Progress In Electromagnetics Research B, Vol. 100, 109–131, 2023

Variational Bayesian Learning for the Modelling of Indoor Broadband Powerline Communication Impulsive Noise

Florence Chelangat* **and Thomas Afullo**

Abstract—Powerline communication (PLC) noise is the main cause of reduced performance and reliability of the communication channel. The major source of these noise bursts, which distort and degrade the communication signal, is the arbitrary plugging in and unplugging of electric devices from the electrical network. It is therefore important to perform statistical modelling of the PLC noise characteristics to enable the development and optimisation of reliable PLC systems. This paper presents the Variational Bayesian (VB) Gaussian Mixture (GM) modelling of the amplitude distribution of the indoor broadband PLC noise. In the proposed model, a fully Bayesian treatment is employed where the parameters of the GM model are assumed to be random variables. Consequently, prior distributions over the parameters are introduced. The VB criterion is used to determine the optimal number of components where the Bayesian information criterion emerges as a limiting case. To find the parameters of the GM components, the variational-expectation maximisation algorithm is employed. Measurements from different indoor PLC environments are then used to validate the model. Thereafter, performance analysis is carried out, and the VB framework is compared to the Maximum Likelihood (ML) estimate method. It is observed that while the ML model performs better when the amplitude distribution contains multiple peaks, the VB framework offers high accuracy and good generalization to the measured data and is thus effective in modelling the amplitude distribution of the PLC noise.

1. INTRODUCTION

Noise in powerline communication (PLC) indoor environments significantly decreases the reliability of data transmission through the electrical network. This is due to their sporadic occurrence with varying amplitudes that impede the efficient transmission of communication signals, burst errors, and corruption of the transmitted data. These errors between the communication links arise from sources within and without the electrical network [1]. From within, there is interference by other communication signals transmitted through amplitude modulation that cause the electrical wires to behave like antennas. External sources include the various connection points in the PLC network and the unpredictable loads plugged in and out of the power outlets. Despite this challenge, the PLC technology is still able to achieve high data rates with the application of multiple-input multiple-output technique in [2] and has gained more popularity with the development of the smart grid system. Its various applications include smart metering, internet access, and home automation just to mention a few. In addition, its already existing vast and well-established infrastructure that covers regions to which other wired and wireless communication technologies are unavailable makes it a great alternative mode of communication.

For PLC to compete with the existing modes of data transmission, the reliability and performance of the channel still need to be improved. This is because the PLC noise is impulsive and much stronger than the noise present in other communication media, thus, cannot be described as an additive white Gaussian noise [1, 3]. Consequently, accurate modelling of the PLC noise is paramount in combating

Received 8 February 2023, Accepted 15 May 2023, Scheduled 31 May 2023
* Corresponding author: Florence Chelangat (florencebiwott7@gmail.com).
The authors are with the University of KwaZulu-Natal, South Africa.

the degradation of data signals and in turn enhancing the reliability of the PLC technology. PLC noise is broadly categorised as background noise and impulsive noise. The background noise is characterised by a lower power spectral density that varies with frequency and lasts between a few seconds and up to hours. It is mainly caused by the transmission of signals over short, medium, and long ranges by radio broadcasters and by low-power noise sources connected to the powerline. On the other hand, impulsive noise occurs over short durations, exhibits higher amplitudes, and is generated by switching transients and power supplies in the powerline network [4, 5]. Impulsive noise has attracted a lot of interest from researchers as it is the main cause of distortion to the transmitted signal in PLC. As such, extensive measurement campaigns have been carried out to identify the main generators of PLC impulsive noise which has also been observed to occur randomly and in bursts, and thus, is also referred to as bursty impulsive noise [1, 6–8]. Due to the variation in the amplitude characteristics, sporadic occurrence, and duration of the PLC impulsive noise, the development of a unified model is still a work in progress.

The level of noise that reaches the PLC receiver, or the receiver's signal-to-noise ratio, is crucial to the design process for a PLC system. The background and impulsive noise produced by all nearby devices, coupled with spurious narrowband interferences from the power line network and broadcast radio stations, together form noise at the receiver. A detailed discussion on the narrowband interference in the broadband PLC is discussed in [9]. In this study, the PLC noise amplitude distribution is assumed to be a superposition of various Gaussian components. As such, the major tasks in the implementation of the Gaussian mixture (GM) model are to determine the number of components in the GM as well as the amplitude density estimation of the PLC impulsive noise. Machine learning provides a powerful tool for determining the number of components as well as the parameters of the GM model due to its ability to learn and adapt to different types of data. Its application in the PLC noise modelling is still a new research area of interest and can be categorised as supervised learning, unsupervised learning, and reinforcement learning [10]. While only the input vectors are present in unsupervised learning, supervised learning involves both the input vectors and the associated target vectors in the training data. The reinforcement learning technique, on the other hand, entails identifying the best outputs through trial and error since in a given state, the outputs are not specified [10, 11].

This paper contributes to the modelling of the instantaneous PLC impulsive noise amplitude characteristics in the low-voltage network by considering an alternative approach known as Variational Bayesian (VB) learning in order to facilitate the design and implementation of more robust PLC protocols to improve the PLC performance and reliability. In this approach, a fully Bayesian treatment is employed where all the parameters are assumed to be random variables, and the number of components of the GM model is estimated using the VB framework referred to as VB criterion. In addition, the performance of the VB model is compared to the Maximum Likelihood (ML) model which is popularly used in parameter estimation.

The subsequent sections are organised as follows. Section 2 describes the various existing models used to characterise the behaviour of PLC noise and theoretical machine learning parameter approximation methods. In Section 3, the data acquisition and measurement setup are discussed, while Section 4, gives a detailed description of the VB GM modelling and model order selection. The results and discussion are then presented in Sections 5 and 6, where the VB and ML models are compared to measurements. Section 7 finalises the paper with concluding remarks.

2. RELATED WORKS

PLC noise models can be divided into two groups: those that take into account the temporal correlation nature of the PLC noise and those that are generated through independent and identically distributed (i.i.d.) realizations. Most of the models based on the temporal correlation employ Markov chains to characterise the duration and inter-arrival time of impulse noise. Various models that take into account the bursty nature of PLC impulsive noise incorporate the memory term, which defines the transition probability between the impulse noise samples. A generalization of the Gilbert-Elliott model, the partitioned Markov chain [6], takes into account both a set of impulse-free states and impulsive states that aim to capture the bursty behaviour of impulsive noise. This method, however, produces binary outputs and is hence more suited for binary communication channels. To mitigate this drawback, a continuous noise model that incorporates the memory term -the Markov-Gaussian model- is developed

in [12], which comprises impulsive and impulse-free time sequences where the noise sample is generated from a Gaussian distribution. The difference between the two states is the variance that will be lower in the impulse-free states than impulsive states. The major limitation of the Markov-Gaussian model is that it is only limited to two states. In order to address the limitations in [6] and [12], the Markov-Middleton characterisation, a continuous noise model with finite states whose variance is a function of the physical parameters of the noise incorporating the temporal correlation was developed in [13].

Although the models based on i.i.d. realizations do not capture the bursty nature of PLC noise, this drawback can be addressed by using multi-carrier modulation. As such, the time domain impulse noise is spread by discrete Fourier transform on all the sub-carriers in the frequency domain such that the form in which the noise occurred either in bursts or randomly is irrelevant [14]. The widely adopted PLC noise models in this category include the symmetric alpha stable, Bernoulli-Gaussian and Middleton Class A models [15]. It is worth mentioning that the noise models that take into account the bursty nature of the PLC noise are developed from the discontinuous impulse noise modelling approaches but with an additional memory term. The heavy tail characteristic of the symmetric alpha stable models, which is comparable to that present in the impulsive component of the PLC noise, is the basis for their application in the modelling of PLC noise. The shortcoming of the alpha-stable distributions, however, is that they have no meaningful analytical or closed form and are defined by solely their characteristic equations [3]. Both the Bernoulli-Gaussian and the Middleton Class A models also exhibit heavy tails, and the PLC amplitude distribution is modelled using a mixture of Gaussian distributions. Their distinction stems from the fact that although the Bernoulli-Gaussian model assumes just two states, namely, the impulse-free and impulsive noise states, the Middleton Class A model considers that impulses occur in a finite number of states [16]. In addition, the rate of occurrence of the Gaussian components in the Middleton Class A model follows the Poisson distribution while it follows the Bernoulli distribution in the Bernoulli-Gaussian model. Nevertheless, it was shown in [8] that while the Middleton class A noise model was created for man-made impulse interference, it does not adequately describe impulsive noise in PLC.

Various other noise models have also been implemented with the current shift towards machine learning. In [17], the bit error rate is modelled using the GM model where the parameters of the GM are estimated using the ML estimation approach. Further improvements on the same model are studied in [18], where the singularity problem present in the ML estimate is addressed, and the PLC impulsive noise is modelled using up to four components in the GM. Artificial neural network (which is a supervised learning approach) has been applied by [19] to detect the noise mixed with messages received in the orthogonal frequency division multiplexer to enhance the quality of noisy signals. The GM exhibits well-defined statistical properties and thus has been widely adopted in science and technology for modelling and approximation purposes. In PLC noise modelling, various noise models are also based on the GM [5, 13, 17, 20, 21].

The main tasks in the deployment of the GM model are the determination of the model parameters and finding the appropriate number of components in the GM. The most adopted technique used in statistical parameter estimation is the ML estimate [22]. In this method, the specific values that correspond to the local maximum of the likelihood function are chosen for specific values for the model parameters [22, 23]. A two-step algorithm known as expectation-maximization is then used to find the ML estimates for incomplete data in parametric models. One problem with this procedure is the presence of singularities in the likelihood function, and as such the ML is strictly not well defined. This occurs where one of the component parameters has the same value as one of the data points thus assigning infinite density at the location of the data point [18, 23]. Additionally, the ML framework is also known for its tendency to overfit data which becomes even worse for complicated models incorporating high-dimensional data [23]. Another limitation is that it cannot be used to optimise the model structure since it prefers complex models such that as the number of parameters increases, a better fit is obtained [22, 23].

To solve the challenges encountered by the ML framework, a fully Bayesian framework is proposed where the parameters are assumed to be random. A Bayesian model takes into account a finite or infinite class of models instead of focusing on a single model [22]. The posterior probability of each model given the dataset is then calculated. Thereafter, all of the individual models' predictions are averaged and weighted by their posteriors to generate predictions for the test data. By integrating the parameters, the Bayesian framework avoids over-fitting and has good generalization capabilities [22, 24].

Additionally, complex models are automatically penalized by being given a lower posterior probability; as a result, optimal structures can be found [22].

The major aims of algorithms based on Bayesian inference are to derive the right algorithms for integrating over the full parameter space as well as to define appropriate distribution functions for modelling the parameters. The latter task can be computationally demanding and leads to integrals that are difficult to solve. Consequently, the Markov Chain Monte Carlo (MCMC), Laplacian method, and variational learning have been adopted to determine the parameters [11, 22, 24]. In order to express the integrals, the Laplacian method uses the Taylor expansion approximation. Nonetheless, this approximation becomes computationally expensive in high dimensions and can produce poor approximation results [24]. The basis of MCMCs is the development of a Markov chain with the desired distribution. The validity of a certain sample draw is then determined through a selection mechanism. The drawback of the MCMC approaches is the difficulty in choosing acceptable distributions and in designing sampling strategies to create appropriate data samples [11, 22, 24]. MCMC algorithms may take a long time to converge because they are stochastic in nature. The VB framework simplifies the analytical computations of posterior distributions over hidden variables, parameters, and structures. The posteriors arise in standard forms and are then computed using an iterative process known as the variational expectation maximisation algorithm, whose convergence is guaranteed. Furthermore, the posteriors result from a free-form optimization procedure that naturally incorporates conjugate priors. Additionally, model selection is done by computing the posterior over structure, where the Bayesian information and minimum description length criteria emerge as a limiting case, and averaging over models to compute predictive values can be done analytically [11, 22, 25].

The VB framework for the GM model has been applied in the speaker clustering [26], modelling non-Gaussian and non-stationary noise processes in [27], as well as colour image segmentation in [24]. This paper presents the modelling of the PLC noise using the VB approach. The PLC noise is interpreted as a superposition of noise components, and as such the GM model is used to model the amplitude distribution. The mixture weights represent the occurrence probability of the impulse noise at the receiver at a particular time from different sources. The parameters of the GM model, which is the mixture weights, the means, and the precision are assumed to be random variables and are thus assigned prior distributions with the same functional form as the Gaussian likelihood function. Consequently, the resulting prior distributions are the Dirichlet distribution for mixture weights and the Normal-Gamma distribution for means and precisions. The variational expectation maximization framework is then used to determine the optimum hyper-parameters of the prior distributions and consequently the best family of variational distributions that maximises the lower bound.

3. MEASUREMENT SETUP

The increasing development in the number of electronic devices leads to the increase in the number of noise sources introduced into the PLC network and consequently the reduction in the reliability of the PLC technology as a mode of data transmission. These electric appliances are the main cause of impulsive noise that dramatically vary with time as they are randomly turned on and off resulting in bits and burst errors on the transmitted signal. As a result, extensive measurement campaigns have been carried out and also need to be performed to adequately capture and understand the behaviour of the noise in the PLC channel. In this work, measurement campaigns were carried out in three locations at the University of KwaZulu-Natal Electrical building namely: the Electronic Laboratory, Machines Laboratory, and Computer Laboratory.

The PLC noise measurements were performed in 1–30 MHz using the measurement setup shown in Fig. 1. The setup consists of a Rigol DS2202A digital storage oscilloscope (DSO), coupling circuitry, and a personal computer for the storage and processing of information. The DSO has the capability of measuring 14 million samples and was set to a sampling rate of 1 Giga samples per second utilising the maximum storage capacity of the oscilloscope with a window length of 14 ms. The coupling circuit isolates the DSO from the low-frequency high-voltage supply preventing damage to the equipment and comprises series capacitors, transient voltage suppressors, Zener diodes, and a 1:1 broadband transformer. The capacitors are used to prevent the transformer from saturating as well as filter out the low-frequency signals while transmitting the high-frequency signals. The transformer on the other

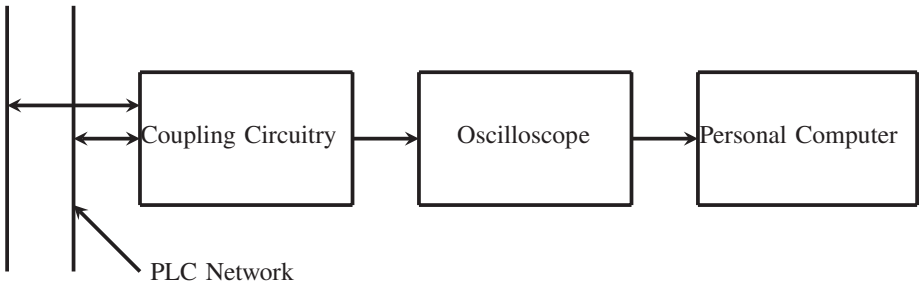


Figure 1. Measurement setup for PLC noise.

hand provides galvanic isolation while the Zener diodes protect the output against power surges by maintaining the output voltage at 5 V commonly used in communication devices. In [28], the coupling circuits have been observed to contribute to the effect of PLC impulsive noise by distorting the inter-arrival time, impulse duration, and amplitude of the impulse signal. This is due to the passive filter components present in the coupling circuitry where as the PLC noise passes through the coupler it excites points of resonance. In this work, the main focus is on the PLC noise as seen by the receiver, and as such, the effect of the coupling circuitry is not considered.

The sample PLC impulsive noise measurements for the Electronic Laboratory are shown in Fig. 2. In this location, the appliances connected to the PLC network include an electronic trainer board with components such as silicon-controlled rectifiers, digital multimeter, and cathode ray oscilloscope that are found in each workstation. The Electronic Laboratory also contains fluorescent lights and air-conditioners that are switched on during the practical sessions. In the Machines Laboratory, the powerline network loading comprises dc generators, ac induction motors, adjustable speed drives, air-conditioners, fluorescent lights, variable resistors, transformers, and measurement devices. The PLC impulsive noise measurements obtained from the Machines Laboratory are depicted in Fig. 3. Finally, the PLC noise measurement results obtained from the Computer Laboratory are shown in Fig. 4 where the main loads are sixty computers connected to the PLC network. In addition, there are also air-conditioners and fluorescent lights in this location. In all three locations, measurements were carried out at the peak hours between 2 pm and 5 pm when the students were undertaking their practicals, and all the electric devices had been switched on. Other than the various loads present in each location, which also serve as sources of noise generation, these locations serve only as representations of actual

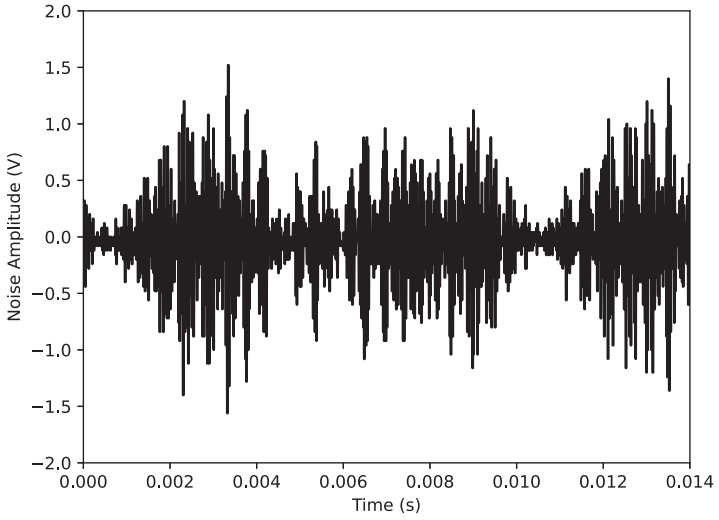


Figure 2. Electronic laboratory PLC noise data.

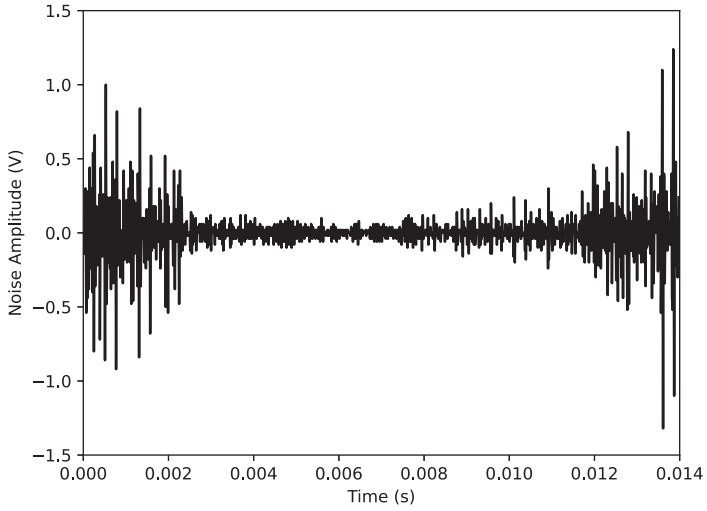


Figure 3. Machines laboratory PLC noise data.

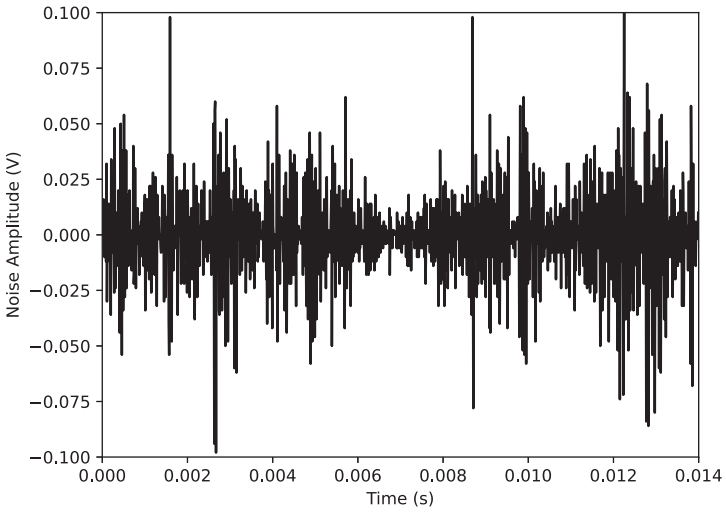


Figure 4. Computer laboratory PLC noise data.

PLC channels.

In previous research studies, the main contributors to the indoor PLC impulsive noise include lightning, switched mode supplies, fluorescent lights, vacuum cleaners, and light dimmers [3]. Further investigations on the sources of PLC noise in the residential indoor environment was carried out by [7]. It was observed that incandescent light dimmers, switching supplies such as silicon-controlled rectifiers, and universal motors were the main sources in the 5–500 KHz frequency range. Similar results were observed in [29], where the main causes of PLC impulsive noise comprised vacuum cleaners, hand-held drilling machines, heaters, and switched-mode power supplies. Although the level of noise produced by the fluorescent lights in both cases, that is at a frequency below 500 KHz, is less than that of the aforementioned electric devices, the interference levels compete with the electromagnetic compatibility levels at a higher frequency range of between 150 KHz and 30 MHz [30, 31]. With regard to the power supplies at the broadband range of up to 30 MHz, [32, 33] found that thermostats, rectifiers within the DC power supplies and power switches were the main source of impulsive noise. From the measurement results obtained in this work, the results follow the same trend as those can be observed in Figs. 2, 3,

and 4, where each environment produces different noise samples as a result of various noise sources. It is seen that the Electronic Laboratory and Machines Laboratory have higher noise levels than the Computer Laboratory.

4. VARIATIONAL BAYESIAN GAUSSIAN MIXTURE MODELLING

In this work, the amplitude distribution of the PLC noise is assumed to be independent and identically distributed (i.i.d) and is modelled as a one-dimensional GM with K components. Thus, all the component parameters of the GM comprising the means $\mu = [\mu_1, \ldots, \mu_k]$, precisions $\phi = [\phi_1, \ldots, \phi_k]$ and mixture weights $\pi = [\pi_1, \ldots, \pi_k]$ can be collectively written as $\theta := \{\pi, \mu, \phi\}$. Let the observed PLC noise amplitude at a particular time be given by x_n for $n = [1, \ldots, N]$, where N is the number of samples at a given observation window, and for each data point x_n there is a latent variable z_n for $z = 1, \ldots, N$ that indicates which component generated the n_{th} data point. Thus, the likelihood of data point x_n is defined by [11, 25]:

$$p(x_n \mid \theta) = \sum_{k=1}^{K} \pi_k \left(\frac{\phi_k}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\phi_k}{2}(x_n - \mu_k)^2\right) \tag{1}$$

where the mixing probabilities are subject to the constraint:

$$\sum_{k=1}^{K} \pi_k = 1 \tag{2}$$

and $0 \leqslant \pi_k \leqslant 1$. However, the mixture weights, means, and precisions of the component Gaussian distributions are unknown. As such, the VB framework introduces prior distributions to the latent variables and parameters that have the same functional form as the Gaussian likelihood function. Consequently, the conjugate prior to the multinomial distribution $p(z_n \mid \pi)$ is Dirichlet given by [11, 25]:

$$p(\pi) = Dir(\pi \mid \lambda_0) = C(\lambda_0) \prod_{k=1}^{K} {\pi_k}^{\lambda_o - 1} \tag{3}$$

where λ_0 is the initial number of samples in each Gaussian component and is chosen to be the same for all the components in the GM, and $C(\lambda_0)$ is the normalization constant for the Dirichlet distribution and is given by:

$$C(\lambda_0) = \frac{\Gamma\left(\sum_{k=1}^{K} \lambda_k\right)}{\prod_{k=1}^{K} \Gamma(\lambda_k)} \tag{4}$$

Similarly, the conjugate prior distribution over the means and precision is a Normal-Gamma distribution obtained as [11, 25]:

$$p(\mu, \phi) = \prod_{k=1}^{K} N(\mu_k \mid t_0, (s_0\phi_k)^{-1}) G(\phi_k \mid \alpha_0, \beta_0) \tag{5}$$

where t_0 and s_0 are the initial hyper-parameters for the means, whereas α_0 and β_0 are the initial hyper-parameter values for the precision, and $G(\phi_k \mid \alpha_0, \beta_0)$ can be expressed as [25, 34]:

$$G(\phi_k \mid \alpha_0, \beta_0) = \frac{\phi_k^{\alpha_0 - 1} exp\left(-\frac{\phi_k}{\beta_0}\right)}{{\beta_0}^{\alpha_0} \Gamma(\beta_0)} \tag{6}$$

Since the measured data is assumed to be i.i.d., the joint distribution of all the random variables conditioned on the number of clusters K can be determined as [11, 25, 34]:

$$p(X, Z, \pi, \mu, \phi) = p(\mu, \phi) \prod_{n=1}^{N} p(\pi) p(X \mid Z, \mu, \phi) p(Z \mid \pi) \tag{7}$$

where:

$$p(Z \mid \pi) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \tag{8}$$

and $p(\mu, \phi)$ in (5) can be rewritten as:

$$p(\mu, \phi) = \prod_{k=1}^{K} \left(\frac{s_0 \phi_k}{2\pi} \right)^{\frac{1}{2}} \exp \left(\left(-\frac{s_0 \phi_k}{2} (\mu_k - t_0)^2 \right) \left(\frac{\phi_k^{\alpha - 1} \exp \left(-\frac{\phi_k}{\beta_0} \right)}{{\beta_0}^{\alpha_0} \Gamma(\beta_0)} \right) \right) \tag{9}$$

In order to compute the marginal likelihood (model evidence) $p(X)$, the joint distribution in (7) is marginalised with respect to π, Z, μ, ϕ.

The main tasks in the application of the GM model are to determine the true posterior distribution and to evaluate the possible number of components. The application of a fully Bayesian treatment implies that the posterior distribution is obtained as [11, 25]:

$$p(Z, \pi, \mu, \phi \mid X) = \frac{p(Z, \pi, \mu, \phi, X)}{p(X)} \tag{10}$$

It can be observed that solving for (10) involves a computationally intensive evaluation of multi-integrals for which expectations are analytically intractable. As such, the variational Bayes method is used to infer the posterior distribution for the mixture weights, means, and precision where an approximate joint variational distribution $q(Z, \pi, \mu, \phi)$ is introduced.

4.1. Variational Approximation Framework

Let the set of all the latent variables and parameters be denoted by $\eta = [Z, \pi, \mu, \phi]$. Thus, the joint variational distribution will be given by $q(\eta)$, and the marginal log-likelihood of the observed data can be obtained as [11, 24]:

$$\log p(X) = \log \int p(X, \eta) d\eta \tag{11}$$

Multiplying both the numerator and denominator of (11) by $q(\eta)$, the marginal log-likelihood becomes:

$$\log p(X) = \log \int q(\eta) \frac{p(X, \eta)}{q(\eta)} d\eta \tag{12}$$

In order to find a solution $q(\eta)$ that provides a tight lower bound to the true posterior distribution, the Jensen's inequality is applied. Therefore, (12) becomes [11, 24]:

$$\begin{aligned} \log p(X) &= \log \int q(\eta) \left(\frac{p(X, \eta)}{q(\eta)} \right) d\eta \\ &\geq \int q(\eta) \log \left(\frac{p(X, \eta)}{q(\eta)} \right) d\eta = L_q(\eta) \end{aligned} \tag{13}$$

where $L_q(\eta)$ denotes the lower bound (or the negative free energy) of the marginal log-likelihood which can also be expressed as:

$$L_q(\eta) = \mathbb{E}_q(\log p(X, \eta) - \mathbb{E}_q(\log q(\eta)) \tag{14}$$

where $\mathbb{E}_q(.)$ denotes the expectation over all the variables. To measure the distance between the true posterior distribution and the approximate variational posterior distribution, the Kullback-Leibler(KL) divergence is employed. As such, the KL divergence between $p(\eta \mid X)$ and $q(\eta)$ is given by:

$$KL(q(\eta) \mid\mid p(\eta \mid X)) = \int q(\eta) \log \left(\frac{q(\eta)}{p(\eta \mid X)} \right) d\eta \tag{15}$$

From the product rule of probability, $p(\eta \mid X)$ can be defined as:

$$p(\eta \mid X) = \frac{p(X, \eta)}{p(X)} \tag{16}$$

Substituting (16) in (15), the KL divergence becomes [11]:

$$\begin{aligned} KL(q(\eta) \,||\, p(\eta \mid X)) &= -\int q(\eta) \log\left(\frac{p(X,\eta)}{q(\eta)}\right) d\eta \\ &\quad + \log p(X) \\ &= -\left(\mathbb{E}_q(\log p(X,\eta) - \mathbb{E}_q(\log q(\eta))\right) \\ &\quad + \log p(X) \end{aligned} \tag{17}$$

Therefore,

$$KL(q(\eta) \,||\, p(\eta \mid X)) = -L_q(\eta) + \log p(X) \tag{18}$$

Consequently, the variational log-likelihood can be obtained as:

$$\begin{aligned} \log p(X) &= L_q(\eta) + KL(q(\eta) \,||\, p(\eta \mid X)) \\ &= \int q(\eta) \log\left(\frac{p(X,\eta)}{q(\eta)}\right) d\eta \\ &\quad - \int q(\eta) \log\left(\frac{p(\eta \mid X)}{q(\eta)}\right) d\eta \end{aligned} \tag{19}$$

Since the marginal log-likelihood $\log p(X)$ is independent of $q(\eta)$, maximizing the lower bound is equal to minimizing the KL divergence which occurs when $q(\eta) = p(X,\eta)$ and hence $KL(q(\eta) \,||\, p(\eta \mid X)) = 0$ [11]. Therefore, the main objective in variational Bayes is to minimise the KL divergence given by (18), which is equivalent to the maximization of the lower bound. As such, the lower bound will be as close as possible to the true posterior. Assuming that the approximating joint distribution is factorised over the parameters and latent variables, $q(\eta)$ becomes [11, 34]:

$$q(Z, \pi, \mu, \phi) = q(Z)q(\pi)q(\mu, \phi) \tag{20}$$

where $q(Z)$, $q(\pi)$ and $q(\mu, \phi)$ are given by:

$$q(Z) = \prod_{k=1}^{K} \prod_{n=1}^{N} \gamma_{nk}{}^{z_{nk}} \tag{21}$$

$$q(\pi) = \Gamma(\sum_{k=1}^{K} \lambda_k) \prod_{k=1}^{K} \frac{\pi_k{}^{\lambda_k - 1}}{\Gamma(\lambda_k)} \tag{22}$$

$$q(\mu, \phi) = \prod_{k=1}^{K} N(\mu_k \mid t_k, (s_k\phi_k)^{-1}) G(\phi_k \mid \alpha_k, \beta_k) \tag{23}$$

Accordingly, the evidence lower bound in (14) can be determined as a functional over the parameters as [11, 34]:

$$L_q(\eta) = \mathbb{E}_q\left(\log \frac{p(Z, \pi, \mu, \phi, X)}{q(Z)q(\pi)q(\mu, \phi)}\right) \tag{24}$$

The explicit hyper-parameter values are then derived from estimating the log of the joint distribution and the lower bound optimised using a coordinate ascent algorithm analogous to the expectation-maximization.

4.2. Lower Bound Maximization

The log of the joint distribution can be obtained as [11, 34]:

$$\log p(Z, \pi, \mu, \phi, X) = \log p(Z \mid \pi) + \log p(\pi) + \log p(\mu, \phi) + \log p(X \mid Z, \mu, \phi) \tag{25}$$

Considering only the terms in the joint distribution that are associated with the latent variable and absorbing those independent of Z into the additive normalization constant, the optimal solution for the latent variables $q(Z)$ can be determined as:

$$\log q^{+}(Z) = \mathbb{E}_{\pi}(\log p(Z \mid \pi) + \mathbb{E}_{\mu,\phi}(\log p(X \mid Z, \mu, \phi)) + constant \tag{26}$$

Substituting for $p(Z \mid \pi)$ and $p(X \mid Z, \mu, \phi)$, Equation (26) can be further simplified as [11]:

$$\log q^{+}(Z) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log \tau_{nk} + constant \tag{27}$$

where:

$$\begin{aligned} \log \tau_{nk} \; = \; & \mathbb{E}_{\pi,\mu\phi} \left(\log \pi_k + \frac{1}{2} \left(\log \phi_k - \log 2\pi \right) \right) \\ & - \mathbb{E}_{\pi,\mu\phi} \left(\frac{\phi_k}{2} (x_n - t_k)^2 \right) \end{aligned} \tag{28}$$

Evaluating (28) yields [25, 34]:

$$\tau_{nk} = \exp \left(\widehat{\pi_k} + \frac{\widehat{\phi_k}}{2} - \frac{1}{2} \left(\frac{1}{s_k} + \alpha_k \beta_k (x_n - t_k)^2 \right) \right) \tag{29}$$

where:

$$\mathbb{E}(\log \pi_k) \; = \; \psi(\lambda_k) - \psi(\bar{\lambda}) = \widehat{\pi_k} \tag{30}$$

$$\mathbb{E}(\log \phi_k) \; = \; \psi(\alpha_k) + \log(\beta_k) = \widehat{\phi_k} \tag{31}$$

$\mathbb{E}(\mu_k) = t_k$, $\mathbb{E}({\mu_k}^2) = {t_k}^2 \alpha_k \beta_k + \frac{1}{s_k}$ and $\bar{\lambda}$ is given by [11, 34]:

$$\bar{\lambda} = \sum_{k=1}^{K} \lambda_k \tag{32}$$

$\psi(.)$ represents the digamma function defined by $\psi(y) = \frac{d \log \Gamma(y)}{dy}$. Thus, the normalized $q^{+}(Z)$ can be obtained as:

$$q^{+}(Z) = \prod_{n=1}^{N} \prod_{k=1}^{K} {\gamma_{nk}^{+}}^{z_{nk}} \tag{33}$$

where γ_{nk}^{+} are the responsibilities for the variational approximation and are given by:

$$\gamma_{nk}^{+} = \frac{\tau_{nk}}{\sum_{j=1}^{K} \tau_{nj}} \tag{34}$$

For a discrete distribution, $\mathbb{E}(z_{nk}) = \gamma_{nk}^{+}$ [11]. To characterize the new posterior parameters, the following quantities are computed from the responsibilities:

$$N_k = \sum_{n=1}^{N} \gamma_{nk}^{+} \tag{35}$$

$$x_k^{+} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}^{+} x_n \tag{36}$$

$$s_k^{+} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}^{+} (x_n - x_k^{+})^2 \tag{37}$$

where (35), (36), and (37) represent the number of data points in component K, weighted data values, and weighted squared data values, respectively. Similarly, considering only the terms in the joint distribution that are associated with the mixture weights and absorbing those independent into the

additive normalization constant, the optimal solution for the mixture weights can then be obtained from (25) as:

$$\begin{aligned}\log q^{+}(\pi) &= \mathbb{E}_{\pi}\left(\log p(\pi) + \log p(Z \mid \pi)\right) + constant \\ &= \sum_{k=1}^{K}\left((N_k + \lambda_0)\log \pi_k - \log \pi_k\right)\end{aligned} \tag{38}$$

Hence, $q^{+}(\pi)$ is a Dirichlet distribution given by:

$$q^{+}(\pi) = Dir(\pi \mid \lambda) \tag{39}$$

and λ is a hyper-parameter given by:

$$\lambda_k = \lambda_0 + N_k \tag{40}$$

Solving for the optimal solution for the means and precision:

$$\log q(\mu, \phi) = \mathbb{E}_{\mu,\phi}\left(\log p(X \mid Z, \mu, \phi) + \log p(\mu, \phi)\right) + constant \tag{41}$$

The corresponding hyper-parameters for the means (t_k, s_k) and precision (α_k, β_k) can then be given by [25, 34]:

$$t_k = \frac{1}{s_k}(t_0 s_0 + x_k^{+} N_k) \tag{42}$$

$$s_k = s_0 + N_k \tag{43}$$

$$\alpha_k = \alpha_0 + \frac{N_k}{2} \tag{44}$$

$$\frac{1}{\beta_k} = \frac{1}{\beta_0} + \frac{1}{2}\left(N_k {s_k}^{+} + \frac{s_0 N_k}{s_k}(x_k^{+} - t_0)^2\right) \tag{45}$$

The detailed derivations for the updated equations are illustrated in Appendix A. To obtain the best family of variational distributions that maximize the lower bound, the variational expectation maximization (VEM) algorithm is employed and involves five steps as follows [22, 25]:

1. Initialize the mixing weights π by the number of clusters which is also a hyper-parameter. Then initialize means and precision from the measured data X. Consequently, compute the initial hyper-parameters a_0, b_0, α_0, β_0, and v_k for the prior distributions.
2. Variational expectation step: Evaluate the posterior probability γ_{nk} using the current hyper-parameters.

$$\gamma_{nk}^{+} = \frac{\tau_{nk}}{\sum_{j=1}^{K}\tau_{nj}} \tag{46}$$

where:

$$\tau_{nk} = \exp\left(\widehat{\pi_k} + \frac{\widehat{\phi_k}}{2} - \frac{1}{2}\left(\frac{1}{s_k} + \alpha_k\beta_k(x_n - t_k)^2\right)\right) \tag{47}$$

3. Variational maximization step: Reevaluate the posterior hyper-parameters in two stages:
   – Compute the intermediate variables;

$$N_k = \sum_{n=1}^{N}\gamma_{nk}^{+} \tag{48}$$

$$x_k^{+} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma_{nk}^{+} x_n \tag{49}$$

$$s_k^{+} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma_{nk}^{+}(x_n - x_k^{+})^2 \tag{50}$$

– Evaluate the new hyper-parameter values ;

$$\lambda_k = \lambda_0 + N_k \tag{51}$$

$$t_k = \frac{1}{s_k}(t_0 s_0 + x_k^+ N_k) \tag{52}$$

$$s_k = s_0 + N_k, \tag{53}$$

$$\alpha_k = \alpha_0 + \frac{N_k}{2} \tag{54}$$

$$\frac{1}{\beta_k} = \frac{1}{\beta_0} + \frac{1}{2}\left(N_k s_k^+ + \frac{s_0 N_k}{s_k}(x_k^+ - t_0)^2\right) \tag{55}$$

4. Evaluate the average log-likelihood at iteration t as [24]:

$$\log_t p(X \mid \theta) = \frac{1}{N}\sum_{n=1}^{N} \log p_t(x_n \mid \theta) \tag{56}$$

5. Check for convergence of the log-likelihood by comparing the variation ϵ in the normalized log-likelihood. If the difference falls below a predefined value (0.01%), the process terminates otherwise return to step 2 [24].

$$\epsilon > \frac{\log_t p(X \mid \theta) - \log_{t-1} p(X \mid \theta)}{\log_t p(X \mid \theta)} \tag{57}$$

4.3. Model Order Selection

Determining the number of components that make up a mixture model in machine learning techniques is crucial because too many or too few components can cause the measured data to be over- or under-fitted, which in turn results in poor generalization, or the model's inability to give a high probability to data that is randomly selected from the same distribution as the training set [24, 23]. Consequently, various studies have been done in the modelling of the PLC impulsive noise using mixture models with different numbers of components. In [15], the Middleton Class A model was modelled using a GM of up to five components. It was observed that as the number of components increases, the accuracy of the model also increases. Similar results were also found in [18] where the GM model which employed the ML estimation method to determine the optimal parameters was used to model the PLC impulsive noise. In this case, two, three, and four-component GM models were used, and the components that do not contribute to the GM were automatically pruned out as their mixture weights were reduced to zero. Further work on the number of components needed to model the GM was done in [17], where the mutual information theory framework was used to determine the number of components in the estimation of the bit error rate of bursty impulsive noise in low-voltage PLC networks.

The most common method of parameter estimation in machine learning is the ML framework in which a specific value of the model parameters is chosen which corresponds to a local maximum of the likelihood function [11]. The expectation-maximization algorithm is then used to find the ML solutions. As a result, it offers no guidance about the selection of the number of component since larger values enable the model to obtain a better fit to the training data and, as a result, assign larger values of the likelihood function for the observed data set.

The VB approach provides a means of finding the number of components similar to the Bayesian Information Criterion (BIC) while training the model. Let M represent a large possible number of model structures and K represent the optimal model for a fixed number of components for $K \in [2, M]$. If the prior distribution of the number of components is given by $p(K)$, then [34]:

$$p(Z, \pi, \mu, \phi, X, M) = \prod_{K=1}^{M} p(Z, \pi, \mu, \phi, X \mid K)p(K) \tag{58}$$

where $p(Z, \pi, \mu, \phi, X \mid K)$ denotes the joint distribution defined in (7) since it is dependent on a fixed number of components K. From the VB methodology, the posterior distribution will be given by:

$$q(Z, \pi, \mu, \phi, K) = q(Z, \pi, \mu, \phi)q(K) \tag{59}$$

where the updates for $q(Z, \pi, \mu, \phi, K)$ conditioned on a specific K have already been defined in Section 4.1. Thus, it can be observed that the normalised posterior distribution $q(K)$ can be obtained as [25, 34]:

$$q(K) = \frac{\exp(Lq(\eta m)p(K))}{\sum_{K'}^{M} Lq(\eta m')p(K')} \tag{60}$$

where $Lq(\eta m)$ is the optimal lower bound for the K_{th} component mixture model. Choosing a uniform prior distribution $p(K) = \frac{1}{M}$, the the optimal posterior distribution becomes [25, 34]:

$$q(K) = \frac{\exp Lq(\eta m)}{\sum_{K'}^{M} Lq(\eta m')} \tag{61}$$

The lower bound in (24) can be reexpressed as [11, 25, 34]:

$$L_q(\eta) = \mathbb{E}_q\left(\log \frac{p(Z, \pi, \mu, \phi, X)}{q(Z)}\right) - KL\left(q(\mu, \phi) \,||\, p(\mu, \phi)\right) - KL\left(q(\pi) \,||\, p(\pi)\right) \tag{62}$$

Thus, the VB criterion in (62) compares to the BIC, where the first term corresponds to the average log-likelihood, and the second and third terms correspond to the KL divergence between the approximate posterior and prior distributions. The KL divergence, therefore, acts as a penalty term that penalises complex models while the lower bound can be used exactly as a model selection criterion. The parameter posterior sharply peaks at the most likely values, which are also the ML values θ^+, as the number of samples increases. Consequently, as $N \to \infty$, the penalty term reduces to $\frac{h}{2}\log N$, where h is the number of parameters in the model, and the lower bound becomes equivalent to the BIC given by [25, 26]:

$$BIC_{(K)} = \log p(X \mid \theta^+) - \frac{h}{2}\log N \tag{63}$$

Hence, the BIC emerges as a limiting case of the VB framework, and in the case of the univariate GM model, $BIC_{(m)}$ becomes [25]:

$$BIC_{(K)} = \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k N(x_n \mid \mu_k, \phi_k)\right) - \frac{3K}{2}\log N \tag{64}$$

The number of components with the highest value of $BIC_{(K)}$ corresponds to the required value of K.

5. RESULTS

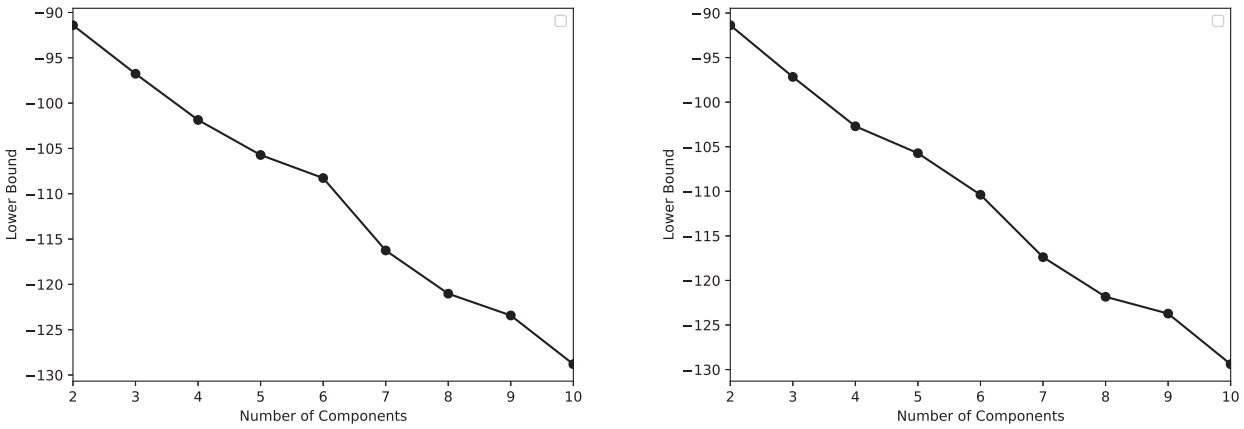
The VB model was validated by measurement results obtained at different locations with different loading conditions. In each venue, two data sets were used so as to ascertain the PLC noise levels as well as the distribution over time. In order to find the best model that fits the data, the optimum number of components is first determined and then the VB framework used to approximate the posterior distribution that provides a tight lower bound. Thereafter, performance analysis is done to check on the accuracy and significance of the proposed model, which is then compared to the common ML estimation approach.

5.1. Selection of Number of Components

The VB model provides a mechanism for the determination of the optimal number of components as summarised in Section 4, where the lower bound is interpreted as the Bayesian information criterion while the KL divergence acts as a penalty term that penalises complex models. As the number of components increases, the KL distance between the priors and posteriors increases, and as a result, the lower bound decreases. Conversely, as the average log-likelihood increases, the lower bound also increases resulting in a higher lower bound value. Thus, for a model to have a high value for the lower

bound, the average log-likelihood must exceed the KL distance. As such, the VB criterion is based on selecting the simplest model, that is, the one with the least number of components, but provides a good fit to the measured data. The optimal model will thus be the one with the highest lower bound after the model complexity has been considered indicating a high average log-likelihood which in turn results in a better fit.

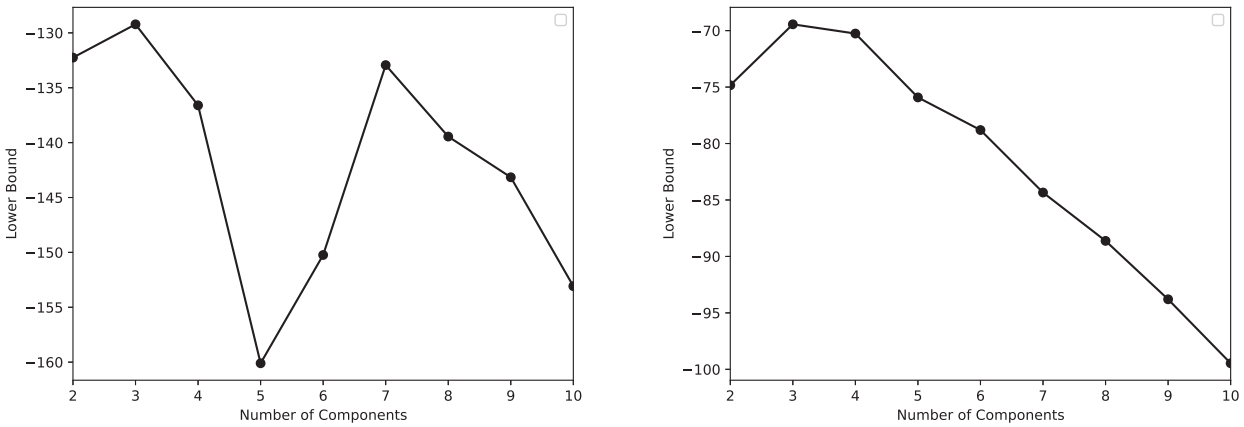
The optimal model selection for the Electronics Laboratory sampled data is shown in Figs. 5(a) and 5(b). It can be observed that as the number of components increases, the value of the lower bound also decreases with the maximum lower bound occurring in the two-component GM model. This is because the increase in the average log-likelihood as the model complexity increases is lower than the increase in the KL distance. Figs. 5(a) and 5(b) are also observed to follow the same trend. This can be attributed to the measurement data having almost similar characteristics to those further seen in Figs. 8 and 9 where the VB model and ML models overlap. As for the Machines Laboratory depicted in Figs. 6(a) and 6(b), the three-component mixture model has the highest value of the lower bound in both cases. In Fig. 6(a), it can be seen that the five-component GM has the lowest lower bound value. This can be attributed to a very small increase in the average likelihood with a high increase in the penalty term. As for Fig. 6(b), the lower bound peaks at the three-component model after which it decreases as the model complexity increases. The maximum lower bounds for the PLC noise samples



(a) Model order selection for data-1.

(b) Model order selection for data-2.

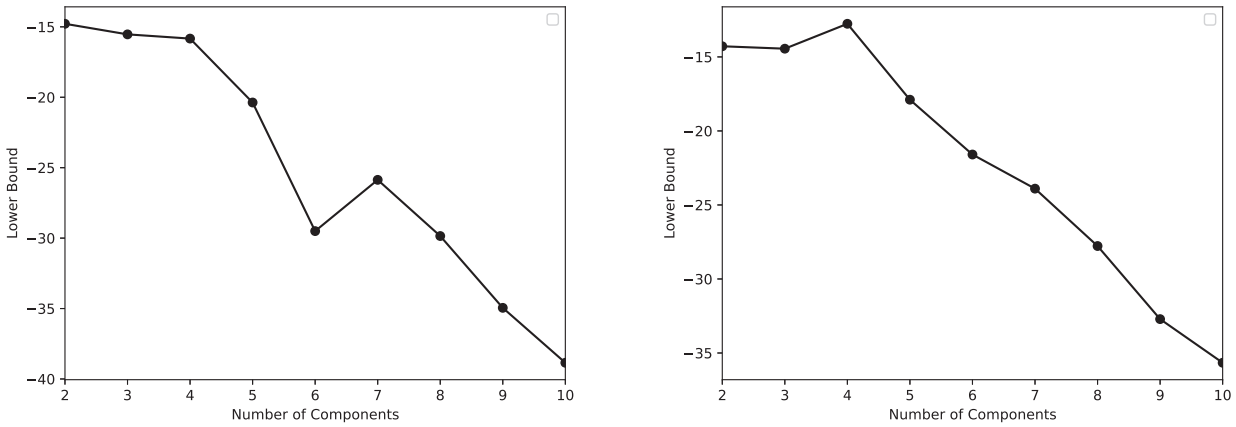
Figure 5. Electronic laboratory.



(a) Model order selection for data-1.

(b) Model order selection for data-2.

Figure 6. Machines laboratory.



(a) Model order selection for data-1.

(b) Model order selection for data-2.

Figure 7. Computer laboratory.

collected from the Computer Laboratory can be seen to occur at the two-component mixture model for the first data and the four-component model for the second data as shown in Figs. 7(a) and 7(b), respectively.

5.2. Impulsive Noise Amplitude Distribution

The optimal number of components determined from Section 5.1 for the various locations was then used to model the PLC noise amplitude distribution using both the VB and the MLE GM models. The predictions from the model were then compared to the measured data. The probability density function (PDF) measurement data for the Electronic Laboratory was observed to exhibit spikes in both the measured data shown in Figs. 8 and 9. The VB and the ML models, in this case, are seen to superimpose and exhibit the same heavy tails as the measurement PDF. This may be attributed to the fact that as the number of samples increases, the VB inference converges to the ML [11]. It is noted in Figs. 8 and 9 that there are outliers in the curves corresponding to measurements. This can be attributed by the silicon-controlled rectifiers present in the electronic trainer board, which have been

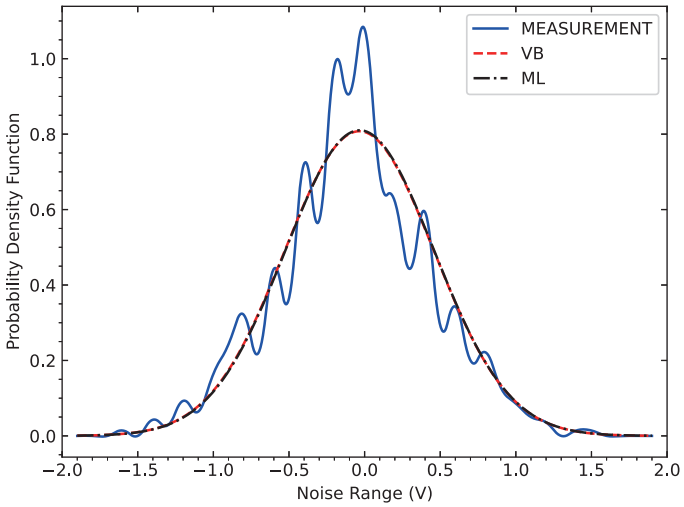


Figure 8. Electronic laboratory PLC noise distribution (data-1).

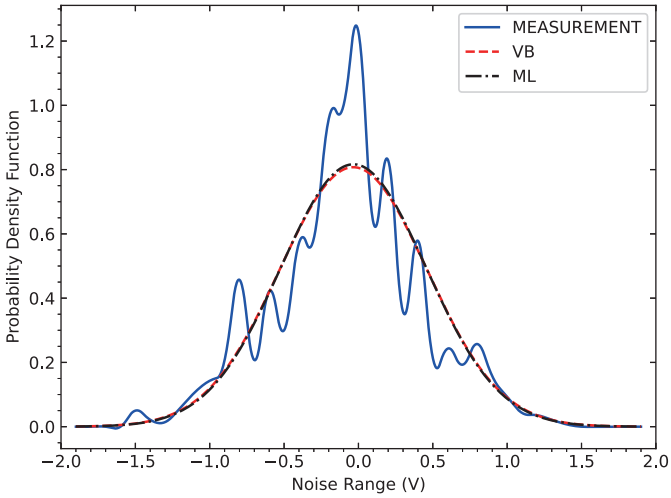


Figure 9. Electronic laboratory PLC noise distribution (data-2).

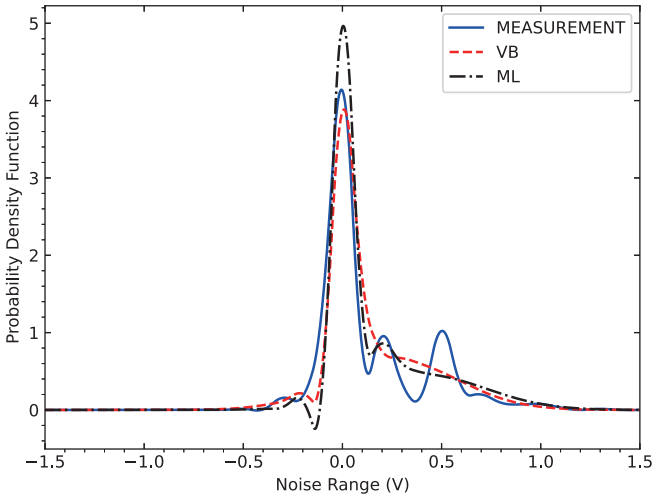


Figure 10. Machines laboratory PLC noise distribution (data-1).

observed to produce significant levels of impulsive noise. The PDF results obtained for the Machines Laboratory are shown in Figs. 10 and 11. It is observed that the measured impulsive noise PDF has multiple peaks with the highest peak occurring at zero. Of the three PDFs the ML estimate has the highest peak while the VB peak at zero is lower than that of the measurement data. In terms of the other peaks, the ML estimate is seen to follow the measurement result closely as compared to the VB. This is due to the highly flexible nature of the ML model. Also, the VB model does not provide the exact solution but rather an approximation that gives the tightest lower bound to the measurement.

Figures 12 and 13 show the PLC noise amplitude distribution for the Computer Laboratory where the peak-to-peak voltage is low and is between $-0.1v \leqslant x \leqslant 0.1v$. In Fig. 12, the VB and the ML estimates peak at almost the same point and are lower than the peak of the measurement data PDF. However, in Fig. 13, the ML estimate has the highest peak width and the smallest width of the curve. The VB, though having a lower peak, has a wider curve that follows closely with the measured data. In both cases, the PDF is smooth and is characterised by low amplitude and thus can be interpreted as background noise.

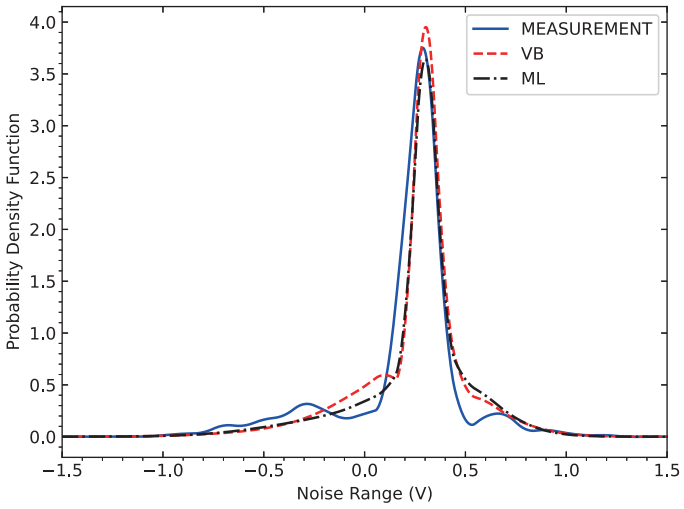


Figure 11. Machines laboratory PLC noise distribution (data-2).

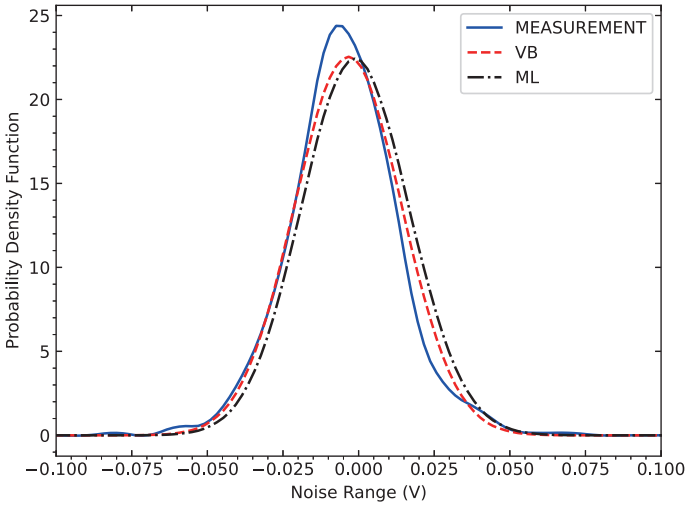


Figure 12. Computer laboratory PLC noise distribution (data-1).

5.3. Perform Analysis

In this study, the degree of dependence between the GM model and the measured data is assessed using Pearson's parametric correlation (R) test. The goodness-of-fit of the GM model is further examined using the root mean square error (RMSE) and χ^2 statistic. The formulations defined in (65), (66), and (67) are for RMSE, correlation coefficient, and the χ^2 statistic which was set to a significance level of 0.05.

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(x_o - x_g)^2}{N}} \tag{65}$$

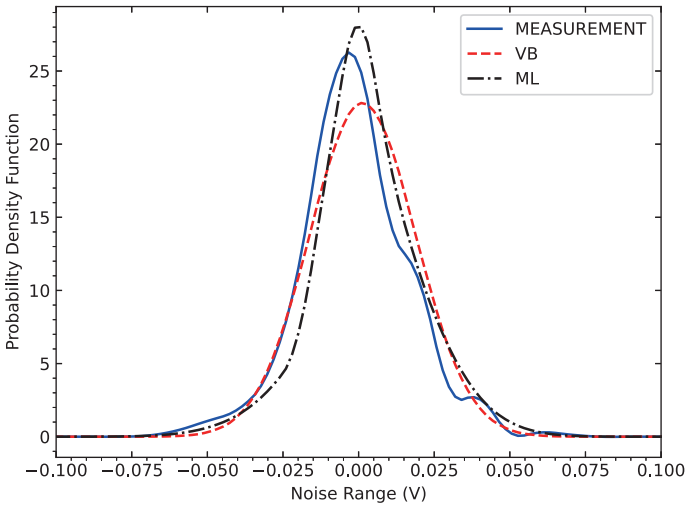


Figure 13. Computer laboratory PLC noise distribution (data-2).

$$R = \frac{\sum_{n=1}^{N}(x_o - \overline{x_o})(x_g - \overline{x_g})}{\sqrt{\sum_{n=1}^{N}(x_o - \overline{x_o})^2(x_g - \overline{x_g})^2}} \tag{66}$$

$$\chi^2 = \sum_{n=1}^{N}\frac{(x_o - x_g)^2}{x_g} \tag{67}$$

where: x_o is the measured value; x_g is the proposed model value; $\overline{x_o}$ and $\overline{x_g}$ are the means of the measured and the proposed model value while N is the measurement sample size. The performance results are summarised in Table 1.

Table 1. Performance analysis.

			VB			ML			
		Model Order	**R**	**RMSE**	χ^2	**R**	**RMSE**	χ^2	**Critical Value**
Electronic	**Data-1**	2	0.9749	0.0146	0.6399	0.9557	0.0193	0.6511	43.773
Laboratory	**Data-2**	2	0.9544	0.0202	0.5746	0.9548	0.02	0.5674	43.773
Machines	**Data-1**	3	0.9503	0.0389	16.1014	0.9572	0.0362	6.793	31.4104
Laboratory	**Data-2**	3	0.958	0.0349	2.361	0.9684	0.0303	1.8832	35.1725
Computer	**Data-1**	2	0.9946	0.1125	8.1073	0.9835	0.1963	9.5279	27.5871
Laboratory	**Data-2**	4	0.9724	0.2554	9.2498	0.9792	0.2225	7.8809	26.2962

From Table 1 it is observed that the correlation coefficient for all the measured data is above 0.95, and thus the VB and the ML models have a good correlation with the data. The RMSE values vary between 0.0146 and 0.2554 whereas all the χ^2 values are below the critical values indicating that both the proposed model and the ML model fit the data well. There is therefore no significant difference between the VB and ML models derived, with the density distribution acquired from the measured

data, and all of the models are consistent with the measured data with a 95% confidence level. As such, the VB model can be thus used to adequately model the amplitude distribution of the PLC noise. In all the measured data under consideration, the ML estimate performs better than the VB model due to its tendency to over-fit data, except for the first data from the Electronic Laboratory where the VB performs slightly better and the Computer Laboratory data-1 where the VB model outperforms the ML model. This is because there was the presence of singularity in the Computer Laboratory data-1, and regression analysis was used to determine the iteration with the parameters that best suited the measurement.

6. DISCUSSION OF RESULTS

It can be seen that the number of components needed to model the indoor PLC noise amplitude distribution ranges between two and four for the sampled data collected at various locations at the University of KwaZulu-Natal. It is observed that as the number of components increases, the accuracy of the model also increases with more complex models performing better where multiple peaks are present in the measured data. The ML estimate model captures the multiple peaks better than the VB model as shown in Fig. 10. This is due to the highly flexible property of the ML technique. However, it is also observed that the VB model provides a good generalisation of the amplitude distribution of PLC noise in the Machines Laboratory.

The PLC noise as seen by the receiver is a superposition of noise from various sources in the PLC network. In this case, the PLC noise is assumed to be independent and identically distributed and to follow a Gaussian distribution. Thus, the VB criterion seeks to find the optimal number of components that may have contributed to the final noise observed. From the noise amplitude distribution of the Machines Laboratory data, there are three peaks depicted in Figs. 10 and 11. This number of peaks corresponds to the VB criterion predicted value as the maximum lower bound is seen to occur at the three-component GM model. The four components chosen for the PLC noise amplitude distribution in the Computer Laboratory may be attributed to an overlap in the various components that comprise the GM. Although the measured amplitude distribution for the Electronic Laboratory has a higher amplitude impulsive noise level and outliers, the two-component GM model is selected as the optimum model. In [18], it was also observed that when the PDF of the amplitude noise has outliers, the likelihood of the model changes minimally as the model complexity increases.

The same was also found in [35], where the first two to three components were found to adequately approximate the PDF of the PLC impulsive noise amplitude using the Middleton Class A model. An extension of the Middleton Class A model four terms was also studied in [13]. Further research was carried out by [15], where five components were used to approximate the PLC impulsive noise amplitude distribution using the Middleton Class A model. It was confirmed that as the number of components in the mixture increases, a more accurate approximation is achieved. The same is also observed in [18], where the two, three, and four-component GM components were employed in the modelling of the PLC impulsive noise where the ML approach was employed in the determination of model parameters. In this study, the VB criterion has been employed to determine the appropriate number of components needed to adequately approximate the PDF of the PLC impulsive noise amplitude for a given observation window from different locations with different loading conditions. It is observed that the two, three, and four components are sufficient in the modelling of the PLC impulsive noise amplitude distribution.

As for the PDFs, the impulsive noise is seen to exhibit outliers in the measurement PDFs observed in the Electronic Laboratory in Figs. 8 and 9. The Machines Laboratory, in Figs. 10 and 11, is also characterised by high impulsive noise levels with multiple peaks. This may be attributed to the silicon-controlled rectifiers coupled with fluorescent lights that have been seen to contribute to high impulsive noise levels in broadband PLC. On the contrary, the Computer Laboratory exhibits low and smooth amplitude distribution levels as compared to those present in the Electronic and Machines laboratories as seen in Figs. 12 and 13. This is because the computers in this location employ switched mode power supply, which has been found to emit low-amplitude impulsive noise [8]. This is also evident in the time series analysis of the impulsive noise where the PLC noise amplitude PDF for the Computer Laboratory, in Fig. 4, is lower than that for the Electronics and Machines Laboratory in Figs. 2 and 3, respectively.

7. CONCLUSION

In this paper, the amplitude characteristic of the PLC impulsive noise, which is the main source of decreased performance and reliability of the PLC channel, has been modelled using the VB GM model. The proposed model provides a tractable and suitable estimation of the amplitude distribution of the PLC noise and does not suffer from the singularity present in the ML framework. Additionally, the proposed model provides a mechanism for model-order selection as well as parameter estimation which is crucial in the application of the GM for modelling of the PLC noise. It is, however limited because it provides an approximate solution rather than the exact solution since the aim is to maximise the lower bound. Despite this minor setback, the performance analysis results indicate that the VB model provides a good correlation as well as a high level of significance to the measured PLC noise data. Although the ML estimate model provides higher accuracy levels, the VB model also provides a good generalization of the measurement data. Therefore, either model, ML or VB, can be utilised as a simulation tool to optimize and develop efficient and reliable transmission schemes for PLC systems.

APPENDIX A. DERIVATION OF UPDATE EQUATIONS

The optimal solution for the k_{th} component means and precisions can be derived from (41) as:

$$\begin{aligned}\log q(\mu_k,\phi_k) \;=\; & \mathbb{E}_{\mu,\phi}\left(\frac{1}{2}\left(\log\phi_k-\log 2\pi-\phi(x_n-\mu_k)^2\right)\sum_{n=1}^{N}z_{nk}\right)\\ & +\mathbb{E}_{\mu,\phi}\left(\frac{1}{2}\left(\log\phi_k+\log s_0-\log 2\pi-\phi_k s_0(\mu_k-t_0)^2\right)\right)\\ & +\mathbb{E}_{\mu,\phi}\left((\alpha_0-1)\log\phi_k-\alpha_0\log\beta_0-\log\Gamma(\alpha_0)-\frac{\phi_k}{\beta_0}\right)\\ & +constant\end{aligned} \tag{A1}$$

Adding all terms not related to μ and ϕ to a constant, (A1) becomes:

$$\begin{aligned}\log q(\mu_k,\phi_k) \;=\; & \mathbb{E}_{\mu,\phi}\left(\left(\frac{1}{2}\sum_{n=1}^{N}z_{nk}+\frac{1}{2}+(\alpha_0-1)\right)\log\phi_k\right)\\ & -\mathbb{E}_{\mu,\phi}\left(\frac{\phi_k}{2}\sum_{n=1}^{N}z_{nk}\left(x_n-x_k^{+}+x_k^{+}-\mu_k\right)^2\right)\\ & -\mathbb{E}_{\mu,\phi}\left(\frac{\phi_k}{\beta_0}-\frac{\phi_k s_0}{2}(\mu_k-t_0)^2\right)+constant\end{aligned} \tag{A2}$$

where x_k^{+} is defined in (19). Solving (A2):

$$\begin{aligned}\log q(\mu_k,\phi_k) \;=\; & \left(\frac{1}{2}\sum_{n=1}^{N}\gamma_{nk}^{+}+\frac{1}{2}+(\alpha_0-1)\right)\log\phi_k\\ & -\frac{\phi_k}{\beta_0}-\frac{\phi_k}{2}\left(\sum_{n=1}^{N}\gamma_{nk}^{+}{x_k^{+}}^2+t_0^2 s_0+\sum_{n=1}^{N}\gamma_{nk}^{+}(x_n-x_k^{+})^2\right)\\ & -\frac{\phi_k}{2}\left(\left(\sum_{n=1}^{N}\gamma_{nk}^{+}+s_0\right){\mu_k}^2-2\left(\sum_{n=1}^{N}\gamma_{nk}^{+}x_k^{+}+t_0 s_0\right)\mu_k\right)\end{aligned} \tag{A3}$$

Completing the square for the last term in (A3) yields:

$$\begin{aligned}\log q(\mu_k,\phi_k) \;=\; & \left(\frac{1}{2}\sum_{n=1}^{N}\gamma_{nk}^{+}+\frac{1}{2}+(\alpha_0-1)\right)\log\phi_k-\frac{\phi_k}{\beta_0}\\ & -\frac{\phi_k}{2}\left(\sum_{n=1}^{N}\gamma_{nk}^{+}(x_n-x_k^{+})^2\right)-\frac{\phi_k}{2}\left(\sum_{n=1}^{N}\gamma_{nk}^{+}+s_0\right)\\ & \left(\mu_k-\frac{\sum_{n=1}^{N}\gamma_{nk}^{+}x_k^{+}+t_0s_0}{\sum_{n=1}^{N}\gamma_{nk}^{+}+s_0}\right)^2\\ & -\frac{\phi_k}{2}\left(\left(\sum_{n=1}^{N}\gamma_{nk}^{+}{x_k^{+}}^2+t_0^2s_0\right)+\frac{\left(\sum_{n=1}^{N}\gamma_{nk}^{+}x_k^{+}+t_0s_0\right)^2}{\sum_{n=1}^{N}\gamma_{nk}^{+}+s_0}\right)\end{aligned}\tag{A4}$$

Simplifying (A4):

$$\begin{aligned}\log q(\mu_k,\phi_k) \;=\; & \left(\frac{N_k}{2}+\frac{1}{2}+(\alpha_0-1)\right)\log\phi_k\\ & -\frac{\phi_k}{2}(N_k+s_0)\left(\mu_k-\frac{N_kx_k^{+}+t_0s_0}{N_k+s_0}\right)^2\\ & -\phi_k\left(\frac{1}{\beta_0}+\frac{1}{2}\left(N_ks_k^{+}+\frac{N_ks_0}{N_k+s_0}(x_k^{+}-t_0)^2\right)\right)\end{aligned}\tag{A5}$$

where N_k is given by (20), and the update parameters can then be obtained from (A4).

REFERENCES

1. Asiyo, M. O. and T. J. Afullo, "Analysis of bursty impulsive noise in low-voltage indoor power line communication channels: Local scaling behaviour," *SAIEE Africa Research Journal*, Vol. 108, No. 3, 98–107, 2017.
2. Benaissa, A., A. Abdelmalek, and M. Feham, "Improved reliability of power line communication under alpha-stable noise," *2017 5th International Conference on Electrical Engineering — Boumerdes (ICEE-B)*, 1–5, 2017.
3. Awino, S. O., T. J. O. Afullo, M. Mosalaosi, and P. O. Akuon, "Time series analysis of impulsive noise in power line communication (plc) networks," *SAIEE Africa Research Journal*, Vol. 109, No. 4, 237–249, 2018.
4. Awino, S., T. Afullo, M. Mosalaosi, and P. Akuon, "Measurements and statistical modelling for time behaviour of power line communication impulsive noise," *International Journal on Communications Antenna and Propagation*, Vol. 9, No. 4, 236–246, 2019.
5. Prakash, S., A. Bansal, and S. K. Jha, "Performance analysis of narrowband plc system under gaussian laplacian noise model," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 3597–3600, 2016.
6. Zimmermann, M. and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 44, No. 1, 249–258, 2002.

7. Vines, R. M., H. J. Trissell, L. J. Gale, and J. B. O'neal, "Noise on residential power distribution circuits," *IEEE Transactions on Electromagnetic Compatibility*, Vol. EMC-26, No. 4, 161–168, 1984.
8. Meng, H., Y. Guan, and S. Chen, "Modeling and analysis of noise effects on broadband power-line communications," *IEEE Transactions on Power Delivery*, Vol. 20, No. 2, 630–637, 2005.
9. Awino, S. O., T. J. Afullo, M. Mosalaosi, and P. O. Akuon, "Empirical identification of narrowband interference in broadband PLC networks at the receiver," *2018 Progress in Electromagnetics Research Symposium (PIERS-Toyama)*, 2160–2164, 2018.
10. Tonello, A. M., N. A. Letizia, D. Righini, and F. Marcuzzi, "Machine learning tips and tricks for power line communications," *IEEE Access*, Vol. 7, 82 434–82 452, 2019.
11. Bishop, C. M. and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, Vol. 4, No. 4, Springer, 2006.
12. Fertonani, D. and G. Colavolpe, "On reliable communications over channels impaired by bursty impulse noise," *IEEE Transactions on Communications*, Vol. 57, No. 7, 2024–2030, 2009.
13. Ndo, G., F. Labeau, and M. Kassouf, "A Markov-Middleton model for bursty impulsive noise: Modeling and receiver design," *IEEE Transactions on Power Delivery*, Vol. 28, No. 4, 2317–2325, 2013.
14. Suraweera, H. A. and J. Armstrong, "Noise bucket effect for impulse noise in OFDM," *Electronics Letters*, Vol. 40, No. 18, 1156–1157, 2004.
15. Shongwe, T., A. J. H. Vinck, and H. C. Ferreira, "A study on impulse noise and its models," *SAIEE Africa Research Journal*, Vol. 106, No. 3, 119–131, 2015.
16. Bai, T., H. Zhang, J. Wang, C. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Fifty years of noise modeling and mitigation in power-line communications," *IEEE Communications Surveys & Tutorials*, Vol. 23, No. 1, 41–69, 2020.
17. Awino, S., T. J. O. Afullo, M. Mosalaosi, and P. O. Akuon, "GMM estimation and BER of bursty impulsive noise in low-voltage PLC networks," *2019 Photonics & Electromagnetics Research Symposium — Spring (PIERS-Spring)*, 1828–1834, 2019.
18. Chelangat, F. and T. Afullo, "Low-voltage plc noise modelling," *International Journal on Communications Antenna and Propagation (IRECAP)*, Vol. 12, 237, August 2022.
19. Baroud, D., A. Hasan, and T. Shongwe, "A study towards implementing various artificial neural networks for signals classification and noise detection in ofdm/plc channels," *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, 1–6, 2020.
20. Nassar, M., K. Gulati, Y. Mortazavi, and B. L. Evans, "Statistical modeling of asynchronous impulsive noise in powerline communication networks," *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011. IEEE*, 1–6, 2011.
21. Liu, S., F. Yang, and J. Song, "An optimal interleaving scheme with maximum time-frequency diversity for PLC systems," *IEEE Transactions on Power Delivery*, Vol. 31, No. 3, 1007–1014, 2014.
22. Attias, H., "A variational Bayesian framework for graphical models," *Advances in Neural Information Processing Systems*, Vol. 12, 1999.
23. Corduneanu, A. and C. M. Bishop, "Variational Bayesian model selection for mixture distributions," *Artificial intelligence and Statistics*, Vol. 2001, 27–34, Morgan Kaufmann Waltham, MA, 2001.
24. Nasios, N. and A. Bors, "Variational learning for gaussian mixture models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 36, No. 4, 849–862, 2006.
25. Penny, W. and S. Roberts, "Variational bayes for 1-dimensional mixture models," *Techn. Rep. PARG–00–2*, Dept. of Engineering Science, University of Oxford, 2000.
26. Valente, F. and C. Wellekens, "Variational Bayesian speaker clustering," *ODYSSEY04 — The Speaker and Language Recognition Workshop*, 2004.
27. Roberts, S. and W. Penny, "Variational Bayes for generalized autoregressive models," *IEEE Transactions on Signal Processing*, Vol. 50, No. 9, 2245–2257, 2002.

28. Samakande, T., T. Shongwe, A. S. de Beer, and H. C. Ferreira, "The effect of coupling circuits on impulsive noise in power line communication," *2018 IEEE International Symposium on Power Line Communications and its Applications (ISPLC)*, 1–5, 2018.
29. Tina, J. A. W., A. J. Snyders, and H. C. Ferreira, "Implementation of a gap recorder for measuring impulsive noise error distributions in power line communications using the Fritchman model," *2012 IEEE International Symposium on Power Line Communications and Its Applications*, 374–379, 2012.
30. Emleh, A., A. S. de Beer, H. C. Ferreira, and A. J. H. Vinck, "The influence of fluorescent lamps with electronic ballast on the low voltage PLC network," *2014 IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014)*, 276–280, 2014.
31. Emleh, A., A. De Beer, H. Ferreira, and A. H. Vinck, "The impact of the cfl lamps on the power-line communications channel," *2013 IEEE 17th International Symposium on Power Line Communications and Its Applications*, 225–229, IEEE, 2013.
32. Antoniali, M., F. Versolatto, and A. M. Tonello, "An experimental characterization of the plc noise at the source," *IEEE Transactions on Power Delivery*, Vol. 31, No. 3, 1068–1075, 2016.
33. Tlich, M., H. Chaouche, A. Zeddam, and F. Gauthier, "Impulsive noise characterization at source," *2008 1st IFIP Wireless Days*, 1–6, 2008.
34. Ying, Y., "A note on variational Bayesian inference," https://www.albany.edu/ yy298919/realvb.pdf, 2009.
35. Vastola, K., "Threshold detection in narrow-band non-gaussian noise," *IEEE Transactions on Communications*, Vol. 32, No. 2, 134–139, 1984.