A Maximum Likelihood Approach for Signal Estimation and Data Quality Assessment for High-Energy Calorimeter Systems

Sarita de Miranda Rimes, Bernardo Sotto-Maior Peralva, Gustavo Barbosa Libotte, Thiago Campos Acácio Paschoalin, Luciano Manhães de Andrade Filho, and José Manoel de Seixas

***Abstract*—In high-energy calorimetry, energy estimation is typically carried out using digital linear filters, from which the amplitude of a conditioned signal is inferred. The linear design is justified when the electronic noise can be approximated by a Gaussian function. However, calorimeters operating at high event rates and under high-luminosity conditions, such as those in modern colliders, experience an undesirable non-Gaussian noise source: signal tails from adjacent collisions that contain no relevant physical information regarding the event of interest. Despite this, the actual noise probability density function (pdf) is often neglected, and only second-order statistics are used to preserve the linear model. In this work, a novel design based on the complete information of the noise distribution, using maximum likelihood estimators (MLE), is presented. An iterative optimization algorithm is proposed to search for the value that maximizes a non-Gaussian multivariate distribution. Additionally, this article also evaluates the use of maximum likelihood estimates for data quality assessment, where it can be useful to discard signals with poor estimation quality. Simulation results, comparing the proposed method with current linear methods, are presented. The efficiency achieved from the non-Gaussian MLE, in terms of estimation error, surpasses that of linear models across the full range of cell occupancy conditions. It is also shown that the final likelihood probability can be used as a performance metric for signal reconstruction, providing a robust quality factor (QF) discriminant for physics analysis with respect to typical strategies.**

Received 19 August 2025; revised 23 September 2025; accepted 28 October 2025. Date of publication 10 November 2025; date of current version 18 December 2025. This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil, under Grant 001. The work of Bernardo Sotto-Maior Peralva was supported by the Carlos Chagas Filho Foundation for Supporting Research in the State of Rio de Janeiro (FAPERJ) under Grant E-26/201.304/2022. The work of Gustavo Barbosa Libotte was supported in part by FAPERJ and in part by the National Council for Scientific and Technological Development (CNPq) under Grant E-26/210.430/2024 and Grant 309784/2025-5. *(Corresponding author: Sarita de Miranda Rimes.)*

Sarita de Miranda Rimes, Bernardo Sotto-Maior Peralva, and Gustavo Barbosa Libotte are with the Polytechnic Institute of Rio de Janeiro State University (UERJ), Nova Friburgo 28625-570, Brazil (e-mail: smrimes@iprj.uerj.br; bernardo@iprj.uerj.br; gustavolibotte@iprj.uerj.br).

Thiago Campos Acácio Paschoalin is with the Federal Center for Technological Education of Minas Gerais (CEFET-MG), Leopoldina 36700-000, Brazil (e-mail: thiago.paschoalin@cefetmg.br).

Luciano Manhães de Andrade Filho is with the Federal University of Juiz de Fora (UFJF), Juiz de Fora 36036-900, Brazil (e-mail: luciano.andrade@ufjf.br).

José Manoel de Seixas is with the Federal University of Rio de Janeiro, Rio de Janeiro 21941-853, Brazil (e-mail: seixas@lps.ufrj.br).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNS.2025.3630987.

Digital Object Identifier 10.1109/TNS.2025.3630987

***Index Terms*—Data quality, high-energy calorimetry, maximum likelihood estimators (MLE), non-Gaussian noise, signal estimation.**

I. Introduction

MODERN high-energy calorimeters [1] are designed to operate in environments with high luminosity. These systems typically consist of multiple layers, each containing thousands of readout channels that sample and estimate the energy of particles at high collision rates (ranging from several tens of megahertz) [2], [3]. Traditional energy reconstruction techniques rely on the stability of a conditioned pulse shape, where the amplitude is directly proportional to the energy deposited by incident particles [4]. However, in high-luminosity scenarios, signal pile-up, or overlapping of signals, compromises the accuracy of energy reconstruction. Standard approaches treat these overlapping signals as noise, using second-order statistics to design a linear optimal filter (OF) that mitigates the effect of pile-up on the central signal amplitude estimation [5], [6], [7]. As a result, when pile-up is considered noise, the signal distribution deviates from Gaussian behavior, causing the linear filter to perform sub-optimally.

Recently, an alternative technique has been proposed that interprets the pile-up as adjacent signals and employs a deconvolution process aiming at separating the superimposed signals. This approach is called the multi-amplitude estimator (MAE) [8] and has proved to be effective at certain signal pile-up levels. However, under severe signal pile-up conditions—that is, with occupancy levels above 50% (as observed in forward calorimeter regions)—attempting to separate individual signals becomes impractical due to the overwhelming number of superimposed signals. In this work, similar to standard OF methods, pile-up signals are treated as a noise source that must be minimized under high occupancy conditions. The key contribution of this article is the use of a maximum likelihood estimator (MLE) technique that uses a suitable statistical model for the pile-up noise.

MLE methods are highly efficient when a model of the random process, expressed as a function of the target parameters, is known in advance and when the process stochasticity is strongly conditioned to those parameters [9]. However, deriving such analytical models is often challenging, and when available, they typically result in complex, non-linear formulations with no trivial solutions. Fortunately, in calorimeters with

unipolar pulse shapes, the pile-up noise distribution exhibits a skewed structure, which can be modeled as a convolution of exponential functions. This results in multivariate Gamma distributions, which, despite their accuracy, are complex to implement. In this work, the use of Log-normal distributions is explored as an alternative model for pile-up noise, leveraging their similarity to Gamma distributions. Furthermore, the parameters of the Log-normal functions are estimated using only first- and second-order statistical moments, similar to the Gaussian case. As a result, the multivariate probability distribution model required for the MLE design can be efficiently determined and is computationally feasible to implement.

Although a suitable analytical model has been proposed, the Log-normal model results in a nonlinear MLE formulation with no closed-form solution. To address this, an iterative algorithm that finds the maximum of the likelihood function using the Golden Section optimization procedure [10] is proposed. The search is constrained by the dynamic range of the ADC electronics, which, for example, could span from 0 to 1023 ADC counts in a 10-bit ADC setup.

Another contribution of this article is the use of the maximum likelihood probability as a quality factor (QF) metric for accepting or rejecting the reconstructed energy. A QF based on this measure is proposed and compared to standard metrics used in calorimetry, which rely on the Euclidean norm [11], [12]. Since the proposed approach takes into account the pile-up information, it could provide a more robust QF assessment under such harsh conditions.

This article is organized as follows. Section II discusses the choice of the Log-normal distribution as a suitable model for pile-up noise. Section III presents the design of the amplitude estimator using MLE under Log-normal noise conditions. Section IV describes the simulation setup and analyzes the results. Finally, the conclusions are presented in Section V.

II. Signal Pile-up Modeling

In this section, an analytical expression for the signal pile-up distribution that promotes a feasible MLE design is developed. Naturally, this distribution is intrinsically related to the energy spectra of the calorimeter conditions. The transverse momentum (p_T) spectrum will depend on the nature of the colliding particles and the characteristics of the accelerator. However, in the low p_T region—dominated by multiple scattering interactions—the p_T distribution exhibits strong exponential characteristics [13] regardless of the collider environment. In the LHC [2], for instance, a proton–proton collider that imposes severe signal pile-up conditions on its calorimeter systems, the p_T spectrum is modeled as a Hagedorn distribution [14]. This function features an exponential behavior in the low p_T region and a power-law structure for high p_T collision byproducts. Since the signal pile-up is primarily generated by low p_T interactions, exponential distributions can effectively model the amplitude of these signals without loss of generality.

The transverse momentum is the component of momentum (p) that is perpendicular to the beam [15]. This component is intrinsically related to the transverse energy (E_T) and is necessary for computing this from the total energy E, which is actually measured by the calorimeters. This comes from the fact that p_T is essential to describe the geometry of the events that come from collisions, and E_T depends on the trajectory of the reconstructed particle [16].

Given that the calorimeter reference pulse shape is unipolar, the samples from superimposed signals add up. Consequently, the pile-up effect corresponds to the sum of exponentially distributed samples, which, by definition, results in a subclass of the Gamma distribution—the Erlang distribution [17]. More precisely, the signal pile-up distribution is further convolved with electronic noise. However, since the energy of the pile-up noise tends to be significantly higher than that of the electronic noise, the latter can be neglected for simplicity.

A useful characteristic of Gamma distributions, often exploited in practice, is their similarity to the Log-normal model, allowing both functions to be used interchangeably for analyzing skewed, non-negative datasets [18], [19], [20]. Although their parameter interpretations differ, the Log-normal distribution possesses properties that make it a suitable and attractive choice for the current application.

A. Log-Normal Distribution

According to the Central Limit Theorem, the sum of independent random variables tends to follow a Gaussian (Normal) distribution [21]. As a result, the product of positive independent random variables results in a Log-normal distribution, since in the logarithmic domain, multiplication transforms into addition. The Log-normal distribution is given by equation [22]

$$f(r) = \frac{1}{r\sigma\sqrt{2\pi}} \exp\left[\frac{-(\ln r - \mu)^2}{2\sigma^2}\right]. \tag{1}$$

In comparison with the Normal distribution, apart from the natural logarithm applied in the input variable r, it is important to note the presence of this variable also in the denominator of the normalization factor, to preserve differential probability. The two parameters of the Log-normal distribution are precisely the mean μ and the variance σ^2 extracted from the logarithm of the data (which, by definition, follows the Gaussian distribution).

Similarly, the multivariate Log-normal distribution, for a vector $\mathbf{r}$ of dimension N, has the form [23]

$$f(\mathbf{r}) = B \exp[D] \tag{2}$$

being

$$B = \frac{1}{\left(\prod_{i=1}^{N} r_i\right) |\mathbf{C}|^{\frac{1}{2}} (2\pi)^{\frac{N}{2}}}$$

and

$$D = \frac{-(\ln \mathbf{r} - \boldsymbol{\mu})' \, \mathbf{C}^{-1} \, (\ln \mathbf{r} - \boldsymbol{\mu})}{2}$$

where $|\mathbf{C}|$ is the determinant of the covariance matrix and the superscript prime character stands for the transpose operation. One should stress that the natural logarithm operating in a vector actually applies the logarithm to each vector component individually

$$\ln \mathbf{a} = [\ln a_1 \;\; \ln a_2 \;\; \ldots \;\; \ln a_N]'. \tag{3}$$

In the multivariate Log-normal equation, the mean of each component (vector μ) and the $\mathbf{C}$ covariance matrix are extracted from the logarithm of the multivariate data as well. The simplicity of obtaining the distribution parameters—by

estimating only the first and second-order statistics—makes the Log-normal particularly attractive for modeling the signal pile-up distribution on the MLE design proposed in this article.

III. Energy Reconstruction Using the MLE

The MLE approach is a class of estimators that uses the actual probability density function (pdf) model of the random process [9] to estimate one or a set of parameters. In this model, the pdf is parametrized by the target variables. Unlike the Bayesian estimators [24], these parameters are considered unknown constants to be estimated. The best guess for those constants is the one that maximizes the pdf function. Therefore, once an acceptable pdf model as a function of the target parameters is determined, the MLE can be found by computing the parameter values that maximize the pdf output.

In regular calorimeter energy reconstruction tasks, the target parameter is the amplitude A of a deterministic signal embedded in noise, whose actual distribution is ignored, and only the second order statistics are taken into account, to achieve a linear implementation [5], [6].

This is consistent with modern information theory: in the absence of higher-order statistics information, the Gaussian distribution is the one that maximizes entropy [25].

A. MLE for Signal Amplitude in Gaussian Noise

Given N samples of a deterministic signal, represented by the $\mathbf{s}$ vector, multiplied by an unknown amplitude A, and embedded in correlated Gaussian noise with zero mean, the pdf modeling this random process is

$$f(\mathbf{r}|A) = \frac{1}{|\mathbf{C}|^{\frac{1}{2}}(2\pi)^{\frac{N}{2}}} \exp\left[\frac{-(\mathbf{r}-A\mathbf{s})'\,\mathbf{C}^{-1}(\mathbf{r}-A\mathbf{s})}{2}\right]. \quad (4)$$

The signal amplitude A that maximizes this probability can be found simply by minimizing the exponent. This leads to a linear equation in A whose solution can be easily determined as

$$\hat{A} = \frac{\mathbf{s}'\mathbf{C}^{-1}\mathbf{r}}{\mathbf{s}'\mathbf{C}^{-1}\mathbf{s}} = \mathbf{g}'\mathbf{r} \quad (5)$$

where the $\mathbf{C}$ noise covariance matrix contains the second-order statistics of the noise. The $\mathbf{g}$ vector represents the coefficients of a linear filter known as the OF weights. During the exponent minimization process, some constraints can be introduced to provide immunity against small random phase shifts and pedestal fluctuations. This is achieved through the implementation of Lagrange multipliers, which appropriately modify the final OF weights [4]. Also, the $\mathbf{s}$ vector refers to the pulse shape of the detector and is usually well known from the calibration system.

Although this linear model leads to a simple implementation, the Gaussian noise does not adequately describe the signal pile-up conditions, as explained in Section II. Therefore, next, a more accurate pdf model for the MLE design of signal amplitude estimators is proposed, using Log-normal modeling for the noise.

B. MLE for Signal Amplitude in Log-Normal Noise

For the Log-normal MLE modeling, the likelihood equation becomes

$$f(\mathbf{r}|A) = B\exp\{D\} \quad (6)$$

Algorithm 1 Amplitude Estimation Under Log-Normal Noise

Parameters:
$\delta = 10^{-6}$ ▷threshold
$\alpha \leftarrow (\sqrt{5}+1)/2$ ▷Golden ratio
Inputs:
$\mathbf{r}$ ▷recorded samples
$\mathbf{s}$ ▷normalized reference pulse shape samples
μ ▷mean vector parameter from Log-normal noise
$\mathbf{C}$ ▷covariance matrix from Log-normal noise

```
while |b − a| > δ do
    c ← b − (b − a)/α
    d ← a + (b − a)/α
    if J(c) > J(d) then
        b ← d
    else
        a ← c
    end if
end while
return (b + a)/2
```

where

$$B = \frac{1}{\left[\prod_{i=1}^{N}(r_i - \mathrm{As}_i)\right]|\mathbf{C}|^{\frac{1}{2}}(2\pi)^{\frac{N}{2}}} \quad (7)$$

and

$$D = \frac{-\left[\ln(\mathbf{r}-A\mathbf{s}) - \boldsymbol{\mu}\right]'\mathbf{C}^{-1}\left[\ln(\mathbf{r}-A\mathbf{s}) - \boldsymbol{\mu}\right]}{2}. \quad (8)$$

It should be noted, by the presence of the $\boldsymbol{\mu}$ vector, that this noise model is not zero-mean. This is expected, as it results from the sum of positive pulse samples. Due to the presence of the parameter A also in the normalization factor (7), and the use of the natural logarithm in the exponent (8), this likelihood equation cannot be simplified as the Gaussian case, and the final solution for A becomes non-linear.

In this work, the signal amplitude A that maximizes the multivariate Log-normal likelihood function is obtained iteratively. This unimodal distribution has a single global maximum [26]. Furthermore, for optimization purposes, it is equivalent to work with the logarithm of the likelihood function

$$\ln f(\mathbf{r}|A) = \ln\left[|\mathbf{C}|^{-\frac{1}{2}}(2\pi)^{-\frac{N}{2}}\right] - \sum_{i=1}^{N}\ln(r_i - \mathrm{As}_i) + D. \quad (9)$$

The cost function to be minimized follows directly from (9):

$$J(A) = \sum_{i=1}^{N}\ln(r_i - As_i) - D. \quad (10)$$

Next, an iterative optimization process can be used to reach the maximum probability. The Golden-section search method is used since it comprises an unimodal objective function, and it avoids the calculation of derivatives. Note that any optimization method for single-variable functions can be used in this task. However, it is important to ensure that, for each estimate of A, the condition $As_i < r_i$ is satisfied in (9).

The iterative process for amplitude estimation of a deterministic signal embedded in Log-normal noise is presented in Algorithm 1. In general, the method performs successive cuts

in the initial search interval given by $[a, b]$, discarding the subinterval that does not contain the maximum point, through iterative comparison of the internal values c and d. When the initial interval becomes small enough, the central value can be taken as a solution to the problem, with an error proportional to δ [10].

IV. Simulation Results

To quantify the method's performance and proceed to efficiency comparisons with other state-of-the-art techniques, an unipolar pulse signal from a typical RCCR shaper circuitry is used in a simulated environment [27], also adopted in previous work [8], [28], [29], where signals are superimposed using a given occupancy factor. The pulse amplitudes related to the signal pile-up noise are extracted from an exponential distribution with a mean value of 50 ADC, reflecting the typical energy of low p_T collisions in modern experiments, as reported in previous studies [30], [31], which represents the simulated noise. To cope with realistic scenarios, several other noise sources are employed: 1) random variation of ± 1 in each pulse sample to simulate electronic distortions; 2) random phase shift between ± 10 ns; and 3) additive Gaussian noise of zero mean and standard deviation equal to 1.5 ADC, simulating typical electronic noise. Readout windows of seven samples containing this pile-up noise are recorded, with the samples digitized every 25 ns. From this dataset, the parameters $\boldsymbol{\mu}$ and $\mathbf{C}$ of a multivariate Log-normal distribution are extracted for each specified occupancy level.

To simulate the target signals, an exponential distribution with a mean of 100 ADC is employed, corresponding to a signal-to-noise ratio (SNR) of 2. Subsequently, the target signals were combined with the previously described noise conditions to emulate the complete received signal corresponding to the calorimeter response.

To assess the statistical aspects of the efficiency analyses, a k-fold cross-validation strategy was applied as follows. Initially, for each occupancy level, a set comprising 2 000 000 noise signals was generated. This dataset was then partitioned into ten subsets (folds), and the amplitude estimation (energy reconstruction) for the target signals was performed by each method individually for every subset.

A. Noise Modeling Procedure

Before delving into the energy reconstruction efficiency, it is important to verify the noise modeling by comparing how well both Normal and Log-normal distributions fit the noise. To this end, incoming time-domain noise samples within the readout windows are used with arbitrarily selected four different occupancy rates: 10%, 30%, 50%, and 80%. These choices encompass a reasonable range of occupancy levels, providing a baseline for comparing approaches under harsh pile-up conditions.

Fig. 1(a) depicts the histogram of the noise time samples corresponding to a 30% occupancy. In addition to the histogram, both the Normal and Log-normal fittings are overlaid with the underlying data. It is evident that the Log-normal fit provides a better model for the noise distribution compared to the Normal distribution, especially along the tail. In contrast, the Normal distribution struggles to represent the noise information, with significant challenges in fitting the data.

To further analyze the problem, Fig. 1(b) presents the corresponding results for noise with 50% of occupancy. In this scenario, it is observed that the Log-normal distribution shows a slight deviation from the peak of the target distribution, which also leads to some error along the right tail. However, the Normal distribution exhibits an even greater deviation, failing to approach the peak of the target distribution.

Finally, Fig. 1(c) presents the noise histograms with 80% occupancy. At this level of pile-up, it is clear that the Log-normal distribution effectively describes the data, whereas the Normal distribution continues to diverge from the dataset. The observed behavior at 80% occupancy suggests that as occupancy increases, the Log-normal distribution provides a more efficient fit compared to the Normal distribution. Indeed, it is expected that the Log-normal distribution will better approximate the data as pile-up increases, since pile-up arises from exponential distributions, and their summation forms a Gamma distribution that the Log-normal can more accurately approximate.

Considering independent, equally sized samples and aiming to compare continuous distributions, Kolmogorov–Smirnov (KS) statistics [32] is estimated. The results are depicted in Fig. 1(d). The error bars in this figure have been scaled by a factor of 20 to improve visibility, as they are initially very small and might not be discernible on the graph. From this figure, it is evident that the observed behavior matches what was noted in the fittings [see Fig. 1(a)–(c)]. The lower the value of the KS statistic, the better the fit. Across all evaluated occupancy levels, even for those where the distributions assume a more Gaussian-like profile, the KS statistic is consistently smaller for the Log-normal distribution, indicating that it provides a better noise model compared to the Normal distribution.

B. Performance Evaluation

The efficiency analyses for each discussed method are now presented. To this end, simulated noise (electronic noise plus pile-up) is combined with simulated signals, referred to as target signals. The applied methods aim to recover these target signals, specifically by estimating their amplitudes. Fig. 2(a) shows the absolute error histogram of the four evaluated methods for an occupancy of 80%. The bin width was determined using the Freedman–Diaconis rule [33], which utilizes the distribution under analysis to estimate a reasonable bin width. Since the bin width needs to be consistent across all distributions, the Log-normal MLE error was arbitrarily employed as a reference for this parameter.

By examining the error distributions [see Fig. 2(a)], it is clear that the Log-normal MLE method appears to exhibit a smaller dispersion compared to the other methods. To verify whether this observation holds, the mean and standard deviation [34] of the absolute error were calculated for all considered occupancy levels. These results are presented in Fig. 2, where the error bars were multiplied by a factor of 5 to enhance visibility.

Fig. 2(b) illustrates that, starting from an occupancy of 30% (among the evaluated values), the average error obtained by Log-normal MLE is lower than that of Gaussian MLE and MAE, with a slightly higher value at 10%. A positive

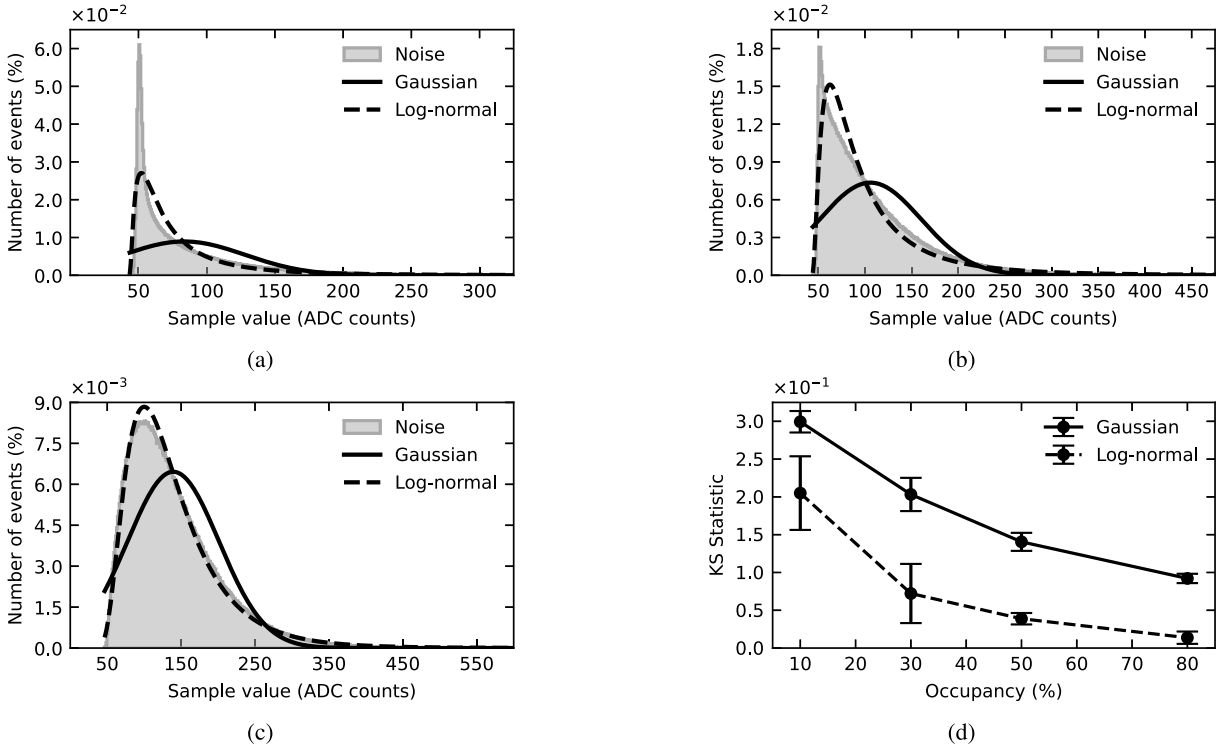

Fig. 1. Comparison of noise data with Normal and Log-normal distributions modeling, to (a) 30%, (b) 50%, and (c) 80% of occupancy and (d) Kolmogorov–Smirnov statistic for different occupancies.

TABLE I

MEAN AND STANDARD DEVIATION OF ERROR AVERAGES ACROSS ALL ITERATIONS, FOR ALL METHODS AND OCCUPANCIES ANALYZED. WHEN INTERPRETING EACH CELL IN THE TABLE, THE UPPER VALUE REPRESENTS THE MEAN, WHILE THE LOWER VALUE DENOTES THE STANDARD DEVIATION

Occupancy	Mean and standard deviation (ADC counts)			
	Gauss MLE	OF	MAE	Logn MLE
10%	4.89 ± 0.07	0.02 ± 0.05	4.71 ± 0.04	5.46 ± 0.06
	22.36 ± 0.19	24.09 ± 0.18	20.53 ± 0.19	18.80 ± 0.20
30%	14.41 ± 0.21	0.03 ± 0.13	14.97 ± 0.09	13.49 ± 0.10
	36.55 ± 0.16	39.28 ± 0.18	34.66 ± 0.17	31.82 ± 0.20
50%	24.38 ± 0.24	0.02 ± 0.14	25.09 ± 0.11	19.83 ± 0.12
	44.64 ± 0.17	48.01 ± 0.19	43.33 ± 0.15	39.76 ± 0.17
80%	40.70 ± 0.36	0.06 ± 0.13	39.05 ± 0.16	25.28 ± 0.24
	51.63 ± 0.29	55.72 ± 0.25	51.50 ± 0.26	47.09 ± 0.26

error mean value is expected, since the noise distribution is a Gamma function. On the other hand, as expected, the OF approach maintains its average values close to zero, due to an additional constraint imposed within its optimization procedure, which forces the sum of the coefficients to be equal to zero [27]. However, this behavior does not imply better performance, since the standard deviation is more meaningful in this case.

In Fig. 2(c), the standard deviation of the error is shown as a function of occupancy. It can be observed that the Log-normal MLE presents the smallest values among the evaluated methods. As occupancy increases, this difference becomes more pronounced. Despite scaling the error bars by a factor of 5, they remain small, indicating the statistical stability of the performance. Numerical results from these analyses are provided in Table I, where methods are referred to by their abbreviated names.

More specifically, the results depicted in Fig. 2(b) demonstrate that, in more substantial occupancies, the means derived from the Log-normal MLE method exhibit lower values compared to those obtained by the MAE and Gaussian MLE methods. At the 10% occupancy, a notable proximity is observed among the results for all three methods, though the Log-normal MLE method positions slightly above the others. At 30% occupancy, the Log-normal MLE method begins to stand out slightly, and this discrepancy becomes more pronounced as the occupancy increases. Throughout the entire occupancy range, the MAE and Gaussian MLE methods exhibit very similar mean values.

The next aspect to consider is the standard deviation of the errors, as illustrated in Fig. 2(c), where the error bars are also scaled up by a factor of 5. It is observed that, for this parameter, all curves exhibit similar behavior, with the standard deviation value increasing proportionally to the occupancy for all methods. The Log-normal MLE method stands out across all occupancies; however, at the 10% occupancy, its standard deviation is very close to that of the MAE. As the occupancy increases, the difference between the standard deviation values of the Log-normal MLE and MAE also grows, highlighting the higher efficiency of the former compared to the latter.

Regarding the Gaussian MLE, values relatively close to those of the MAE are observed, especially for the 10% and 80% occupancy, where they are nearly identical at the latter. The OF, on the other hand, follows a similar pattern to the

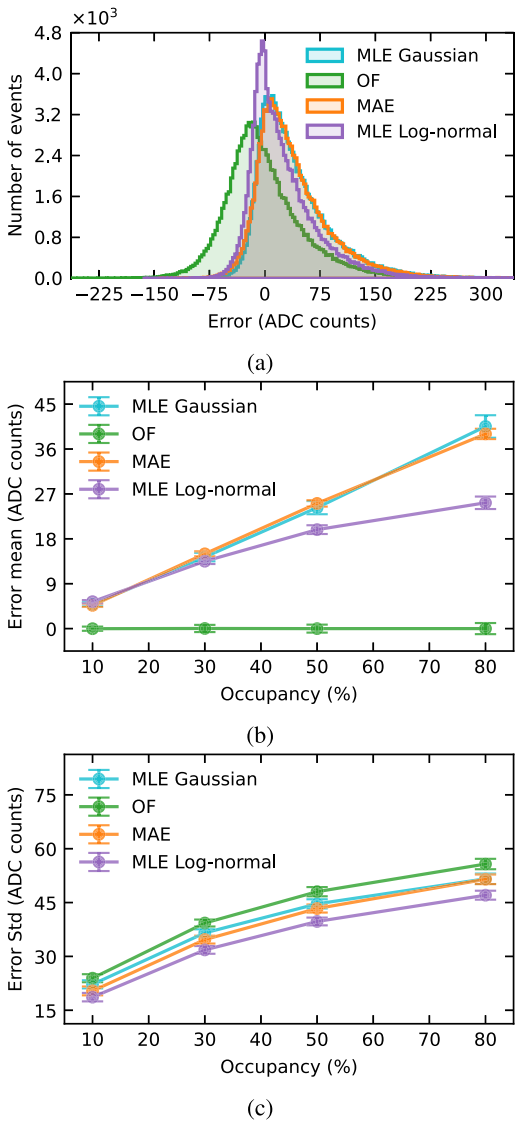

Fig. 2. (a) Error histograms for the four methods at 80% occupancy, (b) mean, and (c) standard deviation of errors for Gaussian MLE, OF, MAE, and Log-normal MLE, at occupancies of 10%, 30%, 50%, and 80%.

TABLE II

IMPROVEMENT (p) OF LOG-NORMAL MLE WHEN COMPARED TO THE OTHER METHODS

Occupancy	Improvement (%)		
	Gauss MLE	OF	MAE
10%	15.89	21.94	8.41
30%	12.94	19.01	8.19
50%	10.92	17.18	8.23
80%	8.79	15.50	8.57

other methods at 10% occupancy, showing considerable proximity to the Gaussian MLE, which also exhibits comparable values. However, as the occupancy increases, the efficiency of the OF method is compromised.

It is also possible to quantify the improvement achieved by Log-normal MLE concerning the other methods. To do so, one can utilize the following equation to estimate how much smaller the standard deviation of the error is for Log-normal MLE relative to the other methods:

$$p = \left(1 - \frac{\sigma_\ell}{\sigma_m}\right) \times 100\% \tag{11}$$

where p represents the percentage of improvement of Log-normal MLE, σ_ℓ denotes the standard deviation of the results associated with the Log-normal MLE approach, and σ_m stands for the standard deviation concerning the method denoted by m, where $m \in \{\text{Gauss MLE}, \text{OF}, \text{MAE}\}$. The outcomes of this analysis are presented in Table II, showing that the improvement of Log-normal MLE remains nearly constant for all evaluated occupancy levels when compared to the MAE method. However, for Gaussian MLE and OF, a decrease in enhancement is noted as the occupancy increases.

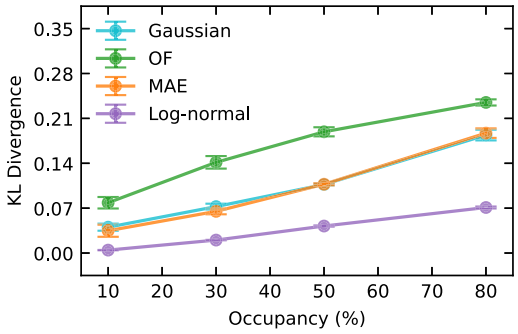

Fig. 3. KL divergence for Gaussian MLE, OF, MAE, and Log-normal MLE, with occupancies at 10%, 30%, 50%, and 80%.

Another analysis used to assess the efficiency of the methods is the comparison of the probability distributions of the estimated amplitudes with the true ones, using the Kullback–Leibler divergence (KL divergence) [35]. This metric measures how different two probability distributions are and provides insights into the amplitude estimation efficiency of each method. A KL divergence value of zero indicates identical distributions, with smaller values indicating greater similarity between the distributions. Analysis of Fig. 3 reveals that the Log-normal MLE method presents distributions closer to the true ones across all occupancy ranges, while Gaussian MLE and MAE appear closer to each other. The OF method, however, stands out as significantly different from the other methods, particularly at higher occupancy levels.

The results indicate that the KL divergence increases almost linearly with the occupancy for Log-normal MLE. Meanwhile, both the Gaussian MLE and MAE approach the Log-normal MLE at lower occupancies, but their distances grow at higher occupancies. Although OF initially appears similar to Gaussian MLE and MAE at lower occupancies, it diverges significantly at higher occupancies, only returning to approximate the others at 80% occupancy.

The relationship between the reference amplitude and the standard deviation of the error is depicted in Fig. 4. This evaluation provides information about the quality of estimation across the entire amplitude range. It presents analysis for all the methods under an occupancy of 80%. It can be observed that the Log-normal MLE method exhibits lower standard deviation values across the amplitude range, indicating superior estimation efficiency.

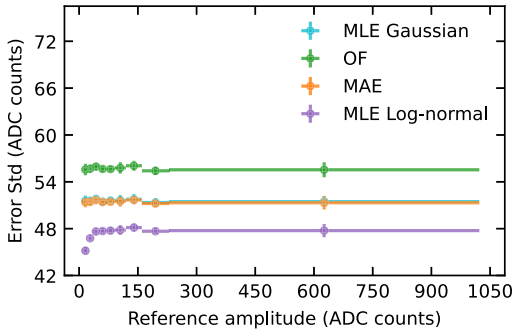

Fig. 4. Efficiency of Gaussian MLE, OF, MAE, and Log-normal MLE, for occupancy of 80%, in the energy range.

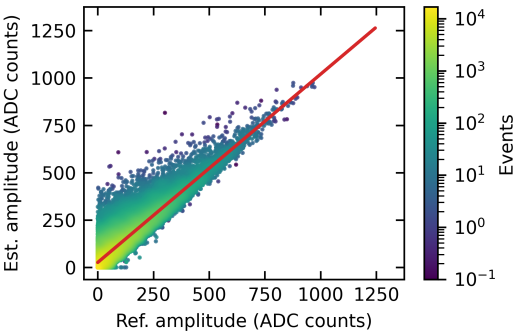

Fig. 5. Reference amplitude versus estimated amplitude for the log-normal MLE, for occupancy of 80%, and linear adjustment using the least squares method.

Similarly, Fig. 5 shows a 45° linear reference line, aligning with the distribution of points corresponding to the reference amplitude versus the estimated. The scatter plot demonstrates a linear trend, which is essential for a reliable estimator.

C. Data Quality Assessment

Since the proposed algorithm maximizes the likelihood estimation probability, this value can be used as a QF instead of the regular χ^2 method, offering a more reliable measure of the energy reconstruction quality under pile-up conditions.

To evaluate the performance, Fig. 6(a) shows the computed χ^2 values as a function of the energy estimation error for an occupancy of 50%, while Fig. 6(b) presents the probability values computed from the Log-normal pdf. The purpose of analyzing the correlation between χ^2 and the estimation error is to quantify their linear relationship, given that high values of the error are expected to have high corresponding χ^2 values. These metrics can be used to flag or discard such bad reconstructed signals, if desired. On the other hand, as a QF measurement, higher probability values related to small values of estimation error correspond to more accurate reconstructions.

From Fig. 6, it can be observed that the χ^2 values exhibit less correlation with the estimation error, whereas the probability-based QF shows larger values for smaller estimation errors, highlighting its more consistent relationship with signal reconstruction quality.

The correlation between estimation error and either χ^2 or probability is also crucial for assessing reconstruction quality:

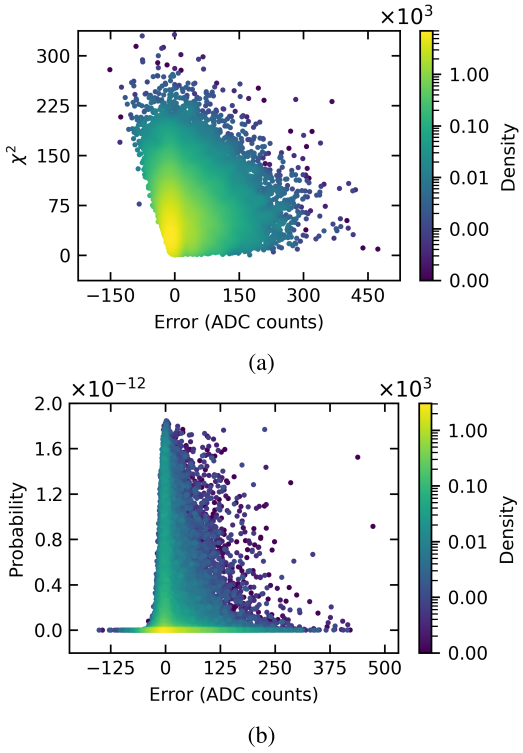

Fig. 6. (a) Error versus χ^2 values and (b) error versus probabilities of the log-normal pdf, for 50% of occupancy.

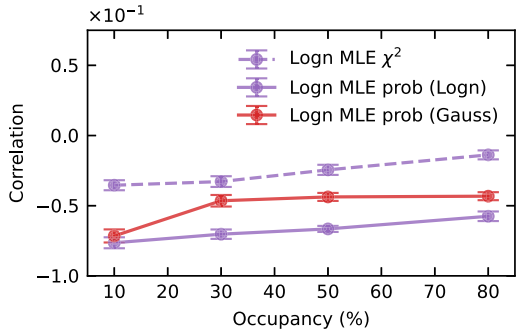

Fig. 7. Correlation between estimation error and χ^2, and estimation error and probability, with estimated amplitudes applied to Gaussian and log-normal PDFs, for all occupancies.

the higher the correlation (absolute value), the better the QF information. In this context, Fig. 7 shows the correlation coefficient when the amplitude estimated by Log-normal MLE is used in the χ^2 equation, Log-normal pdf, and Gaussian pdf. It is observed that the highest correlation is achieved with the Log-normal pdf, as the key factor is the magnitude of the values.

V. Conclusion

Signal estimation in modern calorimeter systems, particularly those operating in high-luminosity conditions, has become a significant challenge. This article introduces a novel approach by modeling the signal pile-up as a multivariate Log-normal distribution, which is evaluated through an MLE approach. The results demonstrate that the Log-normal model better captures the characteristics of the noise when compared

to the Gaussian model, regardless of the signal pile-up condition. Furthermore, the Log-normal MLE method provides a more accurate energy measurement with respect to traditional linear models in terms of both energy estimation error and KL divergence metrics, considering the whole range of pile-up conditions.

Additionally, the Log-normal likelihood value proved to be a useful quantity to flag good and bad energy estimates. Under signal pile-up conditions, this new QF metric exhibits a significantly stronger linear correlation with the estimation error compared to the commonly used χ^2 method. This provides a more robust and reliable metric for assessing energy reconstruction quality, particularly in high-luminosity environments where pile-up noise plays a crucial role.

Finally, to further improve the MLE performance, future work will focus on developing a pre-processing step designed to minimize the mutual information between the time-domain random variables and digitally ensure the pulse integrity [36], potentially enhancing the overall estimation accuracy. Furthermore, the use of real data to evaluate the performance of the proposed methods is also planned.

ACKNOWLEDGMENT

The authors are thankful to CAPES, CNPq, Faperj, Fapemig, and RENAFAE for the support.

REFERENCES

[1] R. Wigmans, *Calorimetry: Energy Measurement in Particle Physics*. London, U.K.: Oxford Univ. Press, 2000, doi: 10.1093/oso/9780198786351.001.0001.

[2] L. Evans and P. Bryant, "LHC machine," *J. Instrum.*, vol. 3, no. 8, Aug. 2008, Art. no. S08001, doi: 10.1088/1748-0221/3/08/s08001.

[3] R. R. Wilson. (1978). *The Tevatron*. Accessed: Jul. 22, 2025. [Online]. Available: lss.fnal.gov/archive/test-tm/0000/fermilab-tm-0763.pdf

[4] G. Bertuccio, E. Gatti, M. Sampietro, P. Rehak, and S. Rescia, "Sampling and optimum data processing of detector signals," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 322, no. 2, pp. 271–279, Nov. 1992, doi: 10.1016/0168-9002(92)90040-b.

[5] W. E. Cleland and E. G. Stern, "Signal processing considerations for liquid ionization calorimeters in a high rate environment," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 338, nos. 2–3, pp. 467–497, Jan. 1994, doi: 10.1016/0168-9002(94)91332-3.

[6] P. Adzic et al., "Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter," *Eur. Phys. J. C*, vol. 46, no. S1, pp. 23–35, Jul. 2006, doi: 10.1140/epjcd/s2006-02-002-x.

[7] B. S. Peralva, J. M. Seixas, L. M. A. Filho, and A. S. Cerqueira, "A matched filter based approach for high-energy estimation in calorimetry," *J. Instrum.*, vol. 16, no. 2, Feb. 2021, Art. no. P02016, doi: 10.1088/1748-0221/16/02/p02016.

[8] L. M. de A. Filho, B. S. Peralva, J. M. de Seixas, and A. S. Cerqueira, "Calorimeter response deconvolution for energy estimation in high-luminosity conditions," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 6, pp. 3265–3273, Dec. 2015, doi: 10.1109/TNS.2015.2481714.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[10] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. Cham, Switzerland: Springer, 2008, doi: 10.1007/978-0-387-74503-9.

[11] M. Delmastro, "Quality factor analysis and optimization of digital filtering signal reconstruction for liquid ionization calorimeters," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 600, no. 3, pp. 545–554, Mar. 2009, doi: 10.1016/j.nima.2008.12.064.

[12] C. Clement and P. Klimek, "Identification of pile-up using the quality factor of pulse shapes in the ATLAS tile calorimeter," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, Oct. 2011, pp. 1188–1193, doi: 10.1109/NSSMIC.2011.6154599.

[13] G. F. Knoll, *Radiation Detection and Measurement*, 4th ed., Hoboken, NJ, USA: Wiley, 2010.

[14] P. K. Khandai, P. Sett, P. Shukla, and V. Singh, "Hadron spectra in p+p collisions at RHIC and LHC energies," *Int. J. Modern Phys. A*, vol. 28, no. 16, Jun. 2013, Art. no. 1350066, doi: 10.1142/s0217751x13500668.

[15] M. Ajaz et al., "Charged particles transverse momentum and pseudo-rapidity distribution in hadronic collisions at LHC energies," *Entropy*, vol. 25, no. 3, p. 452, Mar. 2023, doi: 10.3390/e25030452.

[16] P. A. Zyla et al., "Review of particle physics," *Prog. Theor. Experim. Phys.*, vol. 2020, no. 8, Aug. 2020, Art. no. 083C01, doi: 10.1093/ptep/ptaa104.

[17] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*. New York, NY, USA: Wiley, 2000, doi: 10.1002/9780470627242.

[18] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed., Hoboken, NJ, USA: Wiley, 1995.

[19] A. Alzaid and K. S. Sultan, "Discriminating between gamma and lognormal distributions with applications," *J. King Saud Univ.-Sci.*, vol. 21, no. 2, pp. 99–108, Jul. 2009, doi: 10.1016/j.jksus.2009.07.003.

[20] H.-K. Cho, K. P. Bowman, and G. R. North, "A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission," *J. Appl. Meteorol.*, vol. 43, no. 11, pp. 1586–1597, Nov. 2004, doi: 10.1175/jam2165.1.

[21] D. A. Darling, "Book review: Limit distributions for sums of independent random variables," *Bull. Amer. Math. Soc.*, vol. 62, no. 1, pp. 50–52, Jan. 1956, doi: 10.1090/s0002-9904-1956-09978-1.

[22] C. C. Heyde, "On a property of the lognormal distribution," *J. Roy. Stat. Soc. Ser. B: Stat. Methodology*, vol. 25, no. 2, pp. 392–393, Jul. 1963, doi: 10.1111/j.2517-6161.1963.tb00521.x.

[23] G. Tarmast, "Multivariate log-normal distribution," in *Proc. 53rd Session ISI*, 2001, pp. 1–2. Accessed: Jul. 22, 2025.

[24] H. L. V. Trees, *Detection, Estimation and Modulation Theory, Part I*. Hoboken, NJ, USA: Wiley, 2001, doi: 10.1002/0471221082.

[25] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2001, doi: 10.1002/0471221317.

[26] Y. Dodge, *Lognormal Distribution*. New York, NY, USA: Springer, 2008, pp. 321–322, doi: 10.1007/978-0-387-32833-1_238.

[27] G. I. Gonçalves, B. S.-M. Peralva, J. M. de Seixas, L. M. de Andrade Filho, and A. S. Cerqueira, "Performance of optimal linear filtering methods for signal estimation in high-energy calorimetry," *J. Control, Autom. Electr. Syst.*, vol. 33, no. 5, pp. 1601–1611, Oct. 2022, doi: 10.1007/s40313-022-00907-0.

[28] D. P. Barbosa, L. M. D. A. Filho, B. S. Peralva, A. S. Cerqueira, and J. M. de Seixas, "Sparse representation for signal reconstruction in calorimeters operating in high luminosity," *IEEE Trans. Nucl. Sci.*, vol. 64, no. 7, pp. 1942–1949, Jul. 2017, doi: 10.1109/TNS.2017.2712420.

[29] J. P. B. S. Duarte, L. M. D. A. Filho, E. F. D. S. Filho, P. C. M. A. Farias, and J. M. de Seixas, "Online energy reconstruction for calorimeters under high pile-up conditions using deconvolutional techniques," *J. Instrum.*, vol. 14, no. 12, Dec. 2019, Art. no. P12017, doi: 10.1088/1748-0221/14/12/p12017.

[30] J. Chapman, "ATLAS Simulation Computing Performance and Pile-Up Simulation in ATLAS," in *Proc. LPCC Detect. Simul. Workshop*, 2011, pp. 1–33. Accessed: Jul. 23, 2025.

[31] S. Banerjee, "CMS simulation software," *J. Phys., Conf. Ser.*, vol. 396, no. 2, Dec. 2012, Art. no. 022003, doi: 10.1088/1742-6596/396/2/022003.

[32] Y. Dodge, "Kolmogorov–Smirnov test," in *The Concise Encyclopedia of Statistics*. New York, NY, USA: Springer, 2008, pp. 283–287, doi: 10.1007/978-0-387-32833-1_214.

[33] D. Freedman and P. Diaconis, "On the histogram as a density estimator: L_2 theory," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, no. 4, pp. 453–476, Dec. 1981, doi: 10.1007/bf01025868.

[34] D. Freedman, R. Pisani, and R. Purves, *Statistics*, 4th ed., New York, NY, USA: W. W. Norton & Company, 2007.

[35] D. I. Belov and R. D. Armstrong, "Distributions of the Kullback–Leibler divergence with applications," *Brit. J. Math. Stat. Psychol.*, vol. 64, no. 2, pp. 291–309, May 2011, doi: 10.1348/000711010x522227.

[36] T. M. Quirino, T. C. A. Paschoalin, G. I. Gonçalves, P. H. B. Lisboa, L. M. de Andrade Filho, and B. S.-M. Peralva, "Online pulse compensation for energy spectrum determination: A pole-zero cancellation and unfolding approach," *Electronics*, vol. 14, no. 3, p. 493, Jan. 2025, doi: 10.3390/electronics14030493.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - ROR identifier: 00x0ma614