DLR Deutsches Zentrum
für Luft- und Raumfahrt
German Aerospace Center

University of Stuttgart
Germany

MASTER'S THESIS

Coherent wind-lidar for optical turbulence measurement

Kohärentes Wind-Lidar zur Messung optischer Turbulenzstärken

Presented by
Hugo Uittenbosch
Faculty 6: Aerospace Engineering and Geodesy
Student ID: 3137556

Reviewer
Prof. Dr. rer. nat. Thomas Dekorsy
Institute of Aerospace Thermodynamics (ITLR)
University of Stuttgart
and
Institute of Technical Physics
German Aerospace Center (DLR)

Supervisor
Dr. rer. nat. Oliver Kliebisch
Institute of Technical Physics
German Aerospace Center (DLR)

December 31st, 2021

Eidesstattliche Erklärung

Hiermit erkläre ich mich damit einverstanden, dass meine Masterarbeit zum Thema

Coherent wind-lidar for optical turbulence measurement

in der Institutsbibliothek des Instituts für Thermodynamik der Luft- und Raumfahrt öffentlich zugänglich aufbewahrt und die Arbeit auf der Institutswebseite sowie im Online-Katalog der Universitätsbibliothek erfasst wird. Letzteres bedeutet eine dauerhafte, weltweite Sichtbarkeit der bibliographischen Daten der Arbeit (Titel, Autor, Erscheinungsjahr, etc.). Nach Abschluss der Arbeit werde ich zu diesem Zweck meinem Betreuer neben dem Prüfexemplar eine weitere gedruckte sowie eine digitale Fassung übergeben. Der Universität Stuttgart übertrage ich das Eigentum an diesen zusätzlichen Fassungen und räume dem Institut für Thermodynamik der Luft- und Raumfahrt an dieser Arbeit und an den im Rahmen dieser Arbeit von mir erzeugten Arbeitsergebnissen ein kostenloses, zeitlich und örtlich unbeschränktes, einfaches Nutzungsrecht für Zwecke der Forschung und der Lehre ein. Falls in Zusammenhang mit der Arbeit Nutzungsrechtsvereinbarungen des Instituts mit Dritten bestehen, gelten diese Vereinbarungen auch für die im Rahmen dieser Arbeit entstandenen Arbeitsergebnisse.

. . . .
Ort, Datum, Unterschrift

Hiermit versichere ich, dass ich diese Masterarbeit selbstständig mit Unterstützung des Betreuers angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit oder wesentliche Bestandteile davon sind weder an dieser noch an einer anderen Bildungseinrichtung bereits zur Erlangung eines Abschlusses eingereicht worden. Ich erkläre weiterhin, bei der Erstellung der Arbeit die einschlägigen Bestimmungen zum Urheberschutz fremder Beiträge entsprechend den Regeln guter wissenschaftlicher Praxis eingehalten zu haben. Soweit meine Arbeit fremde Beiträge (z.B. Bilder, Zeichnungen, Textpassagen, etc.) enthält, habe ich diese Beiträge als solche gekennzeichnet (Zitat, Quellangabe) und eventuell erforderlich gewordene Zustimmungen der Urheber zur Nutzung dieser Beiträge in meiner Arbeit eingeholt. Mir ist bekannt, dass ich im Falle einer schuldhaften Verletzung dieser Pflichten die daraus entstehenden Konsequenzen zu tragen habe.

. . . .
Ort, Datum, Unterschrift

Acknowledgements

I would like to thank **Prof. Dr. Thomas Dekorsy** for the opportunity to undertake this thesis and for agreeing to be the reviewer for this work. It has given me a valuable insight into the world of lidar technology and the practical applications of lasers in aerospace.

I would like to thank **Dr. Johann Thurn** for his support throughout the course of this work. He provided me with valuable advice for the alignment of optical systems and reviewed multiple chapters of this thesis.

The same is true for **Peter Mahnke**. His assistance in setting up the electrical components of the system made it possible to achieve an operational state within this short period of time.

I would like to thank **Luis Mandl** for his reviews and constructive criticism regarding neural networks and deep learning. The discussions helped to improve the respective parts of this thesis. I would also like to thank my parents for spell checking this document.

Finally, I want to thank my supervisor, **Dr. Oliver Kliebisch**, for guiding this thesis towards success. He was always eager to help and teach. His expertise in many fields has improved my understanding of programming, data evaluation and experimental operation in general. Without his help, my first steps into the world of optics would not have been possible. His reviews helped to improve the overall quality of this work.

Licenses:

I would also like to thank the Deutscher Wetterdienst for publicly providing the radiosonde data used in Ch. 6.

Abstract

Coherent wind-lidar for optical turbulence measurement

Time-accurate information on the turbulent parameters of the atmospheric boundary layer is relevant for a variety of technical applications, i.e., the turbulent kinetic energy dissipation rate ε or the turbulent index of refraction C_n^2. For instance, these include the real-time detection of clear-air turbulence for turbulence avoidance or the atmospheric propagation of laser beams.

In the context of this work, a lidar system was built to provide this data in near real-time based on preprocessing hardware developed prior to this thesis. Further, the necessary control software and optimized evaluation algorithms were developed. An essential part was the completion of an optical system for the transmission and reception of laser pulses and backscattered light, respectively, in the near infrared spectrum. Mono-, as well as bistatic designs were tested and alignment processes developed. The robustness and operability of the monostatic design proved to be superior. A coherent detection method was used due to the significant downshift in bandwidth that can be achieved.

Field tests were performed to demonstrate the capabilities of the lidar system. Although only static measurements were conducted, that is the wind speed in the direction of the line-of-sight was recorded, interesting measurement data was obtained. An ε-estimator based on the analysis of the velocity density spectrum was used to determine height-resolved values for the turbulent kinetic energy dissipation rate. From this, values for C_n^2 could be derived. An improved and automated algorithm will be presented for the estimation of ε, which can operate on velocity data even from short time intervals. For this purpose, a method was developed that can ignore the influence of statistically dependent effects and achieve a minimization of the relative error for ε. Another novel approach, the use of a convolutional neural network for the velocity estimation from lidar spectrum data, will be investigated. In this context, the distributions of the parameters for a synthetic data generation model are examined in more detail.

Selected field campaigns will be presented and optical turbulence parameters compared to the Hufnagel-Valley model. Qualitatively, it will be shown on this real-world data that the ε-estimator indeed finds optimized solutions in the velocity density spectrum.

Keywords: wind-lidar, coherent detection, velocity estimation, turbulent kinetic energy dissipation rate, turbulent index of refraction, optical turbulence, real-time processing

Kurzzusammenfassung

Kohärentes Wind-Lidar zur Messung optischer Turbulenzstärken

Zeitnahe Informationen über die turbulenten Parameter der atmosphärischen Grenzschicht sind für eine Vielzahl von technischen Anwendungen relevant, z.B. die turbulente kinetische Energiedissipationsrate ε oder der turbulente Brechungsindex C_n^2. Dazu gehören zum Beispiel die Echtzeiterkennung von Klarluftturbulenzen zur Turbulenzvermeidung oder die atmosphärische Propagation von Laserstrahlung.

Im Rahmen dieser Arbeit wurde ein Lidar-System aufgebaut, das diese Daten nahezu in Echtzeit auf der Grundlage einer in Vorfeld entwickelten Prozessierungshardware liefert. Weiterhin wurden die notwendige Steuerungssoftware und optimierte Auswertealgorithmen entwickelt. Ein wesentlicher Bestandteil war die Fertigstellung eines optischen Systems zum Senden und Empfangen von Laserpulsen bzw. Rückstreulicht im nahen Infrarotspektrum. Es wurden sowohl mono-, als auch bistatische Designs getestet und Justageverfahren umgesetzt. Die Robustheit und Funktionsfähigkeit des ersteren Designs erwies sich als überlegen. Aufgrund der signifikanten Bandbreitenminderung wurde eine kohärente Detektionsmethode gewählt.

Um die Leistungsfähigkeit des Lidar-Systems zu demonstrieren, wurden Feldtests durchgeführt. Obwohl nur statische Messungen durchgeführt werden konnten, das heißt die Windgeschwindigkeit in Richtung der Sichtlinie aufgezeichnet wurde, konnten interessante Messdaten gewonnen werden. Mit einem ε-Schätzer, der auf der Analyse des Geschwindigkeitsdichtespektrums basiert, wurden höhenaufgelöste Werte für die turbulente kinetische Energiedissipationsrate ermittelt. Daraus konnten Werte für C_n^2 abgeleitet werden. Für die Schätzung von ε wird zudem ein verbesserter und automatisierter Algorithmus vorgestellt, der auch mit Geschwindigkeitsdaten aus kurzen Zeitintervallen arbeiten kann. Dazu wurde eine Methode entwickelt, die den Einfluss statistisch abhängiger Effekte ignorieren kann und eine Minimierung des relativen Fehlers für ε erreicht. Ein weiterer neuartiger Ansatz, die Verwendung eines neuronalen Netzes zur Geschwindigkeitsschätzung aus Lidar-Spektrumsdaten, wird untersucht. In diesem Zusammenhang werden die Verteilungen der Parameter für ein synthetisches Datengenerierungsmodell näher betrachtet.

Ausgewählte Feldkampagnen werden vorgestellt und optische Turbulenzparameter mit dem Hufnagel-Valley-Modell verglichen. Qualitativ wird an diesen realen Daten gezeigt, dass der ε-Schätzer tatsächlich optimierte Lösungen im Geschwindigkeitsdichtespektrum finden kann.

Contents

List of Figures

List of Tables

Symbols and Acronyms

Symbols

Latin Symbols

f	Hz	Frequency (electrical)
c	$\mathrm{m\,s^{-1}}$	Speed of light
Δf_D	Hz	Doppler frequency shift
P_0	W	Average transmitted pulse power
u, v, w	$\mathrm{m\,s^{-1}}$	Spatial velocity components
v_{LOS}	$\mathrm{m\,s^{-1}}$	Line-of-sight velocity
$\overline{U}$	$\mathrm{m\,s^{-1}}$	Average interval velocity
E	$\mathrm{m^3\,s^{-2}}$	Spectral energy
$\hat{S}$	$\mathrm{m^2\,s^{-1}}$	Spectral density
k	$\mathrm{m^2\,s^{-2}}$	Turbulent kinetic energy
B	-	Mean normalized error
E	-	Relative random error
A	-	Total estimation error
g	$\mathrm{m\,s^{-2}}$	Gravitational acceleration
c_p	$\mathrm{J\,kg^{-1}\,K^{-1}}$	Specific heat capacity
R	$\mathrm{J\,kg^{-1}\,K^{-1}}$	Gas constant
FQ	-	Fluctuating quantity
N	Hz	Brunt–Väisälä frequency
T	K	Temperature
p	Pa	Pressure
n	-	Index of refraction
C_n^2	$\mathrm{m^{-2/3}}$	Turbulent index of refraction

Greek Symbols

λ	m	Wavelength
τ	s	Pulse duration
β	$\mathrm{m^{-1}}$	Backscatter coefficient
σ	$\mathrm{m^{-1}}$	Extinction coefficient
ρ	$\mathrm{kg\,m^{-3}}$	Density
γ	$\mathrm{^{\circ}C\,m^{-1}}$	Dry adiabatic lapse rate
ν	Hz	Frequency (optical)
ε	$\mathrm{m^2\,s^{-3}}$	Turbulent kinetic energy dissipation rate
$\hat{\varepsilon}$	$\mathrm{m^2\,s^{-3}}$	Turbulent kinetic energy dissipation rate estimate
κ	-	Heat capacity ratio

Acronyms

H_2O water. 63

H_2SO_4 sulfuric acid. 63, 65

O_2 molecular oxygen. 63

SO_2 sulfur dioxide. 63–65

SO_3 sulfur trioxide. 63

ADAM adaptive moment estimation. 52

ADC analog-to-digital converter. 16

AMSL above mean sea level. 57, 67, 69

AOM acousto-optic modulator. 13

APD avalanche photodiode. 14–16, 25, 58, 64

AWG arbitrary waveform generator. 13

CDF cumulative distribution function. 43

CDWL coherent detection wind lidar. 4, 7, 32

CNN convolutional neural network. 47, 50, 52–55, 57

COG center-of-gravity. 28, 30, 39, 43–46, 53

CW continuous wave. 4, 14, 21, 22

DALR dry adiabatic lapse rate. 11

DDL direct detection lidar. 3, 5

DFT discrete Fourier transform. 9, 28

DLR German Aerospace Center. 57, 69, 71

DMA direct memory access. 16

DOF degree of freedom. 19–21, 47

DWD Deutscher Wetterdienst. 57–59, 64, 67, 68, 72

ECMWF European Centre for Medium-Range Weather Forecasts. 57, 58, 63

EDFA erbium-doped fiber amplifier. 13, 18

EDFL erbium-doped fiber laser. 13

FFT fast Fourier transform. 16, 42

FOV field of view. 18, 19, 48, 60

FPGA field programmable gate array. 13, 14, 16

GAN generative adversarial network. 50

GUI graphical user interface. 23, 24

HV Hufnagel-Valley. 11, 57–60, 65, 68, 69

I2C inter-integrated circuit. 23

IMU inertial measurement unit. 23, 57

InGaAs indium-gallium-arsenid. 14, 20, 22

ISA International Standard Atmosphere. 11, 38, 40, 58, 59, 64, 67

ISR interrupt service routine. 23

L-BFGS limited memory Broyden–Fletcher–Goldfarb–Shanno. 52

lidar light detection and ranging. 3–6, 13, 15, 16, 23, 25–32, 39, 41, 42, 45–47, 55, 57–60, 63, 65, 67–69, 71, 72

LMS local maxima scalogram. 44–47

LOS line of sight. 3, 6, 15, 23, 26, 30–32, 39, 41, 44, 48, 50, 53, 55, 57, 59, 61, 62, 66–70

LPF low-pass filter. 15, 64

MCU microcontroller unit. 23

NA numerical aperture. 18

NaN not a number. 42, 45, 58

NOAA National Oceanic and Atmospheric Administration. 41

PBL planetary boundary layer. 7, 8, 11, 25, 26, 28–30, 32, 42, 44, 46, 49, 50, 52, 55, 58–60, 63, 65, 68

PFA pulsed fiber amplifier. 14, 23, 24

PFL pulsed fiber laser. 14, 23, 24

PIN positive-intrinsic-negative. 14, 15, 30, 67

PLL phase-locked loop. 16

PM polarization-maintaining. 13, 14

radar radio detection and ranging. 46

RANS Reynolds-averaged Navier-Stokes. 9

ReLU rectified linear unit. 53

REPL read-eval-print loop. 25

SNR signal-to-noise ratio. 15, 25, 27, 41, 42, 52, 62, 66, 68, 70

TDR turbulent kinetic energy dissipation rate. 9, 10, 32, 34–40, 58, 59, 61, 65, 68, 69

TKE turbulent kinetic energy. 9, 32, 38, 60

UART universal asynchronous receive-transmit. 23

UML Unified Modeling Language. 24

USB Universal Serial Bus. 23

UV ultraviolet. 58, 63

1 Introduction

Fluid turbulence remains to this day one of the most complex phenomena in the field of mathematics, physics and engineering. The following quote is attributed to Werner Heisenberg:

> When I meet God, I am going to ask him two questions: why relativity? And why turbulence? I really believe he will have an answer for the first.

An analytical description of the emergent and chaotic behavior that is characteristic for turbulent flows eludes itself even up to the present day, necessitating complex numerical simulations and semi-empirical equations in order to predict its formation and properties.

To elaborate upon the relevance of turbulence for the aviation sector, the following issues are referred to: the dimensioning of safety margins in the structural design of aircraft to accommodate for aerodynamic forces caused by turbulence, the danger posed to passenger aircraft in cruise flight by sudden changes in the flow field due to clear-air turbulence [1] and further, changes in the aerodynamic behavior of an aircraft under turbulent free-stream conditions [2]. Novel real-time prediction methods could provide remedy in this case and help to increase safety for passenger flight in the future.

Turbulent effects also play a role in the atmospheric propagation of radiation. Density fluctuations caused by turbulence lead to a local change in the refractive index of air [3, 4]. This spatially inhomogeneous distribution leads to an aberration for transiting radiation, which is detrimental for both Earth-based space observation [5] and atmospheric propagation of laser beams [6]. Hence, to counteract this effect, the measurement of atmospheric turbulence parameters along the propagation direction of the radiation is necessary.

In addition, another affected field of research is the analysis of flow effects in the planetary boundary layer. A good understanding of this relatively thin layer of the atmosphere is crucial for the improvement of weather and climate models, since it is in this near ground layer where the bulk of mixing effects occur. Turbulent eddies are an important transport mechanism between the vertical layers of the stratosphere [7].

Thus, due to the above circumstances, a determination of turbulent atmospheric parameters is of great interest. One possible approach to this is the spatially resolved measurement of wind speeds using lidar technology. The history of this method is almost as old as that of the laser itself. Since the 1960s, research has been carried out on the measurement of atmospheric parameters using laser beams [8].

The main mechanism hereby is the measurement of the velocity-dependent Doppler shift imparted on the frequency of backscattered light from aerosols in Earth's lower atmosphere. Spatial resolution can be achieved by considering the time of flight. Further, data on the occurrence and type of aerosols can be obtained using variants of this method [9]. In addition, coherent methods allow a significant downshift in the signal bandwidth, making possible a more cost-efficient implementation of this technology.

The aim of this work is to integrate a coherent wind lidar system based on the preprocessing hardware and software developed by Kliebisch et al. (2020) [10]. Although systems already exist that pursue comparable objectives [11], this lidar setup is intended to provide quantitative data for the turbulent kinetic energy dissipation rate ε and the turbulent index of refraction C_n^2 in near real-time and in the form of a highly integrated, compact system that can be easily relocated. For this reason, an optical setup must be implemented, the software and algorithms for data evaluation have to be developed, and field campaigns to obtain said data need to be conducted.

The Navier-Stokes equations, derived from the conservation of energy, momentum and mass for an infinitesimal small fluid volume, are the governing equations for the atmospheric environment under consideration and fully describe the resulting laminar and/or turbulent flow given a complete set of initial conditions. However, solving these equations can quickly become an immense computational challenge. Therefore, the approach was introduced to decompose the flow field variables into a mean and fluctuating part. Representative scalar variables are defined to replace the fluctuating quantities and additional equations are modeled to describe their behavior. These relations are commonly known as the Reynolds-averaged Navier-Stokes equations. The scalar turbulent kinetic energy k hereby is used to represent the high-frequency fluctuations of the velocity field for each point in space [12].

Measuring this quantity would require knowledge of all three spatial velocity components. The temporal derivative of k, named the turbulent kinetic energy dissipation rate ε, however, can be obtained by analyzing the velocity density spectrum of any component of the velocity vector at a given height, based on the Fourier transform of the according velocity time interval. This is possible due to the, in theory, characteristic course of the spectrum, referred to as the Kolmogorov law [13, 14].

Thus, the calculation of certain turbulent quantities can also be performed with a static lidar. This approach, together with a correlation between ε and C_n^2, will be used in the context of this work. One of the main goals will be to develop an algorithm to minimize the estimation error for ε once the underlying time interval becomes small or includes significant statistically dependent effects.

This thesis is divided into the following chapters: First, the literature relevant to this project will be reviewed and the utilized coherent lidar concept will be explained. Furthermore, the characteristics of the planetary boundary layer will be summarized. The spectrum analysis method mentioned above will be explained in detail and a simple model for the prediction of C_n^2 over height will be presented. Following this chapter, the experimental setup will be elaborated. Here, the preprocessing hardware designed by Kliebisch et al. will be briefly explained in the context of the overall system design. Two different optical setups will be considered, as well as the methods used to align them.

A significant part of the work is devoted to explain the developed algorithms, since they are essential for the near-time calculation of the turbulent quantities. The methods that were used to denoise and drift-correct the data, in addition to the velocity estimation algorithms, will be presented. A separate chapter is provided for the latter, as three different estimators are introduced, including the novel use of a neural network for the context-based velocity estimation from lidar spectrum data.

The second last chapter features the exemplary evaluation of data that was obtained from field campaigns. These were carried out throughout the course of this work once the system integration had been completed. The performance of the ε-estimator will be evaluated and the quantitative data obtained for C_n^2 will be compared with the predictions made by a C_n^2-model. The final chapter summarizes the contents of this work and gives an overview of the achieved results. At the end, possible additions and improvements of the system are discussed.

2 Literature Review

The idea for the use of laser light in detection and ranging dates back to the early 1960s. Since then, the development of light detection and ranging (lidar) has enabled a wide range of technical applications, including: range finders, laser altimeters, laser designators, coherent and incoherent wind measurement and aerosol detection [8].

Within the scope of this work, the fundamentals of coherent lidar will be briefly explained. In addition, the fundamentals of the planetary boundary layer are discussed and some important turbulence parameters defined from literature.

2.1 Principles of wind lidar

The aim of this lidar method is to determine wind speeds using optical measuring techniques. The basis for this is the backscattering of light by molecules and aerosols within the atmosphere. For low mass aerosols it can be assumed that their velocity is equal to that of the flow field. When light scatters from a moving target, its frequency is altered according to the Doppler effect. This slight change in frequency of the scattered light can be measured to derive information about the wind speed component in the direction of the line of sight (LOS).

Several concepts exist for the arrangement of the optics, as well as the evaluation of the Doppler shift. The design of the optics can be either a mono-, or bistatic system. In the first case, a single optical system is used for the transmission and reception of the light. The separation of these channels can be achieved, for instance, by means of an optical circulator. In the second case, dedicated optics are used for transmitting and receiving the light signal. This can improve sensitivity at the cost of a higher overall system complexity [9, p. 483].

In order to be able to record the wind speed over height with spatial resolution, the laser light is emitted as pulses at clearly defined intervals. The maximum measuring range, without aliasing, is defined by the speed of light and the period between two light pulses. As a result of the time difference between sending out a pulse and receiving scattered photons, a distance relative to the observer can be determined. Information correlation becomes nontrivial after the second pulse has been emitted [9, p. 485]

2.1.1 Direct detection lidar

If the entire received signal is spectrally analyzed directly, the method is referred to as direct detection lidar (DDL). With the aid of an optical spectrometer the Doppler shift can be determined from both aerosol and molecular scattering. The operation of a DDL system entails the evaluation of a very broad frequency spectrum. This increases the requirements for the used hardware. The system is further complicated by the reception of undesirable photons from secondary sources, i.e., those emitted by the sun. This can reduce the dynamic range of the receiver and increase the noise level [9, p. 505]

2.1.2 Coherent wind lidar

In practice, the frequency change that scattered light undergoes due to the Doppler shift is very small compared to the frequency of the laser light (THz range). This depends on the utilized laser wavelength and the magnitude of the wind speed, but usually the frequency change lies in the MHz range (see Chapter 2.1.3).

For coherent detection wind lidar (CDWL), the received light is mixed with a local reference oscillator. This field superposition contains a beat signal at the difference frequency of the signal and local oscillator field, thus lowering the required detection bandwidth. In addition, incoherent photons (i.e., ambient light) can not contribute to the beat signal and only influence the overall photon shot noise. Fig. 2.1 demonstrates the basic design of a monostatic coherent lidar system.

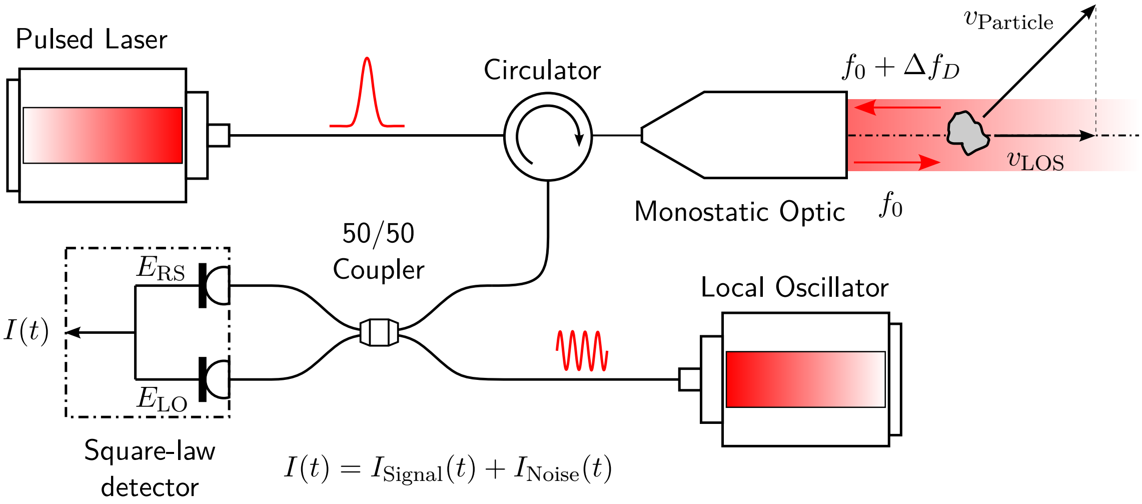


Figure 2.1: Principle diagram of a monostatic CDWL: a laser pulse is sent through a transceiver into the atmosphere where light is backscattered by aerosols moving with the velocity v_{Particle}. The return signal is imparted with a Doppler shift that is proportional to v_{LOS} and sent to a 3 dB coupler via a circulator. There, the signal is mixed with a continuous wave (CW) local oscillator before being applied onto the detector. The local oscillator is usually derived from the pulse laser system (e.g. in a chopped master-oscillator-power-amplifier configuration).

A derivation of the mathematical model of the coherent lidar frequency downshift is described according to DeLange (1968). Assume an optical setup where the incoming signal with a spectral bandwidth S and a local oscillator beam are mixed, before being applied onto a detector with a finite bandwidth $B \ll S$. For the case of an ideal square-law detector it is assumed that the detector output signal is proportional to the sum of all input signals squared as

$$I_{\text{Signal}}(t) \propto \left[\sum_i E_i(t)\right]^2 . \tag{2.1}$$

The coherent, time-dependent signals for the local oscillator E_{LO} with the frequency ω_{LO} and the return signal E_{RS} with the frequency ω_{RS} are modeled as

$$E_{\text{LO}}(t) = A_{LO} \cdot \cos(\omega_{LO} \cdot t) , \tag{2.2}$$

$$E_{\text{RS}}(t) = A_{RS} \cdot \cos(\omega_{RS} \cdot t) . \tag{2.3}$$

It is known that $A_{LO} \gg A_{RS}$ and that $\omega_{RS} \sim \omega_{LO}$. Substituting the equations above into the model for the ideal square-law detector yields

$$I(t) \propto A_{LO}^2 \cos^2(\omega_{LO} \cdot t) + 2A_{LO}A_{RS} \cos(\omega_{LO} \cdot t) \cos(\omega_{RS} \cdot t) + A_{RS}^2 \cos^2(\omega_{RS} \cdot t)\,. \quad (2.4)$$

Using the addition and power laws for trigonometric functions, the equation can be further separated into individual cosine terms and

$$\begin{aligned} I(t) \propto \frac{1}{2}\Big[& A_{LO}^2 \cos(2\omega_{LO}t) + A_{RS}^2 \cos(2\omega_{RS}t) + 2A_{LO}A_{RS} \cos([\underbrace{\omega_{LO} - \omega_{RS}}_{\ll \omega_{LO}}] \cdot t) \\ & + 2A_{LO}A_{RS} \cos([\omega_{LO} + \omega_{RS}] \cdot t) + A_{LO}^2 + A_{RS}^2)\Big] \end{aligned} \quad (2.5)$$

is obtained. The frequencies ω_{LO} and ω_{RS} are known to be outside the bandwidth of the detector (i.e., THz range) and are therefore neglected, especially if they add or multiply. This simplifies the above equation to be

$$I_{\text{Signal}}(t) \propto \frac{1}{2}\left[A_{LO}^2 + A_{RS}^2\right] + A_{LO} \cdot A_{RS} \cdot \cos([\omega_{LO} - \omega_{RS}] \cdot t)\,. \quad (2.6)$$

The resulting time-dependent signal is downshifted in its spectrum significantly and can be spectrally analyzed more simply. Further, note that the amplitude of the return signal A_{RS} is multiplied with the local oscillator amplitude A_{LO}. This amplification corresponds to a mathematical gain. The primary limitation of this method is the so-called shot noise caused by the local oscillator photo current since $A_{LO} \gg A_{RS}$ [15].

This noise signal can be seperately modeled as

$$I_{\text{Noise}}(t) \propto \sqrt{2 \cdot e \cdot |E_{\text{LO}}(t)|^2 \cdot B}\,, \quad (2.7)$$

where e is the elementary charge and B is the detector bandwidth. This type of noise is a fundamental physical effect and a general lower limit in optical signal detection. The Signal to Noise Ratio (SNR) is defined as

$$SNR = \frac{I_{\text{Signal}}}{I_{\text{Noise}}}\,. \quad (2.8)$$

Within coherent detection, two further cases are distinguished. If the local oscillator frequency ω_{LO} is close, but not equal to the laser frequency, the technique is called heterodyne. If ω_{LO} is equal to the laser frequency, the technique is called homodyne. The following relations hold for the frequency of the signal ω_{RS}:

$$\begin{aligned} \text{heterodyne: } & \omega_{RS} = \omega_{LO} + \Delta f_D + \Delta f_{\text{int}}\,, \\ \text{homodyne: } & \omega_{RS} = \omega_{LO} + \Delta f_D\,. \end{aligned}$$

Δf_D is the difference in frequency due to the Doppler shift. In the heterodyne case, Δf_{int} is some systematic offset frequency resulting from system properties. In both cases, the received light is converted into an analog voltage signal via a detector and then digitized. Henderson et al. (2005) state that “[...] coherent detection lidar is less sensitive to intensity distortion and bias and more sensitive to phase distortion, whereas the opposite is true for DDL sensors.” [9, p. 484]

2.1.3 Doppler shift

It is known that for a stationary observer and a moving source the following relation for the non-relativistic Doppler shift holds and

$$f_D = f_0 \cdot \left(\frac{1}{1 + \frac{v}{c}} \right) . \tag{2.9}$$

In common definition the observed frequency is lowered if the source is moving away (redshift) and vice versa (blueshift). The Doppler shift must be considered both from the aerosol's system of reference, as well as from the observer's system. The relative velocity calculated from a measured Doppler shift can be formulated as

$$v_{\text{LOS}} = -\frac{\Delta f_D \cdot \lambda}{2} , \tag{2.10}$$

where Δf_D is the difference in frequency between the emitted and received light, v_{LOS} is the relative velocity and λ is the wavelength of the emitted light [9, p. 484].

If it is defined that the LOS velocity is positive if the particle is moving away from the observer. It follows from the above definition of signs that

$$\begin{aligned} \Delta f_D \leq 0 &\implies v_{\text{LOS}} \geq 0 , \\ \Delta f_D > 0 &\implies v_{\text{LOS}} < 0 . \end{aligned}$$

A derivation of Eq. 2.10 that also takes relativistic effects into account can be found in a publication by Mahnke et al. [16].

2.1.4 Lidar equation

The lidar equation describes the signal power returned to a detection system at a given range r and is characteristic for such measurements. As wind speeds can only be calculated as long as there are sufficient aerosols available to determine the Doppler shift, return signals up to high altitudes are desirable. For a general single-wavelength pulsed lidar the following equation can be assumed to hold for the return signal power

$$P(r) = P_0 \cdot \frac{c \cdot \tau}{2} \cdot A \cdot \frac{\beta(r)}{r^2} \cdot \exp\left[-2 \cdot \int_0^{r_m} \sigma(r') dr' \right] , \tag{2.11}$$

where P_0 is the average transmitted pulse power, τ is the pulse duration and A is the receiver area. The backscattering coefficient β and the extinction coefficient σ are atmospheric parameters and direct indicators for the amount of aerosols present [17].

It is anticipated that the following relationship holds between β and σ:

$$\beta(r) = \frac{P}{4\pi} \cdot \sigma(r) , \tag{2.12}$$

where P is the scattering phase function from Mie theory [18]. For the coaxial or noncoaxial bistatic case with central obstruction Eq. 2.11 can be modified to account for the additional effects on the return signal, especially the overlap function and antenna efficiency. This is covered in detail by Harms (1979) [19].

Table 2.1: Approximations for the five largest aerosol sources per annum [20]

Aerosol Type	Emission flux (per year) [kg]
Sea Spray	1×10^{12} - 6×10^{12}
Desert Dust	1×10^{12} - 3×10^{12}
Terrestrial Biogenic Aerosols	$\sim 1 \times 10^{12}$
Volatile Organic Compounds	1×10^{11} - 6×10^{11}
Terpenes	4×10^{10} - 4×10^{11}

2.2 Planetary boundary layer

Tampieri (2017) defines the planetary boundary layer (PBL) as the lower part of the troposphere in which interactions occur with Earth's surface. In classical fluid mechanics the boundary layer thickness δ is defined at the point where the mean velocity $\bar{u}(y)$ of the boundary layer velocity profile matches the free stream velocity u_0. Due to practical considerations the PBL height h is defined at the point where the aerosol density sharply drops. The measurability of h varies greatly depending on the given aerosol distribution. The flow within this region is highly turbulent and influences the transport of pollutants [7].

To emphasize the importance of this comparatively thin layer of the atmosphere, Tampieri states that "[understanding] these interactions is an important step for the proper modeling of weather and climate, and, in general, of the dynamics of the atmosphere."

2.2.1 Aerosols

As mentioned in section 2.1, the presence of aerosols is essential for CDWL, since the scattering signal strength of aerosols is many times greater than that of molecules and the spectral width is substantially smaller. In addition, the Brownian motion of aerosol particles is negligible [9, pp. 492-496].

Boucher (2015) gives a compiled estimation for the yearly emission fluxes from different aerosol sources. He defines the following categories: desert dust, marine, volcanic and biogenic aerosols, aerosols from the burning of biomass and aerosols from fossil fuel combustion [20]. Tab. 2.1 lists the five largest contributors to annual aerosol emissions. The effective backscatter properties of an aerosol depend on its physical properties. For simple, spherical scatterers whose diameter corresponds approximately to the wavelength of the used light, conclusions about the optical properties can be made using Mie's theory.

There are methods to estimate macroscopic values such as the atmospheric extinction factor σ. The Koschmieder equation can be used to approximate a value for said parameter at ground level based on the atmospheric visibility[1]

$$x = \frac{3.912}{\sigma}\,, \tag{2.13}$$

where x is the visibility range in m. This equation holds for a wavelength range between 520 nm and 550 nm [20, p. 45]. The Kruse formula tries to cover a larger wavelength range and can be used in the infrared domain as well:

$$\Gamma(x, \lambda) = \frac{17}{x} \cdot \left(\frac{550}{\lambda}\right)^{q(x)}\,, \tag{2.14}$$

[1] Range at which a high contrast object can still be identified

where Γ is the atmospheric attenuation coefficient and λ is the wavelength of the used light in relation to a reference wavelength at 550 nm [21]. The exponent $q(x)$ is a heuristic function and defined by Kruse to be

$$q(x) = \begin{cases} 0.585 \cdot x^{1/3} & x \leq 6000\,\mathrm{m} \\ 1.3 - 1.6 & x > 6000\,\mathrm{m} \end{cases} . \tag{2.15}$$

2.2.2 Velocity profile

There are known solutions for the velocity profile of the boundary layer for viscous and compressible fluids in the laminar, as well as the turbulent case. A comprehensive deduction for the plane plate case can be found by Nieuwstadt et al. [22, pp. 89-95]. In the case of the PBL, analogous equations exist.

Assuming steady, horizontally homogeneous conditions and constant fluxes a velocity profile for the surface layer can be derived from dimensional analysis and the boundary layer equations. Several dimensionless quantities are defined for this purpose [7, p. 56]. The friction velocity u^* is defined as

$$u^* = \left(\langle uw \rangle^2 + \langle vw \rangle^2 \right)^{\frac{1}{4}} , \tag{2.16}$$

where $\langle uw \rangle$ and $\langle vw \rangle$ represent the horizontal velocities. The dimensionless height z^+ is defined by the above velocity and the viscosity ν to be

$$z^+ = \frac{z \cdot u^*}{\nu} . \tag{2.17}$$

Tampieri (2017) defines here, analogously to the plane plate boundary layer, the area of the viscous sublayer and a surface region. In the viscous sublayer the profile is largely influenced by viscosity and increases linearly with z^+ as

$$U(z^+) = u^* \cdot z^+ . \tag{2.18}$$

In the surface area ($z^+ \lesssim 1000$) the profile adheres to the so-called log-law

$$U(z^+) = \frac{u^*}{\kappa} \cdot \ln \frac{z^+}{z_0} + U(z_0) . \tag{2.19}$$

Here, κ is the von Karman constant and z_0 is an offset height. The transition between the profiles takes place in the so-called mixing region at $z^+ \approx 11$. For the rough surface case (i.e., tall buildings) only the latter formula is applied with $U(z_0) = 0$ [7].

2.3 Turbulent atmospheric conditions

In general, the assumption that the flow within the PBL is laminar cannot be made. A laminar flow can quickly turn turbulent due to pressure gradients, surface roughness and temperature fluctuations. Turbulent flows are characterized by spatial and temporal inhomogeneity, strong mixing processes and a stochastic behavior at various levels of scale. In turbulent flows, the propagation of light is made substantially more difficult, as the beam is broadened, defocused and even deflected from its original path [6]. Furthermore, as stated in section 2.2, turbulent effects are fundamental to the dynamics of the atmosphere. Therefore, it is important to define parameters for the description of turbulent flows.

Commonly, the scalar turbulent kinetic energy (TKE) k serves as the main measure for the turbulence of a flow field [22, p. 115]. For the exact steps of time averaging the Navier-Stokes equations, the reader is referred to Nieuwstadt (2016) [22] or Pope (2000) [23]. In the Reynolds-averaged Navier-Stokes (RANS) equations the energy k is defined via the fluctuating parts of the flow velocity components $u = \overline{u} + u'$, $v = \overline{v} + v'$ and $w = \overline{w} + w'$ as

$$k = \frac{1}{2} \cdot \left(\overline{(u')^2} + \overline{(v')^2} + \overline{(w')^2} \right) . \tag{2.20}$$

The governing differential equation for the temporal evolution of the TKE in a flow field is

$$\frac{Dk}{Dt} = P_k + T_k + \Pi_k + D_k - \nu \cdot \overline{\left(\frac{\partial u_i'}{\partial x_j} \right)^2} \tag{2.21}$$

and definitions for the transport term $T_k + \Pi_k + D_k$ and production term P_k can be found in Nieuwstadt et al. (2016). $\frac{Dk}{Dt}$ denotes the material derivative of the turbulent kinetic energy. u_i' and x_j denote all fluctuating velocity and spatial components in Einstein notation. The term

$$\varepsilon = -\nu \cdot \overline{\left(\frac{\partial u_i'}{\partial x_j} \right)^2} \tag{2.22}$$

is referred to as the turbulent kinetic energy dissipation rate (TDR) and reflects the local change of k as a scalar value, where ν represents the viscosity of the fluid [22, p. 128].

2.3.1 Kolmogorov law

The idea behind the dissipation of turbulent energy is that turbulent structures, so-called “eddies” are unstable and break down repeatedly into smaller structures, again and again until they are completely dissolved by viscous dissipation. This concept was mathematically introduced by Kolmogorov in 1941. His 5/3-power-law states that there is a characteristic curve in the energy spectrum of the flow field which can be defined as

$$E(f) = C_0 \cdot \varepsilon^{\frac{2}{3}} \cdot f^{-\frac{5}{3}} . \tag{2.23}$$

The spectral energy E has the unit $[E] = \frac{m^3}{s^2}$. The calibration constant $C_0 \approx 0.5$ was determined in experiments to be universal for a wide variety of experimental and geophysical flows [14]. An important assumption made for the validity of the equation above is that $\varepsilon \neq \varepsilon(f)$.

2.3.2 Dissipation estimation from wind data

In order to be able to determine a value for ε from a temporal wind speed measurement, a spectral analysis approach was introduced by Smalikho (1997). Furthermore, they performed a theoretical consideration of the achievable accuracy [13].

From an ensemble of wind speed measurements u_i with $i = 0, 1, ..., M-1$ during the interval $T = \Delta t \cdot M$, the spectral density of the measurements can be determined by

$$\hat{S}\left(\frac{k}{T}\right) = \frac{2 \cdot \Delta t}{M} \cdot \left| Z\left(\frac{k}{T}\right) \right|^2 , \tag{2.24}$$

where $Z\left(\frac{k}{T}\right)$ denotes the discrete Fourier transform (DFT) and $k = 0, 1, ..., M/2$. Now given the spectral density $\hat{S}$ and assuming that k is transported along the flow field without significant statistical change, one can choose an interval $I \in [0, M-1]$ with length n from

$\hat{S}(f)$ for which the spectrum is unbiased, that is

$$\hat{S}(f = I) \simeq \frac{1}{n}\sum_{l=1}^{n} \hat{S}(f_l)\,, \tag{2.25}$$

where n is the length of the chosen interval and all values within lie close to the mean. With the help of this interval, a value for the dissipation can be estimated via a correlation function $Q(f)$ and therefore,

$$\hat{\varepsilon} = \left(\frac{\hat{S}(I)}{Q(I)}\right)^{\frac{3}{2}}\,. \tag{2.26}$$

In the case that the measurements are made at at fixed point in longitudinal direction, Smalikho proposes the following function for

$$Q(f) = 0.15 \cdot \overline{U}^{\frac{2}{3}} \cdot f^{-\frac{5}{3}}\,, \tag{2.27}$$

where $\overline{U}$ is the average velocity over the interval I. The mean normalized error for a given interval length $n \in \mathbb{N}$ is

$$B(n) = \overline{\left(\frac{\hat{\varepsilon}}{\varepsilon}\right)} = \begin{cases} \frac{\Gamma\left(\frac{3}{2}+n\right)}{n^{\frac{3}{2}}\cdot\Gamma(n)} & n \leq 10 \\ 1 + \frac{0.375}{n} & n > 10 \end{cases}\,. \tag{2.28}$$

$\Gamma(n)$ denotes the real-valued gamma function, $\hat{\varepsilon}$ is the estimated value for the TDR and ε is the true value. For $n > 10$ the approximation must be used due to computational issues. The relative random error can be estimated by

$$E(n) = \left[\overline{\left(\frac{\hat{\varepsilon}}{\varepsilon}\right)^2} - B^2\right]^{\frac{1}{2}} = \left[\left(1+\frac{1}{n}\right)\cdot\left(1+\frac{2}{n}\right) - B(n)^2\right]^{\frac{1}{2}}\,. \tag{2.29}$$

For $n \gg 1$ the random error can be approximated by $E(n) \approx \frac{3}{2}\frac{1}{\sqrt{n}}$. The total estimation error $A(n)$ is equivalent to a relative standard deviation for the estimate of the mean TDR $\hat{\epsilon}$ and defined as

$$A(n) = \frac{\sigma_{\hat{\epsilon}}}{\hat{\epsilon}} = \left[E(n)^2 + (B(n)-1)^2\right]^{\frac{1}{2}}\,. \tag{2.30}$$

2.3.3 Turbulent index of refraction

The structure constant $[C_n^2] = \mathrm{m}^{-2/3}$ is a measure for the turbulence-induced spatial variations of the refractive index in the atmosphere. These variations affect the quality of a laser beam during atmospheric propagation. To determine this parameter, the work of Hauchecorne et al. (2016) and Vrancken et al. (2016) is referred to.

A fluctuating quantity FQ must be determined that "[...] represents correctly the square of relative fluctuations of atmospheric density and temperature [...]" [3]. The turbulence-induced density fluctuations are suitable for this purpose and can be calculated as follows [4]:

$$FQ = \frac{\sigma_\rho}{\rho} = \sigma_w \cdot \frac{N}{g}\,. \tag{2.31}$$

Here, N is the Brunt–Väisälä frequency and $g = 9.81\,\mathrm{m\,s^{-2}}$. Using a longitudinal length scale $L_i \approx 762\,\mathrm{m}$, the vertical gust velocity σ_w can be approximated by

$$\sigma_w \approx \sqrt{\frac{(\varepsilon \cdot L_i)^{\frac{2}{3}}}{0.78}}\,. \tag{2.32}$$

Now C_n^2 can be determined as

$$C_n^2 = L_i^{-\frac{2}{3}} \cdot (n-1)^2 \cdot FQ^2 \tag{2.33}$$

with the refractive index of air $n = 1 + 0.776 \times 10^{-6}\,\mathrm{K\,Pa^{-1}} \cdot \frac{p(h)}{T(h)}$. Values for the pressure p and temperature T at altitude h can be taken from the International Standard Atmosphere (ISA) database.

The Brunt–Väisälä frequency is a measure for the oscillations of a fluid caused by buoyancy forces due to temperature gradients in the vertical direction. The corresponding frequency N can be calculated with

$$N = \sqrt{\frac{g}{T(h)} \cdot \left(\frac{\Delta T(h)}{\Delta h} + \Gamma_d\right)}\,, \tag{2.34}$$

where Γ_d is the dry adiabatic lapse rate (DALR) which is defined for the ISA atmosphere as

$$\Gamma_d = \frac{g}{c_p} = \frac{\kappa - 1}{\kappa} \cdot \frac{g}{R}\,, \tag{2.35}$$

in which the specific heat capacity c_p can also be expressed by the heat capacity ratio $\kappa = 1.405$ and gas constant $R = 287.1\,\mathrm{J\,kg^{-1}\,K^{-1}}$ for air. The DALR hence yields $\Gamma_d \approx 9.850 \times 10^{-3}\,\mathrm{K\,m^{-1}}$ [24, p. 137].

2.3.4 Hufnagel-Valley model

Several models exist that attempt to model the course of C_n^2 over altitude. In the context of this work, the Hufnagel-Valley (HV) model is used. This model is a modification of the original Hufnagel-model, proposed in 1978. The HV-model provides values for C_n^2 from the ground level up to the tropopause and beyond. The value for C_n^2 at altitude h in km is modeled as

$$C_n^2(h) = c_0 \cdot W^2 \cdot h^{10} \cdot e^{-h} + c_1 \cdot e^{-\frac{h}{1.5}} + A \cdot e^{-\frac{h}{0.1}}\,, \tag{2.36}$$

$$c_0 = 8.2 \times 10^{-26}\,, \tag{2.37}$$

$$c_1 = 2.7 \times 10^{-16}\,. \tag{2.38}$$

Here, W is defined as the root mean square wind speed in $\mathrm{m\,s^{-1}}$ between 5 km to 20 km above ground, that is

$$W^2 = \frac{1}{15} \int\limits_{5\,\mathrm{km}}^{20\,\mathrm{km}} v^2(h)\,dh \tag{2.39}$$

and A is an adjustable parameter to account for the effects of the PBL. For nighttime conditions, these parameters are commonly set to $W = 21\,\mathrm{m\,s^{-1}}$ and $A = 1.7 \times 10^{-4}$. Nevertheless, for heights below 3 km it must be assumed that the modeling of boundary layer effects is still insufficient. Therefore, inaccurate predictions for C_n^2 are expected in this domain [25, p. 222].

3 Experimental setup

In the context of this thesis, a newly developed coherent lidar system was put into operation. This novel system, with the capability to process wind spectra in real time, is based largely on the work of Kliebisch and Mahnke (2020) [10]. Within the scope of this work, the system integration was essentially completed, the optical system finalized and the control software written within the *Hyperion* framework developed by Dr. Oliver Kliebisch. In the following section, an overview over the used hardware and software for this custom lidar device will be given.

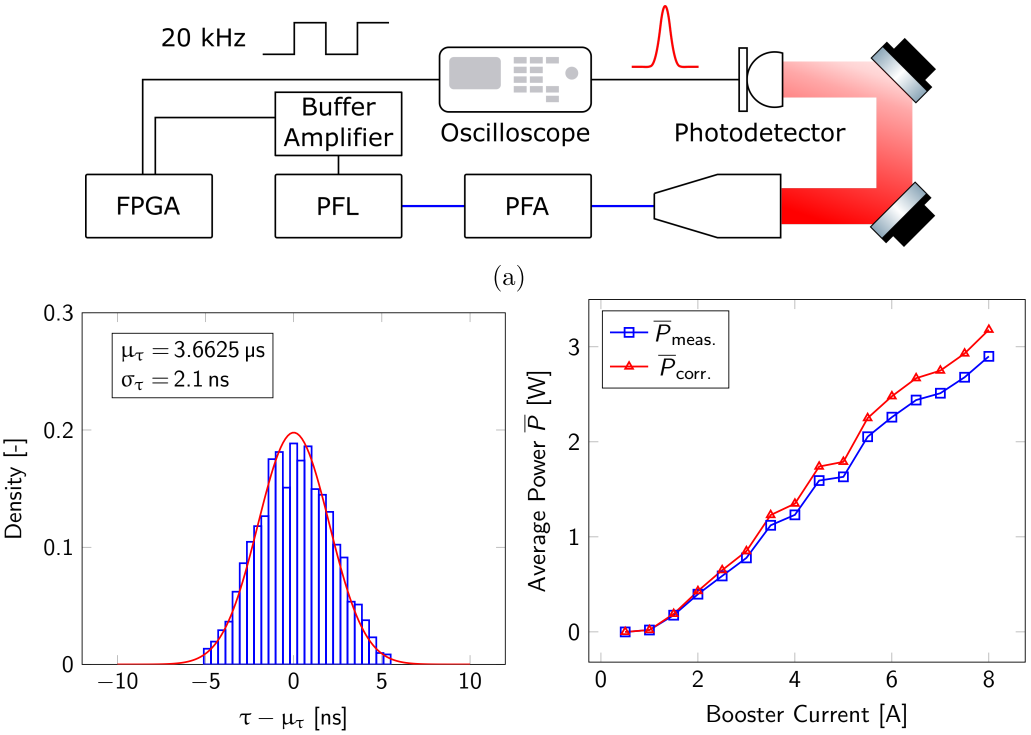
3.1 Hardware Overview

The hardware that was used for this system will be described in detail in the following subsections. The elements will be described in the order of the different process steps that occur in reality, beginning at the pulse generation, through the optical detection up to the data processing step. The optical setup will be described in its own chapter (Ch. 3.2). At the end of this section, a schematic system overview is given.

3.1.1 Laser source

A pulsed erbium-doped fiber laser (EDFL) and erbium-doped fiber amplifier (EDFA) form the backbone of the system, with the utilized hardware being listed in Tab. 3.1., in addition to the nominal operation parameters. The laser source reliably delivers pulsed laser light in the infrared spectrum at 1548 nm. This specific wavelength was chosen due to eye safety concerns, atmospheric attenuation considerations and its technological maturity, as it is extensively employed in the telecommunication sector. An initial laser pulse is generated from a continuous wave tunable diode laser that is shaped into 300 ns long pulses by an acousto-optic modulator (AOM). The shaping process also imparts a systematic frequency offset of 80 MHz that is later used for heterodyne detection. In turn, the AOM is driven by an arbitrary waveform generator (AWG), which itself is controlled by an external trigger pulse. After the laser pulse has been generated in the AOM, it is then amplified by an internal EDFA to 0.6 µJ. Following this step, the light pulse is passed to the external EDFA via a polarization-maintaining (PM) fiber. Here, the pulse energy is increased step by step via three inline EDFAs, called the Pre-Amplifier, Pre-Booster and Booster respectively, until it reaches 180 µJ. Controlling the emitted output power can be achieved by changing the booster pump diode currents. After this step, the pulse is passed to an output port through a circulator [26].

During the system assembly, a series of characterization tests were performed in order to determine the exact laser source parameters. Fig. 3.1a illustrates the setup that was used to measure the latency and jitter of the entire system. A latency of 3.663 µs was measured and used as an offset compensation coefficient for the field programmable gate array (FPGA). Fig. 3.1b also shows that the system jitter is in the nanosecond range and can be neglected. The average pulse power was measured using an optical power meter in place of the photodetector and found to be within the manufacturer specifications (see Fig. 3.1c).



(a)

(b)

(c)

Figure 3.1: **(a)** Experimental setup: the FPGA trigger output is attached to a $50\,\Omega$ buffer amplifier and connected to an oscilloscope, as well as the pulsed fiber laser (PFL). When triggered, the PFL emits a pulse that is amplified by the pulsed fiber amplifier (PFA) and spatially widened by a beam expander. Two gold-coated mirrors redirect the beam onto a photodetector. The detector is attached to the oscillator via a separate input channel. **(b)** Pulse latency histogram of $n = 2017$ samples [26]. The latency is defined as the time interval between the falling edge of the trigger signal and the rising edge of the pulse response on the detector. The number of bins ($b = 26$) was chosen according to Sturges' rule [27] **(c)** Average power as a function of the PFA input current. The booster current refers to the Pre-Booster, as well as the Booster current. The Pre-Amplifier current was set to be 2.3 A as per the PFA data sheet. $\overline{P}_{\text{meas.}}$ represents the raw measured values. $\overline{P}_{\text{corr.}}$ was corrected to include the losses at the optical interfaces in the beam expander and on the surface of the mirror devices ($N_{\text{Loss}} \approx 0.9131$)

3.1.2 Optical detector

Once the scattered light is collected by the optical system and coupled into a PM fiber via the optical circulator, the return signal is mixed with the CW local oscillator source. This is achieved using a 3 dB fused fiber optical coupler before applying the beat signal to a detector. The local oscillator power is decreased to one percent of its original power beforehand, in order to avoid detector oversaturation and damage to the photodiodes. The detector converts the optical signal into an electronic signal.

For this purpose, two different photodetectors were tested. At first, an indium-gallium-arsenid (InGaAs) avalanche photodiode (APD) was used. This detector leverages the photoelectric effect to generate an analog electric signal from incoming photons. In addition, the electrical avalanche effect is used to amplify the signal. This device was later replaced with a positive-intrinsic-negative (PIN) photodiode. This type of detector uses an undoped

Table 3.1: Laser Source Hardware & Specifications [28]

Hardware	Type
Pulsed Fiber Laser	Lumibird PEFL-EOLA
Pulsed Fiber Amplifier	Lumibird PEFA
Parameter	**Specification**
Wavelength	1550 nm
Linewidth	5 kHz
Offset Freq.	80 MHz
Rep. Rate	20 kHz
Pulse Length	300 ns
Pulse Energy	180 µJ
Average Power	3.6 W
Peak Power	600 W

semiconductor region surrounded by p-type and n-type doped layers on either side.

While the APD amplifies the amount of photoelectrons and therefore the signal strength, this process also introduces additional stochastic noise known as excess noise. The signal-to-noise ratio (SNR) is, in general, not improved. PIN detectors are designed in an inherently different way and do not suffer from this issue. A theoretical comparison of these two different sensor types with respect to their performance can be found in a publication by Mahnke et al. (2019) [16]. In Tab. 3.2 the hardware used for this lidar system is listed.

Table 3.2: Photodiode Hardware & Specifications

Hardware	Type	
APD Diode	Thorlabs PDB570C [29]	
PIN Diode	Thorlabs PDB770C [30]	
Parameter	Specification	
	APD	PIN
Type	InGaAs	InGaAs
Bandwidth	400 MHz	400 MHz
Wavelength Range	1200 nm to 1700 nm	1260 nm to 1625 nm
Saturation Power	200 µW	440 µW
Gain Factor	2.5 to 10.0	1
Noise Factor (est.)	1.9 to 7.0	1

3.1.3 Real time data processing

The time-dependent detector signal is filtered by a low-pass filter (LPF) before being digitized at 250 Megasamples per second. From signal theory it is known that the Nyquist frequency at a given sampling rate is equal to

$$f_{\text{Nyq.}} = \frac{1}{2} \cdot f_{\text{Sampling}} = 125\,\text{MHz}\,. \tag{3.1}$$

This represents the upper limit of the frequency domain before undersampling occurs. In heterodyne detection, the systematic offset frequency corresponds to a Doppler shift of 0 Hz. Given an offset value of 80 MHz, frequency shifts between -80 MHz and 45 MHz can be identified. Using Eq. 2.10 with $\lambda \approx 1550\,\text{nm}$ it is calculated that LOS wind velocities between $-62.0\,\mathrm{m\,s^{-1}}$ and $34.9\,\mathrm{m\,s^{-1}}$ can be resolved. This is sufficient for all realistically expected scenarios.

However, at this sampling rate and with a resolution of 16 bits per sample, the data volume becomes quite large at

$$\frac{16\,\mathrm{bit} \cdot 250 \times 10^{6}\,\mathrm{s}^{-1}}{8 \cdot 1024^{2}\mathrm{bit/MiB}} \approx 476.8\,\mathrm{MiB/s}\,. \tag{3.2}$$

Storing and analyzing such an amount of data can be nontrivial. Therefore, a FPGA-based pre-processing solution was developed by Kliebisch et al. (2020). The filtered APD output is passed to an analog-to-digital converter (ADC) that is clocked by a phase-locked loop (PLL) clock chip. The digitized signal and the clock signal are then passed to the FPGA internal logic. The raw signal data is sliced into range bins by a programmable window function and Fourier-transformed using a 128 point fast Fourier transform (FFT). The frequency spectra are then accumulated over 16384 measurements at 20 kHz in memory. Assuming that the system is ergodic, this process is equivalent to averaging over the individual measurements. The 20 kHz trigger signal is synchronously derived from the ADC clock and is also used to trigger the laser pulses. This synchronizes all processes against a common clock. The integration time for a single output spectrum lies at 0.8192 s. The averaged spectrum is then passed into system memory by a direct memory access (DMA) controller and transferred to the onboard computer via an Ethernet connection [10].

Using a resolution of 96 range and 64 frequency bins and the Float64 data type for storage, a data reduction to 48 KiB per spectrum can be achieved. Tab. 3.3 includes the used hardware.

Table 3.3: FPGA & ADC Hardware [10]

Hardware	Type
FPGA	Xilinx Zynq 7020 SOC
Carrier Board	Trenz TE0701
ADC	Analog Devices AD9467
Clock Chip	Analog Devices AD9517
Parameter	**Specification**
Sampling Rate	250 Megasamples/s
Bit Depth	16 bits
Counter	$2^{14} = 16384$
Trigger	20 kHz
No. range gates	96
No. freq. bins	64

3.1.4 Onboard computer and power supply

The lidar system is controlled by a central onboard computer. For this purpose a Nvidia Jetson Nano is used. This microcomputer performs all high level calculations needed during system operation and is responsible for the hardware, as well as software interfaces. The power supply is ensured by a single 24 V power supply unit. To put the system into operation, it only needs to be connected to a common Type F wall connection (230 V).

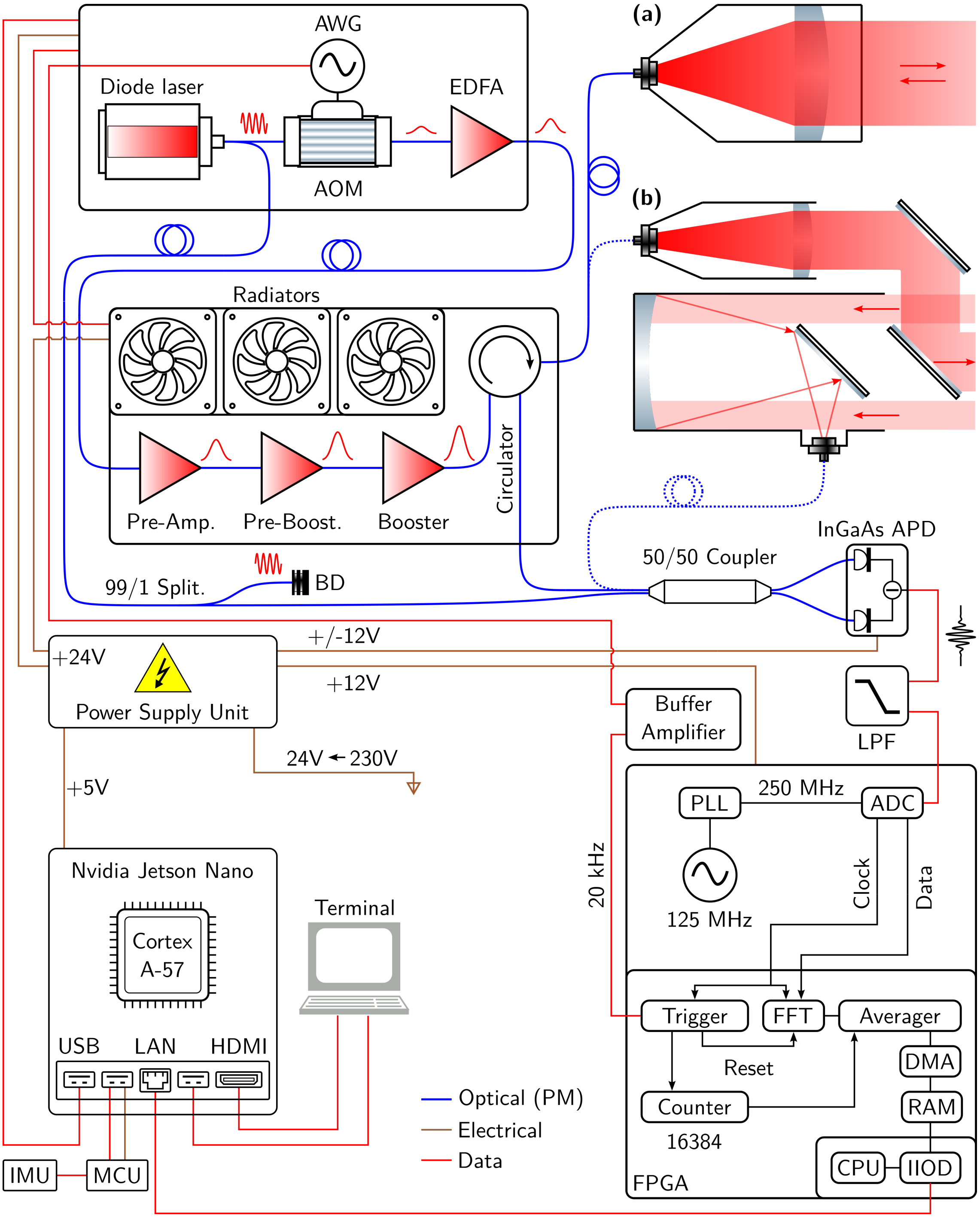


Figure 3.2: Schematic system overview. **AWG**: Arbitrary waveform generator, **AOM**: Acousto-optic modulator, **EDFA**: Erbium-doped fiber amplifier, **BD**: Beam dump, **APD**: Avalanche photodiode, **LPF**: Low pass filter, **PLL**: Phase-locked loop, **ADC**: Analog-to-digital converter, **FFT**: Fast Fourier transform, **DMA**: Direct memory access, **IIOD**: Industrial input/output daemon, **FPGA**: Field programmable gate array, **MCU** Microcontroller unit, **IMU** Inertial measurement unit, **PM**: Polarization-maintaining, **(a)** Monostatic setup, **(b)** Bistatic setup

3.2 Optical setups

As can be seen in Fig. 3.2, two optical setups were considered. In the following section, both the monostatic and bistatic optical setups will be presented, their advantages and disadvantages discussed and alignment techniques showcased.

3.2.1 Monostatic Setup

The monostatic setup uses a single optical system for both the transmission and reception of the laser pulses and the scattered light. This is referred to as a transceiver setup. Since this process occurs in a single fiber for in- and output, a mechanism needs to be used in order to distinguish both channels. The optical circulator that is integrated into the EDFA is used for this purpose. This device outputs light to different ports depending on the direction in which it entered, thereby achieving the channel separation. However, an immediate drawback of this method is its susceptibility to internally reflected light from optical interfaces (i.e., fiber connections) and hence the necessity for high dynamic range detection. In practice, this leads to some detector saturation in the first few range bins. Fig. 3.3 illustrates the principle setup of such an optical, fiber-coupled system.

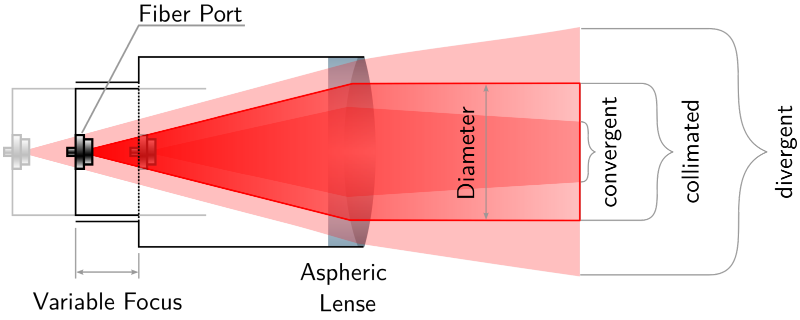


Figure 3.3: Principle monostatic setup: light exits the fiber port under some numerical aperture (NA) and is refracted when passing through the lens. Depending on the distance from the lens, the beam is either converged, collimated or diverged.

For every measurement cycle, a light pulse is emitted through the fiber port and sent through an aspheric lens. Depending on the distance of the fiber port from the lens, different beam alignments can be set. This is mainly adjusted by changing the focusing distance. Ideally, the beam is aligned in a way such that it is collimated when exiting. An advantage of the monostatic method is the automatic alignment of both the receiver and transmitter fields of view (FOVs), since the returned light must take the same path to couple back into the fiber. This makes a monostatic system very easy to use and sturdy against vibrations or other mechanical disturbances. Furthermore, a large beam diameter increases the Rayleigh length of the laser beam and thus minimizes the effects of divergence. A disadvantage of this setup is the need for a large-size lens in order to achieve a relevant beam diameter, which is a driver of system cost. In addition, the internally reflected light can quickly become a major issue once the pulse power exceeds the low kW range.

For this thesis, a monostatic setup with a focal length of 500 mm and a beam diameter of 100 mm was used. Due to the relatively low maximum peak power of 600 W for this system and the high dynamic range of the detector, this approach was feasible. However, the lens was originally optimized for a 780 nm design wavelength with a different application in mind, leading to some aberrations at the edge of the beam.

3.2.2 Bistatic Setup

In the bistatic case, dedicated optics for sending and receiving light are used. The FOVs of the optics must be adjusted accordingly, but optimized optical systems can be used for transmission and reception. If the emitter and transmitter FOVs are aligned to a common optical axis, the setup is referred to as coaxial bistatic. This particular layout is illustrated in Fig. 3.4.

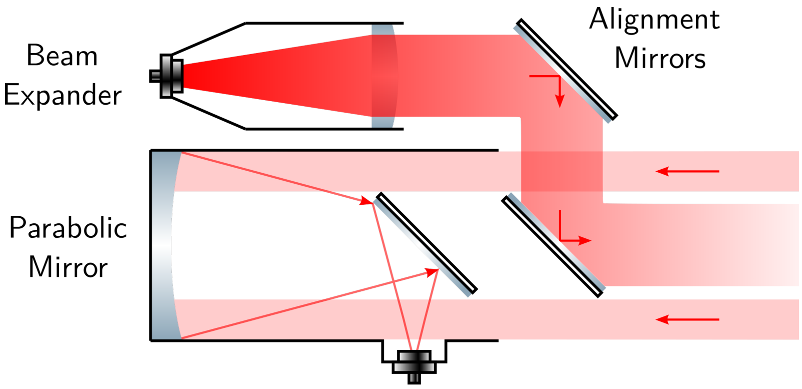


Figure 3.4: Principle coaxial bistatic setup: the emission beam is redirected onto a common optical axis via a pair of alignment mirrors. A Newton telescope is used as the receiver. The respective beams need to be aligned and collimated.

Firstly, the amplified pulse is expanded and then collimated as described for the monostatic setup. Then, two alignment mirrors are used in order to redirect the beam onto the optical axis of the receiver. A Newton telescope is used as the receiving optic. Incoming light is collected at the end of the telescope by a parabolic mirror and focused into a fiber port via an alignment mirror. This design allows for the easy implementation of a large diameter achromatic receiver. Due to the mirror setup, the receiver FOV is partially blocked which leads to a low sensitivity in the near field.

One advantage of this method is that very large receiver diameters can be achieved cost-effectively. In addition, both channels can be aligned independently of each other. However, this is also the main disadvantage, due to the significant increase in the number of degrees of freedom (DOFs). As an example, assume that each mirror has two DOFs for its orientation and three DOFs for its spatial positioning. Furthermore, the transmitting and receiving optics have one DOF for collimation each. Finally, an additional rotational DOF at the receiver fiber port to account for the effects of polarization must be considered. This leads to a total of 23 DOFs that need to be aligned and which are susceptible to mechanical disturbances. Alignment techniques will be presented in the following section. Tab. 3.4 summarizes the used hardware for this setup.

The following theoretical approximation is used to calculate the beam diameter

$$d \approx 4B_\mathrm{e}\lambda \cdot \left(\frac{f}{\pi \cdot MFD}\right) = 80 \cdot 1550\,\mathrm{nm} \cdot \left(\frac{11.32\,\mathrm{mm}}{\pi \cdot 25\,\mu\mathrm{m}}\right) \approx 17.87\,\mathrm{mm}\,, \tag{3.3}$$

where f is the focal length of the fiber collimator, λ is the wavelength, MFD is the mode field diameter of the fiber and $B_\mathrm{e} = 20$ is the expansion factor of the beam expander [31].

Table 3.4: Bistatic setup parameters

Hardware	Type
Telescope	TS-Photon 6" F4 Newton
Beam Collimator	Thorlabs F220APC-1550
Beam Expander	Thorlabs GBE20-C
Beam Mirrors	Thorlabs PFE20-M01
Parameter	**Specification**
Receiver Diameter	152.4 mm
Receiver Focal Length	600 mm
Receiver Obstruction	70 mm
Beam Diameter	17.9 mm

3.3 Alignment Techniques

The techniques that were used to align the optical systems (mono- and bistatic) will be presented in the following section. In general, the monostatic optic is inherently aligned and must only be collimated. In the bistatic case, the alignment of the mirrors must be carefully adjusted in a multi-stage process. It is important to remember that small errors during the initial alignment can cause serious issues when considering the measuring range of up to several kilometers.

3.3.1 Collimation

For the beam collimation a shearing interferometer pattern was leveraged in order to avoid a time intensive beam walk process. Fig. 3.5a illustrates the experimental setup. The optical system in question is aimed onto a thick shear plate that is orientated at an angle to the common axis. Due to the inclination, two separate interface reflections occur. The first one occurs at the surface of the shear plate, the second one at the internal interface. The latter reflection is refracted twice when entering and exiting the shear plate, causing both beams to be deflected upwards through a diffusion plate. Due to the optical path difference, a fringe interference pattern can be measured at the intersection between both reflections. This pattern is parallel to the optical axis of the incident beam when the beam is collimated [32].

In order to avoid chromatic aberrations, the collimation process was performed directly at the correct wavelength of 1550 nm. However, this meant that an InGaAs infrared-capable camera had to be used to observe the fringe pattern (Tab. 3.5). Fig. 3.5b and 3.5c show the fringe pattern of a collimated and uncollimated beam respectively. The collimation of the beam expander was set in real time by adjusting the focusing length.

3.3.2 Alignment

The alignment of the bistatic setup is carried out in separate stages due to the substantial amount of DOFs. Firstly, the internal mirrors of the Newton telescope are adjusted with the help of a Cheshire eyepiece. The eyepiece allows the alignment of the parabolic and secondary mirrors by means of a crosshair in a beam walk-like procedure. Hereby, the alignment of the mirrors is adjusted in an alternating fashion, until the crosshair reflections of the Cheshire overlap.

Once this stage has been completed, the collimation of the receiver can be set using the procedure explained in Ch. 3.3.1. The beam expander and the alignment mirrors can then be mounted. A beam walk over a distance of 5 m is performed to pre-align the receiver and expander. For this purpose, laser light in the visible range is utilized for illumination, where a 532 nm light source is fed into the receiver fiber port and a 650 nm source is connected to the emitter port, similar to the setup shown in Fig. 3.5d.

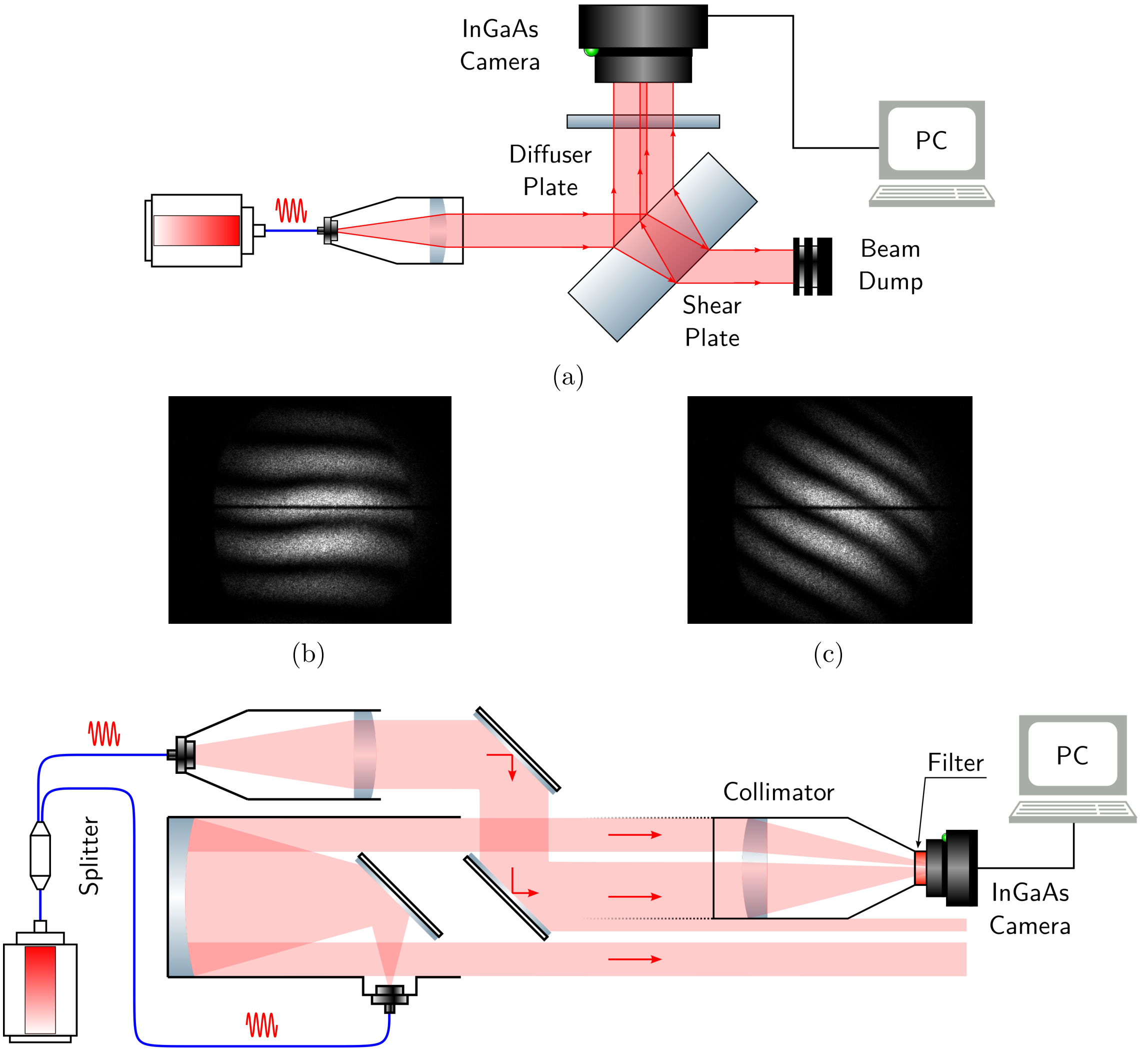
The receiver is achromatic and at this point, chromatic errors in the expander beam alignment are negligible. A crosshair target serves as the point of reference. The target is aligned with the outer ring of the receiver beam and then the alignment mirror DOFs are adjusted until the focused expander beam lies at the center of the receiver beam for two distances of the target from the telescope. During this process, a good near-field alignment can be achieved. The main alignment technique can then follow.

Fig. 3.5d depicts the alignment setup. The optical ports of the bistatic setup are connected to a single 1550 nm CW fiber laser source that is split by a 50/50 fiber splitter. Both beams are collimated and pre-aligned. The setup is facing a camera that is attached to a collimator. The distance between both setups can be smaller than one meter. The collimator faces the bistatic system in such a way that parts of the receiver and emitter beams are captured. The collimator is focused to infinity, de facto being collimated itself, and projects the beams onto the sensor of the camera. An optical density filter is interconnected to protect the sensor from overexposure. As long as the setup is roughly aligned, two spots will appear on the camera sensor, belonging to the emitter and receiver respectively. By adjusting the alignment mirror DOFs, the emitter spot can be overlapped with the receiver spot. In this case, the far-field alignment has been achieved and the laser beams are now collinear.

In practice, this setup requires a suitable collimator and camera for the desired wavelength range, i.e., near-infrared. The *IK1523* from *EHD imaging* was used for this purpose. Further device information is listed in Tab. 3.5. The available collimator lens that was used for the far field alignment of the bi-static setup in this thesis, was originally designed for a nominal wavelength of 780 nm. However, it was evaluated that the chromatic aberrations at 1550 nm were only significant at the edge of the lens.

Table 3.5: Camera Hardware & Specifications [33]

Hardware	Type
Camera	EHD imaging IK1523
Parameter	**Specification**
Wavelength Range	900 nm to 1700 nm
Resolution	640 × 512px
Quantum Efficiency	$> 70\,\%$(1550 nm)
Gain Factor	27
Frame Rate	30 Hz



(a)

(b)

(c)

(d)

Figure 3.5: **(a)** Collimation technique: a low-power 1550 nm CW fiber laser is connected to the beam expander. The expander is aligned with respect to a shearing interferometer plate. The interferometer reflections are projected through a diffuser plate which is observed by an InGaAs infrared camera. The adjustment is performed in real time against the alignment guide of the interferometer. **(b)** Shearing interferometer pattern of a collimated beam. Note that the pattern is parallel to the alignment guide. **(c)** Shearing interferometer pattern of an uncollimated beam. The pattern is orientated at an angle to the alignment guide. **(d)** Alignment technique: the emission and reception ports of the bistatic setup are connected to a 1550 nm CW fiber laser via a 3 dB fiber splitter. The setup is roughly aligned with respect to a collimator-camera setup that is focused to infinity. The collimator collects parts of the emitter and receiver beams and projects them onto the infrared camera. An attenuation filter is used to protect the camera sensor. The spot image is then displayed on a PC. Overlapping, parallel spots indicate an aligned setup.

3.4 Software Overview

The main system control software named *Hyperion* was written in Python by Dr. Oliver Kliebisch and expanded upon within the scope of this thesis. It powers a graphical user interface (GUI) frontend based on the PyQt5 QML package and runs an object-orientated backend that handles the hardware interfacing and high-level data algorithms. In addition, an ATmega328-based Arduino Nano microcontroller unit (MCU) was programmed to manage a Bosch BNO055 inertial measurement unit (IMU) for spatial orientation data.

3.4.1 Hyperion

The Hyperion software runs on the Nvidia Jetson Nano and features a real-time visualization of the wind lidar data, GUI-based hardware control and data recording in the HDF5 format using the multithreading capabilities of the Jetson Nano. Hardware devices can be modularly added since the backend is based on a devicetree approach inspired by the Linux kernel. The serial communication is handled by the Python Serial package. Abstract classes exist to model generic hardware behavior [26].

Within this framework, additional modules for the laser source and MCU control were developed. Fig. 3.6 illustrates the laser control class hierarchy as an example of the modular approach. On program launch, a PEFLEOLA hardware and a QtPEFLEOLA UI object are created. The GUI object uses a QtTimer to periodically call the backend methods that are provided by the hardware object. On user GUI interactions, linked callback methods can be used to i.e., change the target current of the Booster amplifier. The user can fully control the laser parameters and read-out values from within the Hyperion software using the low-level interface provided by the laser manufacturer. The data is transferred via a RS232 connection to the PFL and daisy chained from there to the PFA (see Fig. 3.2).

3.4.2 MCU and IMU

Knowledge about the orientation of the lidar is important in order to correctly evaluate the LOS wind velocity component. Therefore, a microcontroller was programmed in C to supervise an inertial measurement unit. For the communication between Hyperion and the MCU a Universal Serial Bus (USB) to universal asynchronous receive-transmit (UART) converter is used. The microcontroller provides an interrupt service routine (ISR) handler that is i.a. triggered when the UART Receive register is overwritten. Single character commands are then utilized to poll IMU orientation data. The IMU provides several operational modes and is run in accelerometer and gyroscope mode. A timer ISR triggers the inter-integrated circuit (I2C) data readout from the IMU at regular intervals. In Hyperion, MCU and QtMCU classes exist in order to display the orientation data within the GUI. This allows the IMU to be used as a digit level device during system alignment.

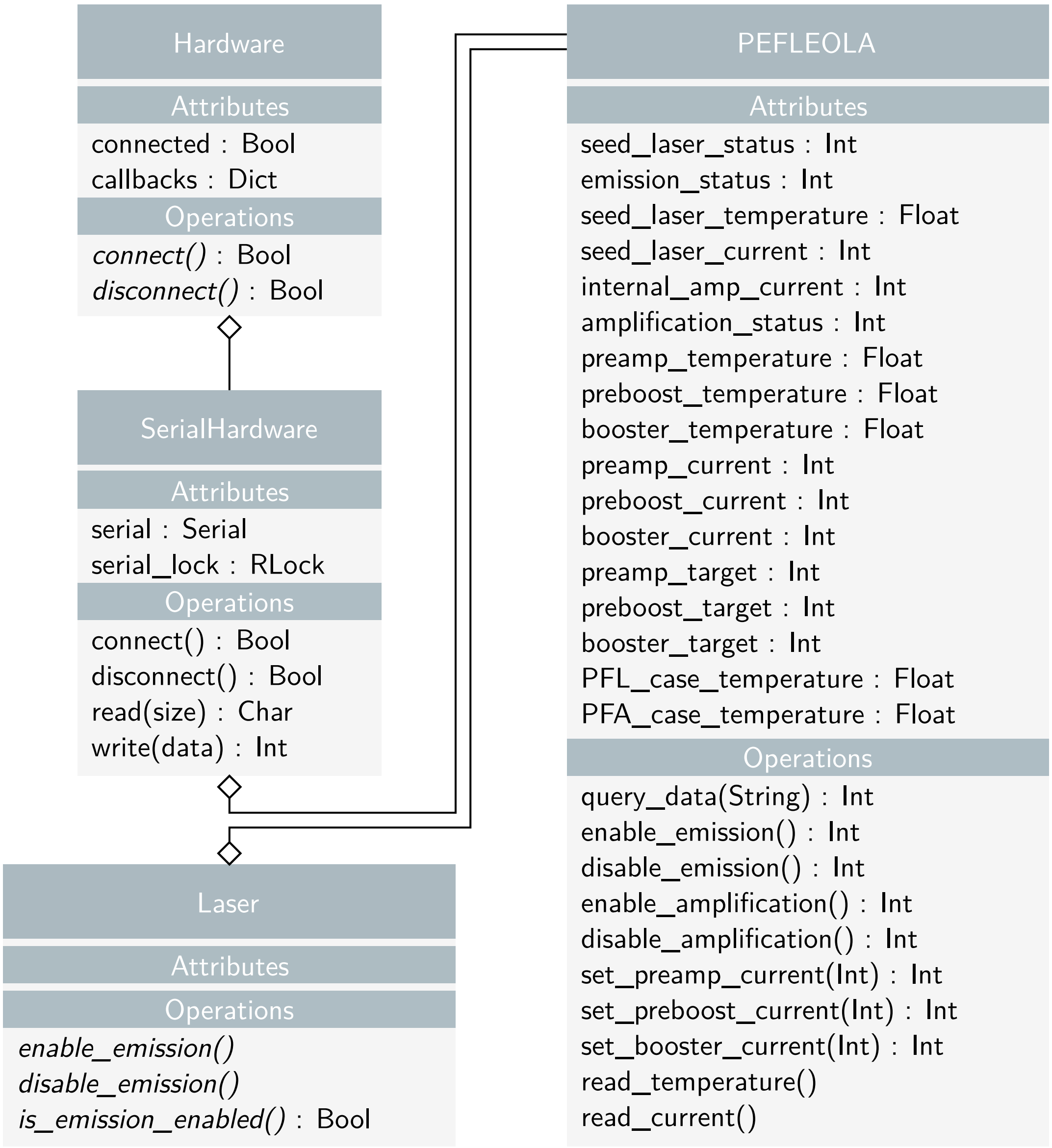


Figure 3.6: Unified Modeling Language (UML) diagram of the laser control backend: the Hardware super class implements the general attributes and methods of an abstract hardware controller. The SerialHardware parent class inherits from Hardware and implements serial communication methods. The Laser class defines methods for the overall control of multiple laser sources [26]. The PEFLEOLA class implements the Laser parent class stub methods and additional surveillance methods to interact with the PFL, PFA and the Hyperion GUI. Some methods have been omitted for clarity.

4 Data Evaluation Algorithms

The averaged raw spectra obtained during measurement can be represented as a tensor L_{ijk} where i is the number of frequency bins, j is the number of range gates and k is the number of spectra saved at time steps $\Delta t \approx 0.8192\,\mathrm{s}$. The data is saved in relation to the noise level as

$$\mathrm{SNR}_{\mathrm{dB}} = 10 \cdot \log_{10}\left(\frac{P_{\mathrm{Signal}}}{P_{\mathrm{Noise}}}\right), \tag{4.1}$$

where P_{Signal} is the signal strength and P_{Noise} is the strength of the background noise for every bin. The noise strength is determined at high altitudes where aerosols are sparse and therefore return signal is unlikely. This ratio is referred to as SNR. The data requires further processing steps before macroscopic flow field values can be extracted. The following section will describe the post-processing steps required for the correction of noise and drift artifacts and the peak detection required for the extraction of time series data. The latter can then be analyzed for the turbulent atmospheric variables ε and C_n^2. In addition, the error propagation will be discussed.

While this lidar system already achieves a significant data reduction due to a high level of preprocessing (see Ch. 3.1.3), the total file size for a field campaign of several hours can be in the order of 1 GiB to 2 GiB. For the purpose of fast post-processing and the long-term preprocessing implementation of the algorithms presented in this chapter, the focus was placed on efficiency and speed. In addition, a novel programming language was used. The Julia Programming Language has been developed by Bezanson et al. (2017) since 2012 and features C-like speed in a read-eval-print loop (REPL) environment [34].

4.1 Post-processing Steps

Due to the temperature-dependent voltage-amplification curve of an APD, significant drifts in the background signal strength can occur over the course of field campaigns lasting several hours. To ensure comparability of the individual spectra, the measurement series must be referenced against a common base level.

Furthermore, the two-dimensional lidar spectra contain irrelevant information and must be reduced into a one-dimensional vector for each time step, containing one relevant signal strength and velocity data point for each range bin. This process will be outlined in Ch. 4.1.2 and expanded upon in Ch. 5

4.1.1 Drift Correction

It is known that the occurrence of atmospheric aerosols approaches nil outside of the PBL and that the lidar sensitivity decrease is proportional to $1/r^2$. The return signal that is registered by the detector at extended ranges (i.e., $\geq 5000\,\mathrm{m}$) thus represents the detector noise itself and can be used to reference the spectrum as described above. However, in practice, clouds at high altitudes can lead to a significant return signal which would distort the noise level if the reference value was calculated at a fixed range interval. Therefore, an algorithm was developed in order to determine the background noise level via an adaptive window that avoids the inclusion of non-noise signals.

The following signal characteristic is exploited for this purpose: the noise level does not significantly change with height and thus a plateau of almost constant signal strength should be discernible. A suitable metric to identify this plateau is the standard deviation σ_N over an interval $f = \{f_k, ..., f_l\}$, since $\sigma_N \approx 0$ if $\mathrm{SNR}[f_i] \approx \mathrm{SNR}[f_{i+1}]$. Even if the rest of the data is not normally distributed, the approximation is correct for the relevant plateau interval. Using a sliding window approach, every frequency bin can be scanned over the range axis with a fixed window width N and σ_N can be estimated as

$$\sigma_N = \sqrt{\frac{1}{l-k-1} \cdot \sum_{i=k}^{l} \left(\mathrm{SNR}[f_i] - \overline{\mathrm{SNR}[f]}\right)^2} . \tag{4.2}$$

An implementation of this approach is provided below. If σ_N is multiplied by a bias function, the resulting curve is referred to as the weight function $\Phi_B = \sigma_N \cdot B(r)$ in the context of this work.

Listing 4.1: Plateau finding algorithm

```
function find_plateau(data, sw::Int, bias(i)=1/i)
        numEl = length(data)-2*sw
        # Start values
        index = 0
        value = 1e9
        # Sliding window loop
        for i = 1:numEl
                k = i + sw
                window = @view data[k-sw:k+sw]
                # Standard deviation combined with bias fct.
                temp = std(window) * bias(i)
                if temp < value
                        # Replace current index with better index
                        index = i
                        value = temp
                end
        end
        # Return index of minimum deviation in data and associated value
        return (index+sw, value)
end
```

The bias function approach was implemented in order to force the selection of the highest plateau if two or more plateaus exist within the weight function. For this, the inverse function $B(r) = \frac{1}{r}$ was chosen, where r denotes the range bin index. It was found that choosing the window with the minimum weight function $\min(\Phi_B)$ yields a good approximation for the location of the signal plateau. The above function (Lst. 4.1) uses an index value buffer in order to avoid storing a vector for Φ_B and the associated memory allocation. In addition, the parallelized application of this algorithm onto the lidar tensor L_{ijk} is trivial.

Fig. 4.1 illustrates the process for a concrete lidar spectrum. The data was taken from the July 27th, 2021 field campaign (index: 15500). The lidar measurements were taken at an elevation angle of 45°. Fig. 4.1a displays the signal course for a constant frequency of 80.078 MHz from the ground region up to the maximum measuring range. The signal strength varies strongly within the PBL that abruptly ends at a range of 2000 m due to a cloud blocking the LOS. A second cloud was detected at a frequency of 67.5 MHz and range of 6500 m, meaning that some light managed to pierce through the first cloud on both ways. The weight function for Fig. 4.1a is displayed in Fig. 4.1b. The function reaches its minimal value above 3000 m. This is due to the left- and right-sided window range of the sliding window approach. A window width of sw = 10 was chosen, meaning that the length of the interval is $2 \cdot \mathrm{sw} + 1 = 21$ range bins. Fig. 4.1c displays the drift corrected raw lidar spectrum. The drift offset value is calculated as the mean signal value over the plateau interval and subtracted equally from all data points in the raw lidar spectrum. This ensures comparability between all lidar spectra. For some purposes, it is useful to redefine the SNR from dB to the number of standard deviations from the noise level

$$\mathrm{SNR}_\sigma = \frac{10^{\mathrm{SNR}_{\mathrm{dB}} \cdot 0.1} - 1}{\sigma_N}. \tag{4.3}$$

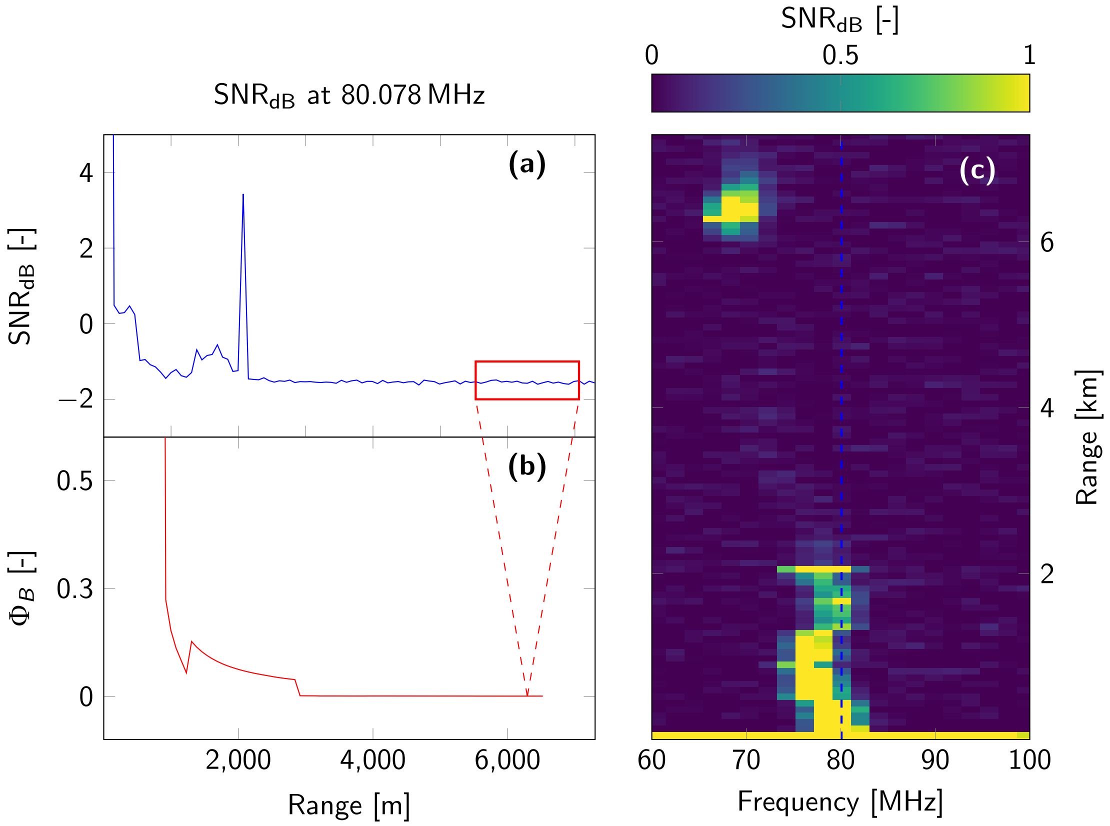


Figure 4.1: **(a)** SNR$_{\mathrm{dB}}$ as a function over range at 80.078 MHz: the signal strength in the first two range bins exceeds 40 dB due to internal reflections in the monostatic setup. A signal spike at a range of 2000 m is caused by cloud reflections. Beyond this range, no further signal is detected and the noise background signal can be assumed. **(b)** windowed weight function Φ_B over range for the data slice in **(a)**: in the ground region, the weight function is dominated by the fiber reflex and the cloud signal. The index and interval that were selected by the algorithm are marked in red in **(a)** and **(b)**. The mean offset value was determined to be −1.533 dB. **(c)** drift corrected wind lidar spectrum: the spectrum was corrected to a reference noise level of 0 dB. Aerosol signal can be distinguished up to a range of 2000 m. Within the ground region, a velocity profile can be recognized. A second cloud signal profile is visible at over 6000 m.

4.1.2 Peak Detection Techniques

Due to the likely presence of multiple scatterers within the measuring volume at any given time, the returned signal spectrum does not show a single distinct peak, but rather some distribution centered around the mean common aerosol velocity. This midpoint is assumed to be the central Doppler shift and therefore proportional to the flow field velocity at a given height. The occurrence of this distributive effect is further amplified by the temporal averaging that is performed during the preprocessing steps (Ch. 3.1.3). As a result, an algorithm is needed that can accurately obtain the central velocities from the lidar spectra.

In Fig. 4.2 the frequency distribution at several ranges has been extracted. The underlying spectrum (Fig. 4.2a) was taken from the July 27th, 2021 field campaign (index: 29300) and the system was angled at 45°. The signal distribution is plotted at a height of 383 m (Fig. 4.2b), 2686 m (Fig. 4.2c) and 6139 m (Fig. 4.2d). It can be seen that the distribution height and width changes drastically within the PBL. Additionally, the return signal of a cloud can be seen to be almost as strong as the signal within the ground region, whilst being almost outside of the maximal measuring range of the system. If a cloud signal were to be returned from outside of the maximum measuring range, it would reappear at the low end of the spectrum as an aliasing artifact. This can further complicate the determination of the mean velocity peak.

As a result, the algorithm must be able to distinguish even weak signals from background noise and be robust against aliasing effects. The peak detection methods investigated within the scope of this thesis are presented in detail in Ch. 5. For the time being, it is assumed that an algorithm exists that can determine the peak of some distribution $\mathrm{SNR}(f)$ (e.g. Fig. 4.2b) with sufficient accuracy. Moreover, it is assumed that $\mathrm{SNR}(f)$ always has a decreasing spectral density around a central mean value. Due to the discrete sampling of the DFT, $\mathrm{SNR}(f)$ may not be fully resolved, that is significant aliasing effects may occur. This effect is described in general by Heinzel et al. (2002) [35] and specifically for the case of lidar spectra by Rhone (2004) [36] and Engelmann (2009) [37].

The simplest approach to determine the central mean would simply be to determine the maximum value. This approach works well if there is only one global maximum which can easily be distinguished from noise. However, if the discretized distribution $\mathrm{SNR}[f]$ does not exactly contain the maximum of the continuous distribution $\mathrm{SNR}(f)$, the error amounts up to $\pm\frac{1}{2}$ of the frequency bin size δf. For the lidar system presented in this work the frequency estimation error Δf would be

$$\Delta f = \pm\frac{\delta f}{2} \approx \pm 0.977\,\mathrm{MHz} \tag{4.4}$$

which corresponds to $\pm 0.757\,\mathrm{m\,s^{-1}}$. For this reason, it is practical to perform an interpolation of the frequency distribution $\mathrm{SNR}[f]$ in order to minimize Δf [36].

While Rhone investigated first and second order interpolation methods, Engelmann proposed a center-of-gravity (COG) approach which is described next. Firstly, a temporary COG frequency is calculated from the distribution $\mathrm{SNR}[f]$ around an initial peak assumption at $\mathrm{SNR}[f_{\mathrm{Peak}}]$ with, i.e., $f_{\mathrm{Peak}} = \max[\mathrm{SNR}[f]]$, as

$$\overline{f} = \frac{\sum\limits_{i=-n}^{n} \mathrm{SNR}[f_i]\cdot f_i}{\sum\limits_{i=-n}^{n} \mathrm{SNR}[f_i]}\,. \tag{4.5}$$

An integer of $n = 4$ is proposed by Engelmann. The value of $\overline{f}$ can lie between the frequency bins of the original spectrum. Therefore, a linear interpolation of $\mathrm{SNR}[f]$ is performed and

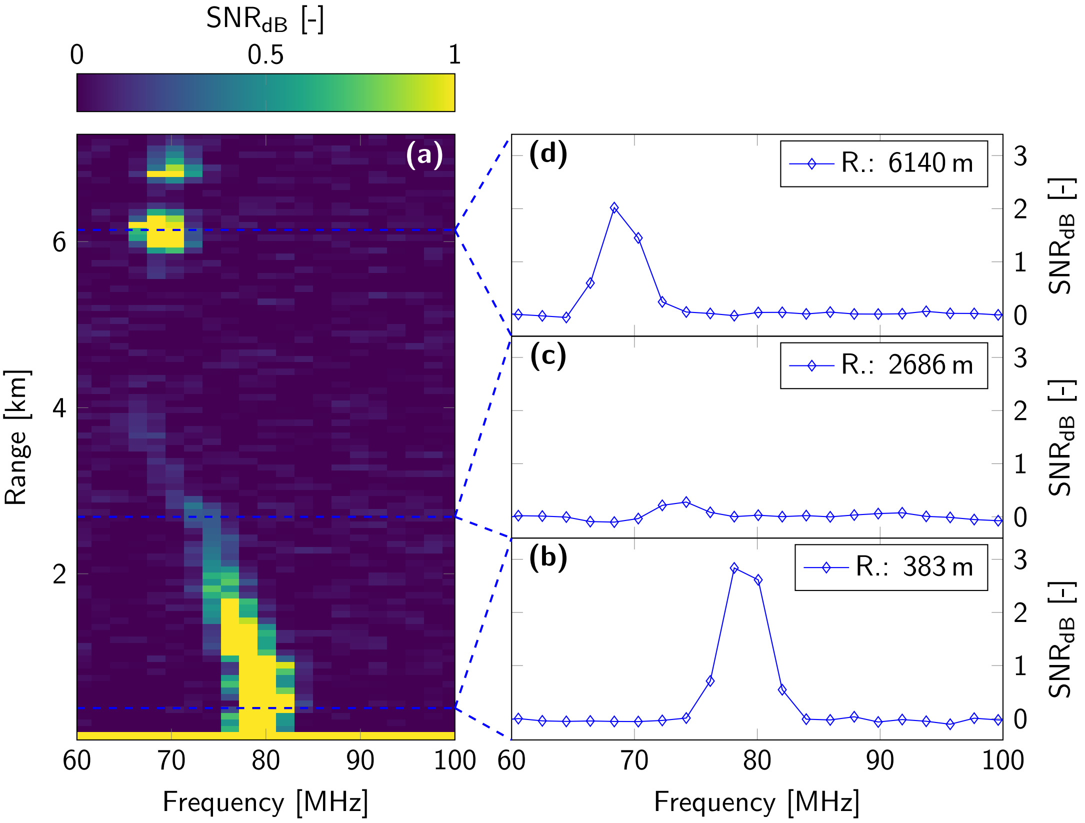


Figure 4.2: **a)** Wind lidar spectrum: ground layer aerosol signal can be seen clearly up to 3000 m with additional cloud signal at over 6000 m. The signal strength SNR$_{\mathrm{dB}}$ is plotted over frequency at three different heights. **b)** near-ground signal: data taken from the surface region shows a distinguishable peak distribution around ~ 78 MHz. Due to the discretization of the frequency bins, aliasing effects can be observed. **c)** boundary layer signal: a significant drop-off in the signal strength compared to the near-ground region is noticeable. However, determining the peak location at ~ 73 MHz is still possible. **d)** cloud signal: return signal becomes increasingly sparse above the PBL. In contrast, the scatter signal of a cloud can be seen to be of the same order of magnitude as the near-ground aerosol signal, albeit exhibiting the same aliasing effects.

the new data set $\mathrm{SNR}'[f']$ obtained. Then, Eq. 4.5 is applied again, this time onto the interpolated data with

$$\overline{f'} = \frac{\sum\limits_{i=-n}^{n} \mathrm{SNR}'[f'_i] \cdot f'_i}{\sum\limits_{i=-n}^{n} \mathrm{SNR}'[f'_i]} \,. \tag{4.6}$$

For this step, the author recommends a value of $n = 3$. In addition to improving the estimate of $f_{\mathrm{Peak}} = \overline{f'}$, the number of significant digits of the peak frequency is raised [37].

An implementation of the routine described above is given below (Lst. 4.2). The signal distribution must be passed in for every range gate of a spectrum. This method can take either predetermined values for the initial value of f_{Peak} or find the maximum within the data vector. It can be used to improve the accuracy of all generic peak detection algorithms while performing some checks to determine the plausibility of the output peaks.

The definition of Eq. 4.5 was changed to

$$f_{\mathrm{COG}} = \left\| \frac{\sum\limits_{i=-n}^{n} \mathrm{SNR}[f_i] \cdot f_i}{\sum\limits_{i=-n}^{n} \mathrm{SNR}[f_i]} \right\| \tag{4.7}$$

in order to avoid unpredictable behavior when the input vector contains $\mathrm{SNR}_{\mathrm{dB}}[f_i] \lesssim 0$, a circumstance that can occur due to floating point rounding errors.

Listing 4.2: COG correction algorithm

```
function COG_correction(SNR_data, n::Int=4, f_cog=argmax(SNR_data))
    # Test if first estimate distr. is out of bounds
    if f_cog < n+1 || f_cog > length(SNR_data)-n
        return (NaN, NaN) # index, value
    end
    f_dat = f_cog-n:1:f_cog+n
    # Calculate new COG
    f_cog_n = abs(sum(SNR_data[f_dat] .* f_dat) / sum(SNR_data[f_dat]))
    # Resample COG over interpolated SNR data vector
    d_itp = LinearInterpolation(1:length(SNR_data), SNR_data, extrapolation_bc=
        ↪ Line())
    f_dat_n = (f_cog_n-n):1:(f_cog_n+n)
    d_itp_n = abs.(d_itp(f_dat_n))
    f_cog_n = abs(sum(d_itp_n .* f_dat_n) / sum(d_itp_n))
    # Check if f_cog_n is out of bounds due to errors
    if f_cog_n < 1 || f_cog_n > length(SNR_data)
        return (NaN, NaN) # index, value
    else
        return (f_cog_n, d_itp(f_cog_n)) # index, value
    end
end
```

Once the peak locations f_{Peak} have been determined, the lidar tensor L_{ijk} can be reduced into two matrices for the return signal and wind velocity time series. These can then be used to determine the aerosol distribution and turbulent properties of the PBL respectively. The LOS wind velocity can be calculated using the relation given for the optical Doppler shift (Eq. 2.10):

$$v_{\mathrm{LOS}} = -\frac{(f_{\mathrm{Peak}} - 80\,\mathrm{MHz}) \cdot \lambda}{2}\,, \tag{4.8}$$

where λ is the laser wavelength. v_{LOS} is stored for each range bin and time step. The return signal $\mathrm{SNR}_{\mathrm{max}}$ is simply extracted at $\mathrm{SNR}_{\mathrm{max}} = \mathrm{SNR}_{\mathrm{dB}}[f_{\mathrm{Peak}}]$ using the interpolated data from the COG correction algorithm. As a result of the operations explained above, the amount of data is further reduced to 1/64 of the original tensor per matrix.

Fig. 4.3 showcases the resulting time series matrices using data from the October 7th, 2021 campaign. A PIN diode was fielded as the detection device. The resulting signal time series can be seen in Fig. 4.3a. A noticeable change in the signal intensity can be seen between 11:00 a.m. to 4:00 p.m., which is proportional to the aerosol distribution. The LOS wind velocity is plotted in Fig. 4.3b. A gradual increase of the velocity throughout the day is apparent. In addition, turbulent structures, i.e., inhomogeneous changes over height and time, are noticeable in the ground region.

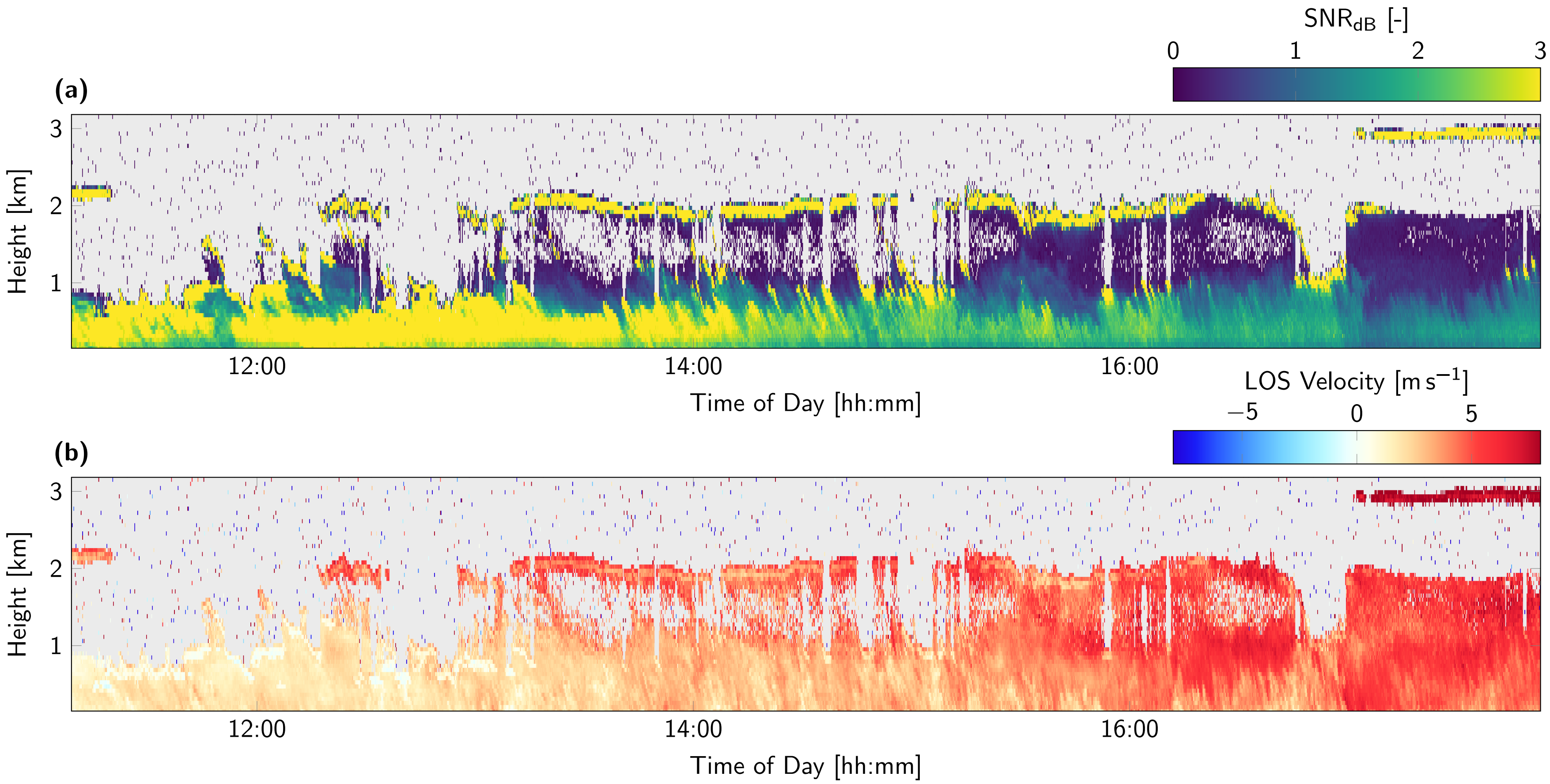


Figure 4.3: Lidar time series: visual representation of the reduced lidar data from the October 7th, 2021 field campaign in Stuttgart-Vaihingen. The y-axis was corrected to account for the 45° inclination of the system and is referenced to height above ground. A threshold algorithm was used for peak detection, random dots are false-positive detections. **(a)** SNR$_{\mathrm{max}}$ time series: displays the signal strength at f_{Peak} and is proportional to the amount of aerosols in the atmosphere. Significant changes in the aerosol distribution can be observed throughout the day. **(b)** LOS wind velocity time series: displays the wind velocity calculated from the optical Doppler shift at f_{Peak}. Turbulent structures can be identified in the near-ground region.

4.2 Wind Velocity Spectrum Analysis

Methods for the estimation of turbulent parameters from CDWL radial velocity data are explained in detail by Banakh et al. (2013). One of the proposed methods intends to use the broadening of the Doppler spectrum (i.e., the width of the distribution in Fig. 4.2b) as a means to analyze the scale of the TDR ε. However, if the radial velocity variations over height become large in relation to the measuring volume, a broadening of the spectrum can occur that is not caused by turbulence alone. Quantifying the contributions of the broadening mechanisms is nontrivial [11, p. 132].

Therefore, the approach of Smalikho (1997) is used within the context of this work. Rather than determining a value for ε for each time step, a "mean" value for the TDR is calculated over a time interval of sufficient length. The estimation quality of ε should improve with the length of the chosen interval. While the TKE k depends on the exact magnitude of the fluctuating spatial velocity components u, v and w, measuring only one component of the overall velocity vector is sufficient to determine ε. This component is referred to as the radial velocity by Banakh and corresponds to the LOS velocity measured by a lidar system. It is known that the vertical velocity fluctuations in the PBL are negligible compared to the horizontal changes, therefore the radial velocity should be measured ideally at an inclination of 0° respective to the horizon at a given height. Nevertheless, information about ε can be obtained even from inclined systems, i.e., 45°, as long as some information about the horizontal velocity component is contained in the radial velocity vector [11, 13].

In Fig. 4.4a an extract from the time series data of Fig. 4.3 can be seen. The velocity variations are extracted at a height of 488 m and plotted from 2:00 p.m. to 3:00 p.m. in Fig. 4.4b. The velocity interval data is processed using a Hampel identifier [38] to remove outliers and a moving-mean filter for noise reduction for the time intervals from 2:15 p.m to 2:30 p.m. and 2:45 p.m. to 2:50 p.m. respectively. A subsequent calculation of the velocity density spectrum is performed according to Eq. 2.24 and plotted in Fig. 4.4c and d. The spectral data was fitted to a power-law function of the form

$$\hat{S}(f) = c \cdot f^k \,, \tag{4.9}$$

or $\hat{S}_{\mathrm{Log}}(f) = \log c + f \cdot k$ on a logarithmic scale, using a least-squares method. The resulting fit parameters and their 95 % confidence intervals are displayed in Tab. 4.1.

Table 4.1: Power-law fit parameters

Figure	c [m^2 s^{-1}]	$c_{95\%}$ [m^2 s^{-1}]	k [-]	$k_{95\%}$ [-]
(c)	1.976×10^{-5}	$[1.635, 2.318] \times 10^{-5}$	-2.291	$[-2.390, -2.192]$
(d)	5.337×10^{-5}	$[3.883, 6.791] \times 10^{-5}$	-1.939	$[-2.093, -1.785]$

It is apparent that the $-5/3$ slope of the Kolmogorov spectrum is only roughly met. In addition, fluctuations in the spectrum are noticeable especially with increasing frequency. This is due to the effect of statistically dependent effects as described by Smalikho and leads to the fact that ε can not be determined accurately by fitting the Kolmogorov law to the energy spectrum that is proportional to the velocity density spectrum [13].

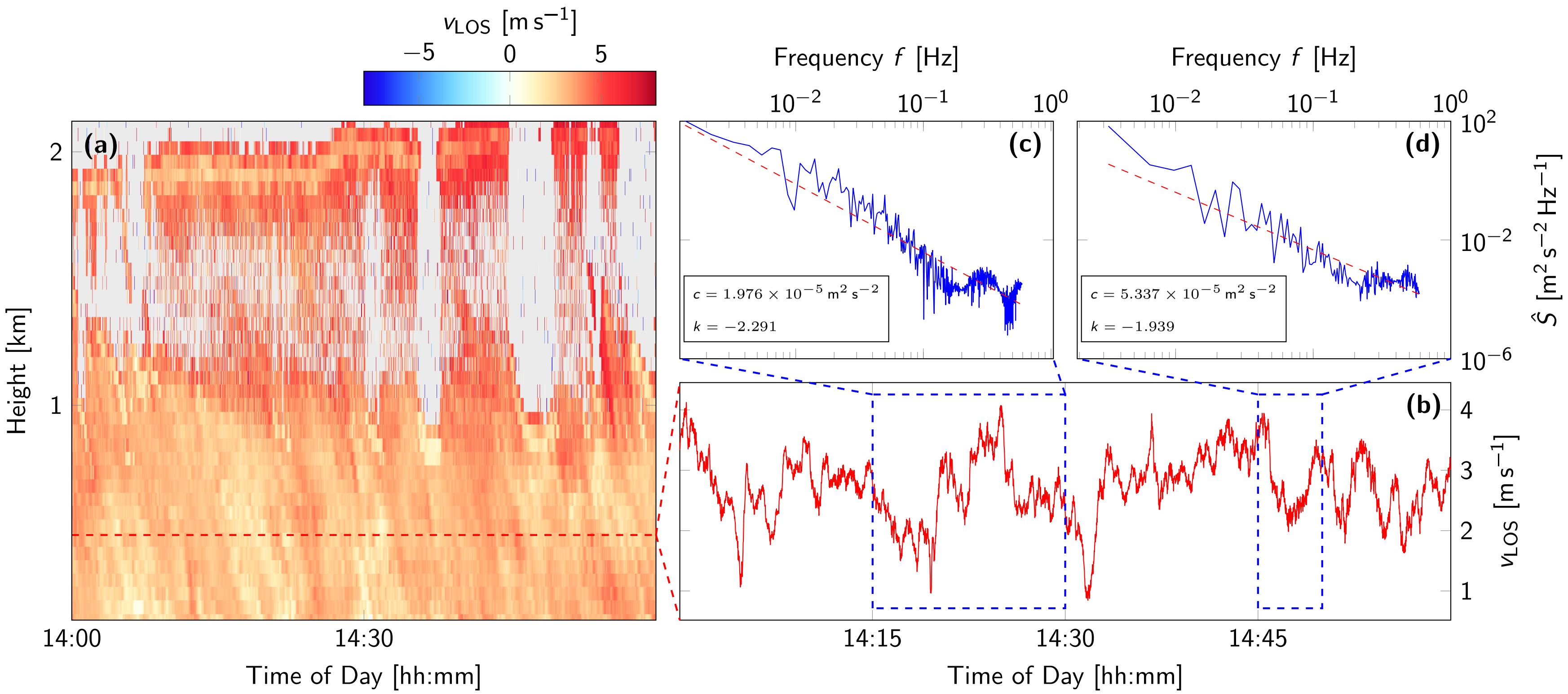


Figure 4.4: **(a)** v_{LOS} time series taken from Fig. 4.3b between 2:00 p.m. and 3:00 p.m. **(b)** Magnitude of v_{LOS} over time at a height of 488 m. Strong gradients due to dependent events (i.e., gusts) can be seen in the velocity course. Two time windows were evaluated for their density spectrum. **(c)**+**(d)** Velocity density spectrum: the spectra were calculated according to Eq. 2.24. Outliers in the velocity interval were removed using a Hampel identifier [38] and the data was smoothed with a moving-mean filter. The quality of the fit depends strongly on the analyzed time interval and the weather conditions. If events occur within that time window which are of significant length and fall out of the energy spectrum assumed by the Kolmogorov law, the agreement between the spectrum and the power-law fit deteriorates. Spectrum **(c)** was calculated over a time period of 15 min. A subsequent least-square power-law fit reveals a rough approximation of the energy cascade. **(d)** illustrates the spectrum over an interval of 5 min.

Hence, a procedure is used where segments of the velocity density spectrum $\hat{S}(f)$ are evaluated that can be assumed to be statistically independent and more representative for the estimation of ε. While Smalikho proposes to find such a segment before applying the correlation function $Q(f)$, a different approach was used for this work. Rather than sampling $I \in [0, M-1]$ such that $\overline{\hat{S}} = \hat{S}_i$, the entire density spectrum $\hat{S}$ is used to obtain the function $\varepsilon = \varepsilon(f)$ as per Eq. 2.26. An example for the resulting function based on experimental data is given in Fig. 4.5a. This function can then be used to find an interval of near constant ε in order to estimate a statistically independent value for the TDR.

4.2.1 Turbulent kinetic energy dissipation estimation

In practice however, the existence of an interval I with $\overline{\varepsilon} = \varepsilon_i$ is not guaranteed or its size over f is small. For the sake of argument, assume that one such interval A exists in $\varepsilon(f)$. Thus the goal is to find the data points in $\varepsilon(f = I)$ such that both the statistical requirements for I are met and as many data points as possible are detected to minimize the error predicted by Eq. 2.29. An example data set is shown in Fig. 4.5a. Note that it is hard to define a suitable range even by hand. The likelihood of meeting the above criteria increases as the underlying time interval, that the density spectrum is derived from, is expanded. As a consequence this leads to the fact that the determined value for the TDR corresponds to a long-term temporal average and its informative value for short-term events decreases.

The following approach was implemented in order to find an optimal estimate $\hat{\varepsilon}$ in data where the above criteria are barely met (i.e., $\overline{\varepsilon} \approx \varepsilon_i$): the spectrum of the TDR is analyzed at increasing length scales using an algorithm akin to the plateau finding algorithm presented in Ch. 4.1.1. Hereby, an interval center and width are selected such that a predefined error function is minimized.

The lower limit for the search window width is 10 bins in order to satisfy the mean normal error $B(n = 10) \approx 1$ from Eq. 2.28. Then, the width is increased successively. For each length scale, a metric over ε is calculated for all frequencies that are within the array bounds. The relative standard deviation Σ was selected as a suitable metric, since it is expected that an ideal interval with $\overline{\varepsilon} = \varepsilon_i$ has $\Sigma = 0$. It is defined by

$$\Sigma = \frac{\sigma_\varepsilon}{\overline{\varepsilon}} = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (\varepsilon_i - \overline{\varepsilon})^2} \cdot \left(\frac{1}{N} \sum_{i=1}^{N} \varepsilon_i \right)^{-1} . \tag{4.10}$$

The code for this function is given in Lst. 4.3. By choosing the relative rather than the absolute deviation, the metric is not biased towards small ε, because the deviation can scale with the order of magnitude of $\overline{\varepsilon}$.

Listing 4.3: Sliding window metric

```
function pmetric(data, sw; metric::Function=wd->std(wd)/mean(wd))
    numEl = length(data)-2*sw
    result = zeros(numEl)
    for i = 1:numEl
        k = i + sw
        window = @view data[k-sw:k+sw]
        # Calculate local value based on metric function.
        result[i] = metric(window)
    end
    return result
end
```

However, the consideration has to be made that the validity of this approach is, strictly speaking, defined such that $\{f \mid \overline{\varepsilon(f)} = \varepsilon(f)\}$. As mentioned before, this condition can not be fulfilled in practice. Therefore, not only does the total estimation error need to be taken into account, but also the violation of the above condition. The total estimation error $A(n)$ for $n > 10$ is defined as per Eq. 2.30 as

$$A(n) = \left[E^2 + (\underbrace{B-1}_{\approx 0})^2\right]^{\frac{1}{2}} \tag{4.11}$$

$$\approx \left[\left(1+\frac{1}{n}\right)\cdot\left(1+\frac{2}{n}\right) - (\underbrace{B(n)}_{\approx 1})^2\right]^{\frac{1}{2}} \quad \text{for } n > 10 \tag{4.12}$$

$$\approx \sqrt{\frac{2}{n^2}+\frac{3}{n}}\,. \tag{4.13}$$

It is obvious that this error term is minimized for increasing n. Combining the terms for the deviation Σ and the total estimation error A into a single error function Φ_ε is nontrivial, since they are both dependent on the interval length. The following functional relationships were considered for Φ_ε:

1. $\Phi_\varepsilon(n) = \max(\Sigma\,,\,A)$
2. $\Phi_\varepsilon(n) = \Sigma + A$
3. $\Phi_\varepsilon(n) = \sqrt{\Sigma^2 + A^2}$

In the first case, by selecting the greater of the errors, both the deviation by violating the plateau condition $\overline{\varepsilon} = \varepsilon_i$ and the inherent deviation caused by the interval length are considered. In the second case, both relative deviations are added to account for the worst case and in the third case, the root mean square of both deviations is formed, which assumes that both errors are uncorrelated. Moving forward, the first definition is used, as it is expected that the second definition overestimates the error and the third equation is not applicable, since it is not known if the the deviations are independent.

For the actual implementation, the error function is imposed by considering only the results of the metric for which $\Sigma \leq A(n)$ holds. By using this cutoff limit, it is ensured that the contributions of the deviation terms are at most equal, or the total estimation error outweighs, which leads to

$$\Phi_\varepsilon(n) = \max(\Sigma\,,\,A) = A(n) \text{ for } \Sigma \leq A\,. \tag{4.14}$$

This in turn allows the negligence of all sections of the ε-spectrum for which the above condition is not fulfilled. This can be easily implemented as an algorithm by summing over the center indices of all intervals that satisfy $\Sigma \leq A$. For each frequency bin, the total length scale count corresponds with its suitability as an estimate for the TDR. In addition, this approach allows for the existence of multiple local plateaus, where the global maximum is the best estimate as long as there are no duplicate counts. Note that this method strongly depends on the error function choice made for Φ_ε. An implementation of this algorithm is given in Lst. 4.4. The center index is determined by applying the previously mentioned condition over all length scales and finding the index with the most counts.

Listing 4.4: $\hat{\varepsilon}$ estimation algorithm

```
function epsilon_estimator(epsilon)
    numEl = length(epsilon)
    # Length scale counter
    lsctr = zeros(Int64, numEl)
    # Iterate over increasing length scale ls=2*sw+1
    for sw = 5:Int(floor(length(epsilon)/2-1))
        temp = fill(NaN, numEl)
        temp[sw+1:end-sw] = pmetric(epsilon, sw)
        # Increase length scale counter if metric <= A(sw)
        lsctr[findall(temp .<= TotlEstmError(2*sw+1))] .+= 1
    end
    # Find index with largest ctr value
    cindx = findmax(lsctr)[2]
    sw = 0
    # Iterate interval width until metric > A(sw)
    while true
        sw += 1
        # Bounds check
        if cindx-sw <= 1 || cindx+sw >= numEl
            break
        end
        buffer = epsilon[cindx-sw:cindx+sw]
        # Error check
        if std(buffer)/mean(buffer) >= TotlEstmError(2*sw+1)
            break
        end
    end
    return (cindx, sw)
end
```

After the center index has been found, the interval width is also determined by gradually increasing the interval width around the center index and calculating the relative standard deviation. As long as $\Sigma \leq A(n)$ holds, the width is increased. Finally, the results are returned. In Fig. 4.5a the TDR spectrum for the data presented in Fig. 4.4c is displayed. Here, the result of $\varepsilon(f)$ varies over several orders of magnitude ($1 \times 10^{-8}\,\mathrm{m^2\,s^{-3}}$ to $1 \times 10^{-3}\,\mathrm{m^2\,s^{-3}}$), but intervals of almost constant ε can be discerned.

In Fig. 4.5b the length scale count metric shows three significant maxima. The associated "plateaus" have been marked in Fig. 4.5a and the respective values for $\hat{\varepsilon}$ and its deviation are given in Tab. 4.2 together with the relative total estimation error. It can be seen that the highest count of the length scale corresponds to the smallest relative deviation, since the interval width is the largest while maintaining $\bar{\varepsilon} \approx \varepsilon_i$. An enlarged view of the relevant data area is given in Fig. 4.5c, together with an indicator of the mean value $\hat{\varepsilon}$ and 95 % confidence intervals. Fig. 4.5d shows that the data in Fig. 4.5c should start to approach a normal distribution.

Table 4.2: TDR estimation results, confidence intervals and total est. error

Interval	$\hat{\varepsilon}$ [$\mathrm{m^2\,s^{-3}}$]	$\hat{\varepsilon}_{95\%}$ [$\mathrm{m^2\,s^{-3}}$]	$A(n)$ [-]
1.	3.610×10^{-7}	$\pm 2.118 \times 10^{-7}$	2.934×10^{-1}
2.	1.782×10^{-6}	$\pm 9.432 \times 10^{-7}$	2.646×10^{-1}
3.	9.080×10^{-6}	$\pm 4.412 \times 10^{-6}$	2.429×10^{-1}

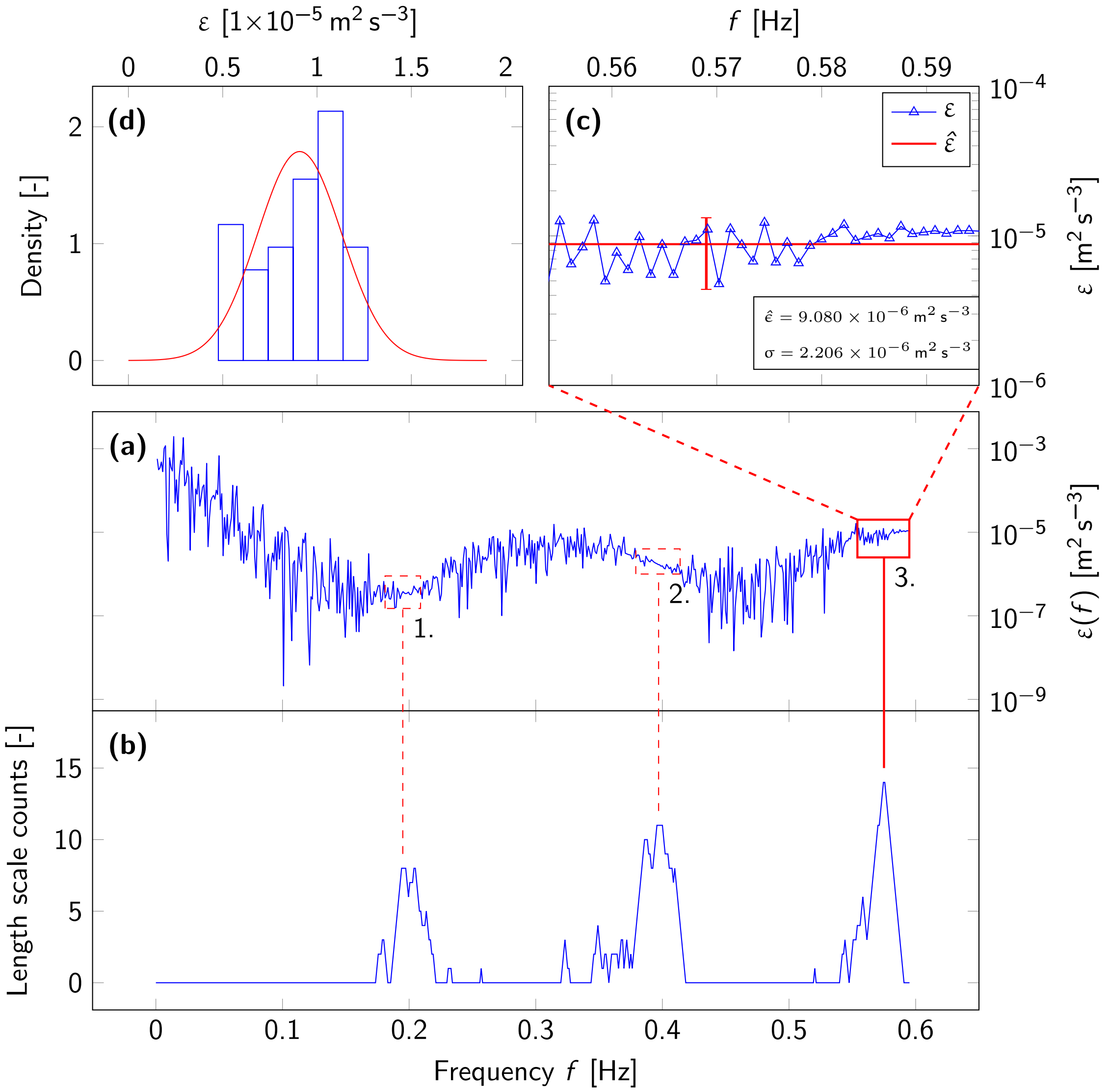


Figure 4.5: **(a)** TDR spectrum of the velocity density spectrum presented in Fig. 4.4c. ε varies heavily over the frequency f. However, intervals of almost constant ε can be seen. **(b)** Visualization of the length scale count metric presented in Lst. 4.4 with three distinguishable local maxima. The corresponding intervals have been marked in **(a)**. Note that the algorithm selects the frequency with the most counts as the center index for the plateau, which corresponds to the lowest total estimation error. The selected plateau is displayed in **(c)**: the mean TDR $\hat{\varepsilon}$ and a 95 % confidence interval resulting from Eq. 2.30 with $n = 39$ are denoted. **(d)** Distribution of ε in **(c)**. A normal distribution with $\mu = 9.080 \times 10^{-6}\,\mathrm{m^2\,s^{-3}}$ and $\sigma = 2.232 \times 10^{-6}\,\mathrm{m^2\,s^{-3}}$ has been fitted and shows that for the selected interval the transition from an exponential to a normal distribution can be assumed.

4.2.2 Turbulent index of refraction estimation

Once a value for the TDR at a given range has been estimated, calculating the turbulent index of refraction C_n^2 is straight forward (ref. Ch. 2.3.3). The gust velocity introduced in Eq. 2.32 is used to estimate the fluctuating quantity from Eq. 2.31 while assuming values for the Brunt-Väisälä frequency N and the longitudinal length scale L_i. For the refractive index of air n at a given height the local pressure and temperature need to be known, since

$$n(h) = 1 + \eta \cdot \frac{p(h)}{T(h)}\,, \tag{4.15}$$

where $\eta = 7.760 \times 10^{-7}\,\mathrm{K\,Pa^{-1}}$. For this purpose, the equations proposed by the ISA model are used. The pressure at height h can be assumed to be

$$p(h) = p_0 \cdot \left(1 - \frac{n-1}{n} \cdot \frac{g}{RT_0} \cdot (h - h_0)\right)^{\frac{n}{n-1}} \tag{4.16}$$

and the temperature T can be calculated with

$$T(h) = T_0 - \frac{n-1}{n} \cdot \frac{g}{R} \cdot (h - h_0)\,, \tag{4.17}$$

where $n = 1.235$ is the polytropic index, $g = 9.81\,\mathrm{m\,s^{-2}}$ is the gravitational acceleration at sea level, $R = 287.1\,\mathrm{J\,kg^{-1}\,K^{-1}}$ is the specific gas constant for air, $T_0 = 288.2\,\mathrm{K}$ is the reference temperature at sea level, $p_0 = 101.3\,\mathrm{kPa}$ is the reference pressure at sea level and $h_0 = 0\,\mathrm{m}$ is the sea level reference height. The full equation for C_n^2 is

$$\begin{aligned} C_n^2 &= L_i^{-\frac{2}{3}} \cdot (n-1)^2 \cdot FQ^2 && (4.18)\\ &= L_i^{-\frac{2}{3}} \cdot \eta^2 \cdot \frac{p^2}{T^2} \cdot FQ^2 && (4.19)\\ &= L_i^{-\frac{2}{3}} \cdot \eta^2 \cdot \frac{p^2}{T^2} \cdot \frac{N^2}{g^2} \cdot \frac{(\hat{\varepsilon} \cdot L_i)^{2/3}}{c_1} && (4.20)\\ &= \frac{1}{c_1} \cdot \left(\frac{\eta p N \sqrt[3]{\hat{\varepsilon}}}{Tg}\right)^2 && (4.21) \end{aligned}$$

with $c_1 \approx 0.78$. Note that the length scale L_i is canceled from the equation and does not need to be considered for C_n^2, which for the purpose of this work, depends mainly on the TKE dissipation rate. For $N = N(T(h))$ it is known from Eq. 2.34 and Eq. 4.17 that

$$\frac{\partial T}{\partial h} = -\frac{n-1}{n} \cdot \frac{g}{R} \approx -6.504 \times 10^{-3}\,\mathrm{K\,m^{-1}} \tag{4.22}$$

and thus

$$N(h) = \frac{c_2}{\sqrt{T(h)}} \tag{4.23}$$

with

$$c_2 = \sqrt{g \cdot \left(\frac{\partial T}{\partial h} - \Gamma_d\right)} \approx 5.729\,\mathrm{K^{1/2}s^{-1}}\,. \tag{4.24}$$

4.3 Error Estimation

The errors that need to be expected when reading out the LOS velocity from the lidar spectra are described in the chapter of the respective detection methods (Ch. 5).

4.3.1 Center of gravity correction error

As has been described in Ch. 4.1.2, when no further corrections are performed, the frequency estimation error would be ±0.977 MHz. However, the COG correction approach can also introduce errors.

For a general weighted sums approach to determine the weighted mean, $\overline{x}_w$ is

$$\overline{x}_w = \frac{\sum\limits_{i=1}^{n} w_i \cdot x_i}{\sum\limits_{i=1}^{n} w_i} . \tag{4.25}$$

The following relations can be used in order to calculate the variance

$$\mathrm{var}(\overline{x}_w) = \frac{\sum\limits_{i=1}^{n} w_i \cdot (x_i - \overline{x}_w)^2}{\sum\limits_{i=1}^{n} w_i} \cdot \frac{n}{n-1} \tag{4.26}$$

and the standard deviation σ for $\overline{x}_w$

$$\sigma = \sqrt{\frac{\mathrm{var}(\overline{x}_w)}{n}} , \tag{4.27}$$

assuming that the random error is normally distributed. However, if the weight terms w_i are distributed such that few terms heavily dominate the mean, a correction must be made in order to account for that bias [39, pp. 336]. The effective number of measurements then is defined as

$$n_{\text{eff}} = \frac{\left(\sum\limits_{i=1}^{n} w_i\right)^2}{\sum\limits_{i=1}^{n} w_i^2} . \tag{4.28}$$

In our case, since the frequency bins are equally spaced, the effective number of measurements will always be the actual number of measurements, that is $n \approx n_{\text{eff}}$ and thus Eq. 4.26 is valid without further correction. There remains some uncertainty as to how the deviations of multiple passes of the COG need to be combined. It is assumed that the last stage of weighting dominates the error and thus

$$\sigma_{\text{COG}} = \sqrt{\frac{1}{n-1} \cdot \frac{\sum\limits_{i=-n}^{n} \mathrm{SNR}'[f'] \cdot (f_i' - \overline{f'})^2}{\sum\limits_{i=-n}^{n} \mathrm{SNR}'[f']}} . \tag{4.29}$$

4.3.2 Dissipation rate estimation error

The deviation that results for the estimate of the TDR $\hat{\varepsilon}$ is detailed in Ch. 4.2.1 and, in essence, is directly connected to the interval width n that is used to calculate the mean of

$\varepsilon(f)$ as per Eq. 2.30. Henceforth, this error is assumed to be primary for $\hat{\varepsilon}$ and therefore errors that take place before this step are not considered further.

4.3.3 Turbulent refractive index estimation error

From Eq. 4.21 it is known that $C_n^2 = C_n^2(\eta, p, N, \hat{\varepsilon}, T, g)$. For simplicity it is assumed that the pressure p and temperature T calculated from the ISA equations are accurate and their errors negligible. The same is assumed for g and the index coefficient η. For this reason the main focus is on the behavior of $C_n^2(N, \hat{\varepsilon}) \propto N^2 \cdot \hat{\varepsilon}^{2/3}$ for large errors of the TDR estimate.

In Fig. 4.6 the equation for C_n^2 (ref. Eq. 4.21) has been rewritten so that only the dependent variables remain on the right hand side. This function is plotted over $\hat{\varepsilon}$ for different values of N. It can easily be seen that linearity can be assumed as long as the relative deviation Σ of $\hat{\varepsilon}$ is not greater than roughly 40 %. Since $C_n^2 \propto N^2$ one can further assume linearity for small deviations of N.

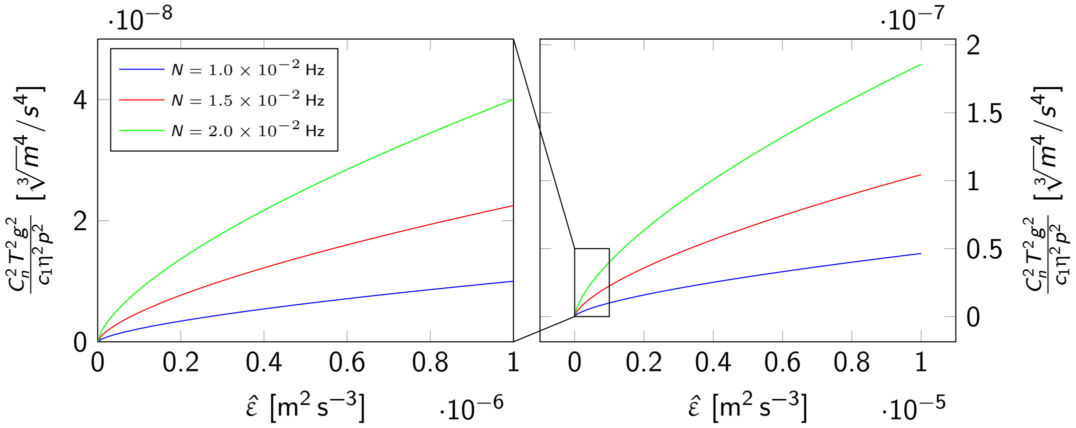


Figure 4.6: Behavior of $C_n^2(N, \hat{\varepsilon})$ plotted for different values of N: it can be seen that as long as the relative deviation Σ of $\hat{\varepsilon}$ remains reasonable ($\sim \pm 40\,\%$), the deviation of the result remains approximately linear. This is also true for small changes of N.

For this reason the propagation of uncertainty onto C_n^2 is assumed to be linear overall and the truncated Taylor series, with linear terms only, is used to derive $\sigma_{C_n^2}$. It is further assumed that N and $\hat{\varepsilon}$ are uncorrelated, however no proof can be given at this point. As per the common equation for the uncorrelated propagation of deviations, the following relation is obtained:

$$\sigma_{C_n^2} = \sqrt{\left(\frac{\partial C_n^2}{\partial N} \cdot \sigma_N\right)^2 + \left(\frac{\partial C_n^2}{\partial \hat{\varepsilon}} \cdot \sigma_{\hat{\varepsilon}}\right)^2}\,, \tag{4.30}$$

where σ_N and $\sigma_{\hat{\varepsilon}}$ are the respective deviations of N and $\hat{\varepsilon}$. The following derivatives are calculated for

$$\frac{\partial C_n^2}{\partial N} = \frac{2}{c_1}\left(\frac{\eta p}{Tg}\right)^2 N \sqrt[3]{\hat{\varepsilon}}^2\,, \tag{4.31}$$

$$\frac{\partial C_n^2}{\partial \hat{\varepsilon}} = \frac{2}{3c_1}\left(\frac{\eta p}{Tg}\right)^2 \frac{N^2}{\sqrt[3]{\hat{\varepsilon}}} \tag{4.32}$$

and accordingly

$$\sigma_{C_n^2} = \frac{2}{c_1}\left(\frac{\eta p}{Tg}\right)^2 \sqrt{\left(N \cdot \hat{\varepsilon}^{\frac{2}{3}} \cdot \sigma_N\right)^2 + \frac{1}{9}\left(N^2 \cdot \hat{\varepsilon}^{-\frac{1}{3}} \cdot \sigma_{\hat{\varepsilon}}\right)^2}\,. \tag{4.33}$$

5 Peak Detection Techniques

An accurate identification of the center offset frequency within the lidar spectrum is essential in order to correctly analyze the LOS wind velocity for each time step. Ideally, such an algorithm will perform well even within low SNR environments and will have a low false-positive detection rate. Correctly identifying spectral peaks can become very challenging once the signal strength starts to approach the background noise strength. Spectral peak analysis is a topic of interest in other research fields as well, i.e., mass spectrometry, audio analysis and radar surveillance [40–42].

Some algorithms that try to approach this problem are, for instance, continuous wavelet transform-based pattern matching [43], the National Oceanic and Atmospheric Administration (NOAA) moment velocity estimator [42], discrete spectral peak estimation [44] and a variety of other algorithms that aim to utilize characteristic properties of the information at and around the peak value [40].

The continuous wavelet transform-based pattern matching algorithm, for instance, tries to solve the issue of false-positive detection by not only analyzing amplitude information, but also the peak shape. This is accomplished by performing a change of basis into wavelet space.

The NOAA moment velocity estimator and the discrete spectral peak estimation algorithm use a fitting approach in order to obtain additional information from the frequency distribution around the maximum amplitude where several statistical moments of increasing order are evaluated. While fitting algorithms can deliver good results for the peak frequency even within aliased data they can be computationally expensive and can become inaccurate for very broad distributions.

The common denominator of the methods presented here is that they each consider only a single spectrum, whilst a wind lidar spectrum is made up of several spectra that are distinguished by range, but are correlated due to the governing equations of fluid motion. Therefore, within this chapter three methods of increasing capability and complexity will be presented, where each one tries to tackle the presented issue from a different point of view.

The Gaussian Threshold algorithm is a simple, fast, low-allocation maximum estimator featuring an improved false-positive detection rate due to an adaptive activation threshold. It represents the simplest implementation of a peak estimator and requires the least amount of computational resources.

The Scalogram Weights algorithm is a biased maximum estimator based on the concept of variable search length scales, as proposed by Scholkmann et al. (2012) for periodic signals [41]. This method aims to improve the estimator performance in low-SNR environments by utilizing the information from the previous range gate.

Finally, spectrum estimation through the use of a neural network will be presented. Novel in this type of application is the implementation of context-sensitive data evaluation. Hereby, the performance is improved by considering the entire information content of a single spectrum at once.

5.1 Gaussian Threshold

The easiest implementation of a frequency estimator would be to simply find the maximum amplitude in the spectrum, i.e.,

$$f_{\max} = \max[\mathrm{SNR}[f]]\,, \tag{5.1}$$

where $f = \{f_1, ..., f_n\}$ are the frequency bins for which amplitude measurements exist. This assumes that the global maximum is the correct estimate for the center offset frequency. Therefore, a frequency interval must be selected such that possible faulty amplitude information, i.e., artifacts that might result from the FFT, are excluded from f.

While the estimate is correct for spectra where the peak shape is distinct, i.e., in the near-ground region, false-positive estimates become abundant once this method is applied to signals with very low or even nil SNR. Outside of the PBL the lidar system will measure white noise that can be modeled as

$$\mathrm{SNR}[f] = \mathcal{N}(\mu_{\mathrm{Noise}}, \sigma_{\mathrm{Noise}})\,, \tag{5.2}$$

where the values for SNR$[f]$ are drawn randomly from a normal distribution $\mathcal{N}$ with mean μ_{Noise} and standard deviation σ_{Noise}. Hence, the resulting estimate for $f_{\max}$ would become random as well. Since it is known that the signal strength decreases with approximately $1/r^2$, one could define a set cutoff limit $\mathrm{SNR}_{\min}$, below which the results of the estimator are rejected. However, accurately defining a fixed limit so that the number of false-positives is minimized is difficult.

The following adaptive threshold is proposed within the context of this work: as has been mentioned above, the noise distribution is known to be normally distributed. Further, assume that the signal distribution around the center offset will appear as outliers in the distribution of SNR$[f]$ but influence the mean only marginally if the peak distribution width is small compared to the total length of f. Then, an adaptive threshold can be defined such that estimates are accepted only when their according signal strength is stronger than the spectrum mean by a few standard deviations $\mu + k \cdot \sigma$, where $k \in \mathbb{R}_0^+$ is a calibration coefficient. The standard deviation in this case is calculated via

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{SNR}[f_i] - \mu)^2}\,, \tag{5.3}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\mathrm{SNR}[f_i]\,. \tag{5.4}$$

The estimator is now defined as

$$f_{\max} = \begin{cases} \max[\mathrm{SNR}[f]] & \mu + k \cdot \sigma \leq \mathrm{SNR}[f_{\max}] \\ \mathrm{NaN} & \mu + k \cdot \sigma > \mathrm{SNR}[f_{\max}] \end{cases}\,, \tag{5.5}$$

where not a number (NaN) is used to express that $f_{\max}$ is undefined. For $k = 0$ the estimator behaves almost equal to Eq. 5.1. An exemplary implementation of Eq. 5.5 is given in Lst. 5.1. The time complexity to calculate the mean and the standard deviation is $\mathcal{O}(n)$. This is also true for finding the maximum element in an unsorted list. The overall time complexity thus is $\mathcal{O}(n)$. Further, given that the standard functions to calculate μ, σ and finding the maximum element have a memory complexity of $\mathcal{O}(1)$, the overall memory complexity is $\mathcal{O}(1)$.

Listing 5.1: Gaussian Threshold peak estimator

```
function gaussian_threshold(SNR, n=3)
    # Estimates peaks if global maximum is outside n*std of the global mean
    mu = mean(SNR)
    sigma = std(SNR)
    peak = findmax(SNR)
    if peak[1] >= mu+n*sigma
        # Return index of peak in SNR vector
        return (peak[2], peak[1]) # index, value
    else
        return (NaN, NaN)
    end
end
```

Since the threshold for acceptance or rejection is now well-defined, the occurrence of false-positive detections can be discussed. Assume that the presented algorithm is used on data outside of the region where the central peak can be reliably identified and white noise starts to make up most of SNR$[f]$.

A false-positive detection will now occur if the random value for any SNR$[f]$ is larger than $\mu + k \cdot \sigma$ calculated over f as previously explained. For the white noise distribution mentioned above, the mean SNR is $\mu \approx 0$ and the deviation will be $\sigma = \sigma_{\text{Noise}}$. In this environment, the probability for the signal strength of a noisy frequency bin to be positive and larger than the activation threshold is

$$\mathbb{P}(\text{SNR}[f_{\text{Noise}}] > k \cdot \sigma + \underbrace{\mu}_{\approx 0}) = \int_{k\cdot\sigma}^{\infty} \mathcal{N}(\mu, \sigma)\, d\text{SNR} \tag{5.6}$$

$$= \underbrace{\Phi\left(\frac{\infty - \mu}{\sigma}\right)}_{=\Phi(\infty)} - \underbrace{\Phi\left(\frac{k \cdot \sigma - \mu}{\sigma}\right)}_{=\Phi(k)} \tag{5.7}$$

$$= 1 - \frac{1}{2}\left(1 + \text{erf}\left(\frac{k}{\sqrt{2}}\right)\right) \tag{5.8}$$

$$= \frac{1}{2} - \frac{1}{2} \cdot \text{erf}\left(\frac{k}{\sqrt{2}}\right), \tag{5.9}$$

where Φ is the cumulative distribution function (CDF) and $\text{erf}(x)$ is the error function of the normal distribution. In practice, for $k = 4$ the probability for a false-positive detection is $\mathbb{P} \approx 3.167 \times 10^{-5}$. However, while increasing k leads to a reduction of false detections, it can also lead to a erroneous rejection of valid data.

Modeling the number of false-negative estimates is a lot harder than assessing the previous case, since the underlying distribution of aerosol-caused peak values in SNR$[f]$ is not exactly known. For the case that a spectrum contains at most one peak with a narrow distribution width, one can assume that its contribution to the deviation and mean remains small, i.e., $\mu_{\text{Signal}} \approx \mu \approx 0$ and $\sigma_{\text{Signal}} \approx \sigma$. Then, the probability that this peak will have an amplitude smaller than the activation threshold is solely dependent on the distribution from which the signal amplitude is drawn. For peak distributions where this is not the case, the issue becomes even more complicated. However, practical testing has shown this estimator to be reliable in general, with low false-positive and false-negative detections for $k \in [3, 4]$, as can be seen in Fig. 5.1.

The frequency resolution of the results of this algorithm can be further enhanced using the COG algorithm presented in Ch. 4.1.2. Hereby, the estimates for $f_{\max}$ are used as the initial center points for the weighted sums correction approach.

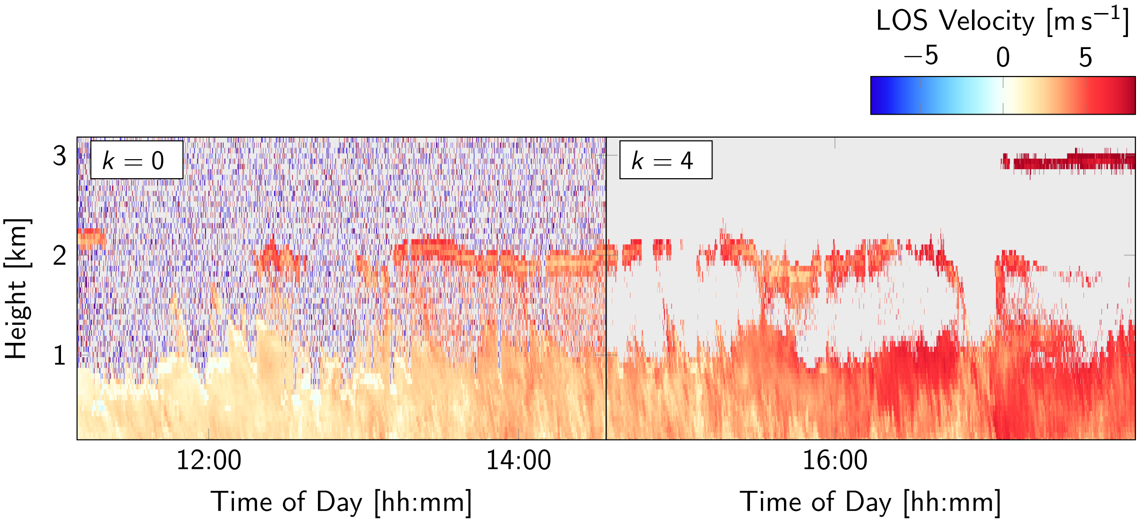


Figure 5.1: Velocity time series: representation of the LOS velocity measured during the October 7th, 2021 field campaign in Stuttgart-Vaihingen. The first half of the time series has been extracted using the Gaussian Threshold algorithm with $k = 0$. Excessive false-positive detections can be noted, as the estimator tries to find a signal peak for every range gate. The second half has been estimated using a value of $k = 4$, meaning the chance for a false-positive is $\mathbb{P} \approx 3.167 \times 10^{-5}$. The amount of false-positives has been significantly reduced and the edge of the measurable PBL can be discerned visually. The results were improved using the COG algorithm.

5.2 Scalogram Weights

The main issue with the previously presented algorithm is its dependency on the existence of a global maximum that is also the correct estimate for $f_{\text{max}} = f_{\text{Peak}}$. However, in noisy data where the peak amplitude is small, there is a probability for the event that $\text{SNR}_{\text{Peak}} < \text{SNR}_{\text{max}}$. In this case, an unbiased estimator that analyzes the entire spectrum at once and searches for the maximum in $\text{SNR}[f]$ would not be able to correctly identify the ground truth.

For this reason, a biased estimator is applied to the SNR-spectrum. This estimator is based on the local maxima scalogram (LMS) approach introduced by Scholkmann et al. (2012). Originally introduced as a method to distinguish peaks in noisy periodic and quasi-periodic data, this type of estimator was not intended for static information like spectral data. Assume that $x = \{x_1, ..., x_N\}$ is some detrended data containing peak information. Then, the local maxima of x are extracted via a moving window approach with increasing window width $\{w_k = 2 \cdot k \mid k = 1, 2, ..., L\}$ with $L = \lceil N/2 \rceil - 1$ where $\lceil x \rceil$ is the ceiling function. The result is the LMS matrix $\mathbf{L} \in \mathbf{M}_{L \times N}(\mathbb{R})$ which is defined as

$$\mathbf{L} = \begin{pmatrix} l_{1,1} & l_{1,2} & \cdots & l_{1,N} \\ l_{2,1} & l_{2,2} & \cdots & l_{2,N} \\ \cdots & \cdots & \ddots & \cdots \\ l_{L,1} & l_{L,2} & \cdots & l_{L,N} \end{pmatrix} \tag{5.10}$$

with the k-th row containing the information for the k-th window w_k. In the original implementation the value $r + \alpha$ is assigned to $l_{k,i}$ for $i = 1, ..., k+1$ and $i = N - k + 2, ..., N$, where r is chosen randomly from a uniform distribution $U \in [0, 1]$ and $\alpha = 1$. For all $l_{k,i}$

that satisfy $i = k+2, ..., N-k+1$ the following check is performed:

$$l_{k,i} = \begin{cases} 0 & x_{i-1} > x_{i-k-1} \wedge x_{i-1} > x_{i+k-1} \\ r+\alpha & \text{else} \end{cases} . \tag{5.11}$$

Then, the LMS for x is summed row-wise and the resulting vector is a measure for the length-scale dependent distribution of local maxima [41].

However, since only the previously presented part of the LMS algorithm is used for the following steps, the reader is referred to the original source for further information. For this work, Eq. 5.11 is changed to

$$l_{k,i} = \begin{cases} 0 & x_{i-1} > x_{i-k-1} \wedge x_{i-1} > x_{i+k-1} \\ 1 & \text{else} \end{cases} \tag{5.12}$$

and $\mathbf{L}$ is now referred to as the binary LMS matrix. This is also true for $i = 1, ..., k+1$ and $i = N-k+2, ..., N$. In general, the LMS is a triangular construct, where activations of $l_{k,i}$ occur if the associated index is a local maximum. In this case, however, $\mathbf{L}$ is not summed row-wise, but rather column-wise, where the total count for each index is interpreted as its suitability to be a local peak.

Since $\mathbf{L}$ is populated triangularly, choosing the center point for the LMS in the spectral data $\text{SNR}[f] = \{\text{SNR}_1[f_1], ..., \text{SNR}_n[f_n]\}$ changes the outcome for the best peak estimate. Ideally, the initial estimate for the center of the LMS is directly at f_{Peak}. The LMS will be especially sensitive for small changes of f_{Peak} around the initial choice, which mimics that the vertical wind velocity course will be contiguous to some degree. With the above definitions set in place, the estimator is applied range gate-wise and defined as follows:

$$f_{\max} = \min[\{w_i = \sum_{k=1}^{L} l_{k,i} \mid i = 1, ..., N\}] \tag{5.13}$$

with the resulting estimate for $f_{\max}$ being the center for the next application of the binary LMS algorithm in the next higher range gate of the lidar spectrum. For the first range gate, an initial estimate must be made. Here, a simple maximum estimator suffices, since for this gate the aerosol density should be high enough so that a clearly distinguishable peak signal exists.

Nevertheless, the issue remains that this estimator is designed such that a peak value must be found even if no physical data is present in $\text{SNR}[f]$. An activation threshold as has been used in Ch. 5.1 is implemented in order to minimize the amount of false-positive detections. In contrast to the threshold used for the Gaussian estimator, the high-altitude background noise strength and its deviation will be used as the cutoff limit in this case. Here, the algorithm that has been introduced in Ch. 4.1.1 can be used in order to find an accurate value for μ_{Noise} and σ_{Noise}. The estimator then is defined as

$$f_{\max} = \begin{cases} \min[\{w_i = \sum\limits_{k=1}^{L} l_{k,i} \mid i = 1, ..., N\}] & \text{SNR}[f_{\max}] \geq \mu_{\text{Noise}} + c \cdot \sigma_{\text{Noise}} \\ \text{NaN} & \text{SNR}[f_{\max}] < \mu_{\text{Noise}} + c \cdot \sigma_{\text{Noise}} \end{cases} , \tag{5.14}$$

where NaN expresses that there exists no valid value for $f_{\max}$ and $c \in \mathbb{R}_0^+$ is a calibration constant. In practice, $\mu_{\text{Noise}} \approx 0\,\text{dB}$ and $\sigma_{\text{Noise}} \ll 1\,\text{dB}$ make the performance of the estimator insensitive to the choice for c. As can be seen in Fig. 5.2, while the LMS approach provides slightly more estimates, it also produces more incorrect choices. An additional improvement of the velocity estimate can be achieved using the COG algorithm.

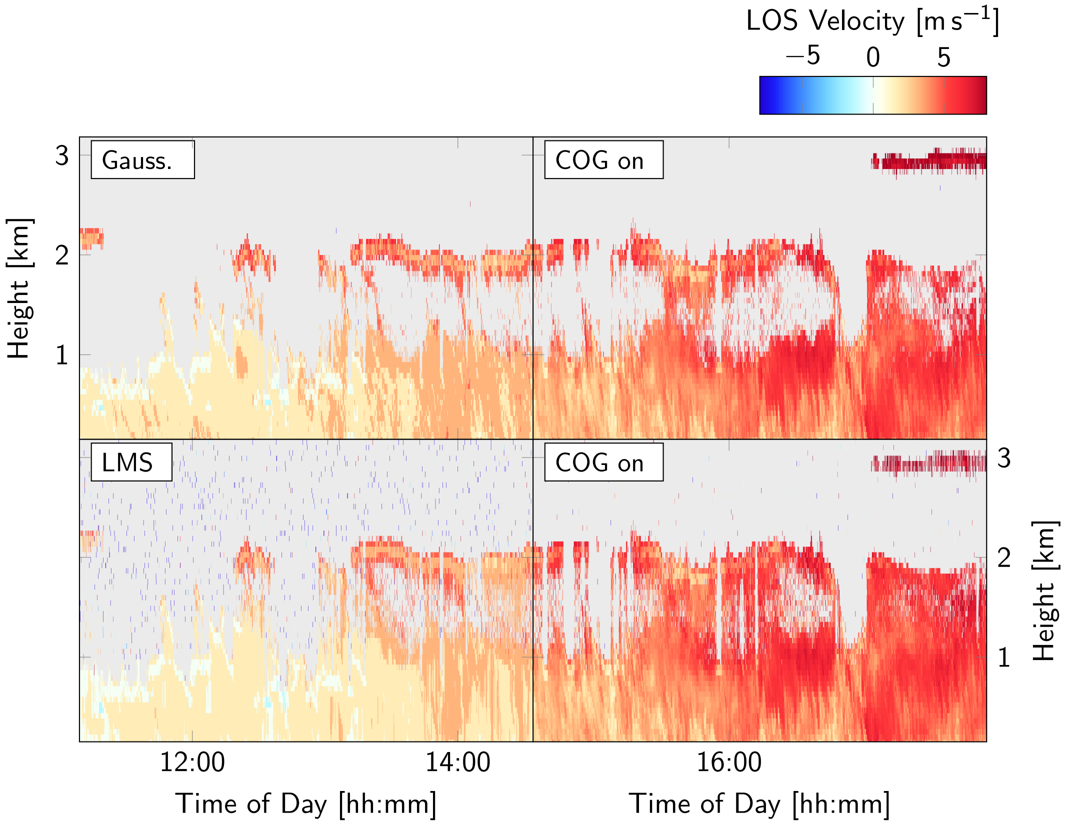


Figure 5.2: Comparison of Gaussian threshold and LMS algorithm: the upper time series was generated using the Gaussian threshold algorithm with and without the COG correction algorithm. The additional detail that can be retrieved, especially in the near ground region, is clearly visible. The lower time series was generated using the LMS algorithm presented previously, also comparing the results with and without the use of the COG interpolation routine. It can be seen that the results of both the Gaussian and LMS algorithm differ slightly. However, note that the frequency resolution is coarse. In the second half, where the COG correction was used, the results are qualitatively the same. The LMS algorithm is able to better estimate the velocity course between the ground region and the clouds, but also produces more false-positive results.

5.3 Neural Network

The main shortcoming of the algorithms presented in the introduction of this chapter and Ch. 5.1 and 5.2 is the strictly local analysis of the spectrum. Hereby, every range gate is analyzed within its own context, while in reality the entire spectrum must be considered. Such approaches can be sufficient for radio detection and ranging (radar) applications where scatterers might be uncorrelated objects, i.e., aircraft, but for the purpose of wind lidar it is assumed that the flow field is continuous to some extent.

This assumption can not be rigorously proven, since it is known that the Navier-Stokes equations can produce non-smooth solutions for compressible flows. In addition, turbulent flows, in general, are highly irregular over a variety of length scales. However, for the cases that were observed in the context of this work, and with the coarse length resolution of the used lidar system, a cohesive velocity course from the ground region up to the edge of the PBL is assumed.

The LMS algorithm tries to consider this to some extent by building upon the peak of the velocity distribution from the previous range gate, from the ground region up to the point where the signal can not be identified anymore. However, this has proven to be insufficient, as gaps can exist in the velocity course after which the signal resumes. Here, the current LMS implementation struggles. Therefore, an estimator that processes all range gates at once and has the ability to inter- and extrapolate the spectrum within physically sound limits would be ideal.

A human observer can easily identify a good approximation for the velocity course by simply looking at the lidar spectrum (i.e., Fig. 4.2). By adjusting the color map, they can also make out even faint traces of the spectrum that classic algorithms struggle to estimate correctly. During this process, the human brain performs the task of extracting, interpolating, and extrapolating data from the spectrum based on prior knowledge of physically reasonable choices for the velocity course.

Due to this observation, the use of a machine learning approach in the form of a neural network was investigated. Since the presented problem showcases a strong similarity with the problem of image recognition and classification, a feed forward network with convolutional kernels was chosen. For a detailed introduction into the topic of neural networks and deep learning, the reader is referred to Goodfellow et al. (2016) [45], Aggarwal (2018) [46] or Mandl (2021) [47].

Prominent convolutional neural network (CNN) architectures are LeNet [48], AlexNet [49], ResNet [50], GoogLeNet [51], SqueezeNet [52] and many more. A very high amount of DOFs is characteristic for these networks (i.e., AlexNet: 6×10^7), which makes them highly accurate in the best case scenario, but also increases the time needed to train them on a given data set. Nevertheless, once trained, the time complexity for input forwarding scales linearly with the amount of DOFs that make up the network, that is $\mathcal{O}(n)$ where n is the number of parameters. In the following sections, the network training strategy, as well as the used architecture for the CNN estimator presented by Kliebisch et al. (2021) [26] will be elaborated upon.

5.3.1 Training strategy

In order to optimize the weights and biases of a chosen network architecture, training data with a known ground truth is required. This data should be selected such that the network is able to find a representation of the training set that allows it to not only correctly classify familiar spectra, but also generalize, to some extent, onto new situations. This implies a high variance data set is desirable and the training therefore requires a large number of spectra, with two ways to obtain them.

5.3.1.1 Labeled data

A common practice for the acquisition of training data is the manual labeling of real-world data. This is a very labor intensive task, but, with a human reviewer in the loop, allows for some error checking. Using the Gaussian estimator, presented in Ch. 5.1 as a way to identify the signal course, is a solid way to speed up this process, but limits the information contained within the resulting data set to the capability of said estimator [26]. This data set is not equivalent to a ground truth, but since the limits of this estimator are well defined, it is a reasonable way to generate a conservative estimate of the ground truth for values of $k \geq 4$, which, in practice, eliminates false-positive detections. Since the Gaussian estimator is susceptible to cloud aliasing, the measured data that was used as the basis for the labeled data set was carefully selected as not to feature such flawed velocity estimates. Otherwise, this can lead to the network misidentifying aliasing effects as correct estimates.

5.3.1.2 Synthetic data

In order to extend the variance of the training data set and not be limited to already observed weather phenomena, the generation of synthetic data was considered. In recent years, the use of synthetic data for network training has become a popular method to generate training sets where the ground truth is exactly known [53].

Here, the word “synthetic” refers to data generated by a known algorithm or equation, rather than originating from real world measurements. The return signal strength is described with good agreement between theory and practice by the lidar equation in Eq. 2.11 and the velocity course is known to follow the general shape dictated by the boundary layer equations in Ch. 2.2.2. Therefore, this method lends itself to generate vast amounts of physically accurate spectra with controllable features.

The following equation was used by Kliebisch et al. in order to generate a synthetic spectrum matrix $\mathbf{S} \in \mathbf{M}_{m\times n}(\mathbb{R})$ with n frequency bins and m range gates:

$$\mathbf{S}[r,f] = \frac{1}{r^2} \cdot \underbrace{A_0 \cdot \eta_A[r]}_{\text{System model}} \cdot \underbrace{\beta[r] \cdot \tau[r]}_{\text{Aerosol model}} \cdot \underbrace{\text{SNR}[r,f]}_{\text{Velocity model}} + \underbrace{\mathcal{N}(0, \sigma_{\text{Noise}})}_{\text{Noise model}}, \tag{5.15}$$

where $r = \{r_1, ..., r_n\}$ represents the range gates, $f = \{f_1, ..., f_m\}$ represents the frequency bins, A_0 is a system factor, which includes the aperture size, laser parameters and electro-optical efficiencies and $\text{SNR}[r,f]$ is the signal strength distribution. In addition, a white noise source is added to the entire spectrum. The antenna function $\eta_A(r)$ describes the divergence of the monostatic FOV for some Rayleigh length z_R as

$$\eta_A(r) = \left[2 + 2\left(\frac{z_R}{r}\right)^2\right]^{-1} . \tag{5.16}$$

The aerosol distribution over range is modeled by the atmospheric volume backscatter coefficient β and the two-way integrated attenuation coefficient τ, where

$$\tau(r) = \exp\left(-2\int_0^r \sigma(r')\, dr'\right) . \tag{5.17}$$

It is assumed that Eq. 2.12 is applicable and thus only $\beta[r]$ needs to be modeled, since β and $\tau = \tau(\sigma)$ are now coupled. The signal strength distribution $\text{SNR}[r,f]$ contains the information about the LOS velocity at a given height and frequency. For each range gate the signal distribution around the offset frequency is modeled to be Gaussian:

$$\text{SNR}(r,f) = \frac{1}{\sqrt{2\pi\sigma_D^2(r)}} \exp\left(-\frac{(f - \Delta f_D(r))^2}{2\cdot\sigma_D^2(r)}\right) , \tag{5.18}$$

where Δf_D is the Doppler shift frequency and σ_D^2 controls the variance of the distribution. Thus, the main control parameters are the backscattering coefficient $\beta = \beta(r)$ and the velocity course, described indirectly by the Doppler frequency shift, $\Delta f_D = \Delta f_D(r)$ [26].

Finding an exact relation for the aerosol backscatter coefficient over range is nontrivial, since the density of aerosols in a given volume of space is of stochastic nature and extremely dependent on local weather, topography and daytime conditions. Hence, the function for $\beta(r)$ is modeled qualitatively, where feasible bounds for the true value of β are assumed. During field campaigns it was observed that β can vary spatially, as well as temporally, in a seemingly random fashion.

Therefore, the following qualitative aerosol model is proposed: the backscatter coefficient has some value β_0 at $r = 0\,\mathrm{m}$ which varies randomly as the range is increased. Strong spatial gradients in the aerosol density are neglected, an assumption that might not hold at low elevation angles, but is assumed to be reasonable for nominal weather conditions. The change of the backscatter coefficient from one range gate to the next is modeled as a random walk, a technique that is primarily used to generate non-deterministic time series data, but is thought to be suitable for this purpose as well and thus

$$\beta[r_i] = \beta[r_{i-1}] \cdot (1 + \Delta\beta)\,, \tag{5.19}$$

where the density step size $\Delta\beta$ is drawn from some normal distribution $\mathcal{N}(0, \sigma_\beta)$. Further, it is known that a sharp drop in the aerosol density occurs at the edge of the aerosol PBL. This in turn is modeled by a decay factor β_δ that enforces $\lim_{r\to r_n} \beta[r] \approx 0$ via

$$\beta[r_i] = \beta[r_{i-1}] \cdot (1 + \Delta\beta) \cdot \frac{1}{\beta_\delta} \tag{5.20}$$

above some random height $r_{\mathrm{PBL}} < r_n$. The aerosol model for the spectrum matrix thus is

$$\beta[r_i] = \begin{cases} \mathcal{N}(\mu_{\beta_0}, \sigma_{\beta_0}) & r_i = r_1 \\ \beta[r_{i-1}] \cdot [1 + \mathcal{N}(0, \sigma_\beta)] & r_1 < r_i < r_{\mathrm{PBL}} \\ \beta[r_{i-1}] \cdot [1 + \mathcal{N}(0, \sigma_\beta)] \cdot \frac{1}{\beta_\delta} & r_i > r_{\mathrm{PBL}} \end{cases}\,, \tag{5.21}$$

where the aerosol density at ground level is also drawn from a normal distribution in order to simulate different visibility conditions. Note that the values for β are fine-tuned to match the observed measurements of the utilized lidar system, rather than trying to achieve the exact quantitative values for β given in the literature [26].

An analysis of the resulting distribution for $\beta(r)$ is visualized in Fig. 5.3a. Here, the relative logarithmic occurrence for 2×10^6 spectra was evaluated with a boundary layer height that varies around 2 km. It can be seen that the most occurrences are located at $\beta < 2.5 \times 10^{-6}\,\mathrm{m}^{-1}$ with a high set variance below 4 km. In addition, some large values for β exist, albeit with a low relative occurrence, at $\beta > 1 \times 10^{5}\,\mathrm{m}^{-1}$ in the lower PBL region. The effect of cloud occurrences was not considered for this analysis

Clouds are known to obstruct the passing of lidar signal for some wavelength ranges and thus act as a hard barrier in the lidar spectrum in this case. Exceptions were observed where thin clouds allowed some light to pass through and return information from above the cloud layer. The synthetic data considers this effect by depicting clouds as hard spikes in the backscatter coefficient course. Thereby, a realistic spectrum behavior can be enforced in the presence of overcast. This is described in detail by Kliebisch et al. [26].

The model for the Doppler shift offset works similarly to the presented random walk approach for $\beta(r)$. The frequency shift starts at the ground with a value that is randomly chosen in a small domain around $\Delta f_D = 0$ since the first range gate is at $r_1 \approx 77\,\mathrm{m}$. The subsequent frequency offset steps are drawn from a normal distribution as follows:

$$\Delta f_D[r_i = r_1] = \mathcal{N}(\mu_0, \sigma_0)\,, \tag{5.22}$$

$$\Delta f_D[r_i > r_1] = \begin{cases} \Delta f_D[r_{i-1}] \cdot [1 + \mathcal{N}(0, \sigma_{\Delta f})] & \Delta f_{\min} < \Delta f_D[r_i] < \Delta f_{\max} \\ \Delta f_{\min} & \Delta f_D[r_i] < \Delta f_{\min} \\ \Delta f_{\max} & \Delta f_D[r_i] > \Delta f_{\max} \end{cases}\,. \tag{5.23}$$

A check is performed for each step and Δf_D is restricted to an upper and lower step size that are imposed if the frequency step is too large or too small, respectively. This allows for some control to avoid statistical outliers in the velocity profile gradient [26].

The resulting relative occurrence density for Δf_D is displayed in Fig. 5.3b. An increase in variance over height can be seen, as the individual Doppler-curves deviate away from their starting point. It is important to note that the symmetry of the distribution plot implies that there is no bias with respect to the LOS wind direction in the synthetic data set. While a single Doppler-curve can lean towards positive or negative offsets, the macroscopic distribution is homogeneous, which is assumed to train the CNN to be impartial to the sign of the Doppler shift.

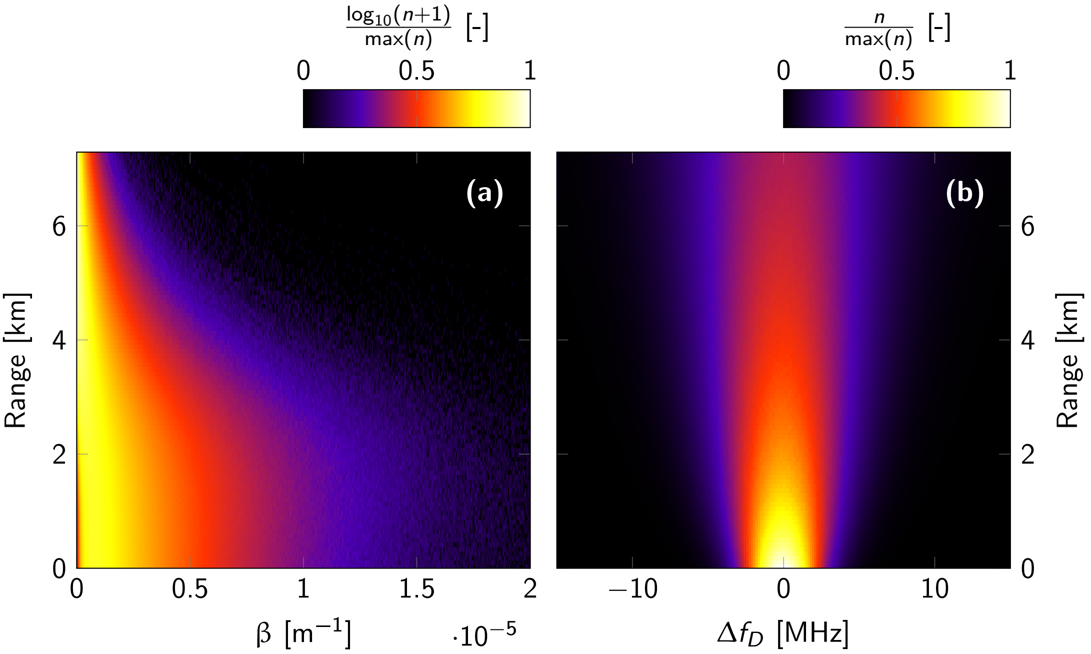


Figure 5.3: **(a)** Backscatter coefficient distribution: relative logarithmic count for β-values over height simulated for 2×10^6 spectra. The intended aerosol boundary layer height varies around $r_{\mathrm{PBL}} = 2\,\mathrm{km}$. In general, values for $\beta \leq 2 \times 10^{-5}\,\mathrm{m}^{-1}$ are represented, but the majority of β-curves does not exceed $2.5 \times 10^{-6}\,\mathrm{m}^{-1}$. A high data set variance is implied by the bulge below 4 km, whereas the variance above r_{PBL} converges against $0\,\mathrm{m}^{-2}$. **(b)** Doppler frequency shift distribution: relative frequency shift count simulated for 2×10^6 spectra. The variance of Δf_D increases with height. The symmetry of the distribution suggests that the network will not be biased towards a specific LOS velocity sign, as long as the generated training data set is large enough.

With the presented model, an arbitrary number of synthetic spectra can be generated in a reasonable amount of time. By adjusting the parameters, a variety of different weather and visibility conditions can be modeled. Fig. 5.4 showcases several generated and real spectra, as to highlight the current capabilities of the model proposed by Kliebisch et al. [26]. The SNR-model was adjusted for an average PBL height of 2 km and up to three cloud occurrences along the LOS. The presented spectra parameters were adjusted to match the signal strength of the real data. The real spectra were selected randomly from the October 7th, 2021 field campaign in Stuttgart-Vaihingen.

Fig. 5.4a, Fig. 5.4j, Fig. 5.4k and Fig. 5.4o visualize real data, while the remaining spectra are synthetic. The easiest way to distinguish between the two groups is by observing the noise patterns. In order to improve the noise model, a generative adversarial network (GAN) or real world noise data might be used in the future.

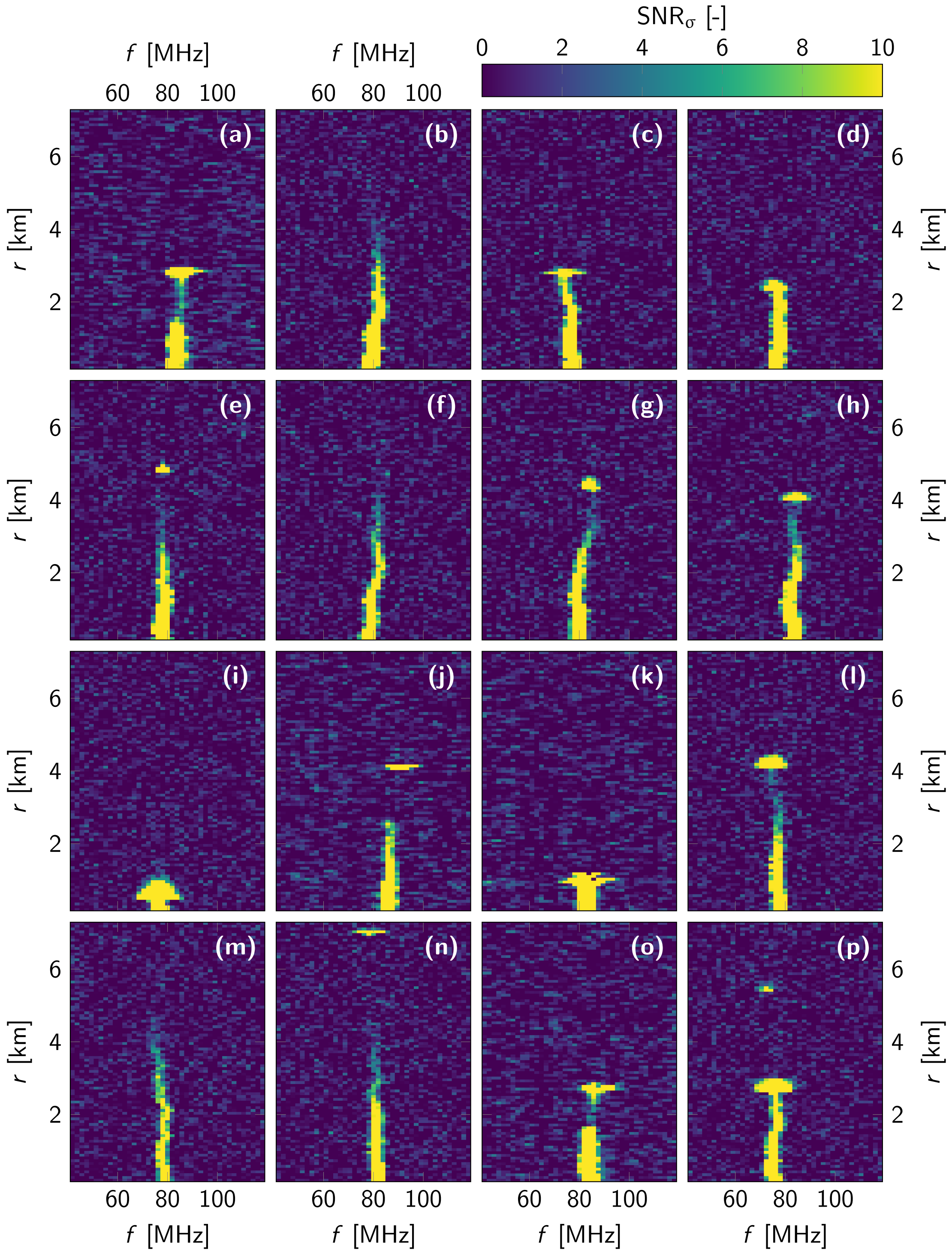


Figure 5.4: Labeled and synthetic data comparison: a set of 16 training spectra is presented, of which four are taken from real-world measurements. Different cloud, wind and signal strength configurations can be seen. The spectra in **(a)**, **(j)**, **(k)** and **(o)** originate from the October 7th, 2021 field campaign, whilst the remaining data has been generated. The main difference hereby can be seen in the noise distribution, an issue that can be fixed in future versions of the noise model. The synthetic spectra were adjusted to match the signal strength SNR$_\sigma$ of the actual measurements for the sake of comparability.

5.3.1.3 Parameter optimization

The model parameters that were introduced in the previous section, as well as the hyperparameters of the specific CNN model and optimization algorithm, were adjusted for estimation accuracy using a continuous training strategy. In Fig. 5.5 a flowchart of the employed strategy is shown. Based on an initial assumption for suitable parameters, the training data set is generated from synthetic and labeled data in a ratio of 20 : 1. The test data set is generated in the same way with a ratio of 10 : 1. The batch size for the training and test data set was chosen to be 1×10^6 and 5×10^4, respectively. After the data has been sourced, the network training is implemented using the adaptive moment estimation (ADAM) optimization algorithm [54] with a logistic cross-entropy loss function [26].

In the future, even more advanced optimization algorithms like limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) might be used [47, 55]. Once the training has been completed, the network output is analyzed and evaluated qualitatively by hand. The considered factors include: the ability to correctly identify low-SNR signals, the ability to abstract signals with missing data, the capacity to correctly estimate data under the influence of overcast, a high resistance against aliasing effects and the proficiency to indicate the absence of aerosols, that is to correctly estimate the edge of the PBL if possible. With these insights, the parameters are modified and a new training cycle is initiated.

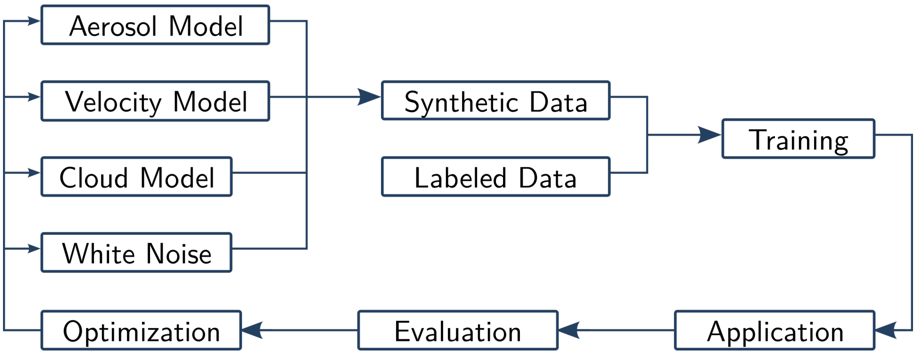


Figure 5.5: Training strategy flowchart: the parameters presented for the fine-tuning of the synthetic spectrum model are optimized by training the network with a given data set, applying the resulting weights and biases to real world measurements and manually evaluating the quality of the estimate. Based on these observations, the parameters for the different models are adjusted and a new training cycle is started.

5.3.2 Network architecture

In general, neural networks consisting of one or more fully-connected hidden layers can be represented as a series of matrix operations, where for each hidden layer the input information can be represented as a matrix that is multiplied by a layer weights matrix and added to a layer-specific bias vector. In contrast, convolutional layers use discrete kernels that are applied to each cell in the input data grid, which can be one- or higher-dimensional. Since the kernel size is usually much smaller than the input data grid, the concept of locality is introduced, as the output value of a specific cell only depends on a certain amount of adjacent data points in the input grid [45, pp. 330-339].

It is assumed that this concept also applies to the wind velocity field, where for a sufficiently short time period the changes at a certain range are only dependent on the immediate surroundings, that is a certain number of neighboring range gates. Thus it is hypothesized that a CNN with trained kernel weights can produce physically sound results.

Several of the CNN designs mentioned at the beginning of Ch. 5.3 were trained for the purpose of spectrum estimation. In this section, SqueezeNet is presented in more detail as a suitable network for the aforementioned purpose. The aim of this architecture, presented by Iandola et al. in 2016, is to achieve performance similar to state-of-the-art convolutional networks, while at the same time reducing the amount of free parameters for increased performance during training and inference. This is a relevant characteristic of SqueezeNet, since comparable CNNs can have up to 50-fold more free parameters [52].

The CNN architecture is detailed in Fig. 5.6. The basic idea is to map an input spectrum to an output spectrum via the SqueezeNet CNN. While the idea was considered to map the input spectrum directly onto an output vector containing the respective LOS velocity for each range gate, it was decided to move ahead with the prior approach, since the network can be trained to also detect the absence of signals and consequently of aerosols. In addition, by increasing the frequency bin resolution of the output spectrum, an interpolation effect similar to that of the COG approach can be achieved.

Ideally, the network would map the input onto the output such that only one grid node is activated for each range gate. That node would be the estimated ground truth for the respective range gate. In practice, however, the network output is slightly blurred, where the width of the activation map can be interpreted as a kind of uncertainty. This effect can be seen in the network output (Fig. 5.6b) for the input spectrum shown in Fig. 5.6a.

At the core of the network architecture sits the so-called Fire module. It consists of kernels $\mathbf{s} \in \mathbf{M}_{1\times1}(\mathbb{R})$ that are collectively called the Squeeze layer, which feeds into a layer of kernels $\mathbf{e} \in \mathbf{M}_{1\times1}(\mathbb{R})$ and $\mathbf{e} \in \mathbf{M}_{3\times3}(\mathbb{R})$ that are applied in parallel and referred to as the Expand layer (Fig. 5.6). The number of Squeeze kernels $s_{1\times1}$ and the number of Expand kernels $e_{1\times1}$ and $e_{3\times3}$ must fulfill the relation $s_{1\times1} < e_{1\times1} + e_{3\times3}$ in order to reduce the number of free parameters. The commonly used rectified linear unit (ReLU) activation function is used for the Fire layers [52].

The first layer in the SqueezeNet architecture is a traditional convolutional layer, with a 3×3 kernel in this case, followed by a pooling layer. Pooling operations, i.e., max pooling, help to make the network resistant against small changes in the input data, that is invariant against local translations of features. This is achieved by defining the network output at these layers to be dependent on some metric that is applied to a defined neighborhood for each node. The stride defines how the pooling kernel is moved over the data [45, pp. 339-345].

Additional Fire modules and pooling operations follow (Fig. 5.6), with the size of the Fire modules being increased for deeper layers. After the input has passed the last Fire layer, a dropout step is performed. Hereby, the trained kernels weights are culled randomly, a common technique that aims to avoid overfitting [56]. A convolutional layer is integrated before a final adaptive pooling step is performed for which the neighborhood size is determined automatically. The last layer (Flatten) is used to reshape the network output into a human-interpretable matrix [26, 52].

In Fig. 5.7b an exemplary CNN-generated time series is displayed. Compared to the results of the Gaussian threshold algorithm in Fig. 5.7a, an increase in range is apparent. Both methods were applied to linearized data for comparability. While the network can produce results with increased range, a kind of blurring effect must be noted. Hereby, the spatial and temporal resolution, especially in the near-ground region, is decreased compared to the Gaussian estimator. The reason for this effect is still under investigation. In order to fix this shortcoming, Kliebisch et al. propose to merge the outputs of both methods such that the increased resolution of the Gaussian estimator and the extended range of the CNN method are combined. This is accomplished by complementing missing estimates of the classical algorithm with the results of the network (Fig. 5.7c) [26].

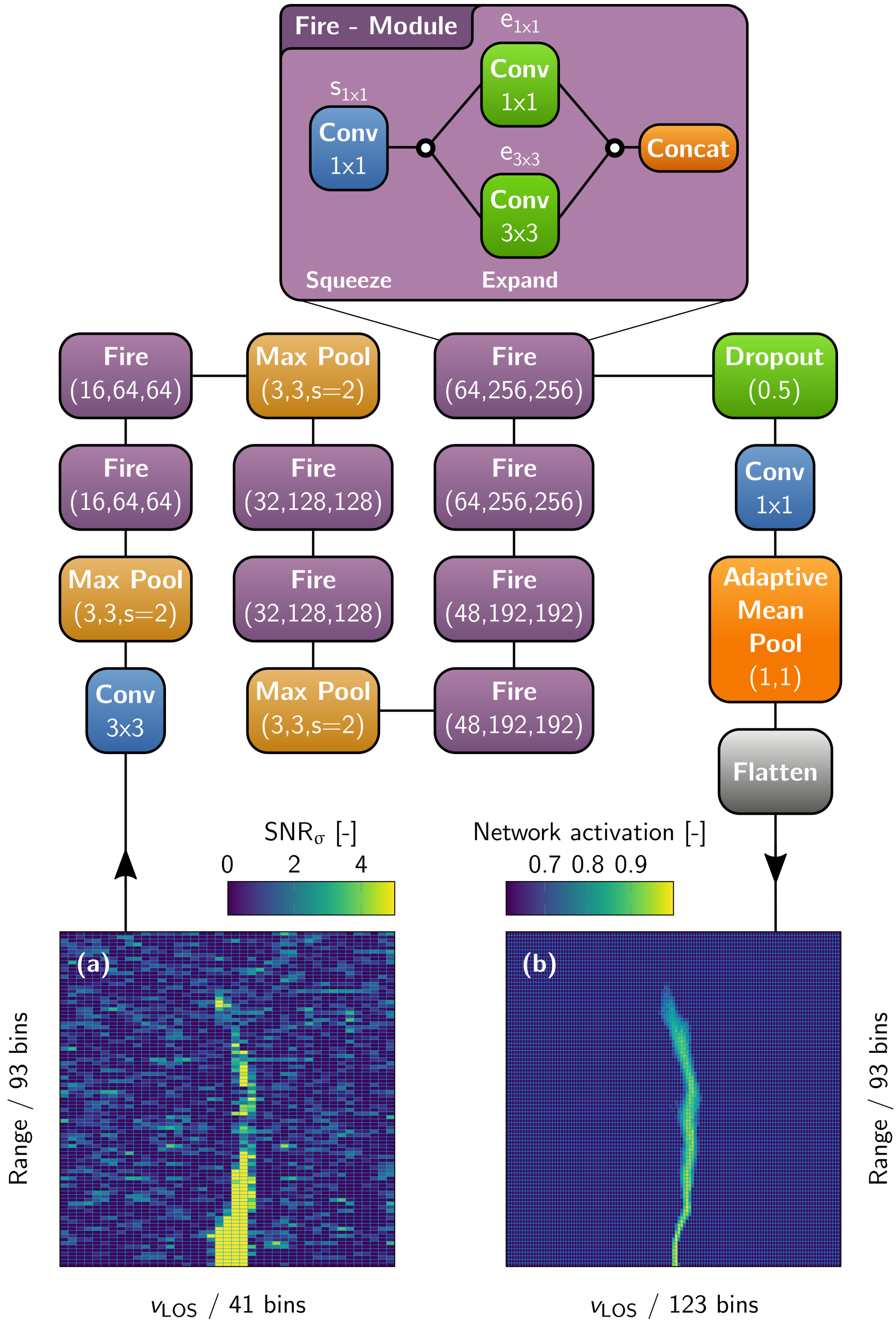


Figure 5.6: CNN schematic: an abstract representation of the SqueezeNet architecture used to analyze wind lidar spectra. The input data is passed through multiple Fire modules with increasing kernel size. Each Fire module consists of a series of kernels that are convolved with the input from the previous layer. The pooling stride is $s = 2$ [26, Fig. 4]. **(a)** Input spectrum: the input spectrum size is limited to a reasonable range around the center velocity at $0\,\mathrm{m\,s^{-1}}$ or 80 MHz in order to simplify the training and inference of the CNN [26, Fig. 3a]. **(b)** Output spectrum: the output spectrum frequency bin resolution is increased by a factor of three. This allows for some velocity interpolation at the end of the network [26, Fig. 3c].

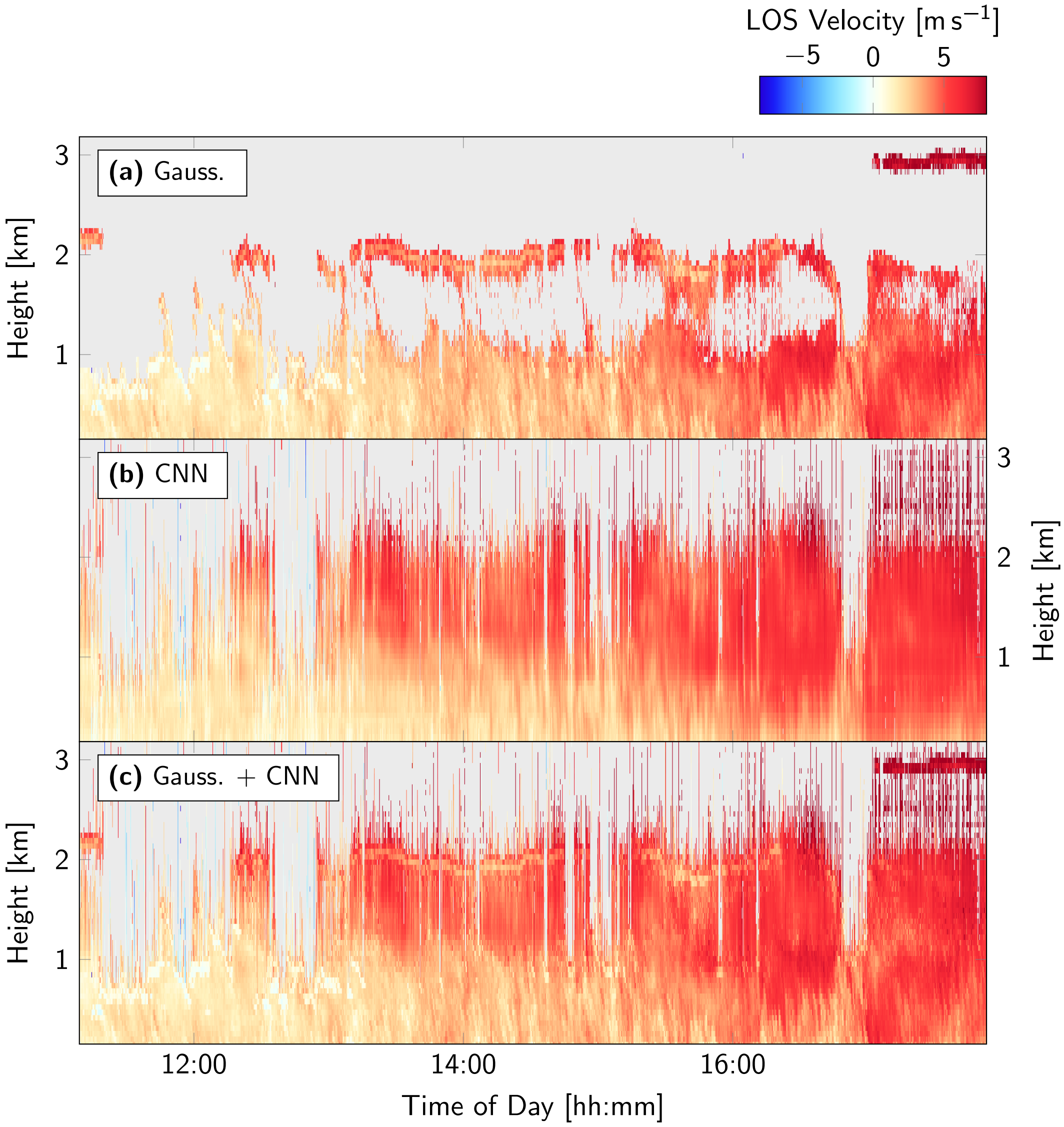


Figure 5.7: **(a)** LOS velocity extracted by the Gaussian threshold algorithm for the October 7th, 2021 field campaign in Stuttgart-Vaihingen. The underlying signal data is given in SNR$_\sigma$ with a threshold value of $k = 8$. **(b)** Velocity time series inferred by the SqueezeNet CNN for the same lidar data as in **(a)**. The output shows that the network is capable to increase the evaluable range for some time-segments. However, some fine-grain detail is lost in the near-ground region, possibly due to smoothing effects. In addition, a streaking effect is visible, where the detection range varies strongly over time. **(c)** A combination of the output produced by the classical algorithm and the network. This approach allows for a high spatial and temporal resolution in the near-ground region, while also extending the output range. The turbulence and cloud structures from **(a)** are complemented by the output of the CNN above the assumed PBL edge at approximately 1.0 km to 1.5 km.

6 Experimental Results

Once a stable operational state of the system presented in Ch. 3 was ensured, a series of measurement campaigns were carried out at the German Aerospace Center (DLR) site in Stuttgart-Vaihingen, Germany. These tests were aimed at obtaining lidar data for different weather conditions, showing the operability for measurements lasting up to several hours and gaining experience in the operation of such a system. These trials were primarily conducted in July, September, and October of 2021. Usually, the measurements were initiated between 9 a.m. and 11 a.m. and completed between 5 p.m. and 6 p.m. during workdays. Nightly measurements were not conducted due to workplace safety regulations. In total, 22 daytime campaigns were carried out at different system orientations and inclinations. For all campaigns the monostatic setup detailed in Ch. 3.2.1 was used.

The general experimental procedure was as follows: after an initial assessment of the weather situation using satellite data and model predictions provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), an alignment of the system was planned such that the lidar LOS and general wind direction would be as parallel as possible. Following this assessment, the system was assembled and orientated using IMU data. Once the laser system was armed and activated, a quick evaluation of the return signal quality was conducted and the initial detector drift offset accounted for. If the laser parameters, i.e. temperature and current, were nominal, the data recording was started. During the course of the measurement, the initially set values and orientation were not changed.

In this section the data obtained by the lidar system will be analyzed using the algorithms presented in Ch. 4 and Ch. 5. For this purpose, the return signal time series is qualitatively investigated and the wind speed time series is evaluated. An analysis of ε and C_n^2 is performed for selected time intervals. The determined wind speed data is compared quantitatively with the radiosonde data provided by the probes launched from the Deutscher Wetterdienst (DWD) station in Stuttgart-Schnarrenberg. This allows for the system to be evaluated against an external, independent measurement device. No validation can be performed for the values determined for ε and C_n^2 at this point, since this data was not available from an independent source at the time of writing. Rather, the values for $C_n^2(r)$ will be compared against the predictions made by the HV-model. However, it is not expected that the agreement over height will be good, as the measurement range for this system falls short of the ranges of relevance for the HV-model.

For the velocity estimation from lidar data the Gaussian threshold algorithm, presented in Ch. 5.1, will be used. While the CNN-method produces promising results, a conservative estimate is made using this well-understood algorithm. The aim is to avoid potential ambiguity of the results due to outstanding questions about the network capabilities. A threshold of $k = 4$ standard deviations above the mean will be used moving forward. Values determined for ε and C_n^2 will be given within a 95 %-confidence interval.

Three campaigns will be presented in the following sections. The system inclination refers to the angle between the lidar system and the horizon and the alignment will represent the cardinal direction in which the system was orientated. Hereby, 0° represents north, 90° represents east and so forth. Wherever height axes are used, range will be adjusted by the inclination angle to reflect height above ground. However, the LOS velocity angle will remain unchanged unless specified otherwise. The site height above mean sea level (AMSL) is approximately 452 m for all campaigns.

6.1 July 27th, 2021

The July 27th, 2021 field campaign was initiated at 10:03 a.m. with an inclination of 45° and northern orientation. Additional system parameters are summarized in Tab. 6.1. Initial predictions made by the ECMWF indicated an average horizontal wind velocity of roughly $5\,\mathrm{km\,h^{-1}}$ to $8\,\mathrm{km\,h^{-1}}$ at ground level throughout the day. Few clouds were present at the beginning of the measurement. At 12 a.m. a temperature of $24.0\,^{\circ}\mathrm{C}$ and air pressure of $1.014 \times 10^{5}\,\mathrm{Pa}$ were measured at the nearby weather station of the University of Stuttgart.

Table 6.1: July 27th field campaign parameters

Date	July 27th, 2021	
Site	Stuttgart-Vaihingen	
	Latitude	48°44’55.9"N
	Longitude	9°06’09.1"E
	Alignment	4°
	Inclination	45°
Duration	Start	10:03
	End	18:00
Setup	Monostatic	
	Beam Diameter	100 mm
	Focal Length	500 mm
	PFA Current	8 A
	Detector Type	APD

A radiosonde was launched by the DWD from Stuttgart-Schnarrenberg at 12:46 a.m. and the recorded flight data is displayed in Fig. 6.1. For comparison, an ISA-model pressure curve with $p_0 = 1.014 \times 10^{5}\,\mathrm{Pa}$ was added to Fig. 6.1a. The same applies to Fig. 6.1b, in which a temperature curve with $T_0 = 297.2\,\mathrm{K}$ was included. It can be easily seen that the agreement between the ISA-model and the measured values for p and T was good for this day, that is below $5\,\mathrm{km}$, which covers the height of the lidar data fully. Above $12\,\mathrm{km}$ a strong deviation of the measured temperature from the model prediction can be observed. This effect is caused due to the absorption of ultraviolet (UV) radiation by various chemical species in the stratosphere [57, p. 53].

Using the algorithm presented in Ch. 4.2.1, the turbulent atmospheric parameters are calculated for the time intervals between 10:55–11:00 a.m. and 11:55–12:00 a.m. over height. The results for the TDR ε and the turbulent index of refraction C_n^2 are depicted in Fig. 6.2a and Fig. 6.2b, respectively. Values for ε were calculated as long as no NaN data exists in the time series. It can be seen for both segments that ε varies over several orders of magnitude as the height increases. A significant spike in the TDR of the first interval is noticeable at a height of $217\,\mathrm{m}$, after which it decreases before increasing in strength again towards the edge of the PBL. An hour later, the overall dissipation rate has increased, whereas now a dip in the near ground region can be seen.

With these results, C_n^2 has been calculated according to the relations presented in Ch. 4.2.2. Note that the Brunt-Väisälä frequency N is calculated using ISA-model predictions for the temperature and pressure over height because of the good match mentioned above. The uncertainty of σ_N is small and therefore neglected, as it is intended to focus the results for C_n^2 on the influence of the uncertainty of ε. In general, the course of $C_n^2(h)$ is proportional to that of $\varepsilon(h)$. In Fig. 6.2 HV-model curves were added with $W = 21\,\mathrm{m\,s^{-1}}$ and $A = 2.5 \times 10^{-16}\,\mathrm{m^{-2/3}}$ to $2.0 \times 10^{-15}\,\mathrm{m^{-2/3}}$ for the sake of comparability. In the first

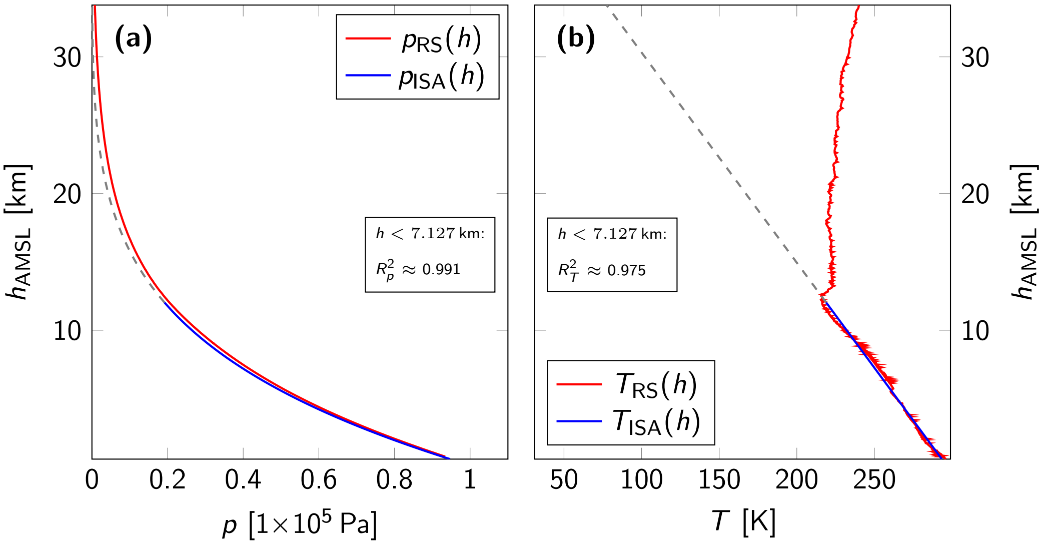


Figure 6.1: DWD radiosonde data: flight data measured during ascent from Stuttgart-Schnarrenberg up to the point of bursting. The analogous ISA-model predictions have been adjusted to fit the ground level measurements in Stuttgart-Vaihingen. **(a)** Static pressure over height. **(b)** Temperature over height.

interval, while the order of magnitude is similar to that of the measurement at a height of 108 m, in between 271 m to 380 m and at 868 m, the match is poor otherwise. With the familiar definition for the coefficient of determination, that is

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{C_{n,i}^2} - \overline{C_n^2} \right)^2}{\sum_{i=1}^{n} \left(C_{n,i}^2 - \overline{C_n^2} \right)^2}, \tag{6.1}$$

a value of $R^2 \approx 0$ is obtained, an indicator for the failure of the model to predict the variance of the data set. Qualitatively, the HV-trend for the second interval seems to fit slightly better, but again $R^2 \approx 0$. However, this is a known shortcoming of the HV-model, since the range under consideration lies deep within the, in this case turbulent, PBL.

In the following, the strong fluctuation of ε over height will be analyzed in more detail. For this purpose, the time series section from 10:55–11:00 a.m. is displayed enlarged in Fig. 6.3a. In addition, the velocity data over time at a height of 162.8 m, 217.0 m and 271.3 m has been plotted in Fig. 6.3b, Fig. 6.3c and Fig. 6.3d, respectively. A striking feature is the strong velocity gradient in Fig. 6.3d between 2 min to 3 min, where the LOS velocity changes sign and scale within a comparatively short time frame. The most likely cause for this is a statistically dependent effect, i.e., a wind gust. Comparable behavior can be seen in Fig. 6.3b at the 3 min mark.

However, all of the showcased plots feature velocity fluctuations. The resulting TDR spectra are displayed next to the respective time data. The spectra resulting from Fig. 6.3b and Fig. 6.3d exhibit a strong dependence of ε over f. The intervals of the spectra selected by the ε-estimation algorithm are marked in red, with the height of the bounding boxes being chosen such that they represent the 95 %-confidence intervals centered around $\hat{\varepsilon}$. It can be seen that, atleast in this case, the proposed approach can be used successfully to determine estimates for the TDR in suboptimal conditions with minimized uncertainty. Relative total estimation errors in these cases lie between 23.7 % to 26.5 %.

The entire lidar time series data for the return signal strength and LOS velocity is

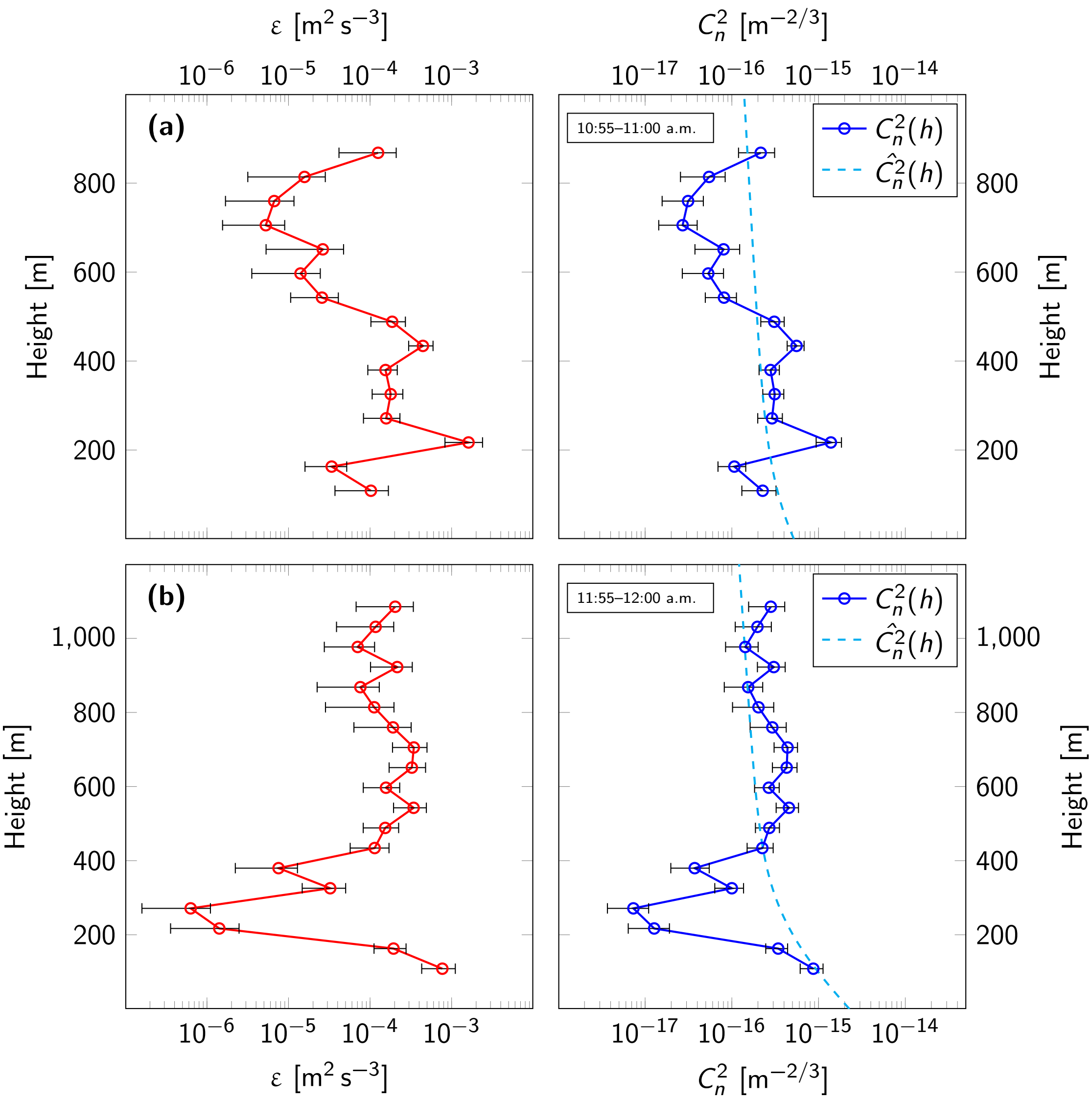


Figure 6.2: TKE dissipation rate ε and turbulent refractive index C_n^2 determined from 5 min velocity time series intervals. Severe changes in the order of magnitude of ε and C_n^2 over height are visible. HV-model curves for $W = 21\,\mathrm{m\,s^{-1}}$ were added for comparison. **(a)** 10:55 a.m. interval: HV-curve with $A = 2.5 \times 10^{-16}\,\mathrm{m^{-2/3}}$. **(b)** 11:55 a.m. interval: HV-curve with $A = 2.0 \times 10^{-15}\,\mathrm{m^{-2/3}}$.

displayed in Fig. 6.4a and Fig. 6.4b. The previously discussed velocity segments have been outlined. A stable edge of the measurable PBL can be discerned, measuring approximately 1.8 km to 2.0 km at the beginning of the campaign and slowly increasing to 2.1 km throughout the day. At the beginning, clouds rarely pass the FOV, but from 1:00 p.m. onward, cloud structures are clearly visible in the data. At this point, aliasing effects from high altitude clouds can be seen, since the velocity and height of these structures is highly irregular. The cloud structures above 4 km are interesting, since the lidar is able to deeply pierce them, meaning that the backscatter coefficient must have been relatively low, a circumstance that was rarely observed in the context of this work.

In the velocity time series, severe changes in the direction and speed of the flow can be noted between 10 a.m. and 2 p.m. These weaken as the day progresses, but never completely subside.

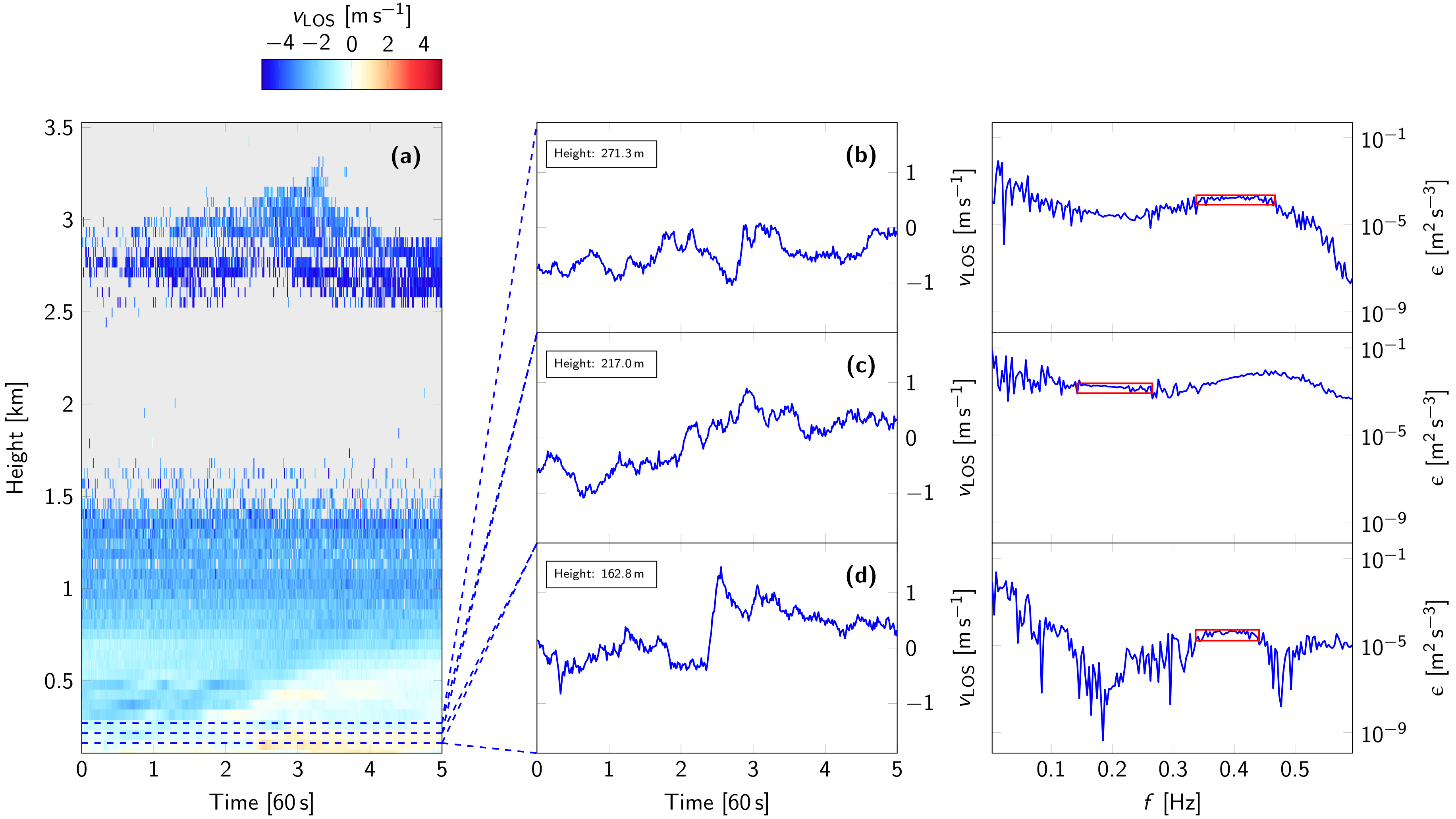


Figure 6.3: TDR spectrum comparison: **(a)** LOS velocity time series. **(b)**, **(c)**, **(d)** Selected velocity plots and TDR spectra.

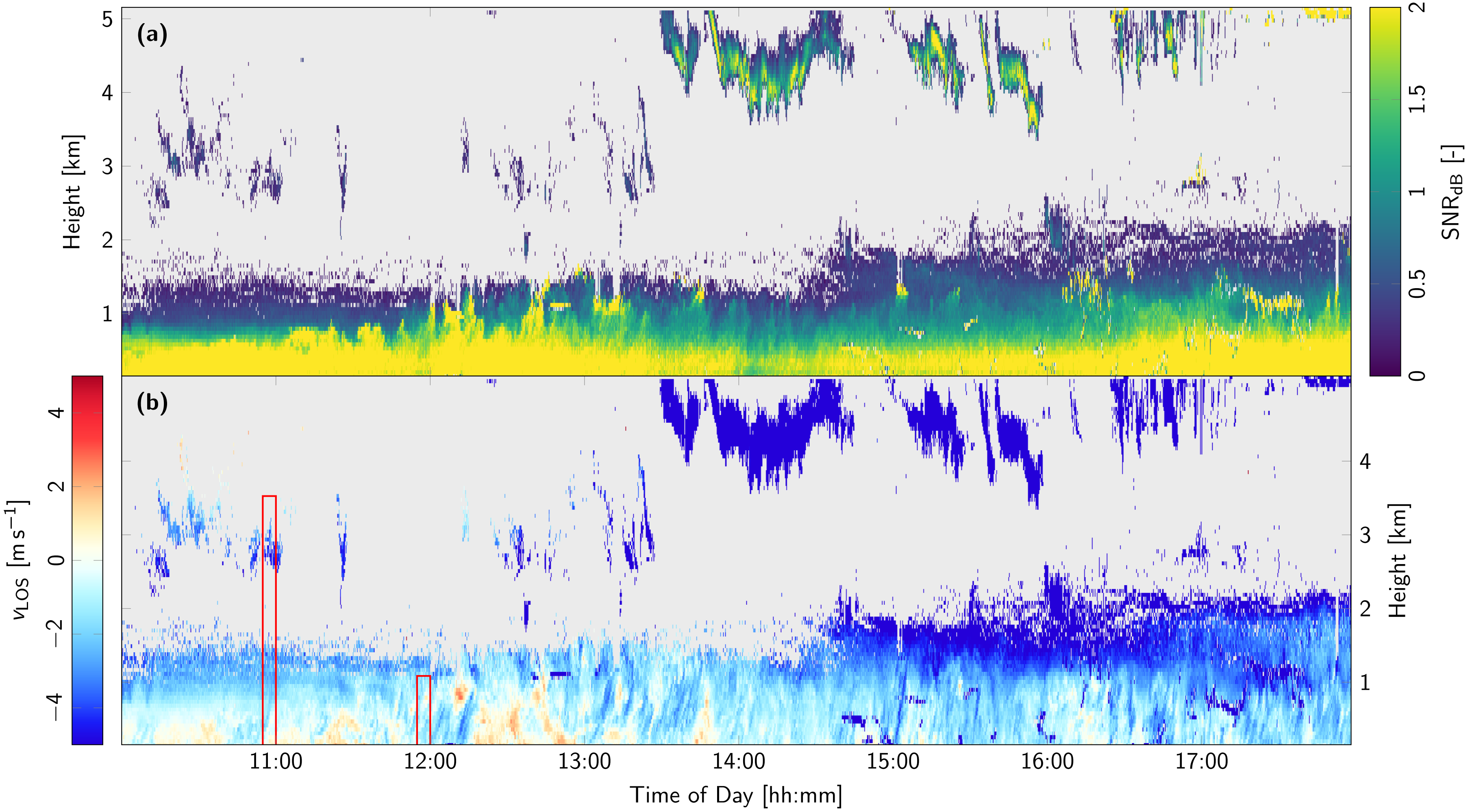


Figure 6.4: July 27th, 2021 field campaign: **(a)** SNR time series. **(b)** LOS velocity time series.

6.2 September 25th, 2021

On September 19th, 2021 the Cumbre Vieja volcano ridge on the island of La Palma erupted after four years of geological unrest. Eight days prior, a significant increase in the earthquake frequency indicated that an eruption had to be imminent. Large quantities of volcanic aerosols and gases were released throughout the event, which lasted 85 days, in addition to considerable quantities of molten lava (Fig. 6.5). Based on data provided by the ECMWF, between 3.2×10^6 kg to 5.4×10^7 kg of sulfur dioxide (SO_2) were released into the atmosphere each day [58].

Figure 6.5: Image of the La Palma volcano eruption captured around 12 p.m. on September 20th, 2021 [59]. The lava can be seen moving towards the village of El Paraiso. Initial eruptions began a day earlier.

The SO_2 itself is too small to be detected by the lidar system used for this work. However, under certain conditions SO_2 molecules can contribute to the formation of sulfuric acid (H_2SO_4) particles. Solar UV radiation can excite SO_2 in its 260 nm to 330 nm absorption band, which causes an effect referred to as photo-oxidation, whereby the excited SO_2 reacts with either adjacent SO_2 molecules or atmospheric molecular oxygen (O_2) to form sulfur trioxide (SO_3). The resulting SO_3 can then react with water (H_2O) vapor to form H_2SO_4 particles of varying size [60].

Under normal conditions, these can range between 0.1 µm to 1 µm, but can grow even larger due to accumulation and condensation effects [61]. In combination with the additional aerosols released by the volcano, i.e., ash particles, an increase in the measurable range of the PBL was expected, if the ejected plumes were to reach Central Europe. In the days following the eruption, a westerly wind carried the ejecta out into the Atlantic, but a change in weather caused a northeastward spread beginning September 23rd, 2021. ECMWF predictions indicated the arrival of the volcanic SO_2-plume in southwest Germany between 12 a.m. and 2 p.m. on September 25th, 2021, accompanied by volcanic aerosols.

It was therefore decided to carry out an unscheduled field campaign in order to record, if possible, the arrival of the SO_2 cloud. A northeasterly wind direction was predicted for the day, and the system was aligned accordingly. In order to maximize the measuring range, an inclination angle of 80° was chosen. Additional system parameters are summarized in Tab. 6.2. At 12 a.m. a temperature of 25.0 °C and air pressure of 1.018×10^5 Pa were measured at the nearby weather station of the University of Stuttgart.

Table 6.2: September 25th field campaign parameters

Date	September 25th, 2021	
Site	Stuttgart-Vaihingen	
	Latitude	48°44'55.9"N
	Longitude	9°06'09.1"E
	Alignment	49°
	Inclination	80°
Duration	Start	10:32
	End	18:00
Setup	Monostatic	
	Beam Diameter	100 mm
	Focal Length	500 mm
	PFA Current	8 A
	Detector Type	APD+LPF

The radiosonde flight data from the DWD probe that was launched from Stuttgart-Schnarrenberg at 12:48 a.m. is displayed in Fig. 6.6 and an ISA-model pressure curve with $p_0 = 1.014 \times 10^5$ Pa , as well as a temperature curve with $T_0 = 299.2$ K were added. Due to the good agreement between the data and model, the turbulent parameters will be calculated using the ISA relations for $p(h)$ and $T(h)$.

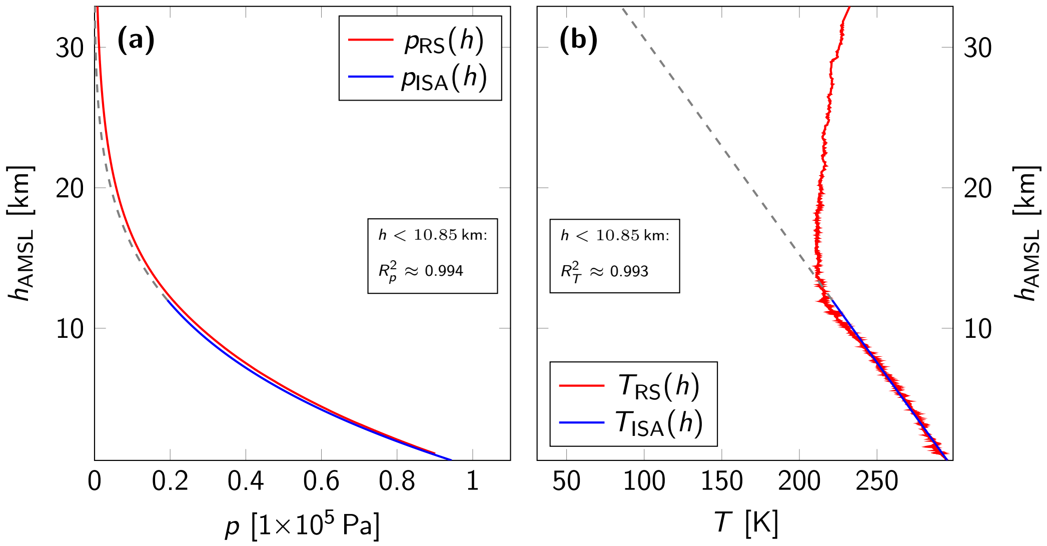


Figure 6.6: DWD radiosonde data: flight data measured during ascent from Stuttgart-Schnarrenberg up to the point of bursting. **(a)** Static pressure over height. **(b)** Temperature over height.

ε and C_n^2 were calculated for the velocity data between 05:55–06:00 p.m., since the evaluable range increased throughout the day and was highest towards the end of the campaign. The results are displayed in Fig. 6.7 with a maximum height of almost 3 km. A HV-model curve was added for comparison, but the match between C_n^2 and the model prediction is poor.

In general, the results for ε and C_n^2 must be interpreted with caution. Due to the high angle of inclination, the velocity data mainly represents upward and downward fluctuations of the flow field. Even though the scalar turbulent parameter ε should not be dependent on direction, it is known that vertical fluctuations in the atmosphere can be caused by different mechanisms than horizontal turbulence, i.e., large scale longitudinal eddies, especially at the vertical length scale resolution of the used system [7, p. 107].

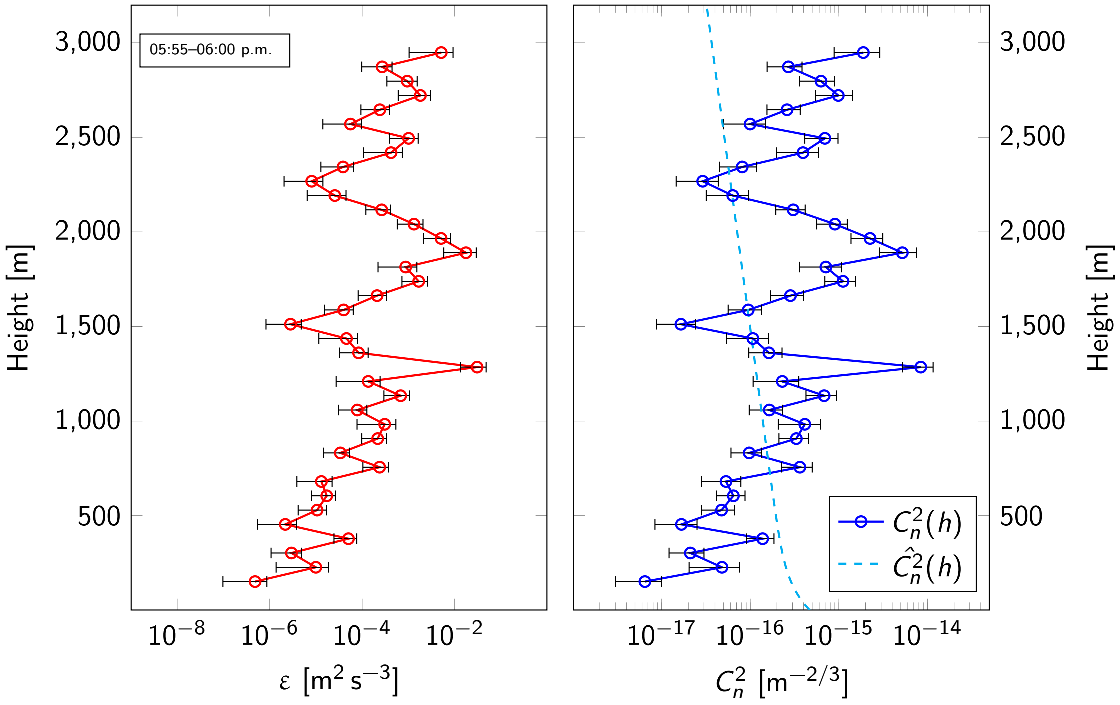


Figure 6.7: TDR ε and turbulent refractive index C_n^2 determined from the velocity interval between 05:55–06:00 p.m. on September 25th, 2021. A HV-model curve for $W = 21\,\mathrm{m\,s^{-1}}$ and $A = 2.0 \times 10^{-16}\,\mathrm{m^{-2/3}}$ was added for comparison. Due to the high angle of inclination the validity of the turbulent parameters must be viewed critically.

The lidar time series data is displayed in Fig. 6.8. The velocity segment analyzed above has been marked in Fig. 6.8b. At the beginning of the campaign a stable PBL edge at a height of 1 km is visible. While the edge height fluctuates slightly throughout the day, it remains stable. More interestingly, multilayered structures begin to appear from 11:00 a.m. onward. During this time, no clouds were visible and the very weak return signal strength in combination with the structure thickness were not observed beforehand. At 2 p.m. the return signal range has doubled, however a sharp drop in the signal strength is visible above 1 km. Between 03:30–06:00 p.m., a secondary aerosol structure appears to push downward into the PBL from above.

From previous eruptions analyzed using satellite data it is known that SO_2 quickly rises to heights of 10 km and above [62]. It is therefore assumed that no H_2SO_4 particles were observed in the presented time series, since the effective range is limited by the pulse frequency of 20 kHz. However, the signal strength of the structure that appears towards the end of the campaign is a clear indicator for a high aerosol density descending towards the surface. Data recorded by the E-PROFILE network on the same day in Stuttgart-Schnarrenberg shows qualitatively similar results [63].

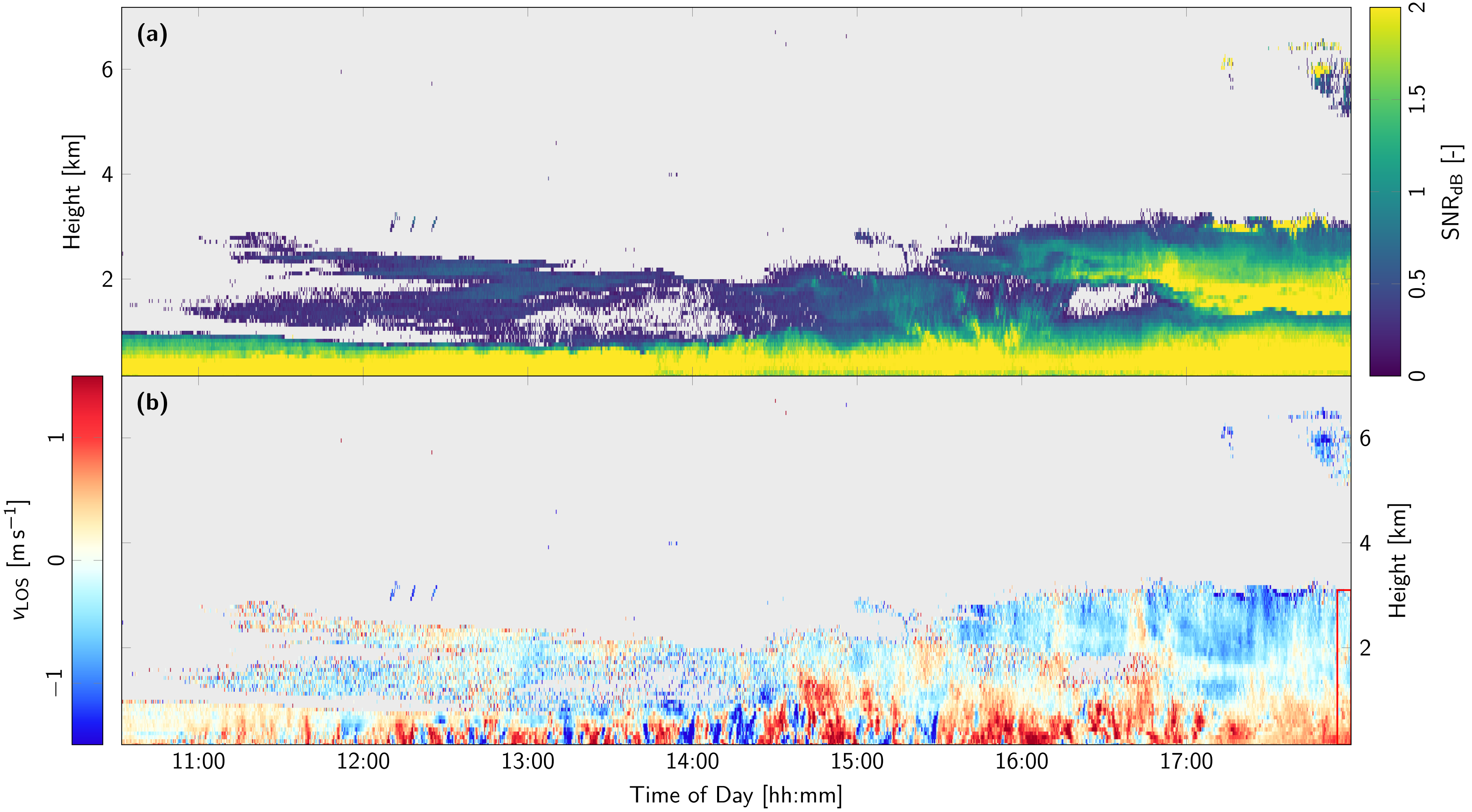


Figure 6.8: September 25th, 2021 field campaign: **(a)** SNR time series. **(b)** LOS velocity time series.

6.3 October 21st, 2021

Towards the end of October, the deep low pressure system Ignatz was responsible for wind speeds of up to 120 km h^{-1} across Germany and France [64]. The data that was obtained during the October 21st field campaign will be presented here to showcase the velocity data from this particular weather event. Initially, the recording was started at 9:18 a.m. with a northeasterly alignment. However, due to strong rainfall, the system had to be relocated and a second recording was started at 10:56 a.m. with a westerly LOS. This allowed for a better alignment with the prevailing flow direction at that point in time. Additional parameters are summarized in Tab. 6.3.

Table 6.3: October 21st field campaign parameters

Date	October 21st, 2021	
Site	Stuttgart-Vaihingen	
	Latitude	48°44’55.9"N
	Longitude	9°06’09.1"E
	Alignment	274°
	Inclination	45°
Duration	Start	10:56
	End	17:17
Setup	Monostatic	
	Beam Diameter	100 mm
	Focal Length	500 mm
	PFA Current	8 A
	Detector Type	PIN

Throughout the recording, the University of Stuttgart weather station reported a temperature of 12.0 °C and average static pressure of 1.014×10^5 Pa. At 12:46 a.m. a radiosonde was launched by the DWD from Stuttgart-Schnarrenberg. The pressure data that was returned matches the previously mentioned ground level value[1] and is displayed in Fig. 6.9a. In contrast to the good pressure correlation between the radiosonde data and the ISA-model, the measured temperature deviates from the model prediction at an unusually low altitude (see Fig. 6.9b). While in previous measurements a good fit was observed up to 12 km, on this day the agreement was satisfactory only below roughly 8 km. However, this is still sufficient to calculate C_n^2 for the altitudes of relevance for this campaign based on the ISA relations.

In addition to the atmospheric parameters, the radiosonde also periodically reports its position and rate of ascent. Using this data, the velocity vector can be calculated. Since the radiosonde has a low mass and the ascent balloon exerts high air resistance, it is assumed that the radiosonde velocity is roughly equal to that of the flow field. If the velocity component is projected onto the LOS of the lidar system, a quantitative comparison of the reported velocities can be made. In Fig. 6.10a the radiosonde ground track is plotted together with the lidar LOS for this field campaign on a map of the greater Stuttgart metropolitan area. It can be seen that the radiosonde ground track runs almost parallel to the lidar LOS while its altitude remains within the maximum lidar range. The velocity data that was recorded by the system between 12:46–12:56 a.m. is compared to the radiosonde velocity measured for the same interval. In Fig. 6.10b the resulting graphs are displayed. For the lidar time interval, the mean velocity and the envelope were calculated, that is

[1] Adjusted for height AMSL.

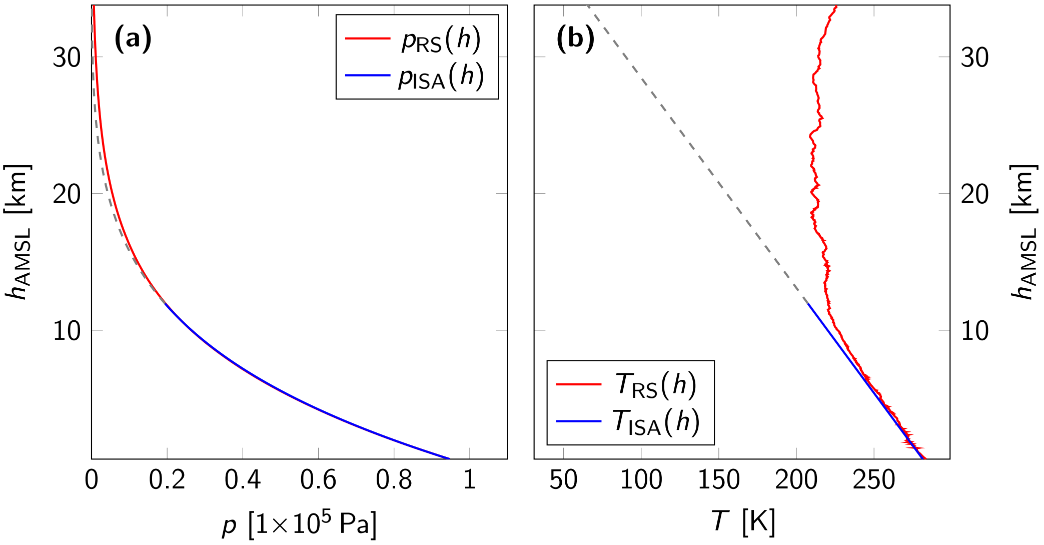


Figure 6.9: DWD radiosonde data: flight data measured during ascent from Stuttgart-Schnarrenberg up to the point of bursting. **(a)** Static pressure over height. **(b)** Temperature over height.

min[v_i] and max[v_i]. A good agreement between both data sets can be seen, where the projected radiosonde velocity course lies within the lidar velocity envelope and follows the general course of the 10 min-mean. Therefore, if it is assumed that the macroscopic flow field was similar for both locations, it can be concluded that the velocities measured by the system are plausible.

Both ε and C_n^2 were calculated exemplary for the 5:12–5:17 p.m. time interval since the evaluable height was highest here. The results are displayed in Fig. 6.11. Remarkably, the TDR magnitude for this interval does not vastly exceed the scales that were observed in previous measurements. This suggests that, even though this was a weather event with strong winds, the flow field must have exhibited some stability that damped major turbulent eddies. In addition, no special trend can be identified over height for this short-scale time interval. The same is true for the C_n^2-curve. Some consistency between the data and the HV-model can be seen below 450 m of height.

The complete time series data of the campaign is shown in Fig. 6.12. It can be seen that, overall, the aerosol density on this day was low. The SNR data in Fig. 6.12a begins with the aforementioned rainfall due to which the system had to be repositioned. It is interesting to note that immediately after the rainfall, the measurable aerosol PBL drastically sinks in height and return signal strength. This is caused due to the removal of aerosols from the atmosphere by rain drops. However, as time progresses, the initial PBL height of approximately 1.0 km is restored. Throughout the day, this height increases up to 1.5 km. Sporadically, signatures of increased SNR can be seen at the edge of the PBL. These represent clouds passing through the lidar LOS. The velocity time series displayed in Fig. 6.12b shows that, while relatively slight perturbations occur, on average, the velocity magnitude did not change in size significantly. The average measured LOS velocity throughout the field campaign was calculated to be 10.03 m s^{-1} or 36.11 km h^{-1}. The lowest and highest measured LOS velocities were 1.727 m s^{-1} and 22.48 m s^{-1}, respectively.

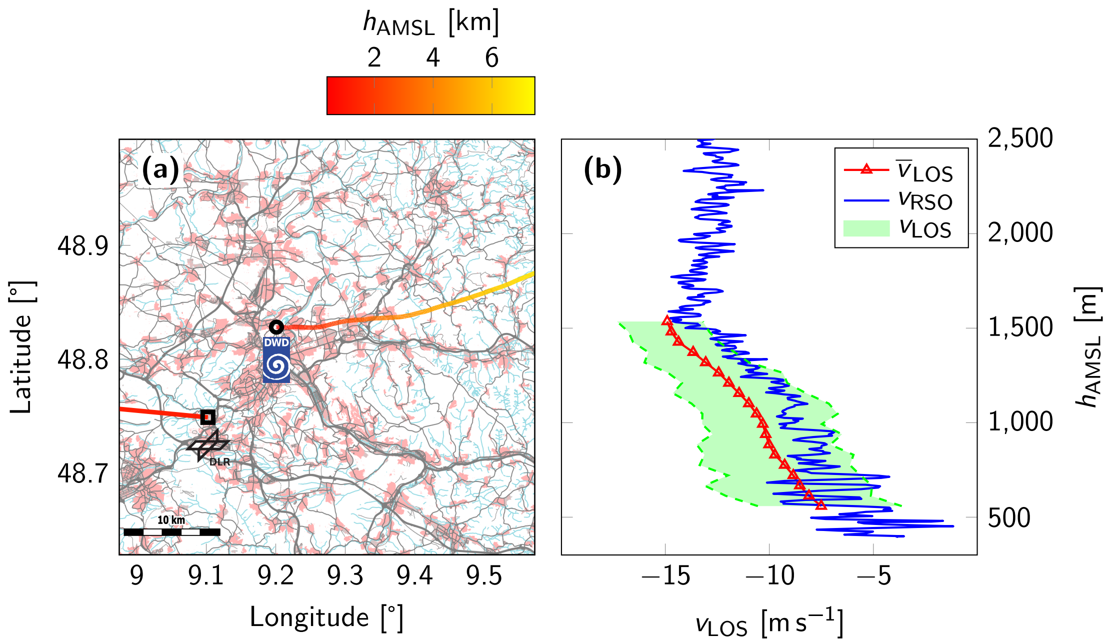


Figure 6.10: Comparison of radiosonde and lidar velocity data: the radiosonde velocity is projected onto the lidar LOS for comparability. **(a)** Map of the Stuttgart metropolitan area featuring the LOS of the lidar at the DLR facility marked in red and ground track of the radiosonde launched from Stuttgart-Schnarrenberg. The path is color-encoded to represent the height AMSL of the probe. **(b)** Comparison of lidar velocity data for the time interval between 12:46–12:56 a.m. and the according radiosonde data.

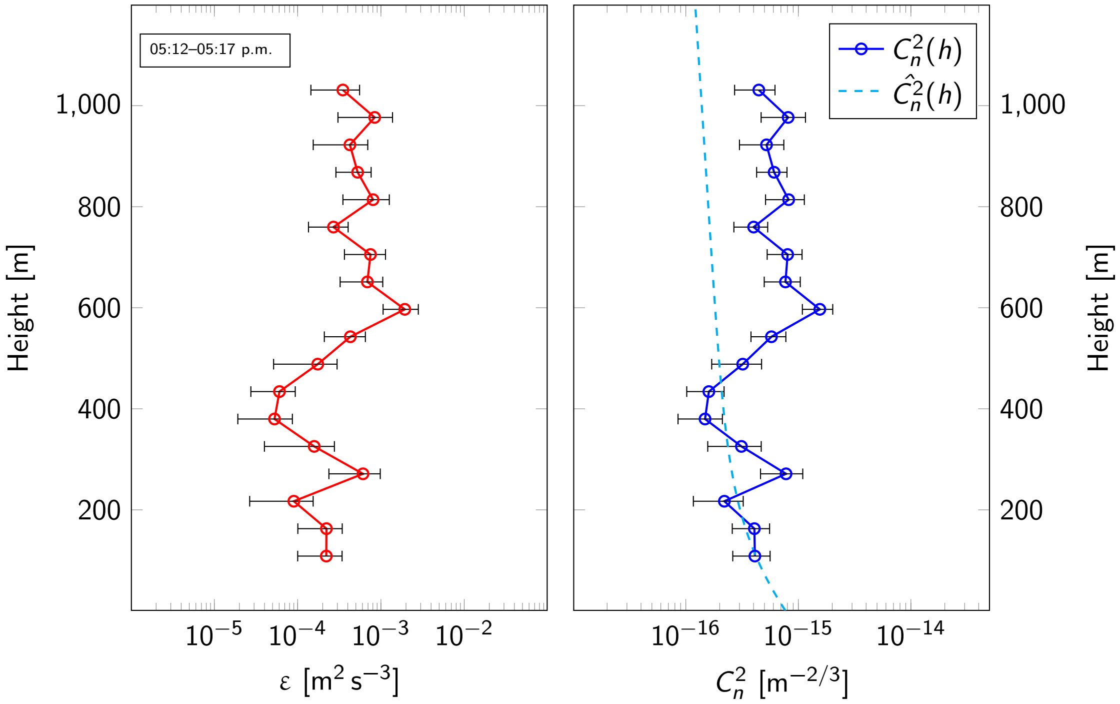


Figure 6.11: TDR ε and turbulent refractive index C_n^2 determined from the velocity interval between 05:12–05:17 p.m. on October 21st, 2021. A HV-model curve for $W = 21\,\mathrm{m\,s^{-1}}$ and $A = 5.0 \times 10^{-16}\,\mathrm{m}^{-2/3}$ was added for comparison.

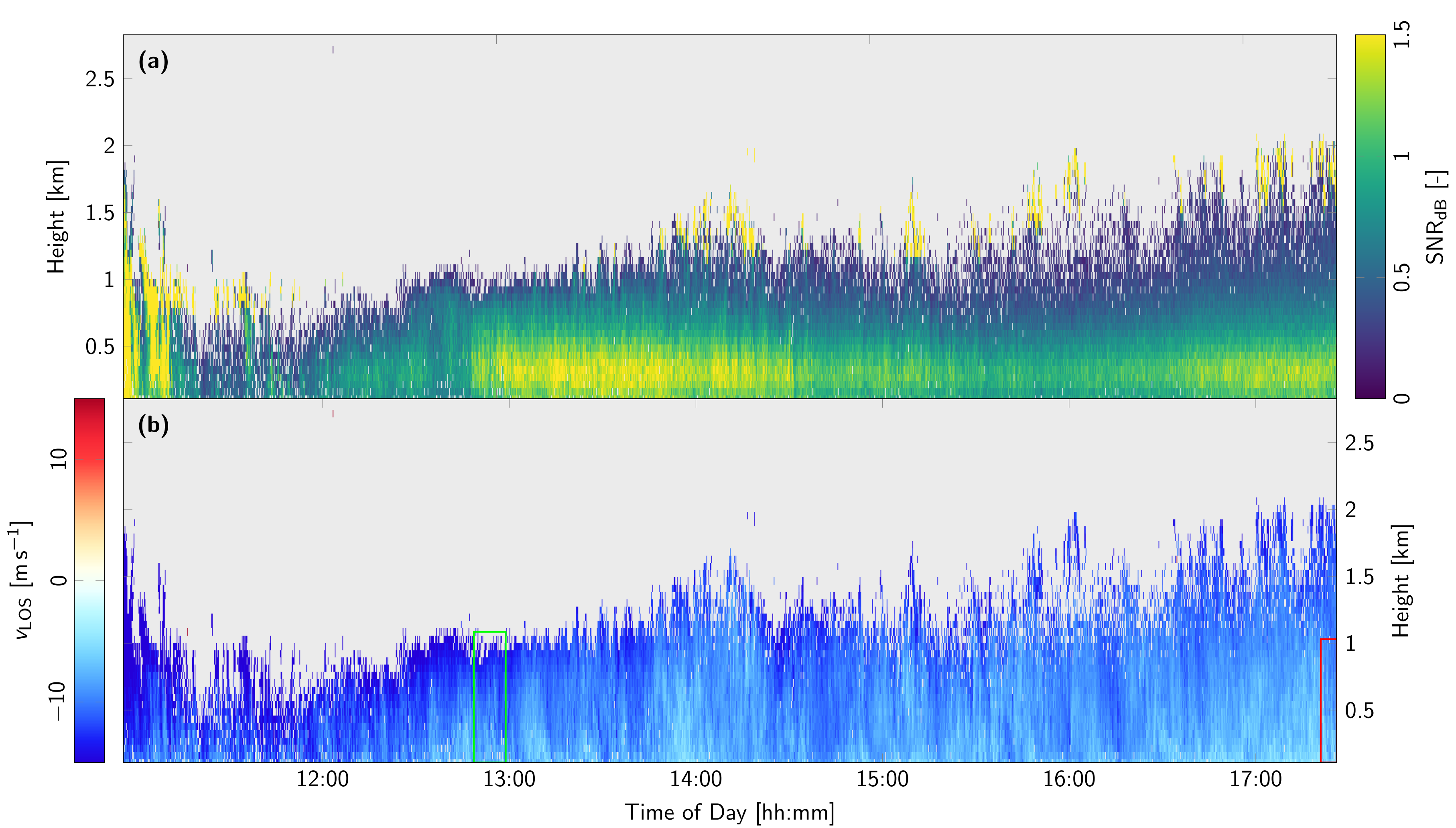


Figure 6.12: October 21st, 2021 field campaign: **(a)** SNR time series. **(b)** LOS velocity time series.

7 Conclusion

One of the main goals of this thesis, the assembly of a lidar system consisting of a transmitting and receiving optical system, as well as the implementation of the necessary software, has been successfully completed. The functionality of the system was proven in a plethora of field campaigns, lasting up to 9 h and totaling in excess of 150 h. A compact lidar system with a range of approximately 7 km is now available, which can be extended in the future by adjusting the pulse repetition rate. The system can be easily relocated and only requires minimal infrastructure at the intended site. The system is shown in action in Fig. 7.1 during the September 25th, 2021 field campaign.

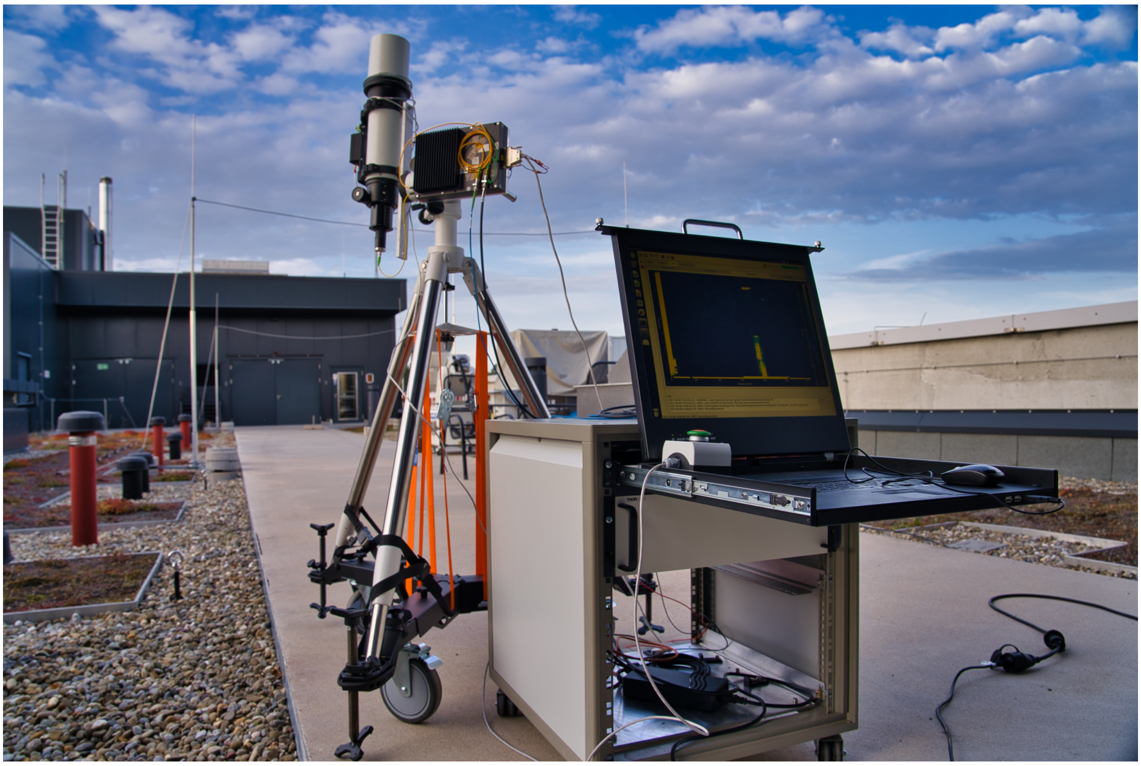


Figure 7.1: Active lidar system at the DLR site in Stuttgart-Vaihingen. This image was taken during the September 25th, 2021 field campaign intended to measure volcanic aerosols. The rack in the foreground contains the evaluation electronics and parts of the optical system. The monostatic setup as well as the pulsed fiber amplifier are mounted on the stand. A live feed of the lidar spectrum updating at about 1.2 Hz can be seen on the integrated monitor.

While a sophisticated process was developed to align the bistatic optics, no successful field tests could be conducted for this thesis. The reasons for this remain unclear and should be definitively further investigated. Although a satisfactory alignment could be achieved under laboratory conditions, in practice no signal was measured with this setup. On the other hand, good results were obtained with the monostatic transceiver. This setup proved to be very robust against vibrations, easy to maintain and initial concerns about detector saturation due to optical interface reflections turned out to be negligible.

In addition, it is possible to further increase the performance of this optical system by installing a custom aspheric lens with aberration optimization for this wavelength and application.

Moreover, the objective of measuring optical turbulence was addressed by implementing the velocity density spectrum method. The approach of the original publication was extended to increase the robustness of the estimate for ε in time-intervals of short length and under the influence of statistically dependent effects. The interval-based approach was automated and extended upon to determine a spectrum segment with optimal relative error. Besides, a global minimum in the error function can be detected even if several local minima are present. The assumptions made to implement this allegedly improved algorithm were put to the test on real-world data obtained during the field campaigns mentioned above. It was shown that the algorithm can ignore statistically dependent effects even if, for example, strong gusts occur in the underlying time interval. Further, the part of the spectrum for which the estimation error of ε is minimized, was reliably found.

Quantitative values for the vertical distribution of ε and C_n^2 were calculated over 5 min-intervals from field campaigns at different inclination angles. The results were compared with Hufnagel-Valley model predictions. Locally, an agreement between the used model and the data could be observed. This is unexpected, since the Hufnagel-Valley model is valid only to some extent for the heights considered in these cases. However, it is accordingly assumed that, in general, the order of magnitude of the derived values for C_n^2 is correct. The error propagation from ε onto C_n^2 was considered to assess the significance of the results. It should be noted that the main driver of uncertainty is the interval length of the usable spectrum. This motivates the algorithm described above.

In addition, different methods for the velocity estimation from lidar spectrum data were presented. Although comparable estimators already exist, the neural network approach by Kliebisch et al., which was further examined, presents a novel method for spectrum estimation in this context. The underlying parameter distributions of the signal-to-noise ratio model used to generate synthetic training data were investigated. This model can be expanded in the future to feature an even more realistic background noise generation procedure. Additionally, the network architecture could be enhanced to not only consider spatial, but also temporal features spanning over several past spectra in order to increase the estimation capabilities.

The software for this thesis was written using the performance-orientated Julia Programming Language and optimized for speed wherever possible and feasible. The package can be used hardware-independently for the evaluation of lidar data and offers a variety of functions to analyze the quantities presented in this work.

To accurately confirm the measured wind speeds, a measurement campaign at the DWD site in Stuttgart-Schnarrenberg is advisable. Due to the spatial proximity to the ascending radiosonde, a verification of the system and the velocity estimators would be achievable with high confidence. Alternatively, a comparison with an acoustic anemometer would be possible, although here the height resolution would not be given.

Moving forward, the capabilities of the system can be increased further. This includes the inversion of the lidar signal through the use of boundary values to obtain measurements for the atmospheric backscattering coefficient β and extinction coefficient σ. The actuation of the optics mount would allow scanning wind lidar techniques to be used to directly determine the full velocity vector for various altitudes instead of being limited to the line-of-sight velocity.

References

[1] Anthony P Tvaryanas. Epidemiology of turbulence-related injuries in airline cabin crew, 1992–2001. *Aviation, space, and environmental medicine*, 74(9):970–976, 2003. 1

[2] Leon Li and R. Jason Hearst. The influence of freestream turbulence on the temporal pressure distribution and lift of an airfoil. *Journal of Wind Engineering and Industrial Aerodynamics*, 209:104456, 2021. 1

[3] Alain Hauchecorne, Charles Cot, Francis Dalaudier, Jacques Porteneuve, Thierry Gaudo, Richard Wilson, Claire Cénac, Christian Laqui, Philippe Keckhut, Jean-Marie Perrin, Agnès Dolfi, Nicolas Cézard, Laurent Lombard, and Claudine Besson. Tentative detection of clear-air turbulence using a ground-based rayleigh lidar. *Appl. Opt.*, 55(13):3420–3428, 2016. 1, 10

[4] Patrick Vrancken, Martin Wirth, Gerhard Ehret, Hervé Barny, Philippe Rondeau, and Henk Veerman. Airborne forward-pointing uv rayleigh lidar for remote clear air turbulence detection: system design and performance. *Appl. Opt.*, 55(32):9314–9328, 2016. 1, 10

[5] Brent L. Ellerbroek. First-order performance evaluation of adaptive-optics systems for atmospheric-turbulence compensation in extended-field-of-view astronomical telescopes. *J. Opt. Soc. Am. A*, 11(2):783–805, Feb 1994. 1

[6] Hugo Weichel. Atmospheric Propagation Of Laser Beams. In Richard Gomez, editor, *Effects of the Environment on Systems Performance*, volume 0547, pages 1 – 15. International Society for Optics and Photonics, SPIE, 1985. 1, 8

[7] Francesco Tampieri. *Turbulence and Dispersion in the Planetary Boundary Layer*. Springer, Cham, 2017. 1, 7, 8, 65

[8] Vasyl Molebny, Paul F. McManamon, Ove Steinvall, Takao Kobayashi, and Weibiao Chen. Laser radar: historical prospective—from the East to the West. *Optical Engineering*, 56(3):1 – 24, 2016. 1, 3

[9] Sammy W Henderson, Philip Gatt, David Rees, and R Milton Huffaker. Wind lidar. In Takashi Fujii and Tetsuo Fukuchi, editors, *Laser remote sensing*, pages 487–740. CRC Press, 2005. 1, 3, 5, 6, 7

[10] Oliver Kliebisch and Peter Mahnke. Real-time laser doppler anemometry for optical air data applications in low aerosol environments. *Review of Scientific Instruments*, 91(9):95 – 106, 2020. 1, 13, 16

[11] Victor Banakh and Igor Smalikho. Estimation of atmospheric turbulence parameters from wind measurements with coherent doppler lidars. In *Coherent Doppler Wind Lidars in a Turbulent Atmosphere*, pages 127–178. Artech House, 2013. 1, 32

[12] Giancarlo Alfonsi. Reynolds-Averaged Navier–Stokes Equations for Turbulence Modeling. *Applied Mechanics Reviews*, 62(4), 06 2009. 040802. 2

[13] I.N. Smalikho. Accuracy of the turbulent energy dissipation rate estimation from the temporal spectrum of wind velocity fluctuations. *Atmospheric and Oceanic Optics*, 10(8), 1997. 2, 9, 32

[14] Katepalli R. Sreenivasan. On the universality of the kolmogorov constant. *Physics of Fluids*, 7(11):2778–2784, 1995. 2, 9

[15] O. E. DeLange. Optical heterodyne detection. *IEEE Spectrum*, 5(10):77–85, 1968. 5

[16] Peter Mahnke, Oliver Kliebisch, and Matthias Damm. Precise characterization of a fiber-coupled laser doppler anemometer with well-defined single scatterers. *Applied Physics B*, 125(7):1–10, 2019. 6, 15

[17] V. E. Zuev. Laser-light transmission through the atmosphere. In E. D. Hinkley, editor, *Laser Monitoring of the Atmosphere*, pages 29–69. Springer, Berlin, Heidelberg, 1976. 6

[18] Jürgen Streicher. Inversion of the lidar signal. `http://www.pa.op.dlr.de/lidar-online/invsign.html`. Accessed: 2021-07-01. 6

[19] Jürgen Harms. Lidar return signals for coaxial and noncoaxial systems with central obstruction. *Applied Optics*, 18(10):1559–1566, 1979. 6

[20] Olivier Boucher. *Atmospheric Aerosols*. Springer, Dordrecht, 2015. 7

[21] Robert Michael Pierce, Jaya Ramaprasad, and Eric C. Eisenberg. Optical attenuation in fog and clouds. In Eric J. Korevaar, editor, *Optical Wireless Communications IV*, volume 4530, pages 58 – 71. International Society for Optics and Photonics, SPIE, 2001. 8

[22] Frans T.M. Nieuwstadt, Jerry Westerweel, and Bendiks J. Boersma. *Turbulence*. Springer, Cham, 2016. 8, 9

[23] Stephen B. Pope. *Turbulent Flows*. Cambridge University Press, 2000. 9

[24] Roland Stull. Atmospheric stability. In *Practical Meteorology - An Algebra-based Survey of Atmospheric Science*, pages 119–158. The University of British Columbia, 2015. 11

[25] Robert R. Beland. Propagation through atmospheric optical turbulence. In Frederick G. Smith, editor, *Atmospheric Propagation of Radiation*, volume 2, pages 159–224. Infrared Information Analysis Center, Michigan and SPIE Optical Engineering Press, Bellingham, 1993. 11

[26] Oliver Kliebisch, Hugo Uittenbosch, Johann Thurn, and Peter Mahnke. Coherent doppler wind lidar with real-time wind processing and low signal-to-noise ratio reconstruction based on a convolutional neural network. Accepted, 2021. 13, 14, 23, 24, 47, 48, 49, 50, 52, 53, 54

[27] Herbert A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, 1926. 14

[28] Lumibird S.A. *Test Report*, 2020. Test Report V.1.0, Specifications V.8.0. 15

[29] Thorlabs Inc. *PDB570C Operation Manual*, 2019. Version 1.2. 15

[30] Thorlabs Inc. *PDB770C OEM Balanced Amplified Detect*, 2020. Version 1.1. 15

[31] Thorlabs Inc. Theoretical approximation of the output beam diameter. `https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=1696`. Accessed: 2021-10-18. 19

[32] M. E. Riley and M. A. Gusinow. Laser beam divergence utilizing a lateral shearing interferometer. *Appl. Opt.*, 16(10):2753–2756, Oct 1977. 20

[33] EHD imaging GmbH. *Data Sheet SWIR Camera IK1523*, 2013. Version 2.1. 21

[34] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017. 25

[35] Gerhard Heinzel, Albrecht Rüdiger, and R Schilling. Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new flat-top windows. *Max Planck Institute for Gravitational Physics*, 12, 2002. 28

[36] Peter Rhone. Development of the data acquisition and analysis systems for a portable raman lidar and a doppler wind lidar. Master's thesis, Faculty of Physics and Geosciences, Leipzig University, Leipzig, Germany, 2004. 28

[37] Ronny Engelmann. *Turbulent particle flux measurements with combined wind and aerosol lidar.* PhD thesis, Faculty of Physics and Geosciences, Leipzig University, Leipzig, Germany, 2009. 28, 29

[38] Frank R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974. 32, 33

[39] Philip R. Bevington and D. Keith Robinson. *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill, 1969. 39

[40] Chao Yang, Zengyou He, and Weichuan Yu. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics*, 10(1):1–13, 2009. 41

[41] Felix Scholkmann, Jens Boss, and Martin Wolf. An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals. *Algorithms*, 5(4):588–603, 2012. 41, 45

[42] BT Lottman and RG Frehlich. Evaluation of doppler radar velocity estimators. *Radio Science*, 32(2):677–686, 1997. 41

[43] Pan Du, Warren A Kibbe, and Simon M Lin. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17):2059–2065, 2006. 41

[44] R Michael Hardesty. Performance of a discrete spectral peak frequency estimator for doppler wind velocity measurements. *IEEE transactions on geoscience and remote sensing*, GE-24(5):777–783, 1986. 41

[45] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning.* MIT Press, 2016. `http://www.deeplearningbook.org`. 47, 52, 53

[46] Charu C. Aggarwal. *Neural Networks and Deep Learning.* Springer, Cham, 2018. 47

[47] Luis Mandl. Physics-informed deep differential equation solvers for coupled problems. Master's thesis, Institute of Mechanics, Structural Analysis and Dynamics, Faculty of Aerospace Engineering and Geodesy, University of Stuttgart, Stuttgart, Germany, 2021. 47, 52

[48] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989. 47

[49] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017. 47

[50] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 47

[51] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. 47

[52] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5MB model size, 2016. 47, 53

[53] Sergey I. Nikolenko. Synthetic data for deep learning, 2019. 48

[54] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. 52

[55] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989. 52

[56] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. 53

[57] M.L. Brusseau, A.D. Matthias, S.A. Musil, and H.L. Bohn. Chapter 4 - physical-chemical characteristics of the atmosphere. In Mark L. Brusseau, Ian L. Pepper, and Charles P. Gerba, editors, *Environmental and Pollution Science (Third Edition)*, pages 47–59. Academic Press, third edition edition, 2019. 58

[58] Marc-Antoine Longpré. Reactivation of cumbre vieja volcano. *Science*, 374(6572):1197–1198, 2021. 63

[59] Eduardo Robaina. Erupción volcánica de la palma de 2021. `https://commons.wikimedia.org/wiki/File:LUN_8401.jpg`, 2021. Licensed under CC BY-SA 3.0 via Wikimedia Commons. 63

[60] R. A. Cox. Particle formation from homogeneous reactions of sulphur dioxide and nitrogen dioxide. *Tellus*, 26(1-2):235–240, 1974. 63

[61] Kenneth T. Whitby. The physical characteristics of sulfur aerosols. *Atmospheric Environment (1967)*, 12(1):135–159, 1978. Proceedings of the International Symposium. 63

[62] N. M. Fedkin, C. Li, N. A. Krotkov, P. Hedelt, D. G. Loyola, R. R. Dickerson, and R. Spurr. Volcanic SO_2 effective layer height retrieval for the Ozone Monitoring Instrument (OMI) using a machine-learning approach. *Atmospheric Measurement Techniques*, 14(5):3673–3691, 2021. 65

[63] A Haefele, M Hervo, M Turp, JL Lampin, M Haeffelin, V Lehmann, et al. The E-PROFILE network for the operational measurement of wind and aerosol profiles over Europe. *Proceeding of TECO, Madrid, Spain*, 2016. 65

[64] Tagesschau. ”Ignatz” bringt orkanartige Böen. `https://www.tagesschau.de/wetter-sturm-107.html`. Accessed: 2021-12-30. 67