds_{1:k} dv_{1:k} dv_0 \end{aligned}$$

where we have abbreviated

$$F_k(x; v_0, v_{1:k}, s_{1:k}) = e^{-\mathbf{I}_{\sigma_{a+s}}(\gamma(x_{0:k+1}))} \mathbf{E}_{\sigma_s f_p}(\zeta(x_{0:k+1})) |\vec{n}(x) \cdot v_k| 1_{\partial V_{(+)}(x)}(v_k) 1_{X^k}(x_k, \dots, x_1)$$

and introduced for any $l = 1, \dots, k$ the implicit quantities $y_l = s_l v_l$, $x_l = x - \sum_{m=l}^k y_m$ as well as $x_{k+1} = x$ and $x_0 = x_1 - \ell_{v_0}^-(x_1) v_0$. The set $\gamma(x_{0:k+1})$ and $\zeta(x_{0:k+1})$ are defined in Definition 5.1.2 along with the E and I shorthand notation. Note that if we had surface scattering F_k would be more complicated as the integration over surface points in ∂X is quite different from that in the interior X . In pursuit of our goal to obtain an expression for \mathbf{m}_k , which implies changing the variables in such a way that the integration is performed w.r.t. $x_1 - \ell_{v_0}^-(x_1) v_0$ and $\sum_{l=1}^k s_l + \ell_{v_0}^-(x_1)$, we first make the change of variables $(s_m, v_m) \mapsto y_m$ followed by $y_m \mapsto x_1$, which is a simple shift, after which we can utilise [27, Lemma 2.1] in order to obtain

$$\int_{\partial V_{(+)}(x)} \mathcal{A}_k[g](x, v, t) dv = \int_{(0, \infty)^k \times (\mathbb{S}^2)^{k-1} \times \partial B_{(-)}} \frac{F_k(x; v_0, v_{1:k}, s_{1:k}) |\vec{n}(x_0) \cdot v_0|}{\|x - x_0 - \sum_{l=0, l \neq m}^k s_l v_l\|_2^2} g\left(x_0, v_0, t - \sum_{l=0, l \neq m}^k s_l - \|x - x_0 - \sum_{l=0, l \neq m}^k s_l v_l\|_2\right) ds_{1:k \setminus \{m\}} dv_{0:k \setminus \{m\}} dv_0 dx_0.$$

Notice that after the change the independent variables are v_0, x_0 along with the directions $v_k, \dots, v_{m+1}, v_{m-1}, \dots, v_1$ and distances $s_k, \dots, s_{m+1}, s_{m-1}, \dots, s_0$ where s_0 was “newly” introduced via the application of [27, Lemma 2.1]. For the implicit variables we have

$$s_m = \left\| x - x_0 - \sum_{l=0, l \neq m}^k s_l v_l \right\|_2, \quad v_m = \frac{x - x_0 - \sum_{l=0, l \neq m}^k s_l v_l}{\|x - x_0 - \sum_{l=0, l \neq m}^k s_l v_l\|_2} \quad (\text{A.1})$$

as the inverse maps of the coordinate transform for the new variables. In order to obtain an expression for the Schwartz kernel \mathbf{m}_k it remains to resolve the time component of the arguments in g . The necessary transformation takes the form

$$s_n \mapsto s = \sum_{l=0, l \neq m}^k s_l + \left\| x - x_0 - \sum_{l=0, l \neq m}^k s_l v_l \right\|_2$$

and maps $(0, \infty) \rightarrow (\sum_{l=0, l \neq m, n}^k s_l + \|x - x_0 - \sum_{l=0, l \neq m, n}^k s_l v_l\|_2, \infty)$. It is invertible

everywhere but on a set of measure zero and the inverse is given by

$$s_n = \frac{(s - \sum_{l=0, l \neq m, n}^k s_l)^2 - \|x - x_0 - \sum_{l=0, l \neq m, n}^k s_l v_l\|_2^2}{2 \left(s - \sum_{l=0, l \neq m, n}^k s_l - v_n \cdot (x - x_0 - \sum_{l=0, l \neq m, n}^k s_l v_l) \right)}. \quad (\text{A.2})$$

It is easily verified that

$$\frac{d}{ds_n} s = 1 - v_n \cdot v_m \implies \frac{d}{ds} s_n = \frac{1}{1 - v_n \cdot v_m}$$

where v_m is as in Eq. (A.1) but should be considered a function of s instead of s_n in the latter expression. Making that change results in

$$\begin{aligned} \int_{\partial V_{(+)}} \mathcal{A}_k[g](x, v, t) dv &= \int_{\partial B_{(-)} \times (0, \infty)} g(x_0, v_0, t-s) |\vec{n}(x_0) \cdot v_0| \\ &\quad \int_{(0, \infty)^{k-1} \times (\mathbb{S}^2)^{k-1}} \frac{G_k(x, t, x_0, v_0, s; \cdot)}{s_m(\cdot)^2 (1 - v_n \cdot v_m(\cdot))} dv_{1:k \setminus \{m\}} ds_{0:k \setminus \{m, n\}} ds dv_0 dx_0 \end{aligned} \quad (\text{A.3})$$

where G_k enforces the integration bounds and can be defined via

$$G_k(x, x_0, v_0, s; \cdot) = 1_{(0, s)} \left(\sum_{l=0, l \neq m, n}^k s_l + \|(x - x_0 - \sum_{l=0, l \neq m, n}^k s_l v_l)\|_2 \right) F_k(\cdot) \quad (\text{A.4})$$

where (\cdot) has been used as a placeholder to indicate that a quantity is dependent on additional variables, i.e. a function of the variables that the integration is performed over. Although this process in principle yields an explicit expression for \mathbf{m}_k this isn't good enough because the change of variables introduces singularities and the integrand is unbounded making it difficult to analyse. However, we don't have to always use the same variables s_m, s_n, v_m for the change but instead we can partition $(0, \infty)^k \times (\mathbb{S}^2)^{k+1}$ and perform the change with respect to different coordinates on each subset. We would like to pick s_m, v_m such that s_m is not too small, e.g. we can take the maximum, and v_n such that it isn't too aligned with v_m which may be accomplished by taking the most misaligned direction. The latter will avoid singularities whenever $s > \|x - x_0\|$. In the situation we will be interested in we have $s \gg \|x - x_0\|$ since $x = x_0$ and $s \gg 0$ for Lidar measurements. First, notice that $s = \sum_{l=1}^k s_l + \ell_{v_0}^-(x_1) > 0$ so taking $\gamma(x_{0:k+1})$

and connecting the endpoints yields at a closed polygon of length $s + \|x - x_0\|$ which must have at least two segments, other than the piece between x and x_0 , that are longer than $\frac{s - \|x - x_0\|}{2k}$. To see that note that due to the triangle inequality no segment of a closed polygon can be longer than half the total length and so we must have

$$\frac{s + \|x - x_0\|}{2k} \leq \frac{s}{k} \leq \max\{s_k, \dots, s_1, \ell_{v_0}^-(x_1)\} \leq \frac{s + \|x - x_0\|}{2}.$$

Taking the total length of the polygon and subtracting $\max(s_k, \dots, s_1, \ell_{v_0}^-(x_1))$ as well as $\|x - x_0\|$ from $s + \|x - x_0\|$ to account for the added segment we obtain the lower bound for the second largest value of $s_k, \dots, s_1, \ell_{v_0}^-(x_1)$ which particularly implies that

$$\max\{s_k, \dots, s_1\} \geq \frac{1}{k}(s - \max\{s_k, \dots, s_1, \ell_{v_0}^-(x_1)\}) \geq \frac{s - \|x - x_0\|}{2k}.$$

Since we also have that

$$s^2 - \|x_0 - x\|^2 = \sum_{l=1}^k \sum_{m=1}^k s_m s_l (1 - v_m \cdot v_l) + 2\ell_{v_0}^-(x_1) \sum_{l=1}^k s_l (1 - v_0 \cdot v_l)$$

it must hold that

$$\max_{l,m=0}^k 1 - v_m \cdot v_l \geq \frac{s^2 - \|x - x_0\|_2^2}{s^2} = 1 - \frac{\|x - x_0\|_2^2}{s^2}.$$

Further we have that $\sqrt{2 - 2a \cdot b} = \|a - b\|_2$ for $a, b \in \mathbb{S}^2$ which means we can apply the triangle inequality to obtain for any $m = 0, \dots, k$

$$\max_{l=0}^k \sqrt{1 - v_m \cdot v_l} \geq \frac{1}{2} \sqrt{1 - \frac{\|x - x_0\|_2^2}{s^2}}.$$

In order to see how we can exploit these findings consider for any pair (m, n) , where $m = 1, \dots, k$ and $n = 0, \dots, k$, the indicator functions

$$J_{m,n}(v_k, \dots, v_0, s_k, \dots, s_1) = \prod_{p=1}^k 1_{[0,\infty)}(s_m - s_p) \prod_{q=0, q \neq n}^k 1_{(0,\infty)}(v_m \cdot v_q - v_m \cdot v_n)$$

and observe that we have $dv_k \dots dv_0 ds_k \dots ds_1$ almost everywhere

$$\sum_{(m,n) \in I_k} J_{m,n} = 1$$

with $I_k = \{(m, n) : m = 1, \dots, k, n = 1, \dots, m-1, m+1, \dots, k\}$. In fact, it is not hard to see that $J_{m,n} = 1$ if $s_m \geq s_p$ for any $p = 1, \dots, k$ and at the same time $1 - v_m \cdot v_n > 1 - v_m \cdot v_q$ for any $q \neq n$. Some s_m is inevitably the largest, which is also unique except on a set of measure zero, while some v_n must be the most misaligned to the corresponding direction v_m . Consequently we may also write

$$\begin{aligned} \int_{\partial V_{(+)}(x)} \mathcal{A}_k[g](x, v, t) dv &= \sum_{(m,n) \in I_k} \int_{(0,\infty)^k \times (\mathbb{S}^2)^{k+1}} J_{m,n}(v_k, \dots, v_0, s_k, \dots, s_1) \\ &\quad g\left(x_1 - \ell_{v_0}^-(x_1)v_0, v_0, t - \sum_{l=1}^k s_l - \ell_{v_0}^-(x_1)\right) F_k(x; v_{0:k} s_{1:k}) dv_{1:k} ds_{1:k} dv_0 \end{aligned}$$

which is nothing but the first identity that we used in a partitioned form with the indicator functions $J_{m,n}$ introduced earlier. By performing the previous transform but choosing carefully the variables depending on the summation index we may arrive at

$$\begin{aligned} \int_{\partial V_{(+)}} \mathcal{A}_k[g](x, v, t) dv &= \int_{\partial B_{(-)} \times (0, \infty)} \sum_{(m,n) \in I_k} \int_{(0,\infty)^{k-1} \times (\mathbb{S}^2)^{k-1}} J_{m,n}(\cdot) \\ &\quad \frac{G_k(x, x_0, v_0, s; \cdot)}{s_m(\cdot)^2(1 - v_n \cdot v_m(\cdot))} dv_{1:k \setminus \{m\}} ds_{0:k \setminus \{m,n\}} g(x_0, v_0, t-s) |\vec{n}(x_0) \cdot v_0| ds dv_0 dx_0 \end{aligned}$$

which means that the Schwartz kernel \mathfrak{m}_k can be expressed as

$$\mathfrak{m}_k(x, x_0, v_0, s) = \sum_{(m,n) \in I_k} \int_{D_k} \frac{J_{m,n}(\cdot) G_k(x, x_0, v_0, s; \cdot)}{s_m(\cdot)^2(1 - v_n \cdot v_m(\cdot))} dv_{1:k \setminus \{m\}} ds_{0:k \setminus \{m,n\}} \quad (\text{A.5})$$

where $D_k = (0, \infty)^{k-1} \times (\mathbb{S}^2)^{k-1}$. Due to our previous developments, we know that there is a constant $C = C(\sigma_a, \sigma_s, f_p)$ that only depends on the optical parameters such that

$$C^k \left(\frac{k}{s - \|x - x_0\|_2} \right)^2 \frac{s^2}{s^2 - \|x - x_0\|_2^2} \geq \frac{J_{m,n}(\cdot) G_k(x, x_0, v_0, s; \cdot)}{s_m(\cdot)^2(1 - v_n \cdot v_m(\cdot))} \quad (\text{A.6})$$

for any pair m, n . This means that for fixed k the integrand is bounded by a polynomial of degree 4 in k multiplied by a term that grows exponentially in k . This shows that \mathbf{m}_k is a function when restricted to the set Z . For continuity it is sufficient to show that each \mathbf{m}_k is continuous on Z . In order to see the claim for the infinite sum \mathbf{m} we may use formulas for the volume of unit balls, see e.g. [145], to obtain the bound

$$\int_{\{z \in (0, \infty)^k : \sum_{j=1}^k z_j \leq 1\}} dz_k \dots dz_1 \leq \int_{\{z \in \mathbb{R}^k : \sum_{j=1}^k |z_j| \leq 1\}} dz_k \dots dz_1 = \frac{2^k}{k!}. \quad (\text{A.7})$$

which means that the polynomial and exponential factors introduced in the coordinate change will not affect summability in any way and the claim follows by the dominated convergence theorem applied to the infinite sum. It remains to obtain continuity of \mathbf{m}_k for each $k \geq 2$. For that take any $(x^*, x_0^*, v_0^*, s_0^*) \in Z$ and observe that due to Eq. (A.6) we can bound

$$\frac{J_{m,n}(\cdot)G_k(x, x_0, v_0, s; \cdot)}{s_m(\cdot)^2(1 - v_n \cdot v_m(\cdot))} = O\left(\frac{C^k k^2 s_0^*}{(s_0^* - \|x^* - x_0^*\|_2)^3}\right)$$

uniformly, i.e. as a function of s_0^* and $\|x^* - x_0^*\|_2$ only, in a sufficiently small neighbourhood of $(x^*, x_0^*, v_0^*, s_0^*) \in Z$. The mappings

$$(x, x_0, v_0, s) \mapsto J_{m,n}(\cdot)G_k(x, x_0, v_0, s; \cdot)$$

is $dv_{1:k \setminus \{m\}} ds_{0:k \setminus \{m,n\}}$ almost everywhere continuous at (x^*, x_0^*, v_0^*, s) because all indicator functions, which are the only sources of discontinuities, are continuous almost everywhere with respect to that measure. This argument holds only for $k \geq 2$, for $k = 1$ there is a discontinuity at the boundary of the domain because no integration takes place. By bounded convergence, we get that \mathbf{m}_k must be continuous in $(x^*, x_0^*, v_0^*, s_0^*)$ and since the point was arbitrary the claim is true on Z . \square

A.1.2 Proof of Lemma 3.1.2 & Lemma 3.1.3

The following facts about the kernel spaces will be useful. Due to the nature of our kernel functions, it must be true that for any $p > 0$ we have

$$1 - \sup\{z \in [0, 1] : (1 - z)^p \leq \phi(z)\} =: \varepsilon(p) > 0.$$

Note that $\varepsilon(p) \rightarrow 0$ as $p \rightarrow \infty$ and by continuity also $\phi(1 - \varepsilon(p)) = \varepsilon(p)^p$. It follows for p sufficiently large that

$$\int_0^{\varepsilon(p)} \phi(1 - z) dz \leq \int_0^{\varepsilon(p)} (1 - z)^p dz = \frac{\varepsilon(p)^{p+1}}{p+1} = \frac{\varepsilon(p)}{p+1} \phi(1 - \varepsilon(p)). \quad (\text{A.8})$$

Also, note that

$$\frac{1}{h} \int_0^\varepsilon \phi\left(1 - \frac{z}{h}\right) dz = \int_0^{\frac{\varepsilon}{h}} \phi(1 - z) dz$$

which, given that ϕ is monotone on $(0, \infty)$, implies for any $h_0 < h_1$ and any $\varepsilon \in (0, h_0)$

$$\begin{aligned} \varepsilon \left(\frac{1}{h_0} - \frac{1}{h_1} \right) \phi \left(1 - \frac{\varepsilon}{h_1} \right) &\leq \int_{\frac{\varepsilon}{h_1}}^{\frac{\varepsilon}{h_0}} \phi(1 - z) dz \\ &= \frac{1}{h_0} \int_0^\varepsilon \phi\left(1 - \frac{z}{h_0}\right) dz - \frac{1}{h_1} \int_0^\varepsilon \phi\left(1 - \frac{z}{h_1}\right) dz. \end{aligned} \quad (\text{A.9})$$

Proof of Lemma 3.1.2. The direction \Rightarrow is trivial so we only have to show \Leftarrow , i.e. that the measurement is different for different differential absorption functions. We may assume without loss of generality that $x_D = 0$, $\phi(0) = 1$ and, because ϕ is monotone, $\phi(z) > 0$ for any $z \in [0, 1]$. Let $\alpha_1 \neq \alpha_2$ be two different perturbations of σ_a as in the lemma. Further let $\tau_{1/2} := \text{dist}(0, \text{supp}(\alpha_{1/2}))$ and $v_{1/2}$ directions such that $\tau_{1/2} v_{1/2} \in \partial \text{supp}(\alpha_{1/2})$. If $\tau_1 > \tau_2$ or $\tau_1 < \tau_2$ then obviously the measurements will differ at times between $2\tau_1$ and $2\tau_2$ in direction v_1 or v_2 as the one corresponding to the smaller first impact time will be influenced by an absorbing perturbation and differ from the offline measurement while the other won't. Therefore we assume $\tau_1 = \tau_2$ and define $\alpha_\delta := \alpha_1 - \alpha_2$ as well as the first impact time $\tau_\delta = \text{dist}(0, \text{supp}(\alpha_\delta))$.

Note that $\tau_\delta \geq \tau_1 = \tau_2$ may have increased and $\tau_\delta > 0$ due to $\alpha_{1/2} \in K(\phi \mid x_D)$. There might be multiple possibilities for the first impact direction, i.e. the direction v such that $v\tau_\delta \in \partial \text{supp}(\alpha_\delta)$. The constant term w_0 in α_i can be identified from any measurement at time $s < \tau_i$ for $i = 1, 2$. Thus, if the constants are different there is nothing more to show. In case that the constant terms are equal we can write

$$\alpha_\delta(x) = \sum_{j=1}^N w_j \phi\left(\frac{\|x - b_j\|}{h_j}\right)$$

with $w_j \in \mathbb{R} \setminus \{0\}$ positive and negative such that $(b_k, h_k) \neq (b_j, h_j)$ for all $j \neq k$ which makes the representation unique. Combining Eq. (A.8) and Eq. (A.9) we can see that it is not possible to replicate the tail of a kernel will tails of larger width which implies that every possibility for v which satisfies $v\tau_\delta \in \partial \text{supp}(\alpha_\delta)$ must point towards a kernel centre b_j . We choose a direction $v_\delta \in \mathbb{S}^2$ with that property such that it points towards the kernel centre b_j with the smallest possible bandwidth parameter h_j . If this isn't unique, any of them is good. Let h_{\min}, b_{\min} and w_{\min} be the kernel parameters that correspond to the kernel that was used in the choice of v_δ and without loss of generality assume that the corresponding weight is positive, i.e. $w_{\min} > 0$. We can define

$$A(\varepsilon) = \int_0^{\tau_\delta + \varepsilon} \alpha_\delta(sv_\delta) ds$$

and now claim that for $\mathbf{m}_{\text{on},i}$ corresponding to α_i , $i \in \{1, 2\}$, we have

$$\limsup_{\varepsilon \searrow 0} \left| \frac{\mathbf{m}_{\text{on},1}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D) - \mathbf{m}_{\text{on},2}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)}{A(\varepsilon)} \right| > 0 \quad (\text{A.10})$$

which would prove our uniqueness statement. Let $h_{2\text{nd}} := \inf\{h_j : h_j > h_{\min}\}$ be the

smallest width parameter larger than h_{\min} . From the above we get for ε sufficiently small,

$$\begin{aligned} A(\varepsilon) - w_{\min} \int_0^\varepsilon \phi \left(1 - \frac{z}{h_{\min}} \right) dz \\ \geq - \int_0^{\tau_\delta + \varepsilon} \left| \alpha_\delta(sv_\delta) - w_{\min} \phi \left(1 - \frac{\|sv_\delta - b_{\min}\|}{h_{\min}} \right) \right| ds. \\ \geq - \sum_{j=1}^N |w_j| \int_0^\varepsilon \phi \left(1 - \frac{z}{h_{2\text{nd}}} \right) dz \end{aligned}$$

where the second inequality holds because for positive weights the integral is maximised when the integration is done in the direction that points from x_D towards b_j and decreases with the kernel width. Consequently,

$$\sum_{j=1}^N |w_j| \int_0^\varepsilon \phi \left(1 - \frac{z}{h_{2\text{nd}}} \right) dz \geq \left| A(\varepsilon) - w_{\min} \int_0^\varepsilon \phi \left(1 - \frac{z}{h_{\min}} \right) dz \right|$$

This is true because ε sufficiently small we have $\|b_j - x_D - (\varepsilon + \tau_\delta)v_\delta\| < h_j$ implies that $b_{\min} = b_j$ and $h_{\min} = h_j$ or $h_{\min} < h_j$ which means that only larger widths than h_{\min} must be taken into account. Taking $\tilde{\varepsilon}(p) = h_{2\text{nd}}\varepsilon(p)$ we see that as $p \rightarrow \infty$

$$\frac{\int_0^{\tilde{\varepsilon}(p)} \phi \left(1 - \frac{z}{h_{2\text{nd}}} \right) dz}{\int_0^{\tilde{\varepsilon}(p)} \phi \left(1 - \frac{z}{h_{\min}} \right) dz} \leq \frac{\int_0^{\tilde{\varepsilon}(p)} \phi(1-z) h_{2\text{nd}} dz}{\varepsilon(p)(1-h_{\min}/h_{2\text{nd}})\phi(1-\varepsilon(p))} \leq \frac{h_{2\text{nd}}^2}{(h_{2\text{nd}} - h_{\min})(p+1)}$$

the first summand in $A(\varepsilon)$ dominates, i.e.

$$\lim_{p \rightarrow \infty} \frac{A(\tilde{\varepsilon}(p))}{w_{\min} \int_0^{\tilde{\varepsilon}(p)} \phi \left(1 - \frac{z}{h_{\min}} \right) dz} = 1 \quad (\text{A.11})$$

Let $\Gamma(\varepsilon)$ be the set of all continuous piece-wise linear curves that start and end in x_D with length $2(\tau_\delta + \varepsilon)$. Pick $\gamma \in \Gamma(\varepsilon)$ and assume that $\gamma : [0, 2(\tau_\delta + \varepsilon)] \rightarrow \mathbb{R}^3$ satisfies $\gamma(0) = \gamma(2(\tau_\delta + \varepsilon)) = x_D$ and $\|\gamma'(s)\| = 1$ almost everywhere. Then we have, again for ε

sufficiently small,

$$\begin{aligned} \left| \int_0^{2(\tau_\delta + \varepsilon)} \alpha_\delta(\gamma(s)) ds \right| &\leq 2 \sum_{j: \|b_j - x_D - (\varepsilon + \tau_\delta)v_\delta\| < h_j} |w_j| \int_0^\varepsilon \phi\left(1 - \frac{z}{h_j}\right) dz \\ &\leq 2 \sum_{j=1}^N |w_j| \int_0^\varepsilon \phi\left(1 - \frac{z}{h_{\min}}\right) dz. \end{aligned} \quad (\text{A.12})$$

We next analyse the decay of first and higher-order scattering when scaled by $A(\varepsilon)$. From Lemma 3.1.1 we know that \mathbf{m}_k is continuous in a neighbourhood of $(x_D, x_D, v_\delta, \tau_\delta)$. Using the same notation, i.e. x_l for $l = 0, \dots, k+1$ as in Lemma 3.1.1, we know that for $d\nu_{m,n} = dv_{1:k \setminus \{m\}} ds_{0:k \setminus \{m,n\}}$

$$\mathbf{m}_{\text{off}}^k(v, s \mid x_D) = \sum_{(m,n) \in I_k} \int_{(0,\infty)^{k-1} \times (\mathbb{S}^2)^{k-1}} R_{m,n}^{\text{off}}(v, s; \cdot) d\nu_{m,n}$$

for functions $R_{m,n}^{\text{off}}$ given by Eq. (A.5) that are bounded via Eq. (A.6) uniformly in a neighbourhood of $(x_D, x_D, v_\delta, 2\tau_\delta)$. Note that the dependence of $R_{m,n}^{\text{off}}$ w.r.t. k has been omitted to simplify notation. If we put

$$R_{m,n}^{\text{on},i} = \exp\left(-\int_{\gamma(x_{0:k+1})} \alpha_i(r) dr\right) R_{m,n}^{\text{off}}$$

for $i = 1, 2$ and set $\Delta\mathbf{m}_{\text{on}}^k = \mathbf{m}_{\text{on},1}^k - \mathbf{m}_{\text{on},2}^k$ to be the difference in the online component of the k -th scattering contribution for all $k \geq 1$, then it is not hard to see that

$$\lim_{p \rightarrow \infty} \left| \frac{\Delta\mathbf{m}_{\text{on}}^1(v_\delta, 2(\tau_\delta + \tilde{\varepsilon}(p)) \mid x_D)}{2A(\tilde{\varepsilon}(p))} \right| = \mathbf{m}_{\text{off}}^1(v_\delta, 2\tau_\delta \mid x_D).$$

We claim that at the same distance τ_δ any higher order contributions $\Delta\mathbf{m}_{\text{on}}^k$ are much less affected by the absorption and we have for any $k \geq 2$

$$\lim_{p \rightarrow \infty} \left| \frac{\Delta\mathbf{m}_{\text{on}}^k(v_\delta, 2(\tau_\delta + \tilde{\varepsilon}(p)) \mid x_D)}{2A(\tilde{\varepsilon}(p))} \right| = 0.$$

Since $\mathbf{m}_{\text{off}}^1(v_\delta, 2\tau_\delta \mid x_D) > 0$ we would obtain Eq. (A.10) by bounded convergence and Eq. (A.12) from the decay rate of $\mathbf{m}_{\text{off}}^k$ which was shown in the proof of Lemma 3.1.1

towards the end. To see this it is enough to show that

$$\frac{R_{m,n}^{\text{on},1}(v_\delta, 2(\tau_\delta + \tilde{\varepsilon}(p)), \cdot) - R_{m,n}^{\text{on},2}(v_\delta, 2(\tau_\delta + \tilde{\varepsilon}(p)); \cdot)}{A(\tilde{\varepsilon}(p))} \xrightarrow{p \rightarrow \infty} 0 \quad (\text{A.13})$$

in $L^1((0, \infty)^{k-1} \times (\mathbb{S}^2)^{k-1}, d\nu_{m,n})$. Thanks to Eq. (A.12) and Eq. (A.11) we know that the left-hand side of Eq. (A.13) is bounded by $CR_{m,n}^{\text{off}}(v_\delta, 2\tau_\delta)$ for some p -independent constant $C > 0$ that may depend on τ_δ . Therefore it is enough to show the limit in Eq. (A.13) $\nu_{m,n}$ almost everywhere. We may also notice that the left-hand side of Eq. (A.13) is non-zero only when $1_{(\tau_\delta, \infty)}(\|x_l(v_\delta, 2(\tau_\delta + \tilde{\varepsilon}(p)); \cdot)\|_2)$ is non-zero for some $l = 1, \dots, k$, i.e. when $\gamma(x_{0:k+1})$ leaves the ball with radius τ_δ around the light source. For $k \geq 3$ there is an index $q = 1, \dots, k$ such that $q \neq m$ and $q \neq n$. If $n \neq 0$ and $s_0 s_q (1 - |v_\delta \cdot v_q|) > z$ for some p -independent $z > 0$ then $1_{(\tau_\delta, \infty)}(\|x_l(v_\delta, 2(\tau_\delta + \tilde{\varepsilon}(p)); \cdot)\|_2) = 0$ for p sufficiently large whereas if $n = 0$ then there are two indices $q_1 \neq q_2$ that differ from m and n and we have the same result for $s_{q_1} s_{q_2} (1 - |v_{q_1} \cdot v_{q_2}|) > z$. Note that both of these conditions are necessary for $\gamma(x_{0:k+1})$ to be sufficiently straight so that it can leave the ball with radius τ_δ around the light source. This shows Eq. (A.13) for $k \geq 3$ because, e.g. in the case of $n = 0$, we have

$$1_{(0,z)}(s_{q_1} s_{q_2} (1 - |v_{q_1} \cdot v_{q_2}|)) \xrightarrow{z \rightarrow 0} 0$$

$\nu_{m,0}$ almost everywhere. Similar conditions can be found when $k = 2$. Given that Eq. (A.13) for all $k \geq 2$ implies Eq. (A.10) this concludes the proof. \square

The proof of Lemma 3.1.3 is similar. Note that it is enough to show the result for $\sigma_a = 0$ since we have

$$\frac{\int_0^\varepsilon \phi(1-z)dz}{\phi(1-\varepsilon)} \rightarrow 0 \quad \text{for } \varepsilon \rightarrow 0$$

and scattering locally at the tails dominates non-differential absorption with respect to its contribution in \mathbf{m}_{off} . Note that the proportionality requirement, which implies that σ_a is uniquely determined by σ_s , can be dropped under certain conditions on the kernel

functions. Similar, although more restrictive, assumptions are made in Theorem 3.1.4 whose proof is the subject of the next section.

Proof of Lemma 3.1.3 - sketch for $\sigma_a = 0$. With notation as in Lemma 3.1.2 we have in the same way as in Eq. (A.10) that

$$\limsup_{\varepsilon \searrow 0} \left| \frac{\mathbf{m}_{\text{off},1}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D) - \mathbf{m}_{\text{off},2}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)}{S(\varepsilon)} \right| > 0 \quad (\text{A.14})$$

where $\mathbf{m}_{\text{off},i}$ correspond to different values $\sigma_{s,i}$ and

$$S(\varepsilon) = \sigma_{s,1}(x_D + v_\delta(\tau_\delta + \varepsilon)) - \sigma_{s,2}(x_D + v_\delta(\tau_\delta + \varepsilon)).$$

As before the above follows from the fact that

$$\limsup_{\varepsilon \searrow 0} \frac{S(\varepsilon)}{w_{\min} \phi \left(1 - \frac{\varepsilon}{h_{\min}} \right)} > 0 \quad (\text{A.15})$$

which can be shown in virtually the same fashion as in the previous proof. \square

A.1.3 Semi-parametric relaxation and proof of Theorem 3.1.4

The proof of Theorem 3.1.4 is somewhat more involved as it only assumes access to the quotient of our data which is equivalent to injectivity in a more general forward model of the form $(\mathbf{y}_{\text{on}}(v, t), \mathbf{y}_{\text{off}}(v, t))$ as per Eqs. (3.12a) and (3.12b) after Eq. (3.12c) has been dropped. Despite the relaxation, much of the previously developed ideas translate to the proof.

Proof. As before we may assume without loss of generality that $x_D = 0$ as well as $\phi_s(0) = \phi_a(0) = 1$ and $\phi_s(z), \phi_a(z) > 0$ for any $z \in [0, 1]$. Let $\alpha_1 \neq \alpha_2$ be two different perturbations of σ_a as in the proof of Lemma 3.1.2 and define $\tau_{1/2}, v_{1/2}$ as well as α_δ and $\tau_\delta = \text{dist}(0, \text{supp}(\alpha_\delta))$ in the same way as before. If $\alpha_\delta = 0$ we set $\tau_\delta = \infty$. Note that If $\tau_1 > \tau_2$ or $\tau_1 < \tau_2$ then obviously the measurements will differ at times between $2\tau_1$ and $2\tau_2$ in direction v_1 or v_2 as the one corresponding to the smaller first impact time

will be influenced by an absorbing perturbation and differ from the offline measurement while the other won't. Therefore we assume without loss of generality that $\tau_1 = \tau_2$. Note that $\tau_\delta \geq \tau_1 = \tau_2$ may have increased and $\tau_\delta > 0$ due to $\alpha_{1/2} \in K(\phi_a \mid x_D)$. There might be multiple possibilities for the first impact direction, i.e. the direction v such that $v\tau_\delta \in \partial \text{supp}(\alpha_\delta)$ and we want to choose a direction $v_\delta \in \mathbb{S}^2$ with that property such that it points towards the kernel centre b_{\min} with the smallest possible bandwidth parameter h_{\min} and without loss of generality assume that the corresponding weight is positive, i.e. $w_{\min} > 0$, and corresponds to a summand in α_1 .

Case 1: $\tau_1 = \tau_2 = \tau_\delta < \infty$. Define

$$Q_1^{(1/2)}(\varepsilon) = \frac{\mathfrak{m}_{\text{on},1/2}^1(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)}{\mathfrak{m}_{\text{off},1/2}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)} \quad (\text{A.16})$$

$$Q_2^{(1/2)}(\varepsilon) = \frac{\sum_{j=2}^{\infty} \mathfrak{m}_{\text{on},1/2}^j(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)}{\mathfrak{m}_{\text{off},1/2}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)} \quad (\text{A.17})$$

and note that the differential absorption, i.e. our data for each set of parameters (which are assumed to result in identical measurements), can be expressed as

$$Q^{(1/2)}(\varepsilon) = Q_1^{(1/2)}(\varepsilon) + Q_2^{(1/2)}(\varepsilon) = \frac{\mathfrak{m}_{\text{on},1/2}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)}{\mathfrak{m}_{\text{off},1/2}(v_\delta, 2(\tau_\delta + \varepsilon) \mid x_D)}.$$

A similar argument that was used to show Eq. (A.10) can be used to show that

$$Q_1^{(1/2)}(\varepsilon) \rightarrow \frac{\mathfrak{m}_{\text{off},1}^1(v_\delta, 2\tau_\delta \mid x_D)}{\mathfrak{m}_{\text{off},1}(v_\delta, 2\tau_\delta \mid x_D)} = \frac{\mathfrak{m}_{\text{off},2}^1(v_\delta, 2\tau_\delta \mid x_D)}{\mathfrak{m}_{\text{off},2}(v_\delta, 2\tau_\delta \mid x_D)} \quad \text{as } \varepsilon \rightarrow 0$$

slower than

$$Q_2^{(1/2)}(\varepsilon) \rightarrow 1 - \frac{\mathfrak{m}_{\text{off},1/2}^1(v_\delta, 2\tau_\delta \mid x_D)}{\mathfrak{m}_{\text{off},1/2}^1(v_\delta, 2\tau_\delta \mid x_D)} \quad \text{as } \varepsilon \rightarrow 0.$$

Further, analogously to the proof of Lemma 3.1.2, we have that $Q_1^{(1)}(\varepsilon)$ converges at a rate of $\int_0^\varepsilon \phi_a(1 - \frac{z}{h_{\min}}) dz$ slower than $Q_1^{(2)}(\varepsilon)$ (due to w_{\min} corresponding to α_1). As such

$$\limsup_{\varepsilon \rightarrow 0} \left| \frac{Q^{(1)}(\varepsilon) - Q^{(2)}(\varepsilon)}{\int_0^\varepsilon \phi_a(1 - \frac{z}{h_{\min}}) dz} \right| > 0$$

which implies $P_1 \neq P_2$ as desired. Note that scattering from the kernels is locally negligible for the limit due to Eq. (3.14) which is even more critical for the next case.

Case 2: $\tau_1 = \tau_2 < \tau_\delta < \infty$ but identical scattering proportionality constants. Essentially identical to case 1 where $\tau_1 = \tau_2 = \tau_\delta < \infty$ except that

$$Q_1^{(1/2)}(\varepsilon) \rightarrow \frac{\mathbf{m}_{\text{on},1}^1(v_\delta, 2\tau_\delta | x_D)}{\mathbf{m}_{\text{off},1}(v_\delta, 2\tau_\delta | x_D)} = \frac{\mathbf{m}_{\text{on},2}^1(v_\delta, 2\tau_\delta | x_D)}{\mathbf{m}_{\text{off},2}(v_\delta, 2\tau_\delta | x_D)}$$

and again we have that $Q_1^{(1)}(\varepsilon)$ converges at a rate of $\int_0^\varepsilon \phi_{\mathfrak{a}}(1 - \frac{z}{h_{\min}}) dz$ slower than $Q_1^{(2)}(\varepsilon)$.

Case 3: Differing proportionality constants. Note that $\tau_\delta > \tau_1 = \tau_2$ or we can fall back to case 1. First assume that the proportionality between σ_s and α is different. Let $\bar{\tau} := \tau_1 = \tau_2$ and note that this is the smallest time where a difference in scattering can be observed but the differential absorption fields $\alpha_{1/2}$ are identical until $\tau_\delta > \bar{\tau}$. In a similar way to v_δ let \bar{v} be the direction pointing to the smallest kernel at a distance of $\bar{\tau}$ from x_D . As before denote this width with h_{\min} which is necessarily (i.e. by assumption of the theorem) identical for both sets of non-differential parameters. Consider $\tilde{\mathbf{m}}_{\text{on/off}}$ which matches both $\mathbf{m}_{\text{on},1}$ and $\mathbf{m}_{\text{on},2}$ in the ambient constants as well as the differential absorption for times $t < 2\bar{\tau} + \varepsilon$ for ε sufficiently small but no scattering from the kernels. In other words non-differential proportionality constants identical to 0. By construction it must be true that

$$\frac{\mathbf{m}_{\text{on},1/2}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)}{\mathbf{m}_{\text{off},1/2}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)} - \frac{\tilde{\mathbf{m}}_{\text{on}}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)}{\tilde{\mathbf{m}}_{\text{off}}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Upon closer inspection it is evident that the dominant term in the numerator takes the form $C w_s \phi_{\mathfrak{s}}(1 - \frac{\varepsilon}{h_{\min}}) \int_0^\varepsilon \phi_{\mathfrak{a}}(1 - \frac{z}{h_{\min}}) dz$ where w_s is the weight parameter of the scattering kernel while $C > 0$ is a constant that only depend on the *known* and strictly positive ambient scattering rate. As such,

$$\frac{\frac{\mathbf{m}_{\text{on},1/2}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)}{\mathbf{m}_{\text{off},1/2}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)} - \frac{\tilde{\mathbf{m}}_{\text{on}}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)}{\tilde{\mathbf{m}}_{\text{off}}(\bar{v}, 2(\bar{\tau} + \varepsilon) | x_D)}}{\phi_{\mathfrak{s}}(1 - \frac{\varepsilon}{h_{\min}}) \int_0^\varepsilon \phi_{\mathfrak{a}}(1 - \frac{z}{h_{\min}}) dz} \quad (\text{A.18})$$

has a non-trivial limit which uniquely determines w_s and thereby the proportionality

constant for σ_s . Similarly, if the proportionality only differs in the σ_a component we additionally match $\tilde{m}_{\text{on/off}}$ in the σ_s related constant and consider

$$\frac{\frac{m_{\text{on},1/2}(\bar{v},2(\bar{\tau}+\varepsilon)|x_D)}{m_{\text{off},1/2}(\bar{v},2(\bar{\tau}+\varepsilon)|x_D)} - \frac{\tilde{m}_{\text{on}}(\bar{v},2(\bar{\tau}+\varepsilon)|x_D)}{\tilde{m}_{\text{off}}(\bar{v},2(\bar{\tau}+\varepsilon)|x_D)}}{\int_0^\varepsilon \phi_{\mathfrak{s}}(1 - \frac{z}{h_{\min}}) dz \int_0^\varepsilon \phi_{\mathfrak{a}}(1 - \frac{z}{h_{\min}}) dz} \quad (\text{A.19})$$

which, as this covers all possible cases, concludes the proof. \square

A.2 Implementation & simulation details

A.2.1 Regularisation terms and hyper-parameter selection

As can be seen in the figures Figs. 5.4 and 5.6, the centre lines are splines with 6 knots, the source, which is anchored to the ground, and five others. Their down-wind displacement is determined by the wind $\vec{\eta}_{\text{amb}} = 1.25 \frac{\text{m}}{\text{s}} \vec{e}_1$, which is assumed known, along with a parametric plume rise vector. We set $t = 75\text{s}$, resulting in a plume of size roughly 120m, and used $\Delta_{\text{step}} = 1.67\text{s}$ which results in an approximation with (theoretical) L_1 error from quadrature of about $\approx 2.5\%$ when measured against exact integration over the spline. That error does however not affect reconstruction in the non-perturbed instance as the ground truth is equally finite/discrete.

In the situation of Section 5.2.2 the parameter θ has 14 components which is to say $\theta \in \Theta \subseteq \mathbb{R}^{14}$. A log-proportionality constant θ_s for σ_s log-release rate θ_0 , a ground-based source term described by its x_1 - and x_2 -coordinates as well as 10 parameters that model the width and height of the centre-line by means of a linear spline with 5 components each, in positions $\theta_{\text{up}} = \theta_{5:9}$ and $\theta_h = \theta_{10:14}$ respectively. Although we have typically adapted a classical, or frequentist, viewpoint in the developments of our theory, the regularisation process is arguably easier understood in a Bayesian setting. In that case the regularisation penalty $R(\theta, \mathbf{m}^{\text{on}}, \mathbf{m}^{\text{off}})$ in Eq. (5.8) can be thought of as the log-prior distribution for the parameter θ of interest and consists of a baseline regulariser R_0 and a data dependent

term as specified in Algorithm 2. We put $R_0(\theta) = (\theta - \theta^{\text{prior}})^\top V_{\text{prior}}(\theta - \theta^{\text{prior}})$ for

$$(V_{\text{prior}})_{5:9,5:9} = (V_{\text{prior}})_{10:14,10:14} = 16 \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \quad (\text{A.20})$$

which is singular and has its null space spanned by the constant vector. The prior used for the logarithm of the five components, which are always positive, is a Gaussian with zero mean and precision matrix whose maximal eigenvalue is $\lambda_5 = 40 + 8\sqrt{5} \leq 60$. The result is a quadratic penalty that promotes smoothness in the sense that the preferred parameters will be roughly constant which is seen in the single scattering reconstructions in Fig. 5.4. There the parameter estimates wouldn't be stable without this regularisation term. The largest eigenvalue λ_5 is the inverse of the a priori variance in the direction of the corresponding eigenvector. All this put together, the prior is such that it corresponds to an assumption that deviations in non-smooth directions of the logarithm have their standard deviation bounded by roughly 15%. This means we are rather confident that the logarithmic deviation, which is very similar to relative change, is within $\pm 30\%$ between one spline knot and the next which corresponds to atmospheric dispersion over a period of 15s.

In the case of wide and multiple FOVs the width component θ_h is also subject to a shrinkage penalty due to reasons described in the main text which is also applied to the logarithmic weight component $\theta_0 = \log(\int_X \alpha(x)dx)$ and is achieved by adding a fraction of the diagonal of $\nabla_\theta^2 \text{Loss}(\theta^{\text{ssct}} | \mathbf{m}, \mathbf{n})$ to V_{prior} . It is done such that the estimated change by means of Eq. (4.60), when averaged over widths and in the weight, is bounded by the estimated standard error and also mustn't exceed 10%. This ensures that the mean squared error is not dominated by bias when there is little scattering. This procedure also has an intuitive Bayesian interpretation, see e.g. [64, sec. 7.2], and can be seen as demanding that the posterior mean be close to the data mean in the case of Gaussian

likelihood and prior, here applied to the logarithm.

Lastly, a non-centred Gaussian prior for the scattering weight $\theta_s = \log(\int_X \sigma_s(x)dx)$ was chosen. It has unit variance and was selected such that the mean optical thickness of the plume is ≈ 2 which results in some low-order scattering. The prior variance means that typical deviations in the logarithm are ± 2 which is almost 2 orders of magnitude on a natural scale. As such, the prior probabilities for the values used in all the simulations are reasonably large and the data does have a noticeable effect on its reconstructed value in the case of wider FOVs. Despite the regularisation we observed that the reconstructed scattering component is in fact more variable and less accurately estimated than the other α -related components. This is most likely due to the low sensitivity of the data to the scattering as suggested in Section 3.2 as well as the proof of Theorem 3.1.4. Note that when multiple FOVs are used and the optical thickness is low, then the likelihood has virtually no sensitivity towards that parameter and it will be very close to the prior mean. Of course, this does no harm since the estimates are in those instances driven by the narrow FOV component and largely independent of the scattering parameter.

As mentioned in the main body, we use MC sampling based algorithms to solve the RTE, approximate the measurement gradients and solve the optimisation problem Eq. (5.8). We recall the quantities Eqs. (5.35) to (5.37) for definitions. The used algorithms can then be described in pseudo-code Algorithms 1 and 2.

The value for scattering strength used in the estimation of uncertainty in the graphics Figs. 5.8 to 5.11 is roughly the one-sided 95% confidence bound for the prior, i.e. the mean plus one and a half standard deviations.

A.2.2 Turbulence process

As mentioned in Section 5.2 we have not only considered smooth plumes that satisfy Eq. (2.11) for a set of parameters resulting in very “smooth” images but also conducted simulations where the ground truth was perturbed in a way to exhibit behaviour that is meant to mimic turbulence. Since the kernel functions are Gaussian, which are numerically very similar to good kernels as in Definition 2.1.1, we can make the following simple

Algorithm 1 RTE-Gradient – MC based computation of derivatives. Note that the algorithm terminates in finite time even for $K_{\max} = \infty$ due to $\mathbb{P}(\prod_{k=1}^{\infty} \mathbf{a}_k = 0) = 1$ so that the sums in lines 15 to 17 equal zero eventually.

- 1: **Input:** $\{t_j : j \in 1:N_t\}$, $\{v_i : i \in 1:N_v\}$, $N_{\vec{p}}$, $K_{\max} \in \mathbb{N}$, $\theta \in \Theta$ along with $\mathcal{G} : \Theta \rightarrow \Xi$, regulariser R , ambient intensity H_a and data $\mathbf{m}^{\text{on/off}}$
- 2: **Output:** Estimates $\hat{D}^\#(\theta)$ and $\hat{H}^\#(\theta)$ of quantities from Eqs. (5.22) and (5.24).
- 3: $k \leftarrow 0$
- 4: **for** $i \in 1:N_v$ **do**
- 5: $\mathbb{X}_i \leftarrow \bigcup_{l=1}^{N_{\vec{p}}} \{(x_D, v_i, 0, 1)\}$ ▷ Multi-sets of $N_{\vec{p}}$ equal starting points
- 6: **end for**
- 7: $\xi[\theta] = (\alpha[\theta], \sigma_s[\theta], \sigma_a[\theta], f_p[\theta]) \leftarrow \mathcal{G}(\theta)$
- 8: **while** $k \leq K_{\max}$ **do**
- 9: $k += 1$
- 10: **for** $i \in 1:N_v$ **do**
- 11: **for** $j \in 1:N_t$ **do**
- 12: $\hat{m}_{\text{on/off}}^k[\theta](v_i, t_j) \leftarrow 0$
- 13: $\nabla_\theta \hat{m}_{\text{on/off}}^k[\theta](v_i, t_j) \leftarrow 0$
- 14: **for** $(\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a}) \in \mathbb{X}_i$ **do** ▷ use Eqs. (5.35) to (5.37)
- 15: $\hat{m}_{\text{on/off}}^k[\theta](v_i, t_j) += M_{\text{on/off}}^k[t_j](\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a} | \xi[\theta])$
- 16: $\nabla_\theta \hat{m}_{\text{on/off}}^k[\theta](v_i, t_j) += \nabla_\theta M_{\text{on/off}}^k[t_j](\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a} | \xi[\theta])$
- 17: $\nabla_\theta \hat{m}_{\text{on/off}}^k[\theta](v_i, t_j) += \frac{\nabla_\theta p_\theta^k(\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a})}{p_\theta^k(\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a})} M_{\text{on/off}}^k[t_j](\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a} | \xi[\theta])$
- 18: **end for**
- 19: **end for**
- 20: **for** $(\mathbf{x}, \mathbf{v}, \mathbf{t}, \mathbf{a}) \in \mathbb{X}_i$ **do**
- 21: $\mathbf{a}_k, \mathbf{t}_k, \mathbf{v}_k, \mathbf{x}_k \leftarrow$ Eqs. (5.31) to (5.33) ▷ Draw & append sample values
- 22: **end for**
- 23: **end for**
- 24: **end while**
- 25: $\hat{m}_{\text{on/off}}[\theta], \nabla_\theta \hat{m}_{\text{on/off}}[\theta] \leftarrow \sum_{k=1}^{K_{\max}} \hat{m}_{\text{on/off}}^k[\theta], \sum_{k=1}^{K_{\max}} \nabla_\theta \hat{m}_{\text{on/off}}^k[\theta]$
- 26: **compute** $\hat{P}_{i,j}(\theta), \nabla_\theta \hat{P}_{i,j}(\theta)$ based on $\hat{m}_{\text{on/off}}[\theta](v_i, t_j), \nabla_\theta \hat{m}_{\text{on/off}}[\theta](v_i, t_j)$ and Eq. (5.4)
 $\hat{D}^\#(\theta), \hat{H}^\#(\theta) \leftarrow$ Eqs. (5.22) and (5.24) using $\hat{P}_{i,j}(\theta), \nabla_\theta \hat{P}_{i,j}(\theta), R$ and H_a
- 27: **return** $\hat{D}^\#(\theta), \hat{H}^\#(\theta)$

Algorithm 2 Optim - MLE estimation for $\theta \in \Theta$ MLE estimation for $\theta \in \Theta$. The estimate $\hat{\theta}^{\text{ssct}}$ uses only single scattering and is used an initial value for the optimisation. The values θ are on a logarithmic scale so that the procedure in lines 17 to 20 limits the relative error (due to shrinkage) on the natural scale through a weighted distance and the choice of regularisation parameter λ depends on data through \hat{H} . For line 18 to have shrinkage effect, the parameters must be larger than 1 on a natural scale which can easily be achieved by a change of unit/scale, e.g. [kg] \rightarrow [g] or [m] \rightarrow [cm]. Step size in lines 7 and 28 is adaptive to ensure decrease in objective value.

```

1: Input:  $\{t_j : j \in 1:N_t\}, \{v_i : i \in 1:N_v\}, N_{\bar{p}}, K_{\text{RTE}} \in \mathbb{N}, \theta^{(0)} \in \Theta$  partitioned per
   Eq. (5.44) along with  $\mathcal{G} : \Theta \rightarrow \Xi$ , tolerances  $\text{tol}_\theta, \text{tol}_{\text{bias}} > 0$ , search step  $\varepsilon > 0$ , base
   regulariser  $R_0(\theta) = (\theta - \theta^{\text{prior}})^\top V_{\text{prior}}(\theta - \theta^{\text{prior}})$ , ambient intensity  $H_a$  and data  $\mathbf{m}^{\text{on/off}}$ 
2: Output:  $\hat{\theta}^{\text{RTE}}$  which maximises Eq. (5.8)
3:  $\Delta_\theta \leftarrow \infty$ 
4:  $i \leftarrow 0$ 
5: while  $\Delta_\theta > \text{tol}_\theta$  do
6:    $\hat{D}^\#(\theta^{(i)}), \hat{H}^\#(\theta^{(i)}) \leftarrow \text{RTE-Gradient}(\theta = \theta^{(i)}, K_{\max} = 1, R = R_0, \dots)$ 
7:    $\theta^{(i+1)} \leftarrow \text{Eq. (5.20) using } \theta^{(i)}, \hat{D}^\#(\theta^{(i)}) \text{ and } \hat{H}^\#(\theta^{(i)})$ 
8:    $\Delta_\theta \leftarrow \max \left\{ \|\theta_j^{(i+1)} - \theta_j^{(i)}\|_2 / \|\theta_j^{(i)}\|_2 : j \in \{0, \text{source, up, } h\} \right\}$ 
9:    $i \leftarrow i + 1$ 
10: end while
11:  $\hat{\theta}^{\text{ssct}} \leftarrow \theta^{(i)}$ 
12:  $V_{\text{reg}}, \Delta_{\text{var}} \leftarrow 0 \in \mathbb{R}^{N_\theta \times N_\theta}$ 
13:  $\text{diag}(V_{\text{reg}})_{(0,h)} \leftarrow \text{diag}(\hat{H}^\#(\theta^{(i)}))_{(0,h)}$   $\triangleright$  Shrinkage in  $\theta_0$  and  $\theta_h$  component only
14:  $\text{diag}(\Delta_{\text{var}})_{(0,h)} \leftarrow \text{diag}(\hat{H}^\#(\theta^{(i)})^{-1})_{(0,h)}$ 
15:  $(\lambda, \Delta_{\text{bias}}) \leftarrow (1, \infty)$ 
16:  $\text{tol}_{\text{bias}} \leftarrow \min\{\text{tol}_{\text{bias}}, \text{mean}(\Delta_{\text{var}})^{1/2}\}$ 
17: while  $\Delta_{\text{bias}} > \text{tol}_{\text{bias}}$  do
18:    $\hat{\theta}^{\text{reg}} \leftarrow [\hat{H}^\#(\hat{\theta}^{\text{ssct}}) + \lambda V_{\text{reg}}]^{-1} \hat{H}^\#(\hat{\theta}^{\text{ssct}}) \hat{\theta}^{\text{ssct}}$   $\triangleright$  Eq. (4.60) for centred prior
19:    $\lambda \leftarrow \frac{\lambda}{1+\varepsilon}$ 
20:    $\Delta_{\text{bias}} \leftarrow \frac{\|(\Delta_{\text{var}}[\hat{\theta}^{\text{reg}} - \hat{\theta}^{\text{ssct}}])_0\|_1 / \|\text{diag}(\Delta_{\text{var}})_0\|_1 + \|(\Delta_{\text{var}}[\hat{\theta}^{\text{reg}} - \hat{\theta}^{\text{ssct}}])_h\|_1 / \|\text{diag}(\Delta_{\text{var}})_h\|_1}{2}$ 
21: end while
22:  $R_{\text{data}}(\theta) = (\theta - \theta^{\text{prior}})^\top [V_{\text{prior}} + \lambda V_{\text{reg}}](\theta - \theta^{\text{prior}})$ 
23:  $i \leftarrow 0$ 
24:  $\Delta_\theta \leftarrow \infty$   $\triangleright$  Restart for full RTE optimiser
25:  $\theta^{(0)} \leftarrow \hat{\theta}^{\text{ssct}}$   $\triangleright$  Initial value from single-scattering approximation
26: while  $\Delta_\theta > \text{tol}_\theta$  do
27:    $\hat{D}^\#(\theta^{(i)}), \hat{H}^\#(\theta^{(i)}) \leftarrow \text{RTE-Gradient}(\theta = \theta^{(i)}, K_{\max} = K_{\text{RTE}}, R = R_{\text{data}}, \dots)$ 
28:    $\theta^{(i+1)} \leftarrow \text{Eq. (5.20) using } \theta^{(i)}, \hat{D}^\#(\theta^{(i)}) \text{ and } \hat{H}^\#(\theta^{(i)})$ 
29:    $\Delta_\theta \leftarrow \max \left\{ \|\theta_j^{(i+1)} - \theta_j^{(i)}\|_2 / \|\theta_j^{(i)}\|_2 : j \in \{0, \text{source, up, } h\} \right\}$ 
30:    $i \leftarrow i + 1$ 
31: end while
32: return  $\hat{\theta}^{\text{RTE}} = \theta^{(i)}$ 

```

observations which will form the basis for our perturbation approach.

- **Additivity:** If $\frac{1}{h_1}\phi\left(\frac{y}{h_1}\right)$ is a centred Gaussian function on the real line with variance h_1^2 and $\mathbf{x} \sim \text{Normal}(0, h_2^2)$ then

$$\mathbb{E}\left[\frac{1}{h_1}\phi\left(\frac{y - \mathbf{x}}{h_1}\right)\right] = \frac{1}{\sqrt{h_1^2 + h_2^2}}\phi\left(\frac{y}{\sqrt{h_1^2 + h_2^2}}\right).$$

This procedure is easily extended in higher dimensions for Gaussian functions with diagonal covariance matrices.

- **Meandering:** If in the previous point one replaces \mathbf{x} with a stationary stochastic process $(\mathbf{x}(t) : t \in [0, \infty))$ with marginal distribution $\mathbf{x}(t) \sim \text{Normal}(0, h_2^2)$ then

$$\mathbb{E}\left[\frac{1}{h_1}\phi\left(\frac{y - \mathbf{x}(t)}{h_1}\right)\right] = \frac{1}{\sqrt{h_1^2 + h_2^2}}\phi\left(\frac{y}{\sqrt{h_1^2 + h_2^2}}\right)$$

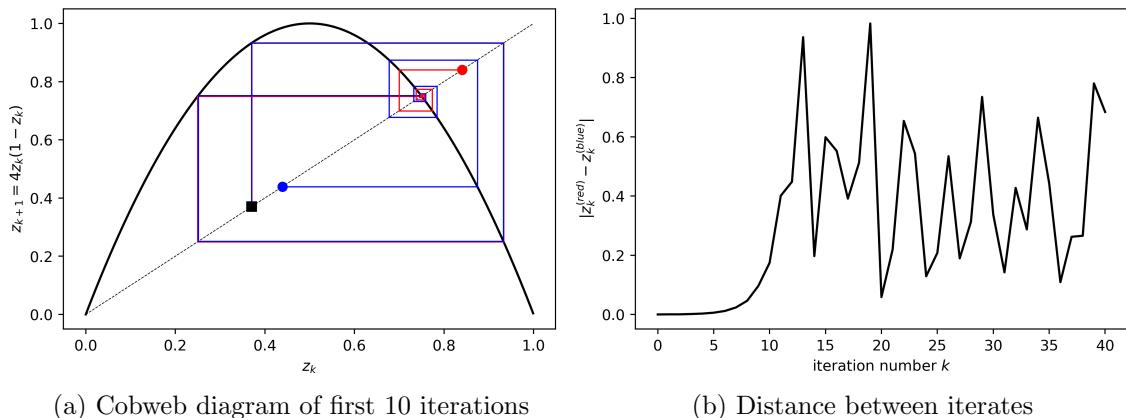
extends the previous observation to any $t \in [0, \infty)$. The correlation structure in $\mathbf{x}(t)$ allows additional degrees of freedom in the model.

- **Multi-scale structure:** Letting $\mathbf{y} \sim \text{Normal}(0, h_1^2 - \tilde{h}_1^2)$ independent of the stochastic process $\mathbf{x}(\cdot)$ we have for all $t \in [0, \infty)$

$$\mathbb{E}\left[\frac{1}{\tilde{h}_1}\phi\left(\frac{y - \mathbf{x}(t) - \mathbf{y}}{\tilde{h}_1}\right)\right] = \frac{1}{\sqrt{h_1^2 + h_2^2}}\phi\left(\frac{y}{\sqrt{h_1^2 + h_2^2}}\right).$$

Sampling multiples and distributing the weight provides additional degrees of freedom and allows the modelling of “eddys” commonly observed in plumes.

Although a random perturbation of the average model can be obtained in many ways we should mention that this would be non-trivial if a Navier-Stokes based turbulence model was used as the non-linearities would mean that matching the parameters, i.e. obtaining identities such as Eq. (A.22), isn’t straightforward. This, together with the computational burden that a simulation of turbulence paired with an RTE solution would bear, is a major factor for the following simplified modelling approach. We want to describe the perturbed



(a) Cobweb diagram of first 10 iterations

(b) Distance between iterates

Figure A.1: Behaviour of logistic map $z_{k+1} = f(z_k) = 4z_k(1 - z_k)$. The black square in the left diagram is at the starting point (z_0, z_0) and can be thought of as a collection of blue and red nearby “particles” $z_0^{(\text{red})} = 0.37069647$ and $z_0^{(\text{blue})} = 0.37028614$. The round dots are at (z_k, z_k) for blue and red “particles” after 10 iterations respectively. Note that the steps near the fixed point after iteration 3 are very small and hard to see before the trajectories split apart. The plots show that despite being close initially, the red and blue “particles” eventually split and become independent. This sensitivity to the initial value, a characteristic trait of chaotic dynamical systems, serves as motivation for the splitting used to model smaller features in the plume. We think of this procedure in a nested form, i.e. each of the resulting individual “particles”, here red and blue, will be subject to another split eventually. Below τ will model the time until the next split, \mathbf{J} the behaviour at the split, \mathbf{c} the behaviour in between splits and \mathbf{W} determines the ratio of particles that go each way.

gas distribution based on the aforementioned ideas. In order to get perturbed plumes \mathbf{u} as in Eq. (5.47) we will consider random functions of the form

$$\mathbf{u}(\cdot) = u[\theta](\cdot \mid \mathbf{T}) := \frac{\sum_{l=1}^{N_T} \sum_{k=1}^{N_l} u[\theta](\cdot \mid \mathbf{T}_{l,k})}{\sum_{l=1}^{N_T} N_l} \quad (\text{A.21})$$

where $\mathbf{T} = ((\mathbf{T}_{l,k})_{k=1}^{N_l})_{l=1}^{N_T}$ will be specified later and $N_T \in \mathbb{N}$ as well as $N_l \in \mathbb{N}$ for $l \in 1:N_T$ are fixed positive integers. In that situation

$$\mathbb{E}[u[\theta](\cdot \mid \mathbf{T}_{l,k})] = u[\theta] \implies \mathbb{E}[u(\cdot \mid \mathbf{T})] = u[\theta] \quad (\text{A.22})$$

We assume that for fixed l the $u[\theta](\cdot \mid \mathbf{T}_{l,1}), \dots, u[\theta](\cdot \mid \mathbf{T}_{l,N_l})$ are i.i.d. random functions and describe a class of distributions over $C(\mathbb{R}^3)$ that achieves this. For some $n_{\text{split}} \in \mathbb{N}$ we define

$$\mathbb{J}(n_{\text{split}}) := \bigcup_{j \in \mathbb{N}} \{(i_1, \dots, i_j) \in \mathbb{N}^j : 1 \leq i_k \leq n_{\text{split}} \forall k = 1, \dots, j\} \cup \{\emptyset\}$$

which means $\mathbb{J}(n_{\text{split}})$ is the set of all finite size integer tuples with values between 1 and n_{split} along with the empty tuple \emptyset . Consider a random element $\mathbf{J} \sim \text{Jump}(n_{\text{split}}, \kappa_h, \kappa_d)$ indexed by $\mathbb{J}(n_{\text{split}})$ such that $\mathbf{J}(i_1), \dots, \mathbf{J}(i_n)$ are independent for all $i_1, \dots, i_n \in \mathbb{J}(n_{\text{split}})$, $n \in \mathbb{N}$ and

$$\begin{aligned} \mathbf{J}(\emptyset) &= 0 \\ \mathbf{J}(i) &\sim \text{Normal}(0, (1 - \kappa_d)(1 - \kappa_h)^{|i|-1} \kappa_h) \end{aligned} \quad (\text{A.23})$$

where $|i| = N \iff i \in \mathbb{N}^N$, i.e. the length of the tuple, and $|\emptyset| = 0$. Eq. (A.23) makes sense for $\kappa_h \in (0, 1)$ in which case an evaluation of the partial sums from a geometric series yields

$$\mathbb{V} \left(\sum_{j=1}^{|i|} \mathbf{J}(i_{1:j}) \right) = \sum_{j=1}^{|i|} \mathbb{V}(\mathbf{J}(i_{1:j})) = (1 - \kappa_d) (1 - (1 - \kappa_h)^{|i|}) < 1 - \kappa_d \quad (\text{A.24})$$

for any $i \in \mathbb{J}(n_{\text{split}})$ and $i_{1:j}$ consists of the first $j \leq |i|$ components of $i \in \mathbb{N}^{|i|} \subseteq \mathbb{J}(n_{\text{split}})$ as introduced in Eq. (2.2) and we put $i_{1:0} = \emptyset$. Further define $\mathbf{W} \sim \text{Weight}(n_{\text{split}})$ indexed by $\mathbb{N} \cup \{0\}$ recursively as

$$\begin{aligned}\mathbf{W}_\emptyset(0) &= 1 \\ \mathbf{W}(n) &\sim (\mathbf{W}(n-1) \otimes_{\text{out}} \mathbb{1}_{n_{\text{split}}}) \odot \text{Dirichlet}^{\otimes^{n-1} 1:n_{\text{split}}}(\mathbb{1}_{n_{\text{split}}}).\end{aligned}\tag{A.25}$$

Here the symbol $\otimes^{n-1} 1:n_{\text{split}}$ denotes the n -fold cartesian product of $1:n_{\text{split}}$, i.e.

$$\otimes^{n-1} 1:n_{\text{split}} = \{(i_1, \dots, i_n) : 1 \leq i_k \leq n_{\text{split}} \forall k = 1, \dots, n\} = \{i \in \mathbb{J}(n_{\text{split}}) : |i| = n\}$$

while \otimes_{out} is the outer product on arrays

$$\otimes_{\text{out}} : (0, 1)^{\otimes^{n-1} 1:n_{\text{split}}} \times (0, 1)^{1:n_{\text{split}}} \rightarrow (0, 1)^{\otimes^{n-1} 1:n_{\text{split}}}, \quad (y \otimes_{\text{out}} x)_i = y_{i_{1:n-1}} x_{i_n}$$

so that $\mathbf{W}(n) \in (0, 1)^{\otimes^{n-1} 1:n_{\text{split}}}$ and the multiplication \odot is to be understood entry-wise.

By construction we have

$$\sum_{j=1}^{n_{\text{split}}} \mathbf{W}_{(i,j)}(n) = \mathbf{W}_i(n-1)\tag{A.26}$$

and therefore it must be true that total mass is preserved at all times $n \in \mathbb{N}$ which means

$$\sum_{i \in \otimes^{n-1} 1:n_{\text{split}}} \sum_{j=1}^{n_{\text{split}}} \mathbf{W}_{(i,j)}(n) = \sum_{i \in \otimes^{n-1} 1:n_{\text{split}}} \mathbf{W}_i(n-1) \implies \|\mathbf{W}(n)\|_1 = \mathbf{W}_\emptyset(0) = 1.$$

Thirdly we consider $\boldsymbol{\tau} \sim \text{Time}(n_{\text{split}}, \tau_{\text{split}})$ also indexed by $\mathbb{N} \cup \{0\}$ and defined via

$$\begin{aligned}\boldsymbol{\tau}_\emptyset(0) &= 0 \\ \boldsymbol{\tau}(n) &\sim (\boldsymbol{\tau}(n-1) + \text{Gamma}^{\otimes^{n-1} 1:n_{\text{split}}}(\tau_{\text{split}}, 1)) \otimes_{\text{out}} \mathbb{1}_{n_{\text{split}}}\end{aligned}\tag{A.27}$$

where $\text{Gamma}^{\otimes^{n-1} 1:n_{\text{split}}}(\tau_{\text{split}}, 1) \in (0, \infty)^{\otimes^{n-1} 1:n_{\text{split}}}$ has independent **Gamma** distributed entries with rate 1 and shape τ_{split} . Note that unlike \mathbf{J} , which has independent entries, $\boldsymbol{\tau}$

has independent increments in the sense of

$$\boldsymbol{\tau}(n+k) - \boldsymbol{\tau}(n) \otimes_{\text{out}}^k \mathbb{1}_{n_{\text{split}}} \perp \boldsymbol{\tau}(n) \quad (\text{A.28})$$

for all $n, k \in \mathbb{N}$ and $\otimes_{\text{out}}^k \mathbb{1}_{n_{\text{split}}}$ means multiplication with $\mathbb{1}_{n_{\text{split}}}$ is executed k times in a row. Given $\boldsymbol{\tau}$ we define a selector process of active indexes

$$\mathbb{A}(t \mid \boldsymbol{\tau}) = \bigcup_{i \in \mathbb{J}(n_{\text{split}}) \setminus \{\emptyset\}} \left\{ i_{1:|i|-1} : \boldsymbol{\tau}_{i_{1:|i|-1}}(|i|-1) < t \leq \boldsymbol{\tau}_i(|i|) \right\} \quad (\text{A.29})$$

An important property of the selector Eq. (A.29) is that it preserves total mass, i.e.

$$\sum_{i \in \mathbb{A}(t \mid \boldsymbol{\tau})} \mathbf{W}_i(|i|) = \lim_{s \searrow 0} \sum_{i \in \mathbb{A}(s \mid \boldsymbol{\tau})} \mathbf{W}_i(|i|) = \mathbf{W}_{\emptyset}(0) = 1. \quad (\text{A.30})$$

To see this note that $\mathbb{A}(t \mid \boldsymbol{\tau}) \neq \mathbb{A}(t+dt \mid \boldsymbol{\tau})$ if and only if $t = \boldsymbol{\tau}_i(|i|)$ for some $i \in \mathbb{J}(n_{\text{split}})$. Let (t', i') be such a pair for given $\boldsymbol{\tau}$. By construction $\boldsymbol{\tau}_{(i'_{1:|i'|-1}, j)}(|i'|) = \boldsymbol{\tau}_{i'}(|i'|)$ for all $j \in 1:n_{\text{split}}$ so that

$$\begin{aligned} \{i'_{1:|i'|-1}\} &\subseteq \mathbb{A}(t' \mid \boldsymbol{\tau}) \setminus \mathbb{A}(t' + dt \mid \boldsymbol{\tau}) \\ \{(i'_{1:|i'|-1}, j) : j \in 1:n_{\text{split}}\} &\subseteq \mathbb{A}(t' + dt \mid \boldsymbol{\tau}) \setminus \mathbb{A}(t' \mid \boldsymbol{\tau}) \end{aligned}$$

and we get from the same argument that led to Eq. (A.26)

$$\sum_{i \in \mathbb{A}(t' \mid \boldsymbol{\tau})} \mathbf{W}_i(|i|) - \sum_{i \in \mathbb{A}(t' + dt \mid \boldsymbol{\tau})} \mathbf{W}_i(|i|) = \mathbf{W}_{i'_{1:|i'|-1}}(|i'| - 1) - \sum_{j \in 1:n_{\text{split}}} \mathbf{W}_{(i'_{1:|i'|-1}, j)}(|i'|) = 0.$$

Consider the Ornstein-Uhlenbeck process $\boldsymbol{\eta} \sim \text{OU}(\kappa_{\text{OU}}, \eta_{\text{OU}}, \eta_0, \tau_0)$, i.e. a process defined via the SDE

$$\begin{aligned} \boldsymbol{\eta}(\tau_0) &= \eta_0 \\ d\boldsymbol{\eta}_j(t) &= -\eta_{\text{OU}} \boldsymbol{\eta}_j(t) dt + \kappa_{\text{OU}} d\mathbf{B}(t) \quad t \in (\tau_0, \infty) \end{aligned} \quad (\text{A.31})$$

where \mathbf{B} is a standard Brownian motion. It is a standard result from stochastic calculus that for η_0 as in Eq. (A.31) we have

$$\mathbf{B} \perp \eta_0 \sim \text{Normal} \left(0, \frac{\kappa_{\text{OU}}^2}{2\eta_{\text{OU}}} \right) \implies \boldsymbol{\eta}(t) \sim \text{Normal} \left(0, \frac{\kappa_{\text{OU}}^2}{2\eta_{\text{OU}}} \right) \quad (\text{A.32})$$

for any $t > 0$. We also define $\mathbf{c}_0 | \boldsymbol{\tau} \sim \text{Drift}(n_{\text{split}}, \kappa_d, \eta_{\text{OU}} | \boldsymbol{\tau})$ indexed by $(0, \infty) \times \mathbb{J}(n_{\text{split}})$

$$\begin{aligned} \mathbf{c}_0(\cdot, \emptyset) &\sim \text{OU} \left(\sqrt{2\kappa_d \eta_{\text{OU}}}, \eta_{\text{OU}}, \text{Normal}(0, \kappa_d), 0 \right) \\ \mathbf{c}_0^*(\cdot, i) | \boldsymbol{\tau} &\sim \text{OU} \left(\sqrt{2\kappa_d \eta_{\text{OU}}}, \eta_{\text{OU}}, \mathbf{c}_0(\boldsymbol{\tau}_i(|i|), i_{1:|i|-1}), \boldsymbol{\tau}_i(|i|) \right) \\ \mathbf{c}_0(t, i) &= \begin{cases} \mathbf{c}_0^*(t, i) & t \in (\boldsymbol{\tau}_i(|i|), \infty) \\ \mathbf{c}_0(t, i_{1:|i|-1}) & t \in (0, \boldsymbol{\tau}_i(|i|)] \end{cases} \end{aligned} \quad (\text{A.33})$$

where $\mathbf{c}_0^*(t, i)$ are independent for each i given $\boldsymbol{\tau}$ and $\mathbf{c}_0^*(t, i)$ makes sense for $t > \boldsymbol{\tau}_i(|i|)$ which is always true if $i \in \mathbb{A}(t | \boldsymbol{\tau})$. Thanks to Eq. (A.32) \mathbf{c}_0 satisfies

$$\mathbf{c}_0(t, i) | \boldsymbol{\tau} \sim \text{Normal}(0, \kappa_d) \quad (\text{A.34})$$

for all $(t, i) \in (0, \infty) \times \mathbb{J}(n_{\text{split}})$ even though it is in general not a stationary OU-process on $(0, \infty)$. In other words, Eq. (A.33) preserves the marginal distribution. Note that κ_d is the same as in Eq. (A.23), i.e. \mathbf{c}_0 and \mathbf{J} have a common parameter $\kappa_d \in (0, 1)$. We are ready to define the turbulence process components as $\mathbf{T}_{l,k} = (\boldsymbol{\tau}_{l,k}, \mathbf{J}_{l,k}, \mathbf{W}_{l,k}, \mathbf{c}_{0,l,k})$ independent in for fixed l and distributed according to

$$\begin{aligned} \boldsymbol{\tau}_{l,k} &\sim \text{Time}(n_{\text{split},l}, \tau_{\text{split}}) \\ \mathbf{c}_{0,l,k} | \boldsymbol{\tau}_{l,k} &\sim \text{Drift}^3(n_{\text{split},l}, \kappa_{d,l}, \eta_{\text{OU},l} | \boldsymbol{\tau}_{l,k}) \\ \mathbf{W}_{l,k} &\sim \text{Weight}(n_{\text{split},l}) \\ \mathbf{J}_{l,k} &\sim \text{Jump}^3(n_{\text{split},l}, \kappa_{h,l}, \kappa_{d,l}) \end{aligned} \quad (\text{A.35})$$

where Jump^3 and Drift^3 are identical to the processes from Eqs. (A.23) and (A.33) except extended from \mathbb{R} to \mathbb{R}^3 with independent components. If we impose the independence

condition

$$(\boldsymbol{\tau}_{l,k}, \mathbf{c}_{0,l,k}) \perp\!\!\!\perp \mathbf{W}_{l,k} \perp\!\!\!\perp \mathbf{J}_{l,k}. \quad (\text{A.36})$$

for all $l = 1, \dots, N_T$ and $k = 1, \dots, N_l$ then Eqs. (A.35) and (A.36) specifies the distribution of $\mathbf{T}_{l,k}$ completely and characterises $\mathbf{T} = ((\mathbf{T}_{l,k})_{k=1}^{N_l})_{l=1}^{N_T}$ under the assumption that $\mathbf{T}_{l,k}$ are mutually independent. Recall the instantaneous solutions $u_\delta[\theta](t, \cdot)$ from Eq. (2.11). We put

$$\mathbf{h}(t | \theta, i) = h[\theta](t) \sqrt{(1 - \kappa_{d,l})(1 - \kappa_{h,l})^{|i|}} \quad (\text{A.37})$$

$$\mathbf{b}(t | \mathbf{T}_{l,k}, \theta, i) = b[\theta](t) + h[\theta](t) \sum_{j=1}^{|i|} \mathbf{J}(i_{1:j}) + h[\theta](t) \mathbf{c}_{0,l,k}(t, i) \quad (\text{A.38})$$

and define the turbulent analog

$$u_\delta[\theta](t, \cdot | \mathbf{T}_{l,k}) = \sum_{i \in \mathbb{A}(t|\boldsymbol{\tau})} \frac{\mathbf{W}_i(|i|)}{(\sqrt{2\pi} \mathbf{h}(t | \theta, i))^3} \exp \left(-\frac{1}{2} \left[\frac{\|\cdot - \mathbf{b}(t | \mathbf{T}_{l,k}, \theta, i)\|_2}{\mathbf{h}(t | \theta, i)} \right]^2 \right)$$

and observe that due to Eq. (A.30) we have for all $t > 0$

$$\mathbb{P} \left(\int_{\mathbb{R}^3} u_\delta[\theta](t, x | \mathbf{T}_{l,k}) dx = 1 \right) = 1 \quad (\text{A.39})$$

which means that the integral is preserved almost surely. We claim that additionally we also preserve the expectation at every point, i.e. for all $t > 0$

$$\mathbb{E}[u_\delta[\theta](t, \cdot | \mathbf{T}_{l,k})] = u_\delta[\theta](t, \cdot). \quad (\text{A.40})$$

After forming the analogous expression to Eq. (2.18) and ignoring ambient constants so

$$u[\theta](t, \cdot | \mathbf{T}_{l,k}) = \int_0^t u_\delta[\theta](s, \cdot | \mathbf{T}_{l,k}) ds \quad (\text{A.41})$$

Eq. (A.40) would imply the property sought in Eq. (A.22) because

$$\mathbb{E} \left[\int_0^t u_\delta[\theta](s, \cdot | \mathbf{T}_{l,k}) ds \right] = \int_0^t \mathbb{E} [u_\delta[\theta](s, \cdot | \mathbf{T}_{l,k})] ds = \int_0^t u_\delta[\theta](t, \cdot) dt = u[\theta](t, \cdot).$$

By construction get from Eqs. (A.24) and (A.34)

$$\mathbf{b}_j(t | \mathbf{T}_{l,k}, \theta, i) \sim \text{Normal} \left(b_j[\theta](t), h^2[\theta] \left((1 - \kappa_d) (1 - (1 - \kappa_h)^{|i|}) + \kappa_d \right) \right).$$

Conditional on $\boldsymbol{\tau}, \mathbf{W}$ we may compute the expectation

$$\mathbb{E} [u_\delta[\theta](t, \cdot | \mathbf{T}_{l,k}) | \boldsymbol{\tau}, \mathbf{W}] = \sum_{i \in \mathbb{A}(t|\boldsymbol{\tau})} \mathbf{W}_i(|i|) \mathbb{E} \left[\frac{\exp \left(-\frac{1}{2} \left[\frac{\|\cdot - \mathbf{b}(t | \mathbf{T}_{l,k}, \theta, i)\|_2}{h(t | \theta, i)} \right]^2 \right)}{\left(\sqrt{2\pi} h(t | \theta, i) \right)^3} | \boldsymbol{\tau}, \mathbf{W} \right]$$

and we see that by construction the expectation happens to be a convolution of Gaussian functions (per dimension) which is again a Gaussian function. Adding the variance of $\mathbf{b}_j(t | \mathbf{T}_{l,k}, \theta, i)$ to the squared width $h^2(t | \theta, i)$ we get

$$h^2[\theta] \left((1 - \kappa_d) (1 - (1 - \kappa_h)^{|i|}) + \kappa_d \right) + \left(h[\theta](t) \sqrt{(1 - \kappa_{d,l})(1 - \kappa_{h,l})^{|i|}} \right)^2 = h^2[\theta](t)$$

which is the variance corresponding to $u_\delta[\theta](t, \cdot)$. Since the mean of $\mathbf{b}_j(t | \mathbf{T}_{l,k}, \theta, i)$ by design matches that of $u_\delta[\theta](t, \cdot)$ we get

$$\mathbb{E} \left[\frac{\exp \left(-\frac{1}{2} \left[\frac{\|\cdot - \mathbf{b}(t | \mathbf{T}_{l,k}, \theta, i)\|_2}{h(t | \theta, i)} \right]^2 \right)}{\left(\sqrt{2\pi} h(t | \theta, i) \right)^3} | \boldsymbol{\tau}, \mathbf{W} \right] = u_\delta[\theta](t, \cdot) \quad (\text{A.42})$$

and the sought claim follows from Eq. (A.30).

Using an Euler-Maruyama discretisation with step Δ_{step} turns the OU-process into an AR(1)-process with auto-correlation parameter $1 - \eta_{\text{OU}} \Delta_{\text{step}} \approx e^{-\eta_{\text{OU}} \Delta_{\text{step}}}$ and marginal variance given by $\frac{\kappa_{\text{OU}}^2}{2\eta_{\text{OU}} - \eta_{\text{OU}}^2 \Delta_{\text{step}}} = \frac{\kappa_d}{1 - \eta_{\text{OU}} \frac{\Delta_{\text{step}}}{2}} \approx \frac{\kappa_{\text{OU}}^2}{2\eta_{\text{OU}}} = \kappa_d$. For the purpose of our simulation the discretisation results in roughly 5% error in the variance, less in terms of standard deviation, relative to its OU counterpart. The specific choices can be found in

Table A.1. The number of kernels is random but is typically between 10^3 and 10^4 and thus considerably larger than what is used in the smooth approximation. Note that $l = 1$ corresponds to a smooth plume while for $l = 2, 3$ we only consider the OU-component and expect no jumps. For $l = 4, 5$ we expect on average 2 and 4 jump events to occur in the relevant time frame respectively.

l	N_l	$\frac{\kappa_{d,l}}{1-\eta_{OU,l}} \frac{\Delta_{step}}{2}$	$1 - \eta_{OU,l} \Delta_{step}$	$\kappa_{h,l}$	$n_{split,l}$	$\tau_{split,l}$
1	1	0	n/a	0	n/a	n/a
2	1	0.89	0.9	0	2	300
3	3	0.67	0.9	0	2	300
4	2	0.67	0.9	0.55	2	33.33
5	1	0.44	0.9	0.40	2	17.65

Table A.1: Choices for parameters in turbulence parametrisation. $\Delta_{step} = 1\text{s}$ was used for integration along with a wind velocity of $1.25\frac{\text{m}}{\text{s}}$ and integration was carried out over $[0\text{s}, 75\text{s}]$. For $\kappa_d = 0$ the OU-process is assumed to be constant 0.

There is no particular reason for these choices other than that the realisations pass an “eye-test” which is to say they look fairly plausible. In the used distributions the kernel centres follow an SDE akin to Eq. (2.17) augmented with a jump-process. The perturbations can be seen as random velocity changes which relates it to the approach taken in [118]. The construction can also be somewhat related to the idea behind Markov-1 models as described in [13]. The primary difference, aside from particle splitting, is that here random perturbations are added directly to the centre $b^0 + \int_0^t \eta[\theta](r)dr = b[\theta](t)$ rather than the velocity $\eta[\theta](t)$. The proposed model for the perturbations should however not be understood as an approach for realistic turbulence modelling but rather it was designed as a noise process that preserves certain statistics as shown in Eqs. (A.39) and (A.40) and has an ergodic property as described in Eq. (4.52). The plume is finite, roughly $120\text{m} \approx 75\text{s} \times 1.25\frac{\text{m}}{\text{s}} + 2h(t)$ in length, because at greater distances the increased width results in absorption strengths that don’t allow any degrees of freedom, and reconstructions would be dominated by the regularisation term.

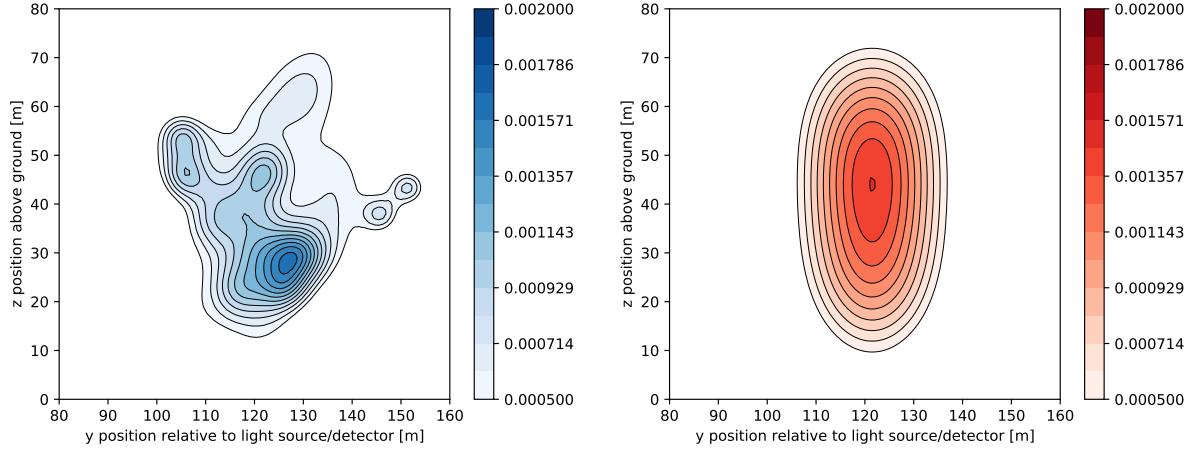


Figure A.2: Cross sections perpendicular to the wind direction of the gas distribution approximately 60m away from the source. Note that the oval shape is due to the upward drift of the gas. The L_1 difference between the smooth and turbulent plume is approximately 50%.

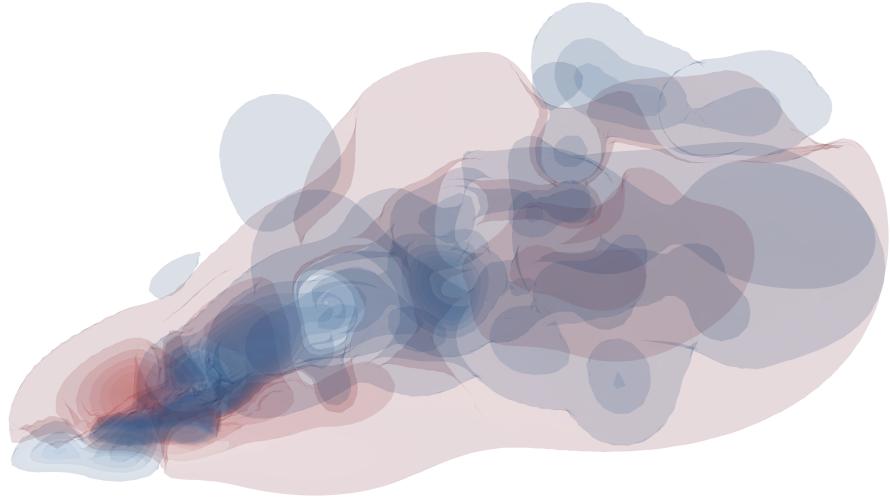


Figure A.3: Qualitative structure of the turbulence: In the blue regions the turbulent gas is larger than its smooth counterpart, red indicates the opposite. The L_1 difference between the smooth and turbulent plume is approximately 50%. Fig. A.2 is the cross-section, roughly down the middle, with values corresponding to those of the differential absorption α in the simulations. In particular, the (differential) loss in intensity for a beam passing through that plume is of the order 10^{-2} .

Bibliography

- [1] I. ABEN, O. P. HASEKAMP, and W. HARTMANN. Uncertainties in the space-based measurements of CO₂ columns due to scattering in the Earth's atmosphere. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 104.3 (2007), pp. 450–459.
- [2] J. B. ABSHIRE et al. Pulsed airborne lidar measurements of atmospheric CO₂ column absorption. In: *Tellus B: Chemical and Physical Meteorology* 62.5 (2010), pp. 770–783.
- [3] S. R. ARRIDGE and J. C. SCHOTLAND. Optical tomography: Forward and inverse problems. In: *Inverse Problems* 25.12, 123010 (2009).
- [4] S. ARTSTEIN et al. Solution of Shannon's problem on the monotonicity of entropy. In: *Journal of the American Mathematical Society* 17.4 (2004), pp. 975–982.
- [5] J. ARVO. Transfer Equations in Global Illumination. In: *Global Illumination, SIGGRAPH '93 Course Notes*. Vol. 42. 1993.
- [6] R. C. ASTER, B. BORCHERS, and C. H. THURBER. *Parameter Estimation and Inverse Problems*. 2nd. Boston: Academic Press, 2013, pp. 93–127.
- [7] M. A. BADRI et al. High performance computation of radiative transfer equation using the finite element method. In: *Journal of Computational Physics* 360 (2018), pp. 74–92.
- [8] G. BAL. Introduction to inverse problems. In: *University of Chicago, Lecture Notes* (2019).

- [9] G. BAL. Inverse transport theory and applications. In: *Inverse Problems* 25.5, 053001 (2009).
- [10] G. BAL and A. JOLLIVET. Time-dependent angularly averaged inverse transport. In: *Inverse Problems* 25.7, 075010 (2009).
- [11] G. BAL, I. LANGMORE, and Y. MARZOUK. Bayesian inverse problems with Monte Carlo forward models. In: *Inverse Problems and Imaging* 7.1 (2013), pp. 81–105.
- [12] M. BENNING and M. BURGER. Modern regularization methods for inverse problems. In: *Acta Numerica* 27 (2018), pp. 1–111.
- [13] P. S. BERLOFF and J. C. MCWILLIAMS. Material Transport in Oceanic Gyres. Part II: Hierarchy of Stochastic Models. In: *Journal of Physical Oceanography* 32.3 (2002), pp. 797–830.
- [14] E. S. F. BERMAN et al. Greenhouse gas analyzer for measurements of carbon dioxide, methane, and water vapor aboard an unmanned aerial vehicle. In: *Sensors and Actuators B: Chemical* 169 (2012), pp. 128–135.
- [15] J. K. BLITZSTEIN and J. HWANG. *Introduction to Probability*. 2nd. Chapman and Hall, 2014.
- [16] A. N. BONDARENKO. Singular structure of the fundamental solution of the transport equation, and inverse problems in particle scattering theory. In: *Doklady Akademii Nauk* 322.2 (1992), pp. 274–276.
- [17] A. N. BONDARENKO. The Structure of the Fundamental Solution of the Time-Independent Transport Equation. In: *Journal of Mathematical Analysis and Applications* 221.2 (1998), pp. 430–451.
- [18] D. C. BROCK, ed. *Understanding Moore's Law: Four Decades of Innovation*. Chemical Heritage Foundation, 2006.
- [19] R. BURDEN and D. FAIRES. *Numerical Analysis*. 9th. Cengage Learning, 2010.
- [20] R. F. CAHALAN et al. THOR—Cloud Thickness from Offbeam Lidar Returns. In: *Journal of Atmospheric and Oceanic Technology* 22.6 (2005), pp. 605–627.

- [21] D. J. CARRUTHERS et al. UK-ADMS: A new approach to modelling dispersion in the Earth's atmospheric boundary layer. In: *Journal of Wind Engineering and Industrial Aerodynamics* 52 (1994), pp. 139–153.
- [22] G. CASELLA and R. BERGER. *Statistical Inference*. 2nd. Duxbury Press, 2001.
- [23] M. CESSENAT. Théoremes de trace L^p pour des espaces de fonctions de la neutronique. In: *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique* 299.16 (1984), pp. 831–834.
- [24] M. CESSENAT. Théoremes de trace pour des espaces de fonctions de la neutronique. In: *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique* 300.3 (1985), pp. 89–92.
- [25] S. CHANDRASEKHAR. *Radiative transfer*. Dover Publications, 1960.
- [26] J. CHENG and B. HOFMANN. Regularization Methods for Ill-Posed Problems. In: *Handbook of Mathematical Methods in Imaging*. Springer New York, 2011, pp. 87–109.
- [27] M. CHOULLI and P. STEFANOV. An inverse boundary value problem for the stationary transport equation. In: *Osaka Journal of Mathematics* 36.1 (1999), pp. 87–104.
- [28] M. CHOULLI and P. STEFANOV. Inverse scattering and inverse boundary value problems for the linear Boltzmann equation. In: *Communications in Partial Differential Equations* 21.5-6 (1996), pp. 763–785.
- [29] A. J. CIMORELLI et al. AERMOD: A Dispersion Model for Industrial Source Applications. Part I: General Model Formulation and Boundary Layer Characterization. In: *Journal of Applied Meteorology* 44.5 (2005), pp. 682–693.
- [30] M. B. COHEN et al. Uniform sampling for matrix approximation. In: *ITCS '15: Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*. 2015, pp. 181–190.
- [31] J. B. CONWAY. *A Course in Functional Analysis*. 2nd. Graduate Texts in Mathematics. Springer New York, 2007.

- [32] D. J. DALEY and D. VERE-JONES. *An Introduction to the Theory of Point Processes – Volume I: Elementary Theory and Methods*. 2nd. Probability and Its Applications. Springer New York, 2003.
- [33] R. DAUTRAY and J.-L. LIONS. *Mathematical analysis and numerical methods for science and technology – Volume 6 Evolution Problems II*. Springer Berlin, Heidelberg, 1993.
- [34] A. C. DAVISON. *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2003.
- [35] A. DEFAZIO, F. BACH, and S. LACOSTE-JULIEN. SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives. In: *Advances in Neural Information Processing Systems*. Vol. 27. 2014.
- [36] E. DÍAZ-FRANCÉS and F. J. RUBIO. On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. In: *Statistical Papers* 54 (2013), pp. 309–323.
- [37] J. T. DOBLER et al. Atmospheric CO₂ column measurements with an airborne intensity-modulated continuous wave 1.57μm fiber laser lidar. In: *Applied Optics* 52.12 (2013), pp. 2874–2892.
- [38] C. L. FEFFERMAN. Existence and smoothness of the Navier-Stokes equation. In: *The millennium prize problems*. Clay Mathematics Institute, 2006, pp. 57–67.
- [39] F. G. FERNALD. Analysis of atmospheric Lidar observations: Some comments. In: *Applied Optics* 23.5 (1984), pp. 652–653.
- [40] F. G. FERNALD, B. M. HERMAN, and J. A. REAGAN. Determination of Aerosol Height Distributions by Lidar. In: *Journal of Applied Meteorology* 11.3 (1972), pp. 482–489.
- [41] T. FOKEN. 50 years of the Monin–Obukhov similarity theory. In: *Boundary-Layer Meteorology* 119 (2006), pp. 431–447.
- [42] U. FRISCH. *Turbulence: The legacy of A. N. Kolmogorov*. Cambridge University Press, 1995.

- [43] M. B. FRISH et al. Progress in reducing size and cost of trace gas analyzers based on tunable diode laser absorption spectroscopy. In: *Advanced Environmental, Chemical, and Biological Sensing Technologies II*. Vol. 5586. 2004, pp. 76–82.
- [44] B. GALPERIN, S. SUKORIANSKY, and P. S. ANDERSON. On the critical Richardson number in stably stratified turbulence. In: *Atmospheric Science Letters* 8.3 (2007), pp. 65–69.
- [45] H. GASK. A proof of Schwartz's kernel theorem. In: *Mathematica Scandinavica* 8.2 (1960), pp. 327–332.
- [46] S. GEMAN, E. BIENENSTOCK, and R. DOURSAT. Neural Networks and the Bias/Variance Dilemma. In: *Neural Computation* 4.1 (1992), pp. 1–58.
- [47] A. GEVA et al. X-ray computed tomography through scatter. In: *Proceedings of The European Conference on Computer Vision (ECCV)*. 2018, pp. 34–50.
- [48] K. GHANNAM and M. EL-FADEL. Emissions characterization and regulatory compliance at an industrial complex: an integrated MM5/CALPUFF approach. In: *Atmospheric Environment* 69 (2013), pp. 156–169.
- [49] I. GKIOULEKAS, A. LEVIN, and T. ZICKLER. An Evaluation of Computational Imaging Techniques for Heterogeneous Inverse Scattering. In: *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, pp. 685–701.
- [50] I. GKIOULEKAS et al. Inverse Volume Rendering with Material Dictionaries. In: *ACM Transactions on Graphics* 32.6, 162 (2013).
- [51] C. S. GOLDENSTEIN et al. SpectraPlot.com: Integrated spectroscopic modeling of atomic and molecular gases. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 200 (2017), pp. 249–257. URL: <https://www.spectraplot.com>.
- [52] L. M. GOLSTON et al. Lightweight mid-infrared methane sensor for unmanned aerial systems. In: *Applied Physics B* 123, 170 (2017).
- [53] I. E. GORDON et al. The HITRAN2020 molecular spectroscopic database. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 277, 107949 (2022).

- [54] L. GREENGARD and J. STRAIN. The Fast Gauss Transform. In: *SIAM Journal on Scientific and Statistical Computing* 12.1 (1991), pp. 79–94.
- [55] E. GREENSMITH, P. BARTLETT, and J. BAXTER. Variance Reduction Techniques for Gradient Estimates in Reinforcement Learning. In: *Advances in Neural Information Processing Systems*. Vol. 14. 2001.
- [56] P. HARRIS et al. Estimation of background gas concentration from differential absorption Lidar measurements. In: *Atmospheric Measurement Techniques* 9.10 (2016), pp. 4879–4890.
- [57] T. HASTIE and R. TIBSHIRANI. Generalized Additive Models. In: *Statistical Science* 1.3 (1986), pp. 297–310.
- [58] T. HASTIE, R. TIBSHIRANI, and J. FRIEDMAN. *The Elements of Statistical Learning*. 2nd. Springer Series in Statistics. Springer New York, 2009.
- [59] A. HATCHER. Notes on introductory point-set topology. In: *Cornell University, Lecture Notes* (2005).
- [60] T. HAWKINS, P. EINARSSON, and P. DEBEVEC. Acquisition of time-varying participating media. In: *ACM Transactions on Graphics* 24.3 (2005), pp. 812–815.
- [61] L. G. HENYET and J. L. GREENSTEIN. Diffuse radiation in the galaxy. In: *The Astrophysical Journal* 93 (1941), pp. 70–83.
- [62] A. E. HOERL and R. W. KENNARD. Ridge Regression: Applications to Nonorthogonal Problems. In: *Technometrics* 12.1 (1970), pp. 69–82.
- [63] A. E. HOERL and R. W. KENNARD. Ridge Regression: Biased Estimation for Nonorthogonal Problems. In: *Technometrics* 12.1 (1970), pp. 55–67.
- [64] P. D. HOFF. *A first course in Bayesian statistical methods*. 1st. Springer Texts in Statistics. Springer New York, 2009.
- [65] R. J. HOGAN. Fast Lidar and Radar Multiple-Scattering Models. Part I: Small-Angle Scattering Using the Photon Variance–Covariance Method. In: *Journal of the Atmospheric Sciences* 65.12 (2008), pp. 3621–3635.

- [66] R. J. HOGAN and A. BATTAGLIA. Fast Lidar and Radar Multiple-Scattering Models. Part II: Wide-Angle Scattering Using the Time-Dependent Two-Stream Approximation. In: *Journal of the Atmospheric Sciences* 65.12 (2008), pp. 3636–3651.
- [67] V. HOLODOVSKY et al. In-situ multi-view multi-scattering stochastic tomography. In: *2016 IEEE International Conference on Computational Photography (ICCP)*. 2016, pp. 1–12.
- [68] H. C. van de HULST. *Light scattering by small particles*. Structure of matter series. Chapman and Hall, 1957.
- [69] D. L. HUTT, L. R. BISSONNETTE, and L. DURAND. Multiple field of view Lidar returns from atmospheric aerosols. In: *Applied Optics* 33.12 (1994), pp. 2338–2348.
- [70] F. INNOCENTI et al. Differential Absorption Lidar (DIAL) Measurements of Landfill Methane Emissions. In: *Remote Sensing* 9.9, 953 (2017).
- [71] E. T. JAYNES. Information Theory and Statistical Mechanics. In: *Physical Review* 106.4 (1957), pp. 620–630.
- [72] R. I. JENNICH. Asymptotic Properties of Non-Linear Least Squares Estimators. In: *The Annals of Mathematical Statistics* 40.2 (1969), pp. 633–643.
- [73] R. JOHNSON and T. ZHANG. Accelerating Stochastic Gradient Descent using Predictive Variance Reduction. In: *Advances in Neural Information Processing Systems*. Vol. 26. 2013.
- [74] J. H. JOSEPH, W. J. WISCOMBE, and J. A. WEINMAN. The Delta-Eddington Approximation for Radiative Flux Transfer. In: *Journal of Atmospheric Sciences* 33.12 (1976), pp. 2452–2459.
- [75] J. T. KAJIYA. The rendering equation. In: *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*. 1986, pp. 143–150.
- [76] O. KALLENBERG. *Foundations of Modern Probability*. 3rd. Springer Cham, 2021.
- [77] M. KINDELAN et al. Application of the RBF meshless method to the solution of the radiative transport equation. In: *Journal of Computational Physics* 229.5 (2010), pp. 1897–1908.

- [78] J. F. C. KINGMAN. *Poisson Processes*. Vol. 3. Oxford Studies in Probability. Oxford University Press, 1993.
- [79] M. KROLL. Concentration inequalities for Poisson point processes with application to adaptive intensity estimation. In: *Stochastics* 95.1 (2023), pp. 54–66.
- [80] S. K. LAM, A. PITROU, and S. SEIBERT. Numba: a LLVM-based Python JIT compiler. In: *LLVM '15: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. Association for Computing Machinery, 2015, 7.
- [81] I. LANGMORE. The stationary transport equation with angularly averaged measurements. In: *Inverse Problems* 24.1, 015024 (2008).
- [82] I. LANGMORE, A. B. DAVIS, and G. BAL. Multipixel Retrieval of Structural and Optical Parameters in a 2-D Scene With a Path-Recycling Monte Carlo Forward Model and a New Bayesian Inference Engine. In: *IEEE Transactions on Geoscience and Remote Sensing* 51.5 (2013), pp. 2903–2919.
- [83] I. LANGMORE et al. Parametric 3D atmospheric reconstruction in highly variable terrain with recycled Monte Carlo paths and an adapted Bayesian inference engine. In: *Radiation Processes in the Atmosphere and Ocean (IRS2012)*. Vol. 1531. American Institute of Physics Conference Series. 2013, pp. 79–82.
- [84] A. LASOTA and M. C. MACKEY. *Probabilistic properties of deterministic systems*. Cambridge University Press, 1985.
- [85] Á. LEELŐSSY et al. Dispersion modeling of air pollutants in the atmosphere: A review. In: *Open Geosciences* 6.3 (2014), pp. 257–278.
- [86] S. LEPICARD, R. HELING, and V. MADERICH. POSEIDON/RODOS models for radiological assessment of marine environment after accidental releases: Application to coastal areas of the Baltic, Black and North Seas. In: *Journal of Environmental Radioactivity* 72.1-2 (2004), pp. 153–161.
- [87] A. LEVIS, Y. Y. SCHECHNER, and A. B. DAVIS. Multiple-Scattering Microphysics Tomography. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 5797–5806.

- [88] A. LEVIS et al. Airborne Three-Dimensional Cloud Tomography. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 3379–3387.
- [89] J. I. LEVY et al. Using CALPUFF to evaluate the impacts of power plant emissions in Illinois: Model sensitivity and implications. In: *Atmospheric Environment* 36.6 (2002), pp. 1063–1075.
- [90] A. LIEMERT and A. KIENLE. Exact and efficient solution of the radiative transport equation for the semi-infinite medium. In: *Scientific Reports* 3, 2018 (2013).
- [91] A. LIEMERT, D. REITZLE, and A. KIENLE. Analytical solutions of the radiative transport equation for turbid and fluorescent layered media. In: *Scientific Reports* 7, 3819 (2017).
- [92] T. M. LIGGETT. *Continuous time Markov processes: An introduction*. Vol. 113. Graduate Studies in Mathematics. American Mathematical Society, 2010.
- [93] J. LOVERIDGE et al. Retrieving 3D distributions of atmospheric particles using Atmospheric Tomography with 3D Radiative Transfer – Part 1: Model description and Jacobian calculation. In: *Atmospheric Measurement Techniques* 16.7 (2023), pp. 1803–1847.
- [94] J. LOVERIDGE et al. Retrieving 3D distributions of atmospheric particles using Atmospheric Tomography with 3D Radiative Transfer – Part 2: Local optimization. In: *Atmospheric Measurement Techniques* 16.16 (2023), pp. 3931–3957.
- [95] M. C. MACKEY. The dynamic origin of increasing entropy. In: *Reviews of Modern Physics* 61.4 (1989), pp. 981–1015.
- [96] M. W. MAHONEY. Randomized algorithms for matrices and data. In: *Foundations and Trends in Machine Learning* 3.2 (2011), pp. 123–224.
- [97] J. MAO and S. R. KAWA. Sensitivity studies for space-based measurement of atmospheric total column carbon dioxide by reflected sunlight. In: *Applied Optics* 43.4 (2004), pp. 914–927.
- [98] P.-G. MARTINSSON and J. A. TROPP. Randomized numerical linear algebra: Foundations and algorithms. In: *Acta Numerica* 29 (2020), pp. 403–572.

- [99] S. R. McDOWALL, A. TAMASAN, and P. STEFANOV. Stability of the gauge equivalent classes in stationary inverse transport. In: *arXiv preprint arXiv:0905.0626* (2009).
- [100] R. M. MEASURES. *Laser remote sensing: Fundamentals and Applications*. Wiley-Interscience, New York, 1984.
- [101] S. MOHAMED et al. Monte Carlo gradient estimation in machine learning. In: *Journal of Machine Learning Research* 21, 132 (2020), pp. 1–62.
- [102] N. MORGAN and H. BOURLARD. Generalization and Parameter Estimation in Feedforward Nets: Some Experiments. In: *Advances in Neural Information Processing Systems*. Vol. 2. 1989.
- [103] Y. MUKAIGAWA, Y. YAGI, and R. RASKAR. Analysis of light transport in scattering media. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 153–160.
- [104] S. A. MURPHY and A. W. VAN DER VAART. On Profile Likelihood. In: *Journal of the American Statistical Association* 95.450 (2000), pp. 449–465.
- [105] S. NA, M. DEREZINSKI, and M. MAHONEY. Hessian averaging in stochastic Newton methods achieves superlinear convergence. In: *Mathematical Programming* 201.1-2 (2022), pp. 473–520.
- [106] S. G. NARASIMHAN et al. Acquiring Scattering Properties of Participating Media by Dilution. In: *ACM Transactions on Graphics* 25.3 (2006), pp. 1003–1012.
- [107] W. K. NEWHEY and D. McFADDEN. Chapter 36: Large sample estimation and hypothesis testing. In: *Handbook of Econometrics*. Vol. 4. Elsevier, 1994, pp. 2111–2245.
- [108] J. NICKOLLS et al. Scalable Parallel Programming with CUDA: Is CUDA the Parallel Programming Model That Application Developers Have Been Waiting For? In: *Queue* 6.2 (2008), pp. 40–53.
- [109] H. C. ÖTTINGER. *Stochastic processes, polymer dynamics, and fluid mechanics*. 1st. Springer Berlin, Heidelberg, 1996.

- [110] L. PARDO. *Statistical Inference Based on Divergence Measures*. 1st. Chapman and Hall, 2005.
- [111] J. O. PAUMIER, D. J. BURNS, and S. G. PERRY. CTDMPLUS: A Dispersion Model for Sources near Complex Topography. Part II: Performance Characteristics. In: *Journal of Applied Meteorology and Climatology* 31.7 (1992), pp. 646–660.
- [112] A. PEDIREDLA, A. VEERARAGHAVAN, and I. GKIOULEKAS. Ellipsoidal path connections for time-gated rendering. In: *ACM Transactions on Graphics* 38.4, 38 (2019).
- [113] L. M. PERLISKI and S. SOLOMON. On the evaluation of air mass factors for atmospheric near-ultraviolet and visible absorption spectroscopy. In: *Journal of Geophysical Research: Atmospheres* 98.D6 (1993), pp. 10363–10374.
- [114] S. G. PERRY. CTDMPLUS: A Dispersion Model for Sources near Complex Topography. Part I: Technical Formulations. In: *Journal of Applied Meteorology and Climatology* 31.7 (1992), pp. 633–645.
- [115] S. G. PERRY et al. AERMOD: A Dispersion Model for Industrial Source Applications. Part II: Model Performance against 17 Field Study Databases. In: *Journal of Applied Meteorology* 44.5 (2005), pp. 694–708.
- [116] U. PLATT and J. STUTZ. *Differential Optical Absorption Spectroscopy*. 1st. Physics of Earth and Space Environments. Springer Berlin, Heidelberg, 2008.
- [117] S. POWELL, B. T. COX, and S. R. ARRIDGE. A pseudospectral method for solution of the radiative transport equation. In: *Journal of Computational Physics* 384 (2019), pp. 376–382.
- [118] J. POZORSKI and J.-P. MINIER. On the Lagrangian turbulent dispersion models based on the Langevin equation. In: *International Journal of Multiphase Flow* 24.6 (1998), pp. 913–945.
- [119] K. PRUEKSAKORN et al. Review of air dispersion modelling approaches to assess the risk of wind-borne spread of foot-and-mouth disease virus. In: *Journal of Environmental Protection* 3.9A (2012), pp. 1260–1267.

- [120] G. RASKUTTI, M. J. WAINWRIGHT, and B. YU. Early Stopping and Non-parametric Regression: An Optimal Data-dependent Stopping Rule. In: *Journal of Machine Learning Research* 15, 11 (2014), pp. 335–366.
- [121] W. G. REES. *Physical Principles of Remote Sensing*. 3rd. Cambridge University Press, 2012.
- [122] T. F. REFAAT et al. Performance evaluation of a 1.6- μm methane DIAL system from ground, aircraft and UAV platforms. In: *Optics Express* 21.25 (2013), pp. 30415–30432.
- [123] P. REYNAUD-BOURET. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. In: *Probability Theory and Related Fields* 126.1 (2003), pp. 103–153.
- [124] R. ROBINSON et al. Infrared differential absorption Lidar (DIAL) measurements of hydrocarbon emissions. In: *Journal of Environmental Monitoring* 13.8 (2011), pp. 2213–2220.
- [125] W. RUDIN. *Functional analysis*. 2nd. International series in pure and applied mathematics. McGraw-Hill, 1991.
- [126] M. SCHMIDT, N. LE ROUX, and F. BACH. Minimizing finite sums with the stochastic average gradient. In: *Mathematical Programming* 162.1-2 (2017), pp. 83–112.
- [127] N. SCHREUDER. Bounding the Expectation of the Supremum of Empirical Processes Indexed by Hölder Classes. In: *Mathematical Methods of Statistics* 29.1 (2020), pp. 76–86.
- [128] J. SECREST, J. CONROY, and H. MILLER. Applying the Maximum Entropy Technique to the Gaussian Dispersion Plume Model. In: *arXiv preprint arXiv:2010.11864* (2020).
- [129] J. SEINFELD and S. PANDIS. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. 3rd. John Wiley & Sons, Incorporated, 2016.

- [130] G. D. SPIERS et al. Atmospheric CO₂ measurements with a 2μm airborne laser absorption spectrometer employing coherent detection. In: *Applied Optics* 50.14 (2011), pp. 2098–2111.
- [131] P. STEFANOV and A. TAMASAN. Uniqueness and non-uniqueness in inverse radiative transfer. In: *Proceedings of the American Mathematical Society* 137.7 (2009), pp. 2335–2344.
- [132] J. M. STOCKIE. The Mathematics of Atmospheric Dispersion Modeling. In: *SIAM Review* 53.2 (2011), pp. 349–372.
- [133] O. N. STRAND. Theory and Methods Related to the Singular-Function Expansion and Landweber’s Iteration for Integral Equations of the First Kind. In: *SIAM Journal on Numerical Analysis* 11.4 (1974), pp. 798–825.
- [134] A. M. STUART. Inverse problems: A Bayesian perspective. In: *Acta Numerica* 19 (2010), pp. 451–559.
- [135] R. B. STILL. *An Introduction to Boundary Layer Meteorology*. 1st. Atmospheric and Oceanographic Sciences Library. Springer Dordrecht, 1988.
- [136] S. SVANBERG. *Atomic and molecular spectroscopy: Basic aspects and practical applications*. 6th. Graduate Texts in Physics. Springer Cham, 2022.
- [137] M. TALAGRAND. A new look at independence. In: *The Annals of Probability* 24.1 (1996), pp. 1–34.
- [138] G. I. TAYLOR. The spectrum of turbulence. In: *Proceedings of the Royal Society of London. Series A – Mathematical and Physical Sciences* 164.919 (1938), pp. 476–490.
- [139] J. A. TROPP. An introduction to matrix concentration inequalities. In: *Foundations and Trends in Machine Learning* 8.1-2 (2015), pp. 1–230.
- [140] J. A. TROPP. User-Friendly Tail Bounds for Sums of Random Matrices. In: *Foundations of Computational Mathematics* 12.4 (2012), pp. 389–434.
- [141] A. B. TSYBAKOV. *Introduction to Nonparametric Estimation*. 1st. Springer Series in Statistics. Springer New York, 2009.

- [142] D. B. TURNER. The long lifetime of the dispersion methods of Pasquill in US regulatory air modeling. In: *Journal of Applied Meteorology and Climatology* 36.8 (1997), pp. 1016–1020.
- [143] P. VIRTANEN et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. In: *Nature Methods* 17 (2020), pp. 261–272.
- [144] R. T. WAINER et al. Scanning, standoff TDLAS leak imaging and quantification. In: *Next-Generation Spectroscopic Technologies X*. Vol. 10210. International Society for Optics and Photonics. SPIE, 2017, 1021006.
- [145] X. WANG. Volumes of Generalized Unit Balls. In: *Mathematics Magazine* 78.5 (2005), pp. 390–395.
- [146] C. WEITKAMP. *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*. Springer Series in Optical Sciences. Springer New York, 2005.
- [147] E. T. WHITTAKER and G. N. WATSON. *A Course of Modern Analysis*. 4th. Cambridge Mathematical Library. Cambridge University Press, 1996.
- [148] W. J. WISCOMBE. The Delta-M Method: Rapid Yet Accurate Radiative Flux Calculations for Strongly Asymmetric Phase Functions. In: *Journal of Atmospheric Sciences* 34.9 (1977), pp. 1408–1422.
- [149] D. P. WOODRUFF. Sketching as a tool for numerical linear algebra. In: *Foundations and Trends in Theoretical Computer Science* 10.1–2 (2014), pp. 1–157.
- [150] C. YANG et al. Improved Fast Gauss Transform and efficient kernel density estimation. In: *Proceedings ninth IEEE international conference on computer vision*. 2003, pp. 664–671.
- [151] Y. YAO, L. ROSASCO, and A. CAPONNETTO. On Early Stopping in Gradient Descent Learning. In: *Constructive Approximation* 26.2 (2007), pp. 289–315.
- [152] Y. ZHOU et al. Estimating population exposure to power plant emissions using CALPUFF: A case study in Beijing, China. In: *Atmospheric Environment* 37.6 (2003), pp. 815–826.