

BACHELOR'S THESIS IN PHYSICS



**Universitat Autònoma
de Barcelona**

FACULTY OF SCIENCE

Maximum likelihood gluing of simultaneous analogue and photon counting LIDAR measurements

June 2022

Òscar Porqueras León

Director: Dr. Markus Gaug

Abstract

In this thesis, a frequently used maximum likelihood method is expanded to take into account a more accurate statistical description of the excess noise, electronic noise and dead-time effects of a data acquisition system commonly used for LIDARs. The method is used to combine signals registered simultaneously in two channels, a process known as gluing. The original method and several proposed improvements are implemented so as to achieve gluing of analogue and photon counting measurements from the Barcelona Raman LIDAR. A complete statistical description of the related processes is derived and the performance of the new methods evaluated with the help of simulations and using real data.

Acknowledgements

Words cannot express my gratitude to my supervisor Dr. Markus Gaug for his continuous support of my study and research, for his patience, enthusiasm and motivation. I would also like to give my warmest thanks to Anna Campoy for her continuous support, enthusiasm and contributions. Their guidance and advice carried me through all the stages of writing my thesis and made this work possible.

I would also like to give special thanks to Roger Grau, who had already worked on this topic with Anna Campoy and Dr. Markus Gaug, for his contributions.

My sincere thanks also go to Dr. Lluís Font for his encouragement, enthusiasm and insightful comments.

A mamá y papá.

Contents

1	Introduction	1
2	Theoretical background	1
2.1	The Cherenkov Telescope Array (CTA)	1
2.2	Light interaction with the atmosphere	2
2.3	Raman LIDARs	2
2.4	The Barcelona Raman LIDAR	2
2.4.1	Components, Design and Characteristics	3
2.4.2	Signal digitisation and gluing	4
2.5	Dead-time-distorted processes	4
3	Analysis & Methods	6
3.1	Likelihood-based method introduced by Veberic	6
3.1.1	Statistical model	6
3.1.2	Initial values	7
3.1.3	Maximum likelihood	8
3.1.4	Corrections	10
3.2	Improvements on the likelihood-based method	11
3.2.1	Modified Gaussian parameters	11
3.2.2	Deviance using W_k	11
3.3	Treatment of summations	12
3.3.1	Sum of W_k -distributed random variables modelled as a W_k -distributed random variable	13
3.3.2	Sum of W_k -distributed random variables modelled as a random variable with other probability distributions	13
4	Results and discussion	14
4.1	Data	14
4.2	Initial values	14
4.3	Original method by Veberic	15
4.4	Modified Gaussian parameters	17
4.5	Simulating and approximating the M_k distribution	18
5	Conclusions and outlook	20
	References	22
A	Appendix	25
A.1	Mathematical expressions	25
A.1.1	Deviance derivatives from section 3.2.1	25
A.1.2	W_k derivatives from section 3.2.2	25
A.1.3	Deviance derivatives from section 3.3.2	26
A.2	Mathematical derivations	26
A.2.1	W_k derivatives	26
A.2.2	V_δ derivatives	29
A.3	Code	31
A.4	Authorship and word count declaration	31

1 Introduction

Light Detection and Ranging (LIDAR) acquisition systems are widely used to read signals in two channels simultaneously: analogue and photon counting. Their combination (gluing) allows for a significant increase in the dynamic range of the system, achieving high linearity for strong signals – thanks to the analogue channel – as well as high sensitivity for weak signals – thanks to the photon-counting mode.

Several gluing methods have been proposed, usually correcting the dead-time effects of the photon counting and then fitting the analogue signal to the photon counting [11, 21, 26, 32]. The main objectives of this thesis are implementing the method introduced by Veberic [32] so as to achieve precise gluing of Barcelona Raman LIDAR data, and proposing and evaluating several improvements and approximations which take into account a full statistical description of the dead-time affected photon-counting signal registration and summation, besides improving the description of the excess and electronic noises for the analogue signal.

2 Theoretical background

2.1 The Cherenkov Telescope Array (CTA)

Much of the radiation that reaches our planet is of thermal origin and under extreme conditions it can reach up to a few keV. However, higher energy radiation also reaches Earth and hence must be of non-thermal origin. The best-known example are the cosmic rays produced by cosmic particle accelerator processes.

When high-energy cosmic rays reach the atmosphere, a cascade of subatomic particles – known as an extensive air shower (EAS) – is produced. Some of the relativistic charged particles in the EAS are sufficiently energetic to travel through the atmosphere – a dielectric medium – faster than the phase velocity of light in it. As a result, electromagnetic radiation known as Cherenkov light is produced, similar to the sonic boom of a supersonic aircraft.

The Imaging Atmospheric Cherenkov Technique (IACT) is a method to detect high-energy gamma rays in the range of 50 GeV to 50 TeV by imaging the Cherenkov radiation produced by the EAS and reconstructing the particle shower. Several IACT systems have been operating and studying high-energy gamma rays for years, such as FACT [8], HESS [16], MAGIC [20] and VERITAS [33].

The Cherenkov Telescope Array (CTA) is a project to build the largest ground-based gamma-ray observatory in the world, which will consist of two arrays of IACT systems in the northern and southern hemispheres with a combined number of almost 100 telescopes [1, 31]. The northern array, located at the Roque de los Muchachos Observatory, will focus on the study of extra-galactic objects at the lowest energy range from 20 GeV to 20 TeV and the southern array, located southeast of the European Southern Observatory's Paranal Observatory in Chile, will cover the full energy range from 20 GeV to 300 TeV and concentrate on galactic sources.

The CTA Consortium is composed of over 1500 members from more than 150 institutes in 25 countries, including the Universitat Autònoma de Barcelona (UAB) in Spain.

The atmospheric conditions at a given moment of observation affect the measured Cherenkov light and are the main source of systematic uncertainties of IACT systems [13]. In order to tackle this issue, CTA opted for the development of Raman LIDARs, with the aim of providing continuous atmospheric monitoring through precise aerosol extinction characterisation within time scales shorter than one minute, during which the CTA telescopes change target or wobble [7, 14, 29, 30, 34].

2.2 Light interaction with the atmosphere

Electromagnetic radiation that travels through the atmosphere is subject to mainly two interactions: absorption and scattering.

Absorption occurs when molecules in the atmosphere absorb incident photons of a given wavelength. Water, carbon dioxide, and ozone are the three main atmospheric constituents which cause absorption in the photon wavelength range of interest for the IACTs (300–700 nm).

Scattering is a physical process where small particles or large gas molecules present in the atmosphere interact with part of the electromagnetic radiation, causing it to be redirected from its original path and diffused in all directions. Scattering can be elastic – when the photon energy is conserved – and inelastic – when a change in photon energy occurs.

There are three main types of elastic scattering according to the size of the particles in the atmosphere causing it: Rayleigh, Mie and non-selective scattering, which occur when particles are much smaller, about the same size and much larger than the wavelength of the incoming radiation, respectively.

Inelastic scattering of photons by matter is known as Raman scattering. Here, the molecules change their vibrational and/or rotational state through a Raman process. In consequence, as opposed to elastic scattering, there is both an exchange of energy and a change in the light's direction, hence the scattered photons are shifted in frequency. This shift is characteristic for the scattering molecule.

2.3 Raman LIDARs

Light Detection and Ranging (LIDAR) is a remote sensing method capable of determining range-resolved optical properties of a medium by using light in form of a pulsed laser and measuring the time it takes to be reflected and return to the receiver. Light is typically reflected through backscattering, as opposed to specular reflection as from a mirror.

A LIDAR's working principle is the same as that of a RADAR, except that instead of radio waves [28] another part of the electromagnetic spectrum is used – from infrared to ultraviolet [5].

LIDAR has numerous applications in multiple fields, including meteorology, archaeology, surveying, geography, geology and robotics. It has even reached our pockets by being featured in Apple's latest products for photography, AR and 3D modelling purposes.

A Raman LIDAR is a type of LIDAR instrument used to measure the vertical profiles of aerosols and water vapour within the atmosphere. The pulsed light is backscattered as a result of its interaction with the atmosphere and then collected with a telescope. These systems mainly consist of a laser, a mirror, receiving optics and detectors, a data acquisition system, and further electronics.

Raman LIDARs can operate at various wavelengths simultaneously if the laser emits at more than one wavelength. In addition to signals at the laser wavelengths due to elastic backscattering, Raman LIDAR systems are also able to detect signals at different wavelengths which emerge from inelastic (Raman) scattering. At photon wavelengths of around 300–700 nm, usually Raman scattering lines of N_2 or O_2 are considered, for they are the dominant scattering particles at this range.

The main advantage of a Raman LIDAR over a simple elastic-backscatter LIDAR is that the extinction coefficient may be measured, rather than inferred under assumptions on the scattering particle size.

2.4 The Barcelona Raman LIDAR

The Barcelona Raman LIDAR pathfinder for CTA-North (BRL) is one of the two prototypes developed for atmospheric characterisation at the future Northern Site of the CTA. BRL is a joint project

between several CTA members: IFAE-BIST, UAB & IEEC-CERES, INFN-Padova and University of Nova Gorica, Slovenia. The BRL design, construction and testing has been going on for several years, with Dr. Markus Gaug as head of the group.

The BRL, built in Barcelona, was installed at the CTA-North site in mid-February 2021 and is currently in the commissioning and evaluation phase. The final CTA-North Raman Lidar is planned to be built at the Observatorio del Roque de los Muchachos (ORM) by the end of 2023.

2.4.1 Components, Design and Characteristics

As shown in Figure 1, the BRL system consists of the following main components: an Nd:YAG pulsed laser, a telescope, a light guide, a polychromator unit, four photomultipliers (PMT) and a data acquisition unit [14].

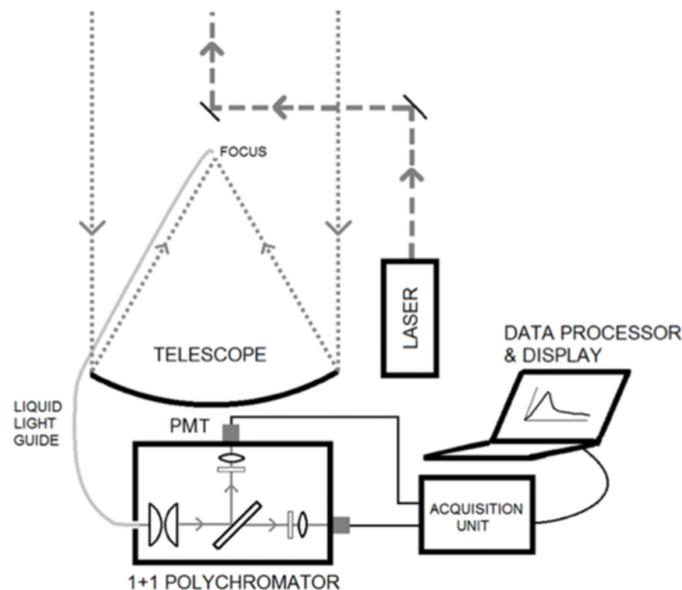


Figure 1: Schematic setup of the BRL. Adapted from [15].

The laser operates at wavelengths of 355 nm and 532 nm, which are frequency doubled and tripled from 1064 nm. The corresponding Raman wavelengths due to N_2 are 387 nm and 607 nm.

The laser beam exits along the optical axis of the receiving mirror in a coaxial configuration thanks to two dichroic guiding mirrors, which also absorb the 1064 nm wavelength.

The telescope consists of a 1.8 m diameter receiving mirror. At its focus, a wide liquid light guide (LLG) transports the backscattered light to a polychromator unit. The polychromator then separates the four wavelengths (the two laser wavelengths due to elastic backscattering, and the two corresponding Raman wavelengths) and light is detected with four PMT units.

Finally, the electric pulses generated by the PMT units are registered by data acquisition units from LICEL,¹ which digitise the signals.

The LICEL units can read the signals in two channels: analogue, useful for the near range because the signal intensity is high, and photon counting (PC), useful for the far range where the intensity is much weaker. Modern acquisition units such as those provided by LICEL combine a dual acquisition mode in which the signal is recorded simultaneously in both analogue and PC modes.

Combining analogue and PC detection, which is often called gluing, allows for a significant increase in the dynamic range of the system, achieving a dynamic range of up to $10^7 : 1$. No other amplifiers with such high dynamic ranges are available.

¹www.licel.com

Although both data records can be analysed separately, gluing of analogue and photon-counting measurements is often advantageous for data analysis. This process however is not trivial due to the fact that both channels present several drawbacks: the photon-counting channel is subject to dead-time effects and the analogue channel is subject to electronic noise, excess noise and saturation.

2.4.2 Signal digitisation and gluing

In the LICEL units, such as those used for the BRL, analogue detection of the PMT current and PC are combined in one acquisition system. The LICEL units consist of a fast transient digitiser, a discriminator for PC detection and a multichannel scalar combined with preamplifiers for both systems. For analogue detection, the signal is amplified according to the input range selected and digitised by an A/D converter [2].

The analogue voltage signal may be converted into a digital signal via an A/D converter with good linearity, but weak signals cannot be efficiently detected [10]. On the contrary, the PC mode is able to detect weak signals by counting the pulses generated by single photons.

Nevertheless, the signal intensity frequently exceeds the maximum counting rate of PC, which results in saturation of the photon-counting rate. In that case, non-linear errors introduced from pulse pile-up appear. The PC channels of the LICEL units can be modelled as non-paralysable or cumulative counters [21, 32], which introduce a so-called non-extending dead time τ . The original process – which in the case of the BRL generates the input photons – is often assumed to be a Poissonian process [6, 32] and for the purposes of this thesis it will be assumed to be so.

The analogue channel also presents several challenges which arise from the photon detection process:

When a photon reaches the PMT, N_e photoelectrons are emitted with a Poissonian probability distribution. The photoelectrons set off a chain of n Poissonian amplification processes which introduce a total gain $\alpha = \alpha_1 \cdots \alpha_n$, where α_i denotes the gain of each amplification step. The total amplification process can be approximately described by a Poisson process with mean αN_e and variance $\alpha^2 F^2 N_e$, where $F^2 = F_1^2 \cdots F_n^2$ is called the excess noise factor and F_i^2 is the excess noise factor introduced by each amplification step [9]. For LICEL units such as those in BRL, F has been found to be around $F \approx 1.06$ [22, 23]. The pedestal or baseline of the signal is dominated by electronic noise, which is approximately Gaussian with variance γ^2 .

The analogue signal may hence be described by the joint probability distribution of the Poissonian amplification process and the Gaussian electronic noise. In consequence, for a high number of photoelectrons, the central limit theorem ensures that the analogue signal converges to a Gaussian with variance $\gamma^2 + \alpha F^2 N_e$ [9]. Removing the original Poissonian contribution to the variance of the emitted photoelectrons N_e yields $\alpha \varepsilon^2 N_e$, where

$$\varepsilon^2 := F^2 - 1. \tag{2.1}$$

The analogue channel is also sensitive to saturation due to the amplifiers involved in the analogue detection process and saturation of the A/D converter when exceeding its maximum input voltage.

2.5 Dead-time-distorted processes

The dead time is the time interval τ which follows a registered event during which a counting system is unable to record further events. In general, two types of dead times are usually distinguished, which differ in the response of the system to pulses that arrive during a dead time: Non-extendable dead times are unaffected, while extendable dead times are prolonged by τ , measured from the last arrival time [19]. In Figure 2, the behaviour of these two types of dead times is illustrated.

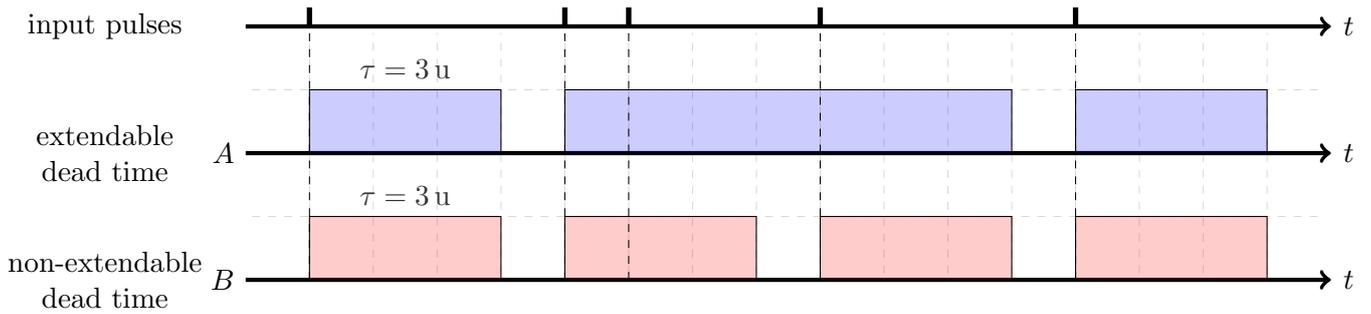


Figure 2: Schematic representation of the output count due to the effect of extendable (system A) and non-extendable (system B) dead times τ . Instead of the original 5 pulses, A would register 3 counts, whereas B would register 4. Time is expressed in arbitrary units “u” represented by the grid.

The counting losses caused by the dead-time effects depend on the statistics of the incoming pulses. In this thesis, we will assume that the input pulses follow a Poisson distribution with a mean of p_i observed photons for each time slot i , distorted by a non-extendable dead time τ . For a time origin choice at random, this process is called an equilibrium process [25] and will be referred to as such henceforth.

The exact distribution, i.e. the probability of observing a specific number of events for a Poisson process modified by dead time, along with its moments, is non-trivial to obtain. Still, it has been found by Müller [24] through renewal process theory and more recently by Omote [27] through complex integration techniques.

For the equilibrium process, the mean number of counts or mean count rate μ in a sampling time Δt is given by

$$\mu = \frac{p_i}{1 + \delta p_i}, \quad (2.2)$$

where $\delta = \tau/\Delta t$ [32].

Defining the truncated mean for k counts, $t_k := p_i(1 - k\delta)$, the probability W_k of observing k counts is given by

$$W_k := W_k(p_i, \delta) = \frac{1}{1 + \delta p_i} [R_{k-1} - 2R_k + R_{k+1} + \Delta_k], \quad (2.3)$$

where

$$\Delta_k = \begin{cases} 0 & \text{if } k \leq K - 1, \\ (K + 1)(1 + \delta p_i) - p_i & \text{if } k = K, \\ p_i - K(1 + \delta p_i) & \text{if } k = K + 1, \end{cases} \quad (2.4)$$

with $K = \lfloor 1/\delta \rfloor$ determining the upper limit on possible counts, where $\lfloor x \rfloor$ denotes the largest integer smaller than **and not equal** to x . $R_k := R_k(t_k)$ is fully expressed as

$$R_k(x) = U(x) \sum_{j=0}^{k-1} (k-j) \mathcal{P}_j(x) = U(x) [(k-x)Q(k, x) + k\mathcal{P}_k(x)], \quad (2.5)$$

where $\mathcal{P}_k(x)$ is the probability distribution for the number of counts k of a Poisson process with mean x , given by

$$\mathcal{P}_k(x) = \frac{x^k e^{-x}}{k!} \quad (2.6)$$

and $Q(k, x) = \Gamma(k, x)/\Gamma(k)$ is the regularised upper incomplete Gamma function, with the upper incomplete Gamma function

$$\Gamma(k, x) = \int_x^\infty u^{k-1} e^{-u} du \quad (2.7)$$

and $\Gamma(k) = \Gamma(k, 0)$ is the Gamma function.

For a positive integer k , the Gamma function may be written as $\Gamma(k) = (k - 1)!$, and the upper incomplete Gamma function may be written as

$$\Gamma(k, x) = (k - 1)! e^{-x} e_{k-1}(x), \quad (2.8)$$

where $e_k(x) = \sum_{j=0}^k (x^j/j!)$ is the exponential sum function [35].

Finally, $U(x)$ is the Heaviside unit step function defined as:

$$U(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases} \quad (2.9)$$

The exact expectation for the equilibrium process is given by expression (2.2) and the exact variance for the equilibrium process is given by

$$V_\delta(p_i) = \frac{2}{1 + \delta p_i} \sum_{k=0}^K [(k - t_k)Q(k, t_k) + k\mathcal{P}_k(t_k)] + H(\mu - K), \quad (2.10)$$

where $H(x) := x(1 - x)$ is the hump function.

When $\delta p_i \ll 1$, the variance asymptotically behaves as

$$V_\delta(p_i) \approx \frac{p_i}{(1 + \delta p_i)^3} \left[1 + \frac{(\delta p_i)^2}{6p_i(1 + \delta p_i)} (6 + 4\delta p_i + (\delta p_i)^2) \right]. \quad (2.11)$$

These expressions may be found in [24] and [25], but the notation introduced by Veberic [32] will be mainly used in this thesis.

3 Analysis & Methods

The implementation of all proposed methods was carried out using Python code developed in conjunction with Dr. Markus Gaug, Anna Campoy and Roger Grau. For more details, see appendix A.3.

3.1 Likelihood-based method introduced by Veberic

In this section, the gluing method introduced by Veberic [32] will be explained.

3.1.1 Statistical model

Suppose we have analogue and photon-counting signals consisting of N data points each. These data points are often called bins. Let i be the subindex denoting the i -th bin.

The transformation of the input photons p_i into the analogue signal a_i is modelled as a linear transformation

$$a_i \sim A(p_i) := \alpha p_i + \beta, \quad (3.1)$$

where α is a parameter related to the PMT and amplifier gain, and β is an offset related to the pedestal or electronic baseline of the analogue signal.

The variance $\text{Var}(a_i)$ of the analogue signal is modelled as being constant,

$$\text{Var}(a_i) = \gamma^2, \quad (3.2)$$

where γ describes the electronic noise.

The probability $P(a_i|p_i)$ of observing an analogue signal a_i given p_i input photons is modelled as a normal distribution with mean $A(p_i)$ and variance (3.2):

$$P(a_i|p_i) = \mathcal{N}_{a_i}(A(p_i), \gamma^2) = \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{(a_i - A(p_i))^2}{2\gamma^2}\right). \quad (3.3)$$

For the PC channel, the mean number of counts μ_i in a sampling time Δt is given by

$$\mu_i = C(p_i) := \frac{p_i}{1 + \delta p_i}, \quad (3.4)$$

where $\delta = \tau/\Delta t$ is the fraction of dead time versus sampling time (see section 2.5).

The probability $P(m_i|p_i)$ of observing a photon count m_i given p_i input photons is given by W_{m_i} – see equation (2.3) – but is approximated by a Poisson distribution (2.6) with parameter $C(p_i)$:

$$P(m_i|p_i) = \mathcal{P}_{m_i}(C(p_i)). \quad (3.5)$$

In LIDAR measurements, data points a_i and m_i are usually obtained through summation of N_s consecutive LIDAR returns, where N_s is a fixed parameter for each LIDAR measurement called number of shots. This is possible thanks to the fast laser-pulse repetition rates, since the atmosphere is assumed not to introduce substantial sources of additional variance during the short data acquisition time.

The models described in equations (3.1)–(3.5) are considered invariant with respect to the summation to some degree if the parameters are transformed in the following way:

$$\alpha \rightarrow \alpha, \quad \beta \rightarrow \frac{\beta}{N_s}, \quad \gamma^2 \rightarrow \frac{\gamma^2}{N_s}, \quad \delta \rightarrow N_s \delta. \quad (3.6)$$

However, for large photon numbers, the variance of the equilibrium process (2.10) depends on N_s in a non-linear way and the approximation loses validity. Veberic proposes to evaluate the variance at $N_s V_\delta(p_s/N_s)$, where p_s is the sum of the arrived photons for N_s photon-counting measurements, although alternatives to this approach will be discussed in the following sections.

3.1.2 Initial values

Gluing parameters $\alpha, \beta, \gamma^2, \delta$ are obtained through likelihood maximisation, but it is first necessary to obtain initial values for the maximisation.

The initial estimates for α, β, γ^2 are obtained by using the data points (a_i, m_i) in the weaker signal region, where a linear relationship between both signals is assumed. The initial values are obtained through least-squares minimisation of

$$\chi_{\min}^2 = \min_{\alpha, \beta} \sum_{i \in I} [a_i - A(m_i)]^2, \quad (3.7)$$

where I denotes the indices of the data points used. About 10% of the data points are usually adequate.

The initial value for the parameter γ^2 is obtained from the residuals of expression (3.7):

$$\gamma_{\text{init}}^2 = \frac{\chi_{\text{min}}^2}{M - 2} \quad (3.8)$$

where $M = n(I)$ is the number of data points used.

The initial estimate for δ is obtained by fitting the photon counts m_i to a constant in the stronger signal region, where the photon counts are saturated to the upper limit due to dead-time effects, i.e.

$$m_{\text{max}} \approx \lim_{p_i \rightarrow \infty} C(p_i) = \frac{1}{\delta} \quad \Rightarrow \quad \delta_{\text{init}} = \frac{1}{\langle m_i \rangle_{i \in J}} \quad (3.9)$$

where J denotes the indices of the data points used. About 30% of the data points are usually adequate.

The resulting initial prediction for m_i as function of a_i is given by $C((A^{-1}(a_i)))$.

3.1.3 Maximum likelihood

In statistics, the parameters of an assumed probability distribution can be estimated with observed data by a method known as maximum likelihood estimation. The method consists in maximising a likelihood function (or simply likelihood) \mathcal{L} so that, under the assumed probability distribution, the observed data is most probable.

The likelihood for Veberic's model is given by

$$\mathcal{L} = \prod_{i=1}^N \mathcal{L}_i(a_i, m_i, p_i), \quad (3.10)$$

where \mathcal{L}_i is the joint probability of observing an analog signal a_i and a photon count m_i given p_i input photons for the i -th data point:

$$\mathcal{L}_i(a_i, m_i, p_i) = P(a_i|p_i) \cdot P(m_i|p_i) = \mathcal{N}_{a_i}(A(p_i), \gamma^2) \cdot \mathcal{P}_{m_i}(C(p_i)). \quad (3.11)$$

Since likelihood functions usually consist of products of other functions, it is often useful to define an alternative function called deviance as $\mathcal{D} = -2 \log \mathcal{L}$, so that the products may be expressed as summations thanks to the properties of logarithms. Likelihood maximisation is therefore equivalent to deviance minimisation.

The corresponding deviance \mathcal{D} for the likelihood (3.10) is given by

$$\mathcal{D} = -2 \log \mathcal{L} = \sum_{i=1}^N (-2 \log \mathcal{L}_i(a_i, m_i, p_i)) = \sum_{i=1}^N \mathcal{D}_i(a_i, m_i, p_i), \quad (3.12)$$

where

$$\begin{aligned} \mathcal{D}_i(a_i, m_i, p_i) &= -2 \log \mathcal{N}_{a_i}(A(p_i), \gamma^2) - 2 \log \mathcal{P}_{m_i}(C(p_i)) \\ &= \log(2\pi\gamma^2) + \frac{(a_i - \alpha p_i - \beta)^2}{\gamma^2} + 2 \log(m_i!) + \frac{2p_i}{1 + \delta p_i} - 2m_i \log\left(\frac{p_i}{1 + \delta p_i}\right) \end{aligned} \quad (3.13)$$

The parameter γ^2 is not subject to minimisation but kept fixed at the initial estimate found through expression (3.8). If minimisation is attempted by including γ^2 , convergence is rarely achieved.

This is a drawback of this procedure which we will attempt to solve through the proposed improvements in section 3.2.

Deviance minimisation is achieved by locating an extremum, which satisfies the following conditions:

$$\nabla \mathcal{D} = 0 \Leftrightarrow \frac{\partial \mathcal{D}}{\partial \alpha} = 0, \quad \frac{\partial \mathcal{D}}{\partial \beta} = 0, \quad \frac{\partial \mathcal{D}}{\partial \delta} = 0, \quad \frac{\partial \mathcal{D}}{\partial p_i} = 0 \quad \forall i \in \{1, \dots, N\}. \quad (3.14)$$

Since only the i -th term of \mathcal{D} depends on p_i , we have that

$$\frac{\partial \mathcal{D}}{\partial p_i} = \frac{\partial \mathcal{D}_i}{\partial p_i}(a_i, m_i, p_i) = 0 \quad \forall i \in \{1, \dots, N\}, \quad (3.15)$$

that is, we obtain N equations each depending only on parameters α, β, δ and p_i .

The proposed minimisation method consists in splitting the procedure in two parts: An inner part dealing with parameters p_i and an outer part dealing with parameters α, β and δ .

For each iteration of the outer part, i.e. for some intermediate values $\tilde{\alpha}, \tilde{\beta}, \tilde{\delta}$, a minimisation of parameters p_i is achieved by solving equations (3.15), which reduce to fourth-order polynomial equations in p_i . An analytical solution for the i -th equation can hence be found as the suitable polynomial root, although for practical reasons a numerical solution is more efficient. Using Newton's method:

$$p_i^{[n+1]} = p_i^{[n]} - \frac{\mathcal{D}'_i(p_i^{[n]})}{\mathcal{D}''_i(p_i^{[n]})} \quad (3.16)$$

where $\mathcal{D}'_i(p_i^{[n]}) := \frac{\partial \mathcal{D}_i}{\partial p_i}(a_i, m_i, p_i^{[n]})$, $\mathcal{D}''_i(p_i^{[n]}) := \frac{\partial^2 \mathcal{D}_i}{\partial p_i^2}(a_i, m_i, p_i^{[n]})$ and $p_i^{[0]}$ is a suitable approximation.

In our case, $p_i^{[0]}$ was found by inverting relationship (3.1) between a_i and p_i , namely

$$p_i^{[0]} = \frac{1}{\tilde{\alpha}}(a_i - \tilde{\beta}). \quad (3.17)$$

The iterative process (3.16) is carried out until the stopping criterion $|p_i^{[n+1]} - p_i^{[n]}| < \epsilon$ is met, where $\epsilon = 1.5 \cdot 10^{-8}$ in our case.

Note that the extreme values of a function may also be found at the boundary and not be a zero of the derivative, and so the extremum over p_i found through Newton's method must be compared to the value of \mathcal{D}_i at the boundary, $p_i = 0$.

The outer part of the minimisation starts with the initial estimates for α, β, δ found in section 3.1.2. For each iteration, the inner part is first performed to determine a value for parameters p_i , so that the deviance is contracted into a form $\mathcal{D}(a_i, m_i)$ which only depends on α, β and δ . The values of these three parameters are then updated by a step of some non-linear minimisation procedure. We found the limited-memory BFGS-B (L-BFGS-B) and Nelder-Mead optimisation algorithms to be adequate in most cases.

The procedure is run until a stopping criterion is met and the final values of α, β, δ and p_i are obtained.

In order to perform the iteration from expression (3.16) the first two partial derivatives of \mathcal{D}_i with respect to p_i need be computed. We have calculated and simplified them to

$$\frac{\partial \mathcal{D}_i}{\partial p_i} = \frac{2(\alpha p_i + \beta - a_i)\alpha}{\gamma^2} - \frac{2m_i}{(1 + \delta p_i)p_i} + \frac{2}{(1 + \delta p_i)^2}, \quad (3.18)$$

$$\frac{\partial^2 \mathcal{D}_i}{\partial p_i^2} = \frac{2\alpha^2}{\gamma^2} + \frac{2(1 + 2\delta p_i)m_i}{(1 + \delta p_i)^2 p_i^2} - \frac{4\delta}{(1 + \delta p_i)^3}, \quad (3.19)$$

for which the usage of an algebraic manipulator such as SageMath² is highly encouraged and will be useful for checking more complicated expressions in the upcoming sections.

Lastly, note that the initial estimates for α and β were obtained through a methodology similar to the gluing procedure proposed by LICEL manufacturers [21] and other researchers [11]. An improvement of these gluing parameters through likelihood maximisation is hence to be expected and will be discussed in section 4.

3.1.4 Corrections

Due to the acquisition system characteristics, two corrections allow for an improvement of the gluing procedure results:

The first correction has to do with the relative offset between the analogue and PC traces: due to the different path of the input signal through the LICEL system for both channels (see [2] for further detail), it is reasonable to expect an offset between the traces. This may be addressed by introducing a shift between the analogue and PC bins before running the likelihood maximisation procedure, i.e. re-indexing the N data points as (a_{i+k}, m_i) , where $k \in \{-(N-1), \dots, N-1\}$ is the shift and $i = 0, \dots, N-k-1$ if $k \geq 0$, $i = -k, \dots, N-1$ if $k < 0$.

The second correction has to do with the data point distribution: in general, the data sets from LIDAR systems contain many more data points in the weaker signal (lower a_i) region, rather than covering the whole range uniformly. As a consequence, the parameters obtained through the likelihood maximisation process are subject to bias. To deal with this issue, Veberic [32] proposes to divide the N data points in M non-overlapping and non-empty groups, which will be weighed accordingly to compute the deviance. The corresponding weights $\{w_j\}_{j=1}^M$ are defined as

$$w_j = \frac{N}{N_j M}, \quad (3.20)$$

where N_j is the number of data points in the j -th group, so that the weights be inversely proportional to the group point density.

A new weighed deviance $\tilde{\mathcal{D}}$ may then be defined as

$$\tilde{\mathcal{D}}(a_i, m_i, p_i) = \sum_{j=1}^M w_j \tilde{\mathcal{D}}_j(a_i, m_i, p_i), \quad (3.21)$$

where $\tilde{\mathcal{D}}_j$ is the contribution of the data points in the j -th group to the deviance of expression (3.12). In other words, if I_j is the set of indices of the data points in the j -th group, then

$$\tilde{\mathcal{D}}_j(a_i, m_i, p_i) = \sum_{i \in I_j} \mathcal{D}_i(a_i, m_i, p_i), \quad (3.22)$$

where $\bigsqcup_{j=1}^M I_j = \{0, \dots, N-1\}$ and \mathcal{D}_i is given by expression (3.13).³

The grouping of the data points may be performed using many techniques, but we will use fan-like groups radiating from the lower right corner of the smallest square that contains all data points.

²www.sagemath.org

³The symbol \sqcup denotes the disjoint union, since the sets of indices of the data points groups form a partition of $\{0, \dots, N-1\}$.

3.2 Improvements on the likelihood-based method

In this section, I will propose several improvements to the gluing method introduced by Veberic.

For all the proposed methods, the initial estimates for parameters α, β, γ^2 and δ are obtained in the same way as in section 3.1.2. The deviance minimisation procedure is also completely analogous to that of sections 3.1.3 and 3.1.4, although two extra parameters γ^2, ε will be added to the outer part of the minimisation.

3.2.1 Modified Gaussian parameters

One of the drawbacks of the original likelihood-based method is the constraint upon γ^2 , which must be kept constant throughout the minimisation process to achieve convergence.

A solution to this obstacle arises from a reformulation of the electronic noise model, which Veberic [32] supposed to be the cause for a constant variance, $\text{Var}(a_i) = \gamma^2$, of the analogue signal. In reality, we have no reason to suspect a constant variance since the analogue signal variance is related to the photon detection and amplification process, whose variance depends upon the number of input photons p_i .

We will therefore model $P(a_i|p_i) = \mathcal{N}_{a_i}(A(p_i), \gamma^2 + \varepsilon^2 \alpha^2 p_i)$, where the variance has been modified to take into account the excess noise factor (see section 2.4.2),

$$\text{Var}(a_i) = \gamma^2 + \varepsilon^2 \alpha^2 p_i. \quad (3.23)$$

The initial estimate for ε^2 emerges from relationship (2.1).

Defining

$$G(p_i) := -2 \log \left[\mathcal{N}_{a_i}(A(p_i), \gamma^2 + \varepsilon^2 \alpha^2 p_i) \right] = \log(2\pi(\gamma^2 + \varepsilon^2 \alpha^2 p_i)) + \frac{(a_i - \alpha p_i - \beta)^2}{\gamma^2 + \varepsilon^2 \alpha^2 p_i}, \quad (3.24)$$

we have that the deviance for this model is given by $\mathcal{D} = \sum_{i=1}^N \mathcal{D}_i(a_i, m_i, p_i)$, where

$$\mathcal{D}_i = G(p_i) - 2 \log \mathcal{P}_{m_i}(C(p_i)). \quad (3.25)$$

We have computed the first two partial derivatives of \mathcal{D}_i with respect to p_i , which can be found in appendix A.1.

3.2.2 Deviance using W_k

The actual probability distribution for an equilibrium process is given by expression (2.3). Thus far, we have only considered its approximation by a Poisson distribution (2.6) with parameter $C(p_i)$.

We shall now consider a model where the probability $P(m_i|p_i)$ is given by (2.3). We also model $P(a_i|p_i) = \mathcal{N}_{a_i}(A(p_i), \gamma^2 + \varepsilon^2 \alpha^2 p_i)$.

First, we will only consider the case of no summation of consecutive LIDAR returns, $N_s = 1$, since the behaviour of the model with respect to summation is complex and will be further discussed and evaluated later.

The deviance \mathcal{D} for this model is given by $\mathcal{D} = \sum_{i=1}^N \mathcal{D}_i(a_i, m_i, p_i)$, where

$$\mathcal{D}_i = G(p_i) - 2 \log W_{m_i}. \quad (3.26)$$

Denoting $k \equiv m_i$, the first two partial derivatives of \mathcal{D}_i with respect to p_i are:

$$\frac{\partial \mathcal{D}_i}{\partial p_i} = \frac{\partial G}{\partial p_i} - \frac{2}{W_k} \frac{\partial W_k}{\partial p_i}, \quad (3.27)$$

$$\frac{\partial^2 \mathcal{D}_i}{\partial p_i^2} = \frac{\partial^2 G}{\partial p_i^2} + \frac{2}{W_k^2} \left(\frac{\partial W_k}{\partial p_i} \right)^2 - \frac{2}{W_k} \frac{\partial^2 W_k}{\partial p_i^2}. \quad (3.28)$$

We have computed the first two partial derivatives of W_k with respect to p_i , which can be found in appendix A.1.

3.3 Treatment of summations

In LIDAR measurements data points a_i and m_i are usually obtained through summation of N_s consecutive LIDAR returns. This has the unfortunate consequence that, for $N_s > 1$, the probability distribution of observing a given photon count m_i given p_i input photons is no longer given by expression (2.3), since the observed photon count is not given by an equilibrium process but by the sum of N_s equilibrium processes.

Therefore, the random variable M_i that describes the observed photon count is given by:

$$M_i = \sum_{j=1}^{N_s} Z_{ij} \quad (3.29)$$

where Z_{ij} are random variables with a probability distribution given by (2.3), i.e. equilibrium processes. In other words, $P(Z_{ij} = k) = W_k$.

The expected value of M_i is given by

$$\mathbb{E}(M_i) = \sum_{j=1}^{N_s} \mathbb{E}(Z_{ij}) \quad (3.30)$$

due to the linearity of the expected value.

It is reasonable to assume that Z_{ij} are independent and identically distributed (iid), since thanks to the fast laser-pulse repetition rates the atmospheric conditions can be expected to remain approximately constant during the short data acquisition time. We will hence assume that Z_{ij} have probability distributions given by (2.3), with parameters p_i and δ .

Under this assumption, the expected value of M_i can be easily computed,

$$\mathbb{E}(M_i) = \sum_{j=1}^{N_s} \frac{p_i}{1 + \delta p_i} = \frac{N_s p_i}{1 + \delta p_i} = N_s C(p_i), \quad (3.31)$$

and the variance of M_i is given by

$$\text{Var}(M_i) = \sum_{j=1}^{N_s} \text{Var}(Z_{ij}) = \sum_{j=1}^{N_s} V_\delta(p_i) = N_s V_\delta(p_i), \quad (3.32)$$

where V_δ is given by expression (2.10), since the variance of the sum of random variables is equal to the sum of each of their variances when the random variables are uncorrelated, which in particular is true for independent random variables.⁴

⁴ $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + \text{Cov}(X, Y)$. X, Y independent $\Rightarrow \text{Cov}(X, Y) = 0$ (i.e. X, Y uncorrelated).

3.3.1 Sum of W_k -distributed random variables modelled as a W_k -distributed random variable

Given that the expectation and variance of M_i scale linearly with the number of shots, ideally we would hope that the resulting probability distribution of M_i would be given by an equilibrium process (2.3) with modified parameters. Unfortunately, proving or disproving this fact happens to be a complex endeavour in which we did not succeed. In spite of this, we have studied the behaviour of the distribution of M_i according to the parameters p_i and δ in order to hypothesise a model.

A first hypothesis could be that the parameters p_i, δ of random variables Z_{ij} also scale linearly with the number of shots and M_i has a probability distribution given by (2.3) with parameters $N_s \cdot p_i$ and δ/N_s .

Nonetheless, although the resulting distribution would clearly have an expectation given by (3.31), it can be checked that its variance would not be (3.32). As a counter-example, for $\delta = 0.1, p_i = 100, N_s = 500$ we have that $N_s V_\delta(p_i) \approx 113.53$ whereas $V_{\delta/N_s}(N_s \cdot p_i) \approx 37.73$.

More complex expressions for the parameters of the presumed probability distribution of M_i should therefore be contemplated.

3.3.2 Sum of W_k -distributed random variables modelled as a random variable with other probability distributions

Another possible approach consists in seeking a better approximation for the distribution of M_i than the Poissonian distribution, for its mean $N_s \cdot C(p_i)$ coincides with the expected value of M_i (3.31) but its variance does not coincide with (3.32).

Despite this, the Poissonian distribution is an adequate approximation for some range of the parameters p_i, δ [32] and so we could consider a mixed approach where M_i be approximated by some other distribution when the Poissonian loses validity.

A candidate for this approach, although not a discrete probability distribution, is the normal distribution, since it offers several benefits: it is easier to work with than the W_k distribution, its mean and variance can be set independently unlike the Poisson distribution, and from the central limit theorem the Poisson distribution can be approximated by the normal distribution when its rate parameter is sufficiently large.

In this case, we model the probability $P(m_i|p_i) = \mathcal{N}_{m_i}(C(p_i), V_\delta(p_i))$.

We also model the probability $P(a_i|p_i) = \mathcal{N}_{a_i}(A(p_i), \gamma^2 + \varepsilon^2 \alpha^2 p_i)$, like in section 3.2.1.

The deviance \mathcal{D} for this model is given by $\mathcal{D} = \sum_{i=1}^N \mathcal{D}_i(a_i, m_i, p_i)$, where

$$\mathcal{D}_i = G(p_i) + \log(2\pi V_\delta) + \frac{(m_i - C(p_i))^2}{V_\delta}. \quad (3.33)$$

We have computed the first two derivatives of \mathcal{D}_i (and in particular of V_δ) with respect to p_i , which can be found in appendix A.1.

Furthermore, the standard ways to handle under-dispersed and over-dispersed count data include using generalised Poisson and negative binomial models, respectively [4, 12, 17, 18]. We could therefore consider an approach where we used a negative binomial distribution when the sample mean is smaller than the sample variance, a generalised Poisson distribution when the sample mean is larger than the sample variance, and a Poisson distribution when the sample mean and variance are approximately equal.

The probability mass function of a random variable X with a negative binomial distribution $\text{NB}(n, p)$ is given by

$$P(X = k) = \binom{k+n-1}{n-1} (1-p)^k p^n. \quad (3.34)$$

Several generalised Poisson models and parametrisations exist. We will consider a model usually called GP-0 or GP-1 [36]. The probability mass function of a random variable X with a generalised Poisson distribution is given by

$$P(X = k) = \frac{\lambda e^{-(\lambda+\alpha k)} (\lambda + \alpha k)^{k-1}}{k!}, \quad (3.35)$$

where $\lambda > 0$ and α is usually restricted to $(0, 1)$ so that the distribution is normalised [3].

4 Results and discussion

4.1 Data

The data used was obtained in two different times and locations:

The first data files were taken while BRL was under testing at UAB, Barcelona. The sampling frequency for the analogue channel was 20 MS/s and the number of shots $N_s = 500$.

The other data files were captured after BRL was installed at the CTA-North site at La Palma during the 2021 Cumbre Vieja volcanic eruption. The sampling frequency for the analogue channel was 40 MS/s and the number of shots $N_s = 1001$.

Each data set consists of a CSV file with two columns, corresponding to the analogue and PC values for each bin (row). The files include a header with information about their corresponding number of shots and sampling frequency.

The analogue data points a_i for each bin i are given in mV units. Admittedly, a_i should have units of voltage \times time, since the digitisation process integrates the input signal over a sampling time Δt . Thus, if the values are multiplied by the sampling time Δt expressed in ns (50 ns and 25 ns for the Barcelona and La Palma data files, respectively), a_i would be properly expressed in pVs. Even so, it is commonplace to express the analogue signal in units of voltage, since the sampling time is constant for all bins, so these will be the units used across this whole section.

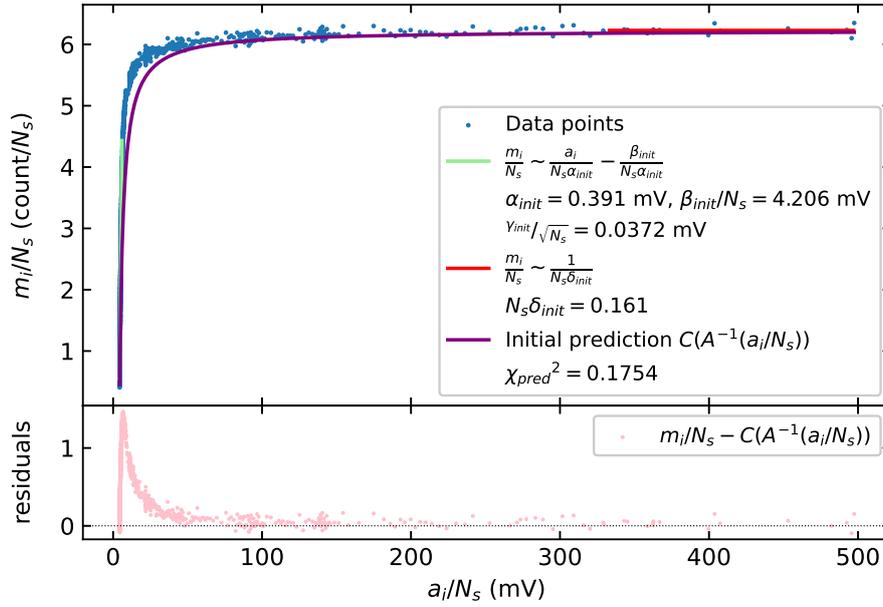
Due to the data acquisition configuration of the BRL team, a custom correction of the relative offset between the analogue and PC traces has already been implemented to obtain the data sets and will hence not be discussed here.

4.2 Initial values

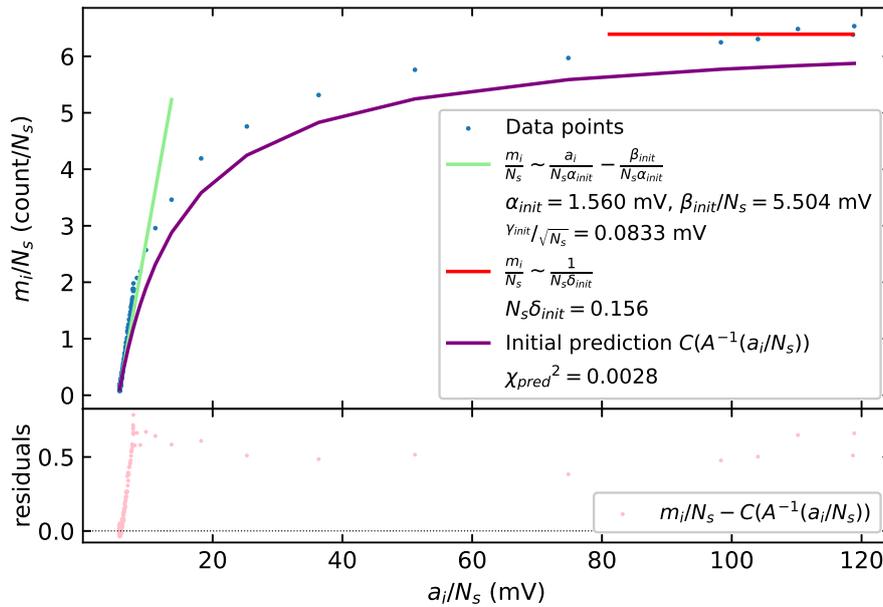
An example of the initial estimates found for one data set of each type can be seen in figure 3. Note that the data points a_i, m_i have been scaled by $1/N_s$ so that the initial values from different data sets may be compared. The initial estimates will be scaled accordingly in their respective likelihood maximisation processes.

The χ^2 of the initial prediction $C(A^{-1}(a_i/N_s))$, along with the residuals $m_i - C(A^{-1}(a_i/N_s))$ are also included in figure 3. These parameters will be used to determine whether and which of the proposed gluing methods improve upon this prediction.

For the models where an initial value of ε is needed, it is found as $\varepsilon_{init}^2 = F_{init}^2 - 1 = 0.1236$, where $F_{init} = 1.06$ is the estimated value of the excess noise of the PMT introduced in section 2.4.2.



(a) Data set "544669_500shots" captured in Barcelona.



(b) Data set "20210921.194500_1001shots" captured in La Palma.

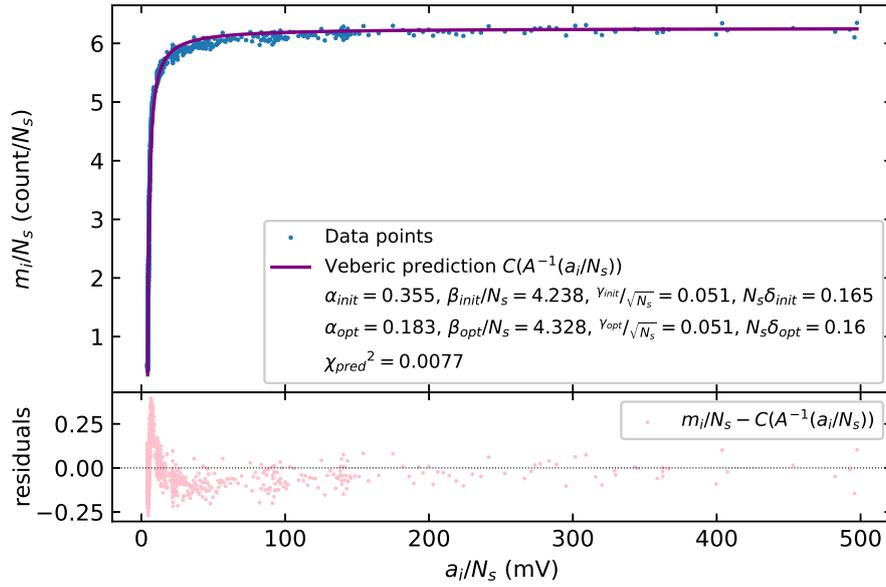
Figure 3: Initial values obtained for a data set of each type.

4.3 Original method by Veberic

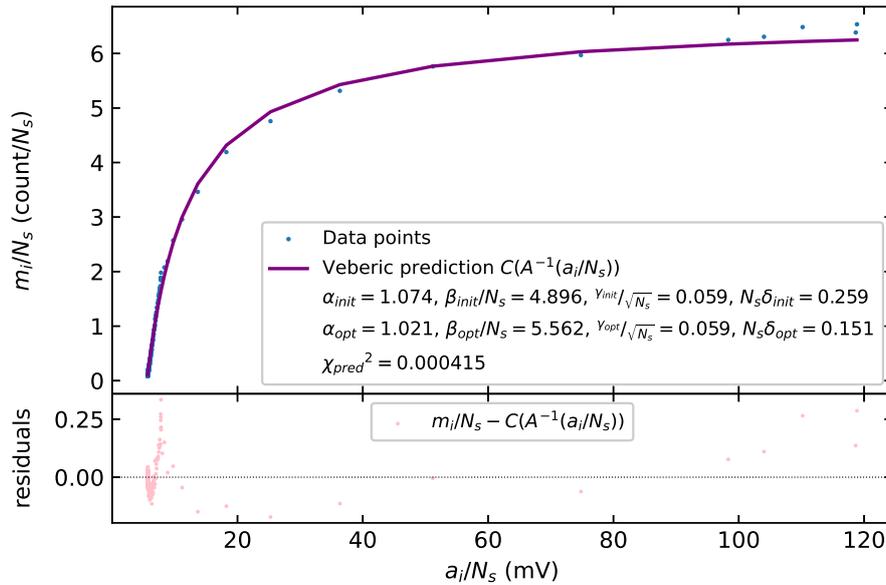
The number of weighed fan-like groups to compute the deviance was $M = 100$ in all cases. We found that a lower number of groups resulted in too many data points per group and either the method did not converge or the resulting fits had higher χ^2 values and residuals, whereas a higher number of groups resulted in too few data points per group and the resulting fits also showed higher χ^2 values and residuals.

An example of the gluing parameters and prediction found using the likelihood-based method by Veberic for one data set of each type can be found in figure 4. The χ^2 value and the residuals of

the prediction are also included.



(a) Data set “544669_500shots” captured in Barcelona.



(b) Data set “20210921.194500_1001shots” captured in La Palma.

Figure 4: Gluing fit obtained through Veberic’s method for a data set of each type. α , β and γ are expressed in mV and δ is dimensionless.

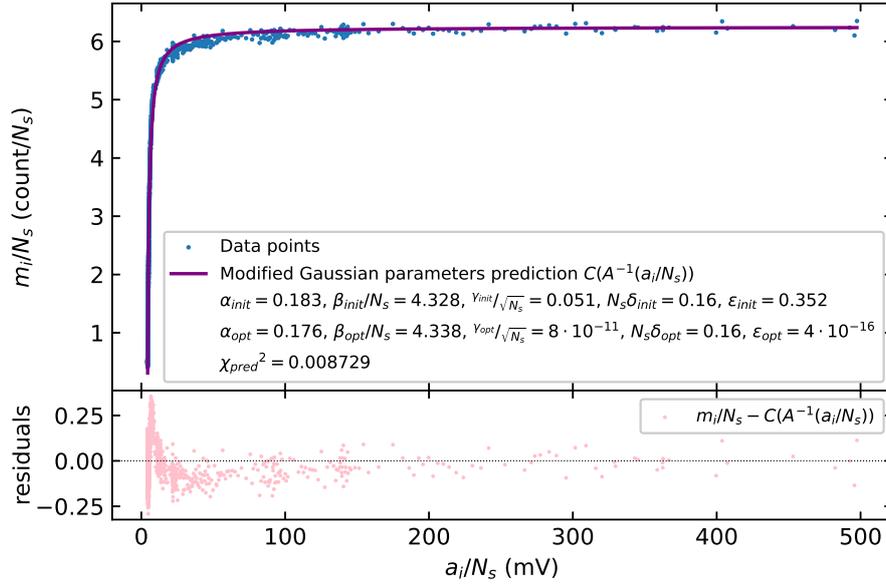
A significant improvement of the fit parameters is achieved through the implementation of this method: in comparison to the initial prediction from figure 3, the χ^2 of the fit is reduced by factors of 23 and 6 and the maximum residual values are reduced by factors of 4 and 2, respectively. Furthermore, the residuals appear to be more symmetrically distributed around zero, instead of showing a clear bias toward positive values as in the initial prediction from figure 3.

These improvements are consistently observed for other data sets: the χ^2 values are reduced by factors ranging 5–25, and the maximum residual values are reduced by factors ranging 2–5.

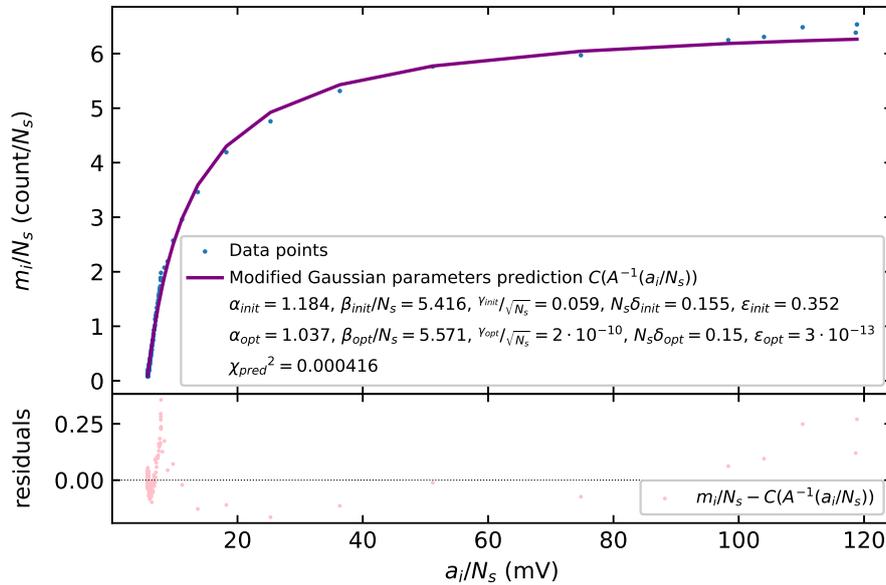
4.4 Modified Gaussian parameters

For this model, the gluing parameters α, β, γ and δ were initialised using the obtained values in section 4.3 in order to accelerate convergence. The number of weighed fan-like groups to compute the deviance was also $M = 100$ in all cases.

An example of the gluing parameters and prediction found using the modified Gaussian parameters from section 3.2.1 for one data set of each type can be found in figure 5. The χ^2 value and the residuals of the prediction are also included.



(a) Data set “544669_500shots” captured in Barcelona.



(b) Data set “20210921.194500_1001shots” captured in La Palma.

Figure 5: Gluing fit obtained through the modified Gaussian parameters method for a data set of each type. α, β and γ are expressed in mV and δ, ϵ are dimensionless.

Again, in comparison to the initial prediction from figure 3, a significant improvement of the fit

parameters is observed: the χ^2 decreases by factors of 20 and 6 and the maximum residual values are reduced by factors of 4 and 2, respectively. The residuals also seem more symmetrically distributed around zero. These improvements are also observed in other data sets: the χ^2 decreases by factors ranging from 5–25, and the maximum residual values decrease by factors ranging from 2–5.

The fit parameters (χ^2 and residuals) obtained for the original method by Veberic and the modified Gaussian parameters method consistently show relative differences below 10%. In addition, the γ_{opt} and ε_{opt} values obtained in the modified Gaussian parameters method are, respectively, 8–9 and 12–15 orders of magnitude lower than those obtained from the initial estimates. These two observations indicate that, in the used data sets, the correlated fluctuations due to the input photons do not have a significant effect on the variance of the signals in comparison to the electronic noise.

As a consequence, the assumption of a constant variance for the analogue signal made by Veberic can be considered an acceptable approximation for these data sets, since the gluing fit parameters are substantially improved with respect to the initial estimates and no further improvement is observed by considering a dependence on the number of input photons.

4.5 Simulating and approximating the M_k distribution

In order to evaluate the behaviour of the probability distribution of the M_k random variable from equation (3.29), we have conducted over 1000 simulations with different parameters p_i, δ and N_s that cover and expand the range of values that appeared in the implementation of the previous sections. Every simulation of M_k was obtained by generating N_s samples of size $S = 10^5$ of a $W_k(p_i, \delta)$ distribution and summing the produced values, which results in a sample of size S of the random variable M_k . The simulation of the probability distribution is then obtained by creating the corresponding density histogram of the M_k sample.

After conducting the simulations, we have attempted to fit the simulated probability distribution by using the distributions described in section 3.3. In figure 6, four examples of the simulations, fits and fit parameters can be seen.

We have observed that the M_k distribution is generally under-dispersed. Consequently, the negative binomial fit has only been possible for a small range of parameters. The generalised Poisson fit overcomes this issue since it may be used for either over-dispersed and under-dispersed data. However, due to the numerical behaviour of the exponential and factorial functions on its probability distribution, the computation of the corresponding probabilities has only been possible for $k \lesssim 100$ approximately. Since the generalised Poisson has been fitted by likelihood maximisation, fitting has been possible for higher k values,⁵ though unfortunately unsuccessful for $p_i\delta \gg 1$.

The W_k and Gaussian fits were possible for the whole parameter range and presented the best fit parameters: χ^2 values between 10^{-4} and 10^{-8} , R^2 values over 0.95. In the extreme example of an almost completely saturated PC channel (figure 6d), the Gaussian fit was also plotted continuously to emphasise that the fitted distribution is in truth a (continuous) probability density function.

Finally, the fitted Poissonians were considered with parameters $N_s C(p_i)$ as in the gluing process. Notice that, as opposed to the generalised Poissonian, the representation was possible for high k values, since the used Python functions have been programmed to consider binomial or normal approximations whenever necessary – we did not find libraries with such approximations for the generalised Poissonian. The relative differences between the χ^2 and R^2 values of the Poissonian fit and the W_k and Gaussian fits have been found to increase as $p_i\delta$ does. In consequence, we have observed that the Poisson distribution may not be considered an appropriate approximation for $p_i\delta \gtrsim 1$, as can be seen in the examples of figure 6(b–d).

⁵Since the factorial $k!$ is a constant that may be disregarded during likelihood maximisation and by using the log-likelihood the exponential behaviour becomes linear.

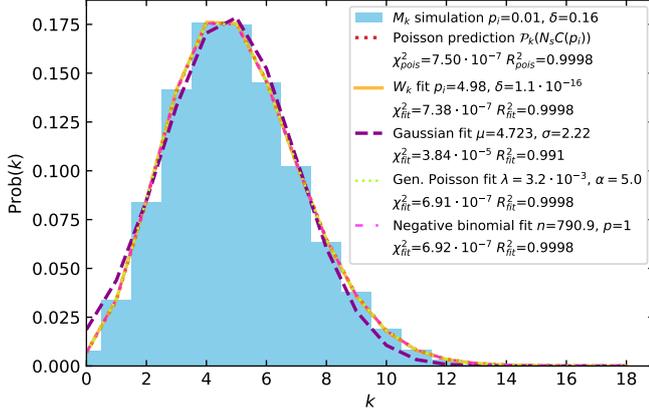
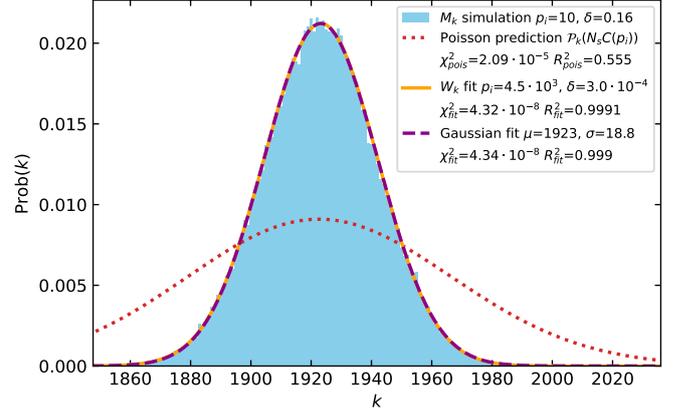
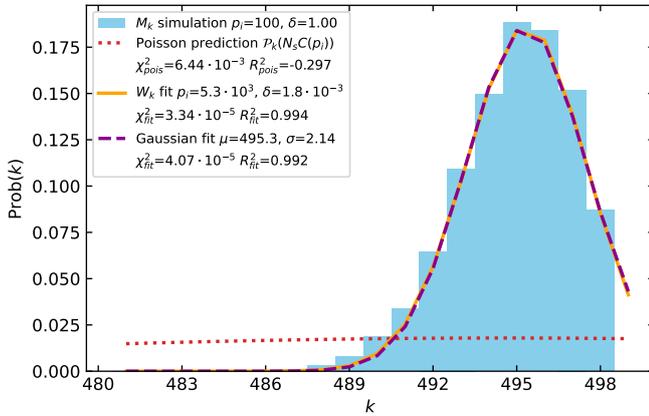
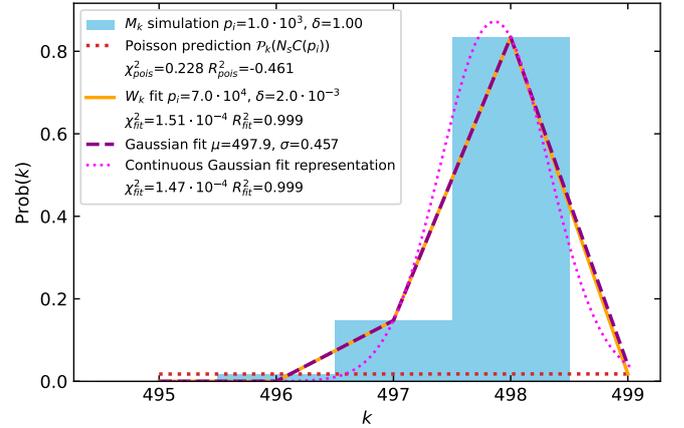

 (a) $p_i = 10^{-2}$ and $\delta = 0.16$

 (b) $p_i = 10$ and $\delta = 0.16$

 (c) $p_i = 10^2$ and $\delta = 1$

 (d) $p_i = 10^3$ and $\delta = 1$

Figure 6: Simulated distributions of the $M_k = \sum_{j=1}^{N_s} Z_{kj}$ random variable for $N_s = 500$ and different values of p_i and δ , along with W_k , Poissonian, Gaussian, negative binomial and generalised Poissonian fit parameters whenever fitting was successful. A sample size of 10^5 values was used for each Z_{kj} .

In light of the above, apart from the Poissonian approximation that was used in the original and the modified Gaussian parameters methods, the other viable approximations to the distribution of M_k are the normal and W_k distributions, since they attain the best fit parameters and were valid for the whole studied parameter range, which includes the N_s , δ and p_i values found in the implementation of sections 4.2–4.4.

Implementation of the likelihood maximisation using W_k has been unsuccessful due to lack of convergence of the optimisation algorithms, which we suspect is caused by a combination of two factors: poor choice of the initial parameters, particularly the initial guess $p_i^{[0]}$ of the Newton's method described in equation (3.16), and non-smooth behaviour of the functions involved in the fit.

Since the distribution of M_k and in consequence the parameters of a W_k fit are unknown, it is most likely that the fit parameters chosen are inaccurate. Considering the resulting W_k fits of the simulations, we have tried to hypothesise models for the p_i and δ values of the fit as function of the initial parameters, but unfortunately were not able to propose a grounded model. We attempted maximisation by using the initial values, scaled accordingly with N_s , and we also tried changing the initial Newton's method guess $p_i^{[0]}$ so that the resulting mean coincides with $N_s C(p_i)$. None of these attempts were successful.

On the contrary, the normal fit parameters may be estimated since they directly relate to the M_k distribution mean and variance, which are known (as seen in section 3.3.2). Unfortunately, the summations involved in the likelihood maximisation procedure (particularly on V_δ and its partial derivatives) highly increase the number of calculations per iteration. For this reason, implementation of the likelihood maximisation by using this method was extremely slow computationally: in the order of hours, as opposed to the other methods which run during few seconds. Naturally, under these circumstances the method becomes impractical.

Nonetheless, we did perform several likelihood maximisations using this normal approximation (see section 3.3.2), but the results were worse than the initial prediction. This indicates that the optimisation algorithms are converging to other local extrema, which was frequent throughout the implementation of the other methods and corrected through tuning of the number of weighed fan-like groups, re-scaling of parameters, tuning or switch of the optimisation algorithm and/or other minor corrections. Unfortunately, performing these corrections was impractical due to the high time scales required to perform each optimisation.

5 Conclusions and outlook

We have successfully implemented the likelihood-based maximisation method introduced by Veberic to achieve gluing of BRL data, improving the procedure proposed by manufacturers. We improve the χ^2 and the maximum prediction residuals of our gluing fit by up to a factor of 25 and 5, respectively.

Additionally, we have improved and expanded the statistical description of dead-time and electronic noise, which I believe to be an important contribution to the field. Some of these improvements have been successfully implemented and yield similar results to the method by Veberic. Conveniently, the improved algorithm allows for optimisation of the electronic noise parameter γ^2 , which was not possible with Veberic's method. In particular, we have noticed that the analogue and photon-counting variances are correlated, which implies that the assumption made by Veberic of a constant analogue signal variance, independent of p_i , is incorrect. Nonetheless, the results obtained indicate that for the BRL data sets Veberic's assumption can be considered a reasonable approximation, since the correlated fluctuations due to the input photons do not have a significant effect on the variance of the signals in comparison to the electronic noise.

Since maximum-likelihood estimation does not minimise fit residuals, but maximises the probability of observing the data under an assumed statistical model, it is consistent to obtain fit residuals with Veberic's method that are not optimal if and where the assumed model is inaccurate. In our analysis, we have found that the worst predictions lie in the transition region between analogue and photon counting where the dead-time effects become significant, but the signal does not yet saturate. This is to be expected, since the Poissonian model employed by Veberic fails to account for the under-dispersion caused by the dead-time count losses.

For this reason, in this thesis, we have studied different possibilities to introduce a complete statistical description of the dead time process using the W_k distribution instead of the Poissonian. We were unable to formally derive the necessary probability distribution of the sum M_k of W_k -distributed random variables, although we have shown through simulations that the normal and W_k distributions are good approximations for the relevant parameter range.

The main limitations of the new methods are computational, and the following optimisation considerations need be considered:

- The choice of optimisation algorithms and initial parameters had great influence on the computation time and convergence. In particular, re-scaling and iterative initialisation of fit parameters are crucial to achieve satisfactory fits to the simulated probability distribution of M_k .

- The used optimisation algorithms have not been provided with an analytically calculated Jacobian to find the extrema and hence make numerical approximations by slightly modifying parameters. This procedure assumes that functions to be optimised are smooth, which is not the case for the W_k distribution.
- The accurate descriptions of dead time, particularly when involving the variance V_δ of W_k and its partial derivatives, increase the number of calculations per iteration significantly, which causes some models to become impractical due to computation times of the order of hours on a single 2.2 GHz CPU.

I believe that our contributions to the description of dead time and further studies in this direction will allow for improvements on the description of dead-time affected acquisition systems, not only in LIDARs but also in other fields.

References

- [1] Actis, M., et al., 2011. Design concepts for the Cherenkov Telescope Array CTA: an advanced facility for ground-based high-energy gamma-ray astronomy. *Experimental Astronomy* 32, 193–316. doi:10.1007/s10686-011-9247-0. arXiv:1008.3703.
- [2] BostonElectronics, 2016. The Licel Transient Recorder. URL: <https://www.boselec.com/wp-content/uploads/PhotonCounting/Licel/DifferentialAbsorbtionLIDARDIALRecordersandDetectorsCatalog2-1-16USB.pdf>.
- [3] Consul, P.C., Jain, G.C., 1973. A generalization of the poisson distribution. *Technometrics* 15, 791–799. URL: <https://www.tandfonline.com/doi/abs/10.1080/00401706.1973.10489112>, doi:10.1080/00401706.1973.10489112, arXiv:<https://www.tandfonline.com/doi/pdf/10.1080/00401706.1973.10489112>.
- [4] Consul, P.C., Shoukri, M.M., 1985. The generalized Poisson distribution when the sample mean is larger than the sample variance. *Communications in Statistics - Simulation and Computation* 14, 667–681. URL: <https://doi.org/10.1080/03610918508812463>, doi:10.1080/03610918508812463, arXiv:<https://doi.org/10.1080/03610918508812463>.
- [5] Cracknell, A.P., 2007. Introduction to remote sensing. CRC press.
- [6] Donovan, D., Whiteway, J., Carswell, A.I., 1993. Correction for nonlinear photon-counting effects in lidar systems. *Applied optics* 32, 6742–6753.
- [7] Doro, M., Gaug, M., Pallotta, J., Vasileiadis, G., Blanch, O., Chouza, F., D’Elia, R., Etchegoyen, A., Font, L., Garrido, D., et al., 2014. Status and motivation of Raman LIDARs development for the CTA Observatory. arXiv preprint arXiv:1402.0638 URL: <https://arxiv.org/pdf/1402.0638.pdf>.
- [8] FACT, 2011. (First G-APD Cherenkov Telescope) URL: <https://www.isdc.unige.ch/fact/>.
- [9] Fegan, S., 2014. Gain estimation with a light pulser. Internal NectarCAM report, Reference: MST-CAM-TN-0060 .
- [10] Feng, C.Z., Liu, B.Y., Liu, J.T., Wu, S.H., 2016. Gluing lidar signals detected in analog-to-digital and photon counting modes. *EPJ Web of Conferences* 119, 25021. doi:10.1051/epjconf/201611925021.
- [11] Gao, F., Veberič, D., Stanič, S., Bergant, K., Hua, D.X., 2013. Performance improvement of long-range scanning Mie lidar for the retrieval of atmospheric extinction. *Journal of Quantitative Spectroscopy and Radiative Transfer* 122, 72 – 78.
- [12] Gardner, W., Mulvey, E.P., Shaw, E.C., 1995. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin* 118, 392. URL: <https://doi.org/10.1037/0033-2909.118.3.392>, doi:10.1037/0033-2909.118.3.392.
- [13] Gaug, M., 2017. CTA Atmospheric Calibration, in: *European Physical Journal Web of Conferences*, p. 01003. doi:10.1051/epjconf/201714401003.
- [14] Gaug, M., Blanch, O., Çolak, M.S., Doro, M., Font, L., Griffiths, S., Maggio, C., Martinez, M., Martinez, Ò., Munar-Adrover, P., Da Deppo, V., 2019. The IFAE/UAB Raman LIDAR for the CTA-North, in: *European Physical Journal Web of Conferences*, p. 02005. doi:10.1051/epjconf/201919702005.

- [15] Gaug, M., Colak, S.M., Maggio, C., 2017. IFAE/UAB Raman LIDAR. CCF - CTA Barcelona Meeting. URL: https://indico.cta-observatory.org/event/1584/contributions/13801/attachments/12088/13872/CCF_LIDAR_Presentation.pdf.
- [16] HESS, 2002. (High Energy Stereoscopic System) URL: <https://www.mpi-hd.mpg.de/hfm/HESS/HESS.shtml>.
- [17] Hilbe, J., 2011. Negative Binomial Regression. Cambridge University Press. URL: https://books.google.es/books?id=0Q_ijx0EBjMC.
- [18] Hilbe, J., 2014. Modeling Count Data. Cambridge books online, Cambridge University Press. URL: <https://books.google.es/books?id=aZLfAwAAQBAJ>.
- [19] ICRU, 1994. ICRU Report 52: Particle Counting in Radioactivity Measurements. International Commission on Radiation Units and Measurements (ICRU). URL: <https://books.google.es/books?id=FFZRAAAAMAAJ>.
- [20] MAGIC, 2004. (Major Atmospheric Gamma Imaging Cherenkov Telescopes or MAGIC Florian Goebel Telescopes) URL: <https://magic.mpp.mpg.de/>.
- [21] Mielke, B., 2005. Analog + photon counting. Licel Tech. Note. URL: https://www.researchgate.net/profile/Bernd-Mielke/publication/265226203_ANALOG_PHOTON_COUNTING/links/55c5d6f608aeb97567422ea8/ANALOG-PHOTON-COUNTING.pdf.
- [22] Mirzoyan, R., Lorentz, E., 1997. On the calibration accuracy of light sensors in atmospheric Cherenkov fluorescence and neutrino experiment, in: Proceedings of the 25th International Cosmic Ray Conference, Durban, pp. 265–268. URL: <http://adsabs.harvard.edu/abs/1997ICRC....7..265M>.
- [23] Müller, D.M., 2016. Evaluation of Photomultiplier Tube Candidates for the CTA and Upgrade of the MAGIC LIDAR System. Master's thesis. Technische Universität München. URL: <https://magic.mpp.mpg.de/backend/publication/show/548>.
- [24] Müller, J.W., 1973. Dead-time problems. Nuclear Instruments and Methods 112, 47–57. URL: <https://www.sciencedirect.com/science/article/pii/0029554X73907738>, doi:[https://doi.org/10.1016/0029-554X\(73\)90773-8](https://doi.org/10.1016/0029-554X(73)90773-8).
- [25] Müller, J.W., 1974. Some formulae for a dead-time-distorted Poisson process: To André Allisy on the completion of his first half century. Nuclear Instruments and Methods 117, 401–404. URL: <https://www.sciencedirect.com/science/article/pii/0029554X74902833>, doi:[https://doi.org/10.1016/0029-554X\(74\)90283-3](https://doi.org/10.1016/0029-554X(74)90283-3).
- [26] Newsom, R.K., Turner, D.D., Mielke, B., Clayton, M., Ferrare, R., Sivaraman, C., 2009. Simultaneous analog and photon counting detection for Raman lidar. Appl. Opt. 48, 3903–3914. doi:10.1364/AO.48.003903.
- [27] Omote, K., 1990. Dead-time effects in photon counting distributions. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 293, 582–588. URL: <https://www.sciencedirect.com/science/article/pii/0168900290903273>, doi:[https://doi.org/10.1016/0168-9002\(90\)90327-3](https://doi.org/10.1016/0168-9002(90)90327-3).
- [28] Parker, M., 2017. Digital Signal Processing 101: Everything you need to know to get started. Newnes.

- [29] Ramos, R.T., 2019. Atmosphere characterization through Barcelona Raman-LIDAR data analysis. Treball de fi de grau. Universitat Autònoma de Barcelona. URL: https://ctan-lidar-pathfinder.ung.si/public/BachelorTheses/Ramos_2019i.pdf.
- [30] Riu Molinero, V., 2020. Posada en marxa del UAB-IFAE Raman LIDAR. Treball de fi de grau. Universitat Autònoma de Barcelona. URL: https://ctan-lidar-pathfinder.ung.si/public/BachelorTheses/RiuMolinero_2020t.pdf.
- [31] The CTA Consortium, 2017. Cherenkov Telescope Array: The Next Generation Gamma-ray Observatory. PoS , 1071. arXiv:1709.05434.
- [32] Veberic, D., 2012. Maximum-likelihood reconstruction of photon returns from simultaneous analog and photon-counting lidar measurements. Applied optics 51, 139–47. doi:10.1364/AO.51.000139.
- [33] VERITAS, 2005. (Very Energetic Radiation Imaging Telescope Array System) URL: <https://veritas.sao.arizona.edu/>.
- [34] Villar, D., 2019. Estudi de l'obtenció del coeficient d'extinció atmosfèrica amb un Raman LIDAR. Treball de fi de grau. Universitat Autònoma de Barcelona. URL: https://ctan-lidar-pathfinder.ung.si/public/BachelorTheses/Villar_2019c.pdf.
- [35] Weisstein, E.W., 2019. Incomplete gamma function. MathWorld – A Wolfram Web Resource. URL: <https://mathworld.wolfram.com/IncompleteGammaFunction.html>.
- [36] Yang, Z., Hardin, J.W., Addy, C.L., 2009. A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. Journal of Statistical Planning and Inference 139, 1514–1521. doi:10.1016/j.jspi.2008.08.018.

A Appendix

A.1 Mathematical expressions

Here we present the longer mathematical expressions, which we computed, that were not included in the text:

A.1.1 Deviance derivatives from section 3.2.1

We have computed the first two partial derivatives of \mathcal{D}_i with respect to p_i and obtained:

$$\frac{\partial \mathcal{D}_i}{\partial p_i} = \frac{\partial G}{\partial p_i} - \frac{2 m_i}{(1 + \delta p_i) p_i} + \frac{2}{(1 + \delta p_i)^2}, \quad (\text{A.1})$$

$$\frac{\partial^2 \mathcal{D}_i}{\partial p_i^2} = \frac{\partial^2 G}{\partial p_i^2} + \frac{2 \delta m_i}{(1 + \delta p_i)^2 p_i} - \frac{4 \delta}{(1 + \delta p_i)^3} + \frac{2 m_i}{(1 + \delta p_i) p_i^2}, \quad (\text{A.2})$$

where

$$\frac{\partial G}{\partial p_i} = -\frac{(\alpha p_i + \beta - a_i)^2 \alpha^2 \varepsilon^2}{(\gamma^2 + \varepsilon^2 \alpha^2 p_i)^2} + \frac{\alpha^2 \varepsilon^2 + 2(\alpha p_i + \beta - a_i) \alpha}{\gamma^2 + \varepsilon^2 \alpha^2 p_i}, \quad (\text{A.3})$$

$$\frac{\partial^2 G}{\partial p_i^2} = \frac{2(\alpha p_i + \beta - a_i)^2 \alpha^4 \varepsilon^4}{(\gamma^2 + \varepsilon^2 \alpha^2 p_i)^3} - \frac{(\alpha^2 \varepsilon^2 + 4(\alpha p_i + \beta - a_i) \alpha) \alpha^2 \varepsilon^2}{(\gamma^2 + \varepsilon^2 \alpha^2 p_i)^2} + \frac{2 \alpha^2}{\gamma^2 + \varepsilon^2 \alpha^2 p_i}. \quad (\text{A.4})$$

A.1.2 W_k derivatives from section 3.2.2

We have computed the first two partial derivatives of W_k with respect to p_i .

Denoting $k \equiv m_i$ and defining $S_k(x) := U(x) [Q(k, x) + k \delta \mathcal{P}_k(x)]$, we obtain:

$$\frac{\partial W_k}{\partial p_i} = -\frac{1}{(1 + \delta p_i)^2} [S_{k-1}(t_{k-1}) - 2S_k(t_k) + S_{k+1}(t_{k+1}) + \Lambda_k] \quad (\text{A.5})$$

where

$$\Lambda_k := \delta \cdot \Delta_k - (1 + \delta p_i) \frac{\partial \Delta_k}{\partial p_i} = \begin{cases} 0 & \text{if } k \leq K - 1, \\ 1 & \text{if } k = K, \\ -1 & \text{if } k = K + 1. \end{cases} \quad (\text{A.6})$$

Defining $T_k(x) := U(x) k(1 - k \delta) \mathcal{P}_k(x)$, we obtain:

$$\frac{\partial^2 W_k}{\partial p_i^2} = -\frac{2 \delta}{1 + \delta p_i} \frac{\partial W_k}{\partial p_i} + \frac{1}{p_i (1 + \delta p_i)} [T_{k-1}(t_{k-1}) - 2T_k(t_k) + T_{k+1}(t_{k+1})]. \quad (\text{A.7})$$

Expressions (A.5)-(A.7) are non-trivial to obtain and the process we followed to deduce them is included in appendix A.2.

A.1.3 Deviance derivatives from section 3.3.2

$$\frac{\partial \mathcal{D}_i}{\partial p_i} = \frac{\partial G}{\partial p_i} + \frac{2 \left(m_i - \frac{p_i}{\delta p_i + 1} \right) \left(\frac{\delta p_i}{(\delta p_i + 1)^2} - \frac{1}{\delta p_i + 1} \right)}{V_\delta} - \frac{\left(m_i - \frac{p_i}{\delta p_i + 1} \right)^2}{V_\delta^2} \frac{\partial V_\delta}{\partial p_i} + \frac{1}{V_\delta} \frac{\partial V_\delta}{\partial p_i} \quad (\text{A.8})$$

$$\begin{aligned} \frac{\partial^2 \mathcal{D}_i}{\partial p_i^2} &= \frac{\partial^2 G}{\partial p_i^2} + \frac{2 \left(\frac{\delta p_i}{(\delta p_i + 1)^2} - \frac{1}{\delta p_i + 1} \right)^2}{V_\delta} - \frac{4 \left(m_i - \frac{p_i}{\delta p_i + 1} \right) \left(\frac{\delta^2 p_i}{(\delta p_i + 1)^3} - \frac{\delta}{(\delta p_i + 1)^2} \right)}{V_\delta} \\ &\quad - \frac{4 \left(m_i - \frac{p_i}{\delta p_i + 1} \right) \left(\frac{\delta p_i}{(\delta p_i + 1)^2} - \frac{1}{\delta p_i + 1} \right)}{V_\delta^2} \frac{\partial V_\delta}{\partial p_i} + \frac{2 \left(m_i - \frac{p_i}{\delta p_i + 1} \right)^2}{V_\delta^3} \left(\frac{\partial V_\delta}{\partial p_i} \right)^2 \\ &\quad - \frac{\left(m_i - \frac{p_i}{\delta p_i + 1} \right)^2}{V_\delta^2} \frac{\partial^2 V_\delta}{\partial p_i^2} - \frac{1}{V_\delta^2} \left(\frac{\partial V_\delta}{\partial p_i} \right)^2 + \frac{1}{V_\delta} \frac{\partial^2 V_\delta}{\partial p_i^2} \end{aligned} \quad (\text{A.9})$$

We hence need to compute the first two derivatives of $V_\delta(p_i)$ with respect to p_i : Denoting $k \equiv m_i$ and defining $\widetilde{S}_k(x) := [Q(k, x) + k\delta \mathcal{P}_k(x)]$, we obtain:

$$\frac{\partial V_\delta}{\partial p_i} = -\frac{2}{(1 + \delta p_i)^2} \left[\sum_{k=0}^K \widetilde{S}_k(t_k) + \widetilde{\Lambda} \right], \quad (\text{A.10})$$

where

$$\widetilde{\Lambda} = 2C(p_i) - 2K - 1. \quad (\text{A.11})$$

Defining $\widetilde{T}_k(x) := k(1 - k\delta) \mathcal{P}_k(x)$, we obtain:

$$\frac{\partial^2 V_\delta}{\partial p_i^2} = -\frac{2\delta}{1 + \delta p_i} \frac{\partial V_\delta}{\partial p_i} + \frac{2}{p_i(1 + \delta p_i)} \sum_{k=0}^K \widetilde{T}_k(t_k) - \frac{4}{(1 + \delta p_i)^4}. \quad (\text{A.12})$$

Expressions (A.10)-(A.12) are non-trivial to obtain and the process we followed to deduce them is included in appendix A.2.

A.2 Mathematical derivations

Here we include details of the mathematical derivations, which we computed, of some results presented in the text.

A.2.1 W_k derivatives

Here is the process we followed to compute the derivatives of W_k :

- First derivative:

Differentiating equation (2.3) yields:

$$\frac{\partial W_k}{\partial p_i} = -\frac{\delta}{(1 + \delta p_i)^2} [R_{k-1} - 2R_k + R_{k+1} + \Delta_k] + \frac{1 + \delta p_i}{(1 + \delta p_i)^2} \left[\frac{1}{1 + \delta p_i} \left[\frac{\partial R_{k-1}}{\partial p_i} - 2 \frac{\partial R_k}{\partial p_i} + \frac{\partial R_{k+1}}{\partial p_i} + \frac{\partial \Delta_k}{\partial p_i} \right] \right] \quad (\text{A.13})$$

Let us first differentiate R_k :

From the definition of t_k , we have that $\partial t_k / \partial p_i = t_k / p_i = 1 - k\delta$.

Recall the definition $Q(k, t_k) = \Gamma(k, t_k) / \Gamma(k, 0)$, where $\Gamma(k, x) = \int_x^\infty u^{k-1} e^{-u} du$. From the Leibniz integral rule,

$$\frac{\partial \Gamma(k, x)}{\partial x} = -x^{k-1} e^{-x}. \quad (\text{A.14})$$

We may now differentiate $Q(k, t_k)$ by using the chain rule and (A.14):

$$\frac{\partial Q(k, t_k)}{\partial p_i} = \frac{1}{\Gamma(k)} \left[\frac{\partial \Gamma(k, t_k)}{\partial t_k} \frac{\partial t_k}{\partial p_i} \right] = -\frac{t_k^k e^{-t_k}}{p_i \Gamma(k)} = -\frac{t_k^k e^{-t_k}}{p_i (k-1)!} = -\frac{k \mathcal{P}_k(t_k)}{p_i}, \quad (\text{A.15})$$

where we have used that $\Gamma(k) = (k-1)!$ for k a positive integer and recalled the definition $\mathcal{P}_k(t_k) = t_k^k e^{-t_k} / k!$.

We may also differentiate $\mathcal{P}_k(t_k)$ by using the chain rule:

$$\frac{\partial \mathcal{P}_k(t_k)}{\partial p_i} = \frac{\partial \mathcal{P}_k(t_k)}{\partial t_k} \frac{\partial t_k}{\partial p_i} = \frac{t_k^k e^{-t_k}}{p_i (k-1)!} - \frac{t_k^{k+1} e^{-t_k}}{p_i k!} = \frac{k \mathcal{P}_k(t_k)}{p_i} - \frac{t_k \mathcal{P}_k(t_k)}{p_i}. \quad (\text{A.16})$$

We may now differentiate R_k by using (A.15) and (A.16):

$$\begin{aligned} \frac{\partial R_k}{\partial p_i} &= U(t_k) \frac{\partial}{\partial p_i} [(k - t_k)Q(k, t_k) + k\mathcal{P}_k(t_k)] \\ &= U(t_k) \left[-\frac{t_k}{p_i} Q(k, t_k) + (k - t_k) \frac{\partial Q(k, t_k)}{\partial p_i} + k \frac{\partial \mathcal{P}_k(t_k)}{\partial p_i} \right] \\ &= U(t_k) \left[-\frac{t_k}{p_i} Q(k, t_k) - \frac{k^2 \mathcal{P}_k(t_k)}{p_i} + \frac{k t_k \mathcal{P}_k(t_k)}{p_i} + \frac{k^2 \mathcal{P}_k(t_k)}{p_i} - \frac{k t_k \mathcal{P}_k(t_k)}{p_i} \right] \\ &= -U(t_k) \frac{t_k}{p_i} Q(k, t_k) \\ &= -U(t_k) (1 - k\delta) Q(k, t_k). \end{aligned} \quad (\text{A.17})$$

Note that we have considered $\partial U(t_k) / \partial p_i = 0$ for simplicity to avoid a discontinuity at $t_k = 0$, since the derivative of the Heaviside function is a Dirac delta, which is an extreme case that may be resolved in the implementation.

Let us now differentiate Δ_k . From equation (2.4), we have that:

$$\frac{\partial \Delta_k}{\partial p_i} = \begin{cases} 0 & \text{if } k \leq K-1, \\ (K+1)\delta - 1 & \text{if } k = K, \\ 1 - K\delta & \text{if } k = K+1. \end{cases} \quad (\text{A.18})$$

Substituting (A.17) and (A.18) into (A.13) and using the definition of R_k yields:

$$\begin{aligned}
 \frac{\partial W_k}{\partial p_i} &= -\frac{\delta}{(1 + \delta p_i)^2} \left[U(t_{k-1}) [(k-1 - t_{k-1})Q(k-1, t_{k-1}) + (k-1)\mathcal{P}_{k-1}(t_{k-1})] \right. \\
 &\quad - 2U(t_k) [(k - t_k)Q(k, t_k) + k\mathcal{P}_k(t_k)] \\
 &\quad \left. + U(t_{k+1}) [(k+1 - t_{k+1})Q(k+1, t_{k+1}) + (k+1)\mathcal{P}_{k+1}(t_{k+1})] + \Delta_k \right] \\
 &\quad - \frac{1 + \delta p_i}{(1 + \delta p_i)^2} \left[U(t_{k-1})(1 - (k-1)\delta) - 2U(t_k)(1 - k\delta) + U(t_{k+1})(1 - (k+1)\delta) + \frac{\partial \Delta_k}{\partial p_i} \right] \\
 &= \dots = -\frac{1}{(1 + \delta p_i)^2} \left[U(t_{k-1}) [Q(k-1, t_{k-1}) + (k-1)\delta\mathcal{P}_{k-1}(t_{k-1})] \right. \\
 &\quad - 2U(t_k) [Q(k, t_k) + k\delta\mathcal{P}_k(t_k)] + U(t_{k+1}) [Q(k+1, t_{k+1}) + (k+1)\delta\mathcal{P}_{k+1}(t_{k+1})] \\
 &\quad \left. + \delta\Delta_k - (1 + \delta p_i) \frac{\partial \Delta_k}{\partial p_i} \right].
 \end{aligned} \tag{A.19}$$

Defining $S_k(x) := U(x) [Q(k, x) + k\delta \mathcal{P}_k(x)]$ and

$$\Lambda_k := \delta \cdot \Delta_k - (1 + \delta p_i) \frac{\partial \Delta_k}{\partial p_i} = \begin{cases} 0 & \text{if } k \leq K-1, \\ 1 & \text{if } k = K, \\ -1 & \text{if } k = K+1, \end{cases} \tag{A.20}$$

we obtain:

$$\frac{\partial W_k}{\partial p_i} = -\frac{1}{(1 + \delta p_i)^2} [S_{k-1}(t_{k-1}) - 2S_k(t_k) + S_{k+1}(t_{k+1}) + \Lambda_k], \tag{A.21}$$

as we wanted to show.

• Second derivative:

Differentiating (A.21) yields:

$$\begin{aligned}
 \frac{\partial^2 W_k}{\partial p_i^2} &= \frac{2\delta}{(1 + \delta p_i)^3} [S_{k-1} - 2S_k + S_{k+1} + \Lambda_k] - \frac{1}{(1 + \delta p_i)^2} \left[\frac{\partial S_{k-1}}{\partial p_i} - 2\frac{\partial S_k}{\partial p_i} + \frac{\partial S_{k+1}}{\partial p_i} + \frac{\partial \Lambda_k}{\partial p_i} \right] \\
 &= -\frac{2\delta}{1 + \delta p_i} \frac{\partial W_k}{\partial p_i} - \frac{1}{(1 + \delta p_i)^2} \left[\frac{\partial S_{k-1}}{\partial p_i} - 2\frac{\partial S_k}{\partial p_i} + \frac{\partial S_{k+1}}{\partial p_i} \right].
 \end{aligned} \tag{A.22}$$

We may differentiate S_k by using (A.15) and (A.16):

$$\begin{aligned}
 \frac{\partial S_k}{\partial p_i} &= U(t_k) \left[\frac{\partial Q(k, t_k)}{\partial p_i} + k\delta \frac{\partial \mathcal{P}_k(t_k)}{\partial p_i} \right] \\
 &= U(t_k) \left[-\frac{k\mathcal{P}_k(t_k)}{p_i} + \left(1 - \frac{t_k}{p_i}\right) \left(\frac{k\mathcal{P}_k(t_k)}{p_i} - \frac{t_k\mathcal{P}_k(t_k)}{p_i} \right) \right] \\
 &= U(t_k) \left[-\frac{t_k\mathcal{P}_k(t_k)}{p_i} - \frac{kt_k\mathcal{P}_k(t_k)}{p_i^2} + \frac{t_k^2\mathcal{P}_k(t_k)}{p_i^2} \right] \\
 &= U(t_k) \frac{t_k^{k+1}e^{-t_k}}{p_i(k-1)!} \left[-\frac{1}{k} - \frac{1}{p_i} + \frac{t_k}{kp_i} \right] \\
 &= -U(t_k) \frac{t_k^{k+1}e^{-t_k}}{p_i(k-1)!} \frac{1 + \delta p_i}{p_i} \\
 &= -U(t_k) \mathcal{P}_k(t_k) k(1 - k\delta) \frac{1 + \delta p_i}{p_i},
 \end{aligned} \tag{A.23}$$

where we used that $k\delta = 1 - t_k/p_i$ and again we considered $\partial U(t_k)/\partial p_i = 0$.

Defining $T_k(x) := U(x) k(1 - k\delta) \mathcal{P}_k(x)$ and substituting (A.23) into (A.22) yields:

$$\frac{\partial^2 W_k}{\partial p_i^2} = -\frac{2\delta}{1 + \delta p_i} \frac{\partial W_k}{\partial p_i} + \frac{1}{p_i(1 + \delta p_i)} [T_{k-1}(t_{k-1}) - 2T_k(t_k) + T_{k+1}(t_{k+1})], \tag{A.24}$$

as we wanted to show.

A.2.2 V_δ derivatives

Here is the process we followed to compute the derivatives of V_δ :

The derivatives of V_δ are computed in a similar way to the derivatives of W_k , and so we will be concise and use similar notation to show the analogy.

- First derivative:

Defining $\tilde{R}_k = R_k/U(t_k) = [(k - t_k)Q(k, t_k) + k\mathcal{P}_k(t_k)]$, we have that:

$$V_\delta = \frac{2}{1 + \delta p_i} \sum_{k=0}^K \tilde{R}_k + H(\mu - K) \tag{A.25}$$

The derivative of \tilde{R}_k is clearly given by:

$$\frac{\partial \tilde{R}_k}{\partial p_i} = \frac{1}{U(t_k)} \frac{\partial R_k}{\partial p_i} = -(1 - k\delta)Q(k, t_k). \tag{A.26}$$

Recalling that $H(x) = x(1 - x)$ and $\mu = p_i/(1 + \delta p_i)$,

$$H(\mu - K) = (\mu - K)(1 - \mu + K) \tag{A.27}$$

Notice now that

$$\frac{\partial \mu}{\partial p_i} = \frac{1}{(1 + \delta p_i)^2} \tag{A.28}$$

and therefore:

$$\frac{\partial H}{\partial p_i} = \frac{1}{(1 + \delta p_i)^2} (1 - 2\mu + 2K) \tag{A.29}$$

We may now differentiate V_δ and we obtain:

$$\begin{aligned}
 \frac{1}{2} \frac{\partial V_\delta}{\partial p_i} &= \frac{1 + \delta p_i}{(1 + \delta p_i)^2} \sum_{k=0}^K \frac{\partial \tilde{R}_k}{\partial p_i} - \frac{\delta}{(1 + \delta p_i)^2} \sum_{k=0}^K \tilde{R}_k + \frac{\partial H}{\partial p_i} \\
 &= -\frac{1 + \delta p_i}{(1 + \delta p_i)^2} \sum_{k=0}^K (1 - k\delta) Q(k, t_k) - \frac{\delta}{(1 + \delta p_i)^2} \sum_{k=0}^K [(k - t_k) Q(k, t_k) + k \mathcal{P}_k(t_k)] + \frac{\partial H}{\partial p_i} \\
 &= \dots = -\frac{1}{(1 + \delta p_i)^2} \sum_{k=0}^K [Q(k, t_k) + k\delta \mathcal{P}_k(t_k)] + \frac{\partial H}{\partial p_i}
 \end{aligned} \tag{A.30}$$

Defining $\tilde{S}_k(x) := S_k(x)/U(x) = [Q(k, x) + k\delta \mathcal{P}_k(x)]$ we obtain:

$$\frac{1}{2} \frac{\partial V_\delta}{\partial p_i} = -\frac{1}{(1 + \delta p_i)^2} \sum_{k=0}^K \tilde{S}_k + \frac{\partial H}{\partial p_i} \tag{A.31}$$

Finally, defining

$$\tilde{\Lambda} := 2\mu - 2K - 1 = 2C(p_i) - 2K - 1 \tag{A.32}$$

we obtain:

$$\frac{\partial V_\delta}{\partial p_i} = -\frac{2}{(1 + \delta p_i)^2} \left[\sum_{k=0}^K \tilde{S}_k(t_k) + \tilde{\Lambda} \right], \tag{A.33}$$

as we wanted to show.

- Second derivative:

Differentiating (A.31) yields:

$$\frac{1}{2} \frac{\partial^2 V_\delta}{\partial p_i^2} = \frac{2\delta}{(1 + \delta p_i)^3} \sum_{k=0}^K \tilde{S}_k - \frac{1}{(1 + \delta p_i)^2} \sum_{k=0}^K \frac{\partial \tilde{S}_k}{\partial p_i} + \frac{\partial^2 H}{\partial p_i^2} \tag{A.34}$$

The derivative of \tilde{S}_k is clearly given by:

$$\frac{\partial \tilde{S}_k}{\partial p_i} = \frac{1}{U(t_k)} \frac{\partial S_k}{\partial p_i} = -\mathcal{P}_k(t_k) k(1 - k\delta) \frac{1 + \delta p_i}{p_i} \tag{A.35}$$

The second derivative of H is given by:

$$\frac{\partial^2 H}{\partial p_i^2} = -\frac{2\delta}{(1 + \delta p_i)^3} (1 - 2\mu + 2K) - \frac{2}{(1 + \delta p_i)^4} \tag{A.36}$$

Defining $\tilde{T}_k(x) := k(1 - k\delta) \mathcal{P}_k(x)$, we have that:

$$\frac{1}{2} \frac{\partial^2 V_\delta}{\partial p_i^2} = \frac{2\delta}{(1 + \delta p_i)^3} \sum_{k=0}^K \tilde{S}_k + \frac{1}{p_i(1 + \delta p_i)} \sum_{k=0}^K \tilde{T}_k + \frac{\partial^2 H}{\partial p_i^2} \tag{A.37}$$

Substituting (A.36) into (A.37), simplifying and rearranging terms yields:

$$\frac{\partial^2 V_\delta}{\partial p_i^2} = -\frac{2\delta}{1 + \delta p_i} \frac{\partial V_\delta}{\partial p_i} + \frac{2}{p_i(1 + \delta p_i)} \sum_{k=0}^K \tilde{T}_k(t_k) - \frac{4}{(1 + \delta p_i)^4}. \tag{A.38}$$

as we wanted to show.

A.3 Code

The gluing scripts we developed have over 10 000 lines of code, which in consequence will not be included here. For further information, kindly contact oscarporqueras@gmail.com.

A.4 Authorship and word count declaration

I, Òscar Porqueras León, hereby declare that the thesis titled “Maximum likelihood gluing of simultaneous analogue and photon counting LIDAR measurements” and the work presented in it is my own, other than the material that is result of joint research or other sources which are fully acknowledged and all quotations properly identified. I ensure that this project submitted for assessment is my own and it is expressed in my own words. A list of the references employed is also included.

The total count of words from the introduction to the conclusions is 8018, using Microsoft Word's word counter after having converted the file from pdf. There are 7 figures considered as 200 words each. There are 29 equations in “math display” (not inline) which were converted to images by Microsoft Word, considered as 20 words each.

The total of words computed using the proposed counting procedure therefore is

$$8018 \text{ words} + 7 \cdot 200 \frac{\text{words}}{\text{figure}} + 29 \cdot 20 \frac{\text{words}}{\text{equation}} = 8018 + 1400 + 580 = 9998,$$

which is within the limits of this type of work.

Signed:



Òscar Porqueras León